Matula
Kornerup
22 July 88

1. Rationally biased rounding can be beneficial.

The larger gaps between representable values in floating slash representation occur only where one of the boundary values is a particularly simple rational. Rounding by the number theoretic concept of best rational approximation appropriately chooses the simple rational. For problems where simple rational values may occur and be meaningful, this feature could be very useful. E.g. linear programming and combinatorial optimization are two of many areas where simple rational output may often arise.

Two references providing more information on this feature are:

Matula, D.W. and Kornerup, P.: "Approximate Rational Arithmetic Systems: Analysis of Recovery of Simple Fractions During Expression Evaluation", *Symbolic and Algebraic Computation*, E.W. Ng, ed., Lecture Notes in Computer Science **72**, Springer-Verlag, Berlin, 1979, 383-397.

Matula, D.W. and Ferguson, W.: "Rationally Biased Arithmetic", Proc. 7th Sym. on Comp. Arith., IEEE Cat #85CH2146-9, 1985, 194-202.

2. A precision fill feature can be added to floating slash representation.

The standard separate numerator and denominator word representation of rational fractions provides limitations in both range and precision features for many desired approximate real applications. Thus for approximation of reals in floating slash we must consider both a range problem and a precision problem.

Range problem: Our previously described extended range floating slash format allows scaled up numerators (implicit denominator unity) and scaled up denominators (implicit numerator unity) to achieve ranges comparable to typical floating point ranges. This provides a reasonably convenient solution to the range problem for rational representation.

Precision problem: We wish to point out a "precision fill" feature can be created in floating slash representation to achieve reasonably uniform relative spacing. This is possible by treating the presence of a non unit valued GCD as an indication that more accuracy is to be provided by interpreting the value of the GCD as giving leading bit information on the next partial quotient in the expansion of the approximate value. I.e. 1000/2000 in such a "denormalized" interpretation would have the GCD of 1000 indicate a small deviation from 1/2 whose precise meaning requires more discussion. We simply summarize here that the current redundancy of a simple fraction (GCD values available) for each fraction in floating slash is of just the right order to appropriately allow filling the gaps on both sides of the simple rational. "Appropriate" means that n bit floating slash can achieve accuracy of about say one part in $2^{(n-k)}$ for k a constant bounded by two or three in an appropriate development of this feature. This idea is currently under investigation by these authors as part of further developments in our investigation of floating slash arithmetic.

3. An alternative continued fraction based rational system is also available.

The use of a scaled version of k-bit LCF bitstrings provides an alternative to floating slash representation with a very similar and competitive arithmetic unit architecture.

Such systems have far less gap variation between k-bit representable values than in standard floating slash and still allow for exact representation of all appropriately simple rational values. We suggest that scaling to achieve a large range is a reasonably achievable extention to the unscaled k-bit LCF format. The unscaled LCF representation and corresponding arithmetic unit are described in the following two references:

Kornerup , P. and Matula, D.W.: "Finite Precision Lexicographic Continued Fraction Number Systems", Proc. 7th Sym. on Comp. Arith., IEEE Cat #85CH2146-9, 1985, 207-214.

"An On-Line Arithmetic Unit for Bit-Pipelined Rational Arithmetic", J. Parallel and Distributed Comp., **5**, 1988, 310-330.

4. Further references.

For further information on floating slash representation we note the following two references:

Kornerup, P. and Matula, D.W.: "Finite Precision Rational Arithmetic: An Arithmetic Unit", IEEE Trns. on Comp., **C-32**, 1983, 378-388.

Matula, D.W. and Kornerup, P.: "Finite Precision Rational Arithmetic: Slash Number Systems", IEEE Trans. on Comp., C-34 1985, 3-18.

David W. Matula
Peter Kornerup
Aarhus, 22 July 1988

Example from:

APPROXIMATE RATIONAL ARITHMETIC SYSTEMS: ANALYSIS OF RECOVERY OF SIMPLE FRACTIONS DURING EXPRESSION EVALUATION*

David W. Matula
Peter Kornerup

Recovery of Exact Simple Fractions: A Computational Example

Recovery of Exactness: Evaluation in fixed-slash or floating-slash approximate rational arithmetic of a rational expression whose true result is a relatively simple fraction will generally yield that exact simple fraction as the final result even though intermediate results are rounded.

We motivate this feature by the following example.

Example: Table 1 illustrates the computation of the determinant

$$D = \det \begin{vmatrix} \frac{10}{13} & \frac{20}{17} & \frac{1}{13} \\ \frac{11}{19} & \frac{7}{11} & \frac{77}{95} \\ \frac{69}{91} & \frac{4}{17} & \frac{56}{65} \end{vmatrix}$$

$= (((\frac{10}{13} \cdot \frac{7}{11}) \frac{56}{65} + (\frac{11}{19} \cdot \frac{4}{17}) \frac{1}{13}) + (\frac{69}{91} \cdot \frac{20}{17}) \frac{77}{95}) - (((\frac{69}{91} \cdot \frac{7}{11}) \frac{1}{13} + (\frac{11}{19} \cdot \frac{20}{17}) \frac{56}{65}) + (\frac{10}{13} \cdot \frac{4}{17}) \frac{77}{95}) = \frac{5}{13}$

in the order indicated by the parenthesis. The approximate computation where all intermediate fractions are rounded by mediant rounding to the last convergent not exceeding 3 decimal digits in numerator or denominator (e.g. ϕ_{999}) is shown in Table I along with the absolute and relative errors accumulated at each stage. Note that the final step of the computation involves the rounding of $\frac{320}{277} - \frac{84}{109} = \frac{11612}{30193}$, whose convergents are $0, \frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{3}{8}, \frac{5}{13}, \frac{1288}{3349}, \frac{2581}{6711}, \frac{11612}{30193}$, so that $\phi_{999}(\frac{11612}{30193}) = \frac{5}{13}$ and the true result is recovered by the final rounding.

The number theoretic foundations of mediant rounding serve to explain our success in this example. From Theorem 1 (vi) and the definition of ϕ_n in (5) it is evident that the absolute error in rounding to $\frac{p_i}{q_i}$ is at most $1/q_i q_{i+1} \leq 1/q_i(n+1)$. From basic properties of Farey series and mediants [Hardy and Wright, 1960, Ch. III] it is further noted that the interval rounding to $\frac{p_i}{q_i}$ must extend at least $1/q_i(n+q_i)$ on either side of $\frac{p_i}{q_i}$. Thus the interval rounding to $\frac{5}{13}$ must extend at least $1/(13 \times 1012) = 7.60 \times 10^{-5}$ on either side of 5/13, and the accumulated computation error had not grown sufficiently to escape this relatively large interval rounding to the value 5/13 in our example.

	exact	computed value	accumulated abs. error	accumulated rel. error
$t_1 = \frac{10}{13} \cdot \frac{7}{11}$	$\frac{70}{143}$	$\frac{70}{143}$	0	0
$t_2 = t_1 \cdot \frac{56}{65}$	$\frac{784}{1859}$	$\frac{229}{543}$	$9.9 \cdot 10^{-7}$	$2.2 \cdot 10^{-6}$
$t_3 = \frac{11}{19} \cdot \frac{4}{17}$	$\frac{44}{323}$	$\frac{44}{323}$	0	0
$t_4 = t_3 \cdot \frac{1}{13}$	$\frac{44}{4199}$	$\frac{7}{668}$	$3.5 \cdot 10^{-7}$	$3.4 \cdot 10^{-5}$
$t_5 = \frac{69}{91} \cdot \frac{20}{17}$	$\frac{1380}{1547}$	$\frac{157}{176}$	$3.7 \cdot 10^{-6}$	$4.1 \cdot 10^{-6}$
$t_6 = t_5 \cdot \frac{77}{95}$	$\frac{3036}{4199}$	$\frac{449}{621}$	$1.9 \cdot 10^{-6}$	$2.6 \cdot 10^{-6}$
$t_7 = \frac{69}{91} \cdot \frac{7}{11}$	$\frac{69}{143}$	$\frac{69}{143}$	0	0
$t_8 = t_7 \cdot \frac{1}{13}$	$\frac{69}{1859}$	$\frac{17}{458}$	$-1.17 \cdot 10^{-6}$	$3.2 \cdot 10^{-5}$
$t_9 = \frac{11}{19} \cdot \frac{20}{17}$	$\frac{220}{323}$	$\frac{220}{323}$	0	0
$t_{10} = t_9 \cdot \frac{56}{65}$	$\frac{2464}{4199}$	$\frac{169}{288}$	$8.2 \cdot 10^{-7}$	$1.4 \cdot 10^{-6}$
$t_{11} = \frac{10}{13} \cdot \frac{4}{17}$	$\frac{40}{221}$	$\frac{40}{221}$	0	0
$t_{12} = t_{11} \cdot \frac{77}{95}$	$\frac{616}{4199}$	$\frac{109}{743}$	$-9.6 \cdot 10^{-7}$	$6.6 \cdot 10^{-6}$
$t_{13} = t_2 + t_4$	$\frac{259524}{600457}$	$\frac{51}{118}$	$7.4 \cdot 10^{-6}$	$1.7 \cdot 10^{-5}$
$t_{14} = t_{13} + t_6$	$\frac{693672}{600457}$	$\frac{320}{277}$	$5.4 \cdot 10^{-6}$	$4.7 \cdot 10^{-6}$
$t_{15} = t_8 + t_{10}$	$\frac{374639}{600457}$	$\frac{73}{117}$	$-8.5 \cdot 10^{-6}$	$1.3 \cdot 10^{-5}$
$t_{16} = t_{15} + t_{12}$	$\frac{462727}{600457}$	$\frac{84}{109}$	$-1.7 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$
$D = t_{14} - t_{16}$	$\frac{5}{13}$	$\frac{5}{13}$	0	0

Table I. Illustration of the recovery of the exact simple fraction result in evaluating a particular determinant although intermediate approximate results were rounded to at most 3 digits in numerator and denominator.