BILL GOSPER (date unknown)
Hackmem



ITEM 97 (Schroeppel):
Simple proofs that certain continued fractions are $\sqrt{2}$, $\sqrt{3}$, etc.
Proof for $\sqrt{2}$:

$X = [1, 2, 2, 2, \ldots]$
$(X-1)(X+1) = [0, 2, 2, 2, \ldots] * [2, 2, 2, 2, \ldots] = 1$
$X^2 - 1 = 1$
$X = \sqrt{2}$

Proof for $\sqrt{3}$:

$Y = [1, 1, 2, 1, 2, \ldots]$
$(Y + 1)(Y - 1) = [2, 1, 2, 1, 2, \ldots] * [0, 1, 2, 1, 2, \ldots]$
$= 2 * [1, 2, 1, 2, 1, \ldots] * [0, 1, 2, 1, 2, \ldots] = 2$
$Y^2 - 1 = 2$
$Y = \sqrt{3}$

Similar proofs exist for $\sqrt{5}$ and $\sqrt{6}$; but $\sqrt{7}$ is hairy.

ITEM 98 (Schroeppel):
The continued fraction expansion of the positive minimum of the factorial function (about 0.46) is
[0, 2, 6, 63, 135, 1, 1, 1, 1, 4, 1, 43, ...].

ITEM 99 (Schroeppel):
The value of a continued fraction with partial quotients increasing in arithmetic progression is

$$[A+D, A+2D, A+3D, \ldots] = \frac{I_{1+(A/D)}(2/D)}{I_{A/D}(2/D)}$$

where the I's are Bessel functions.

A special case is $[1, 2, 3, 4, \ldots] = \frac{I_0(2)}{I_1(2)}$.

ITEM 100 (Perron):

$$\prod_{k=1}^{n} (1 + 1/Ak) =$$

$$1 + \frac{1}{A1 -} \quad \frac{(A1 + 1)A1}{A1+A2+1 -} \quad \frac{(A2 + 1)A2}{A2+A3+1 -} \quad \cdots \quad \frac{(A(n-1) + 1)A(n-1)}{A(n-1)+An+1}.$$

RWG at SAIL on Arpanet



ITEM 101A (Gosper):
On the theory that continued fractions are underused, probably because of their unfamiliarity, I offer the following propaganda session on the relative merits of continued fractions versus other numerical representations. For a good cram course in continued fractions, see Knuth, volume 2, page 316. (In what follows, "regular" means that all numerators are 1, and any radix can be read in place of decimal.)

0) $\pi$ is 3. But not really 3, more like 3 + 1/7. But not really 7, more like 7 + 1/15. But not really 15, ... . So the regular continued fraction for $\pi$ is written 3 7 15 1 292 1 1 ... .

1) The continued fractions for rational numbers always come out even, and rather quickly. Thus, the number of inches per meter is exactly 100/2.54 or 39 2 1 2 2 1 4. The corresponding decimal fraction 39.3700787... has period 42, making it almost impossible to tell if the number is rational. (But if our data are ALL rational, the ordered pair 5000/127 is even more concise.)

2) Quadratic surds, which are of course inexpressible as rationals, are generally unrecognizable in decimal. Their continued fractions, on the other hand, are periodic. Nth roots of $e^2$, ratios of Bessel functions, and ratios of linear functions of these all have regular continued fractions formed by interleaving one or more arithmetic sequences. These special properties will show up regardless of number base. You might recognize 5.436563... as 2e, but even Schroeppel might not notice that 6.1102966796... was $(4e^{2/3} - 2)/(e^{2/3} - 1)$ until he wrote it as 6 9 15 21 27 33 ... .

The familiar transcendental functions of rational arguments also have simple continued fractions, but these are generally not regular and cannot be reconstructed from numerical values by a simple algorithm, since nonregular representations aren't unique. The point is, however, that numbers like e, $\pi$, $\sqrt{2}$, sin .5, $\sqrt{7}$ arctan $\sqrt{7}$, etc. can be expressed to unlimited precision by simple programs which produce the terms on demand.

3) If we define a rational approximation to be "best" if it comes closer than any other rational with such a small denominator, then continued fractions give the complete set of best rational approximations to the value which they represent. That is, if you truncate a (regular) continued fraction at any point, then the resulting rational number is a best approximation. Furthermore, this remains true if the last term of this approximation is replaced by any smaller positive integer other than 1. All best approximations can be generated in this manner, in order of increasing denominators (or numerators). For example, the approximants to $\pi$ = 3 7 15 1 292 ... are:

```
 3:    1/1, 2/1, 3/1
 7:    (4/1), 7/2, 10/3, 13/4, 16/5, 19/6, 22/7
15:    (25/8), 47/15, 69/22, 91/29, 113/36, ... 311/99, 333/106
 1:    355/113
...      ...
```

Note that they are all automatically in lowest terms. The size of a denominator is greater than the product of the terms involved and less than the product of the numbers 1 greater than the terms. The approximations are low if the number of terms is odd, high if it's even. (Note that if a 1 ends a continued fraction, it should be added in to the previous term. Thus, to "round off" a continued fraction after a certain term, add in the next term iff it is ±1. In the above, 4/1 and 25/8 correspond to termination with a 1 and are not "best"; 355/113 is "best" because the corresponding term really should be 1.) The error is smaller than 1 over the product of the denominator squared and the first neglected term, so that the total number of digits (numerator and denominator) is usually slightly smaller than with equally accurate decimal fractions. 355/113 is good to 7.5 places instead of 5.5, due to the unusually large term (292) which follows.

4) Numerical comparison of continued fractions is slightly harder than in decimal, but much easier than with rationals -- just invert the decision as to which is larger whenever the first discrepant terms are even-numbered. Contrast this with the problem of comparing the rationals 113/36 and 355/113.

5) Regular continued fractions are in 1 to 1 correspondence with the real numbers, unlike decimal (.5 = .49999...) or rationals (2/3 = 6/9, $\sqrt{6}$ = ?). Even infinity has a continued fraction, namely, the empty one! (Minus and plus infinity are the same in continued fraction notation.)

6) Each representation favors certain operations. Decimal favors multiplication by powers of 10. Rationals favor reciprocation, as do continued fractions. To reciprocate a regular continued fraction, add (or if possible, delete) an initial 0 term. To negate, negate all the terms, optionally observing that -a, -b, -c, -d ... = -a-1, 1, b-1, c, d ... .

7) The strongest argument for positional (e.g., decimal or floating) representation for non-integers is that arithmetic is easy. Rational number arithmetic often loses because numerators and denominators grow so large as to require icky multiprecision. Algorithms for arithmetic on continued fractions seem generally unknown. The next items describe how to arithmetically combine continued fractions to produce new ones, one term at a time.

Unfortunately, the effort required to perform these operations manually is several times that for decimal, but the rewards for machine implementation are considerable (which can also be said of floating point). Specifically, these rewards will be seen to be: unlimited significance arithmetic without multiprecision multiplication or division, built in error analysis, immorally easy computation of algebraic functions, no unnecessary computations, no discarding of information (as with roundoff and truncation), reversibility of computations, and the terms of the answer start to come out right away and continue to do so until shut off.

ITEM 101B (Gosper):
Continued Fraction Arithmetic

Continued fractions let us perform numerical calculations a little at a time without ever introducing any error, such as roundoff or truncation. As if this weren't enough, the calculations provide automatic error analysis, and obviate most forms of successive approximation. This means we can start with an arithmetic expression like $\sqrt{3/\pi^2 + e} / (\tanh \sqrt{5} - \sin 69)$ and immediately begin to produce the value as a sequence of continued fraction terms (or even decimal digits, if we should be so reactionary), limited only by time and storage. If there are quantities in the expression which are known only approximately, the calculation can provide error bounds on the answer as well as identify the quantity that limited the significance.

All this is possible because each operation (+, /, -, $\sqrt{\ }$) in the arithmetic expression requests terms from the continued fractions of its operands only when necessary, and consequently produces terms of its own value as soon as possible. Numbers like $\pi$ and e and functions like sin and tanh have continued fraction terms in simple sequences which can be produced by short programs. Imprecise quantities can also be programs which deliver terms until they run out of confidence, whereupon they initiate special action. By then, the last guaranteeable term of the overall expression will have already been produced.

We see then that no calculation is performed unnecessarily, so that, for example, a subexpression which happened to be multiplied by zero would never be evaluated. Also, an operation detecting a deficiency in two or more of its operands provides a natural mechanism for allocating multiprocessor resorces, should you have some.

Here are the algorithms for the elementary arithmetic operations on continued fractions.

Let x be a continued fraction $p_0 + q_0/(p_1 + q_1/( \ldots = p_0 + q_0/x'$ where x' is again a continued fraction and the p's and q's are integers. We shall call a (p q) pair a "term" of the continued fraction for x. Often, only the p's are mentioned, in which case the q's are implicitly all 1, and x is called a "regular" continued fraction.

Instead of a list of p's and q's, let x be a computer subroutine which produces its next $p_0$ and $q_0$ each time it is called. Thus on its first usage x will "output" $p_0$ and $q_0$ and, in effect, change itself into x'. Similarly, let y be another procedurally represented continued fraction $r_0 + s_0/y'$. Our problem will be solved if we can write such subroutines for $z(x,y) = x+y$, $x-y$, $xy$, and $x/y$. When called upon to output a term of z, the subroutine might in turn call for (or "input") terms from x and y until it is satisfied that the unread portions of x and y cannot affect the pending term of z. Then it would output this term and change itself into z', so that it could produce the next term next time. Unfortunately, when we try to do this, our expressions quickly complicate. Let us preempt this complication by computing instead the more general function

$$z(x,y) = (axy+bx+cy+d)/(exy+fx+gy+h)$$

(or (a b c d)/(e f g h) for short) where a through h are integer variables whose initial values we are free to choose. Various choices express

addition: $x+y = (0\ 1\ 1\ 0)/(0\ 0\ 0\ 1)$,
subtraction: $x-y = (0\ 1\ -1\ 0)/(0\ 0\ 0\ 1)$,
multiplication: $xy = (1\ 0\ 0\ 0)/(0\ 0\ 0\ 1)$,
division: $x/y = (0\ 1\ 0\ 0)/(0\ 0\ 1\ 0)$.

As we shall see, the process of inputting terms of x and y and outputting terms of z will reduce to replacing the eight integers a through h with linear combinations of each other.

When z inputs a term of x, z becomes a new function of x'. To see how this happens, substitute $p + q/x'$ for every occurrence of x in the expression for $z(x,y)$, then multiply numerator and denominator through by x':

$$z(x',y) = (pa+c\ pb+d\ qa\ qb)/(pe+g\ pf+h\ qe\ qf).$$

If x was rational and has run out of terms, it has in effect become infinite:

$$z(\infty,y) = (0\ 0\ a\ b)/(0\ 0\ e\ f)$$

If instead we input a term of y by substituting r + s/y' for every occurrence of y:

$$z(x,y') = (ra+b\ sa\ rc+d\ sc)/(re+f\ se\ rg+h\ sg).$$

If y runs out of terms:

$$z(x,\infty) = (0\ a\ 0\ c)/(0\ e\ 0\ g)$$

To output the term (t u), so that z = t + u/z' (i.e., z' = u/(z-t)):

$$z'(x,y) = (ue\ uf\ ug\ uh)/(a-te\ b-tf\ c-tg\ d-th).$$

Thus this basic eight variable form is preserved by all three operations, which can be performed in any order since they represent independent substitutions.

For simplicity, let us assume that z will output in standard form, that is, every u = 1 (regular) and every output term $t \geq 1$ except perhaps the first. This means that z' will always exceed 1 and thus $0 \leq u/z' < 1$, so that the integer t = z - u/z' must = [z], the greatest integer $\leq$ z.

Since z generally varies with x and y, it should not output unless [z] is constant for the range of possible x and y. We can easily compute the range of z given the ranges of x and y if we represent each range by the endpoints of an interval (in either order), along with a bit indicating Inside or Outside. Thus if z is in standard form, we can say that z will always be (Inside 1 $\infty$) (or (Outside $-\infty$ 1)) after the first term. If z were to always output its nearest integer instead of its greatest, then none of the terms after the first would be 1, although they would probably vary in sign. In this case, z would be (Outside -2 2).

Now hold y fixed and examine the behavior of z with x. If x is (Inside a b) then z is (Inside z(a) z(b)) unless the denominator of z changes sign between a and b (i.e., z has its pole in this interval), whereupon z is (Outside z(a) z(b)). Symmetrically, when x is (Outside a b) then z is (Outside z(a) z(b)) unless the signs of the denominators of z(a) and z(b) differ, whereupon z is (Inside z(a) z(b)). This argument still holds with x and y interchanged.

Now suppose that with y fixed at one of its endpoints, x constrains z (Inside 1 2), and at y's other extreme, z(x) is (Outside 0 3). Suppose further that at the two extremes of x, z(y) is (Inside 1 3) and (Outside 0 2). Then z(x,y) is (Outside 0 1), the union of the four ranges. (Outside 0 2) is the widest, indicating that z will probably get more information from a term of y than a term of x. (Topology hackers should recognize this Inside-Outside nonsense as ordinary intervals in toroidal space. The clue is that both plus and minus infinity are denoted by the empty continued fraction.)

Due to the basically monotonic behavior of z, we can guarantee that the actual range of z will be the union of these four ranges, and that this range will be Inside or Outside some interval. If it is (Inside z1 z2) and [z1] = [z2], z can output the term t = [z1]. Otherwise, z must input a term from x or y, whichever was associated with the widest of the four ranges of z. (Outside narrowness) is wider than (Outside wideness) is wider than (Inside wideness) is wider than (Inside narrowness).

Evaluating z on these endpoints may be facilitated by keeping estimates for the integer variables in floating point.

Even if z doesn't produce a term, narrowing the range of possible z will still help in computing the range of a function of z, especially if z gets stuck trying to output the last term of a rational number resulting from irrational x and y. (There is no way to guarantee that x or y won't eventually deviate, whereupon z would egest a gigantic term.)

z can produce its value as decimal digits by multiplying by 10 instead of reciprocating, after outputting t = [z]:

z'(x,y) = (10(a-te) 10(b-tf) 10(c-tg) 10(d-th))/(e f g h).

Strange to say, it is not serious if z for some reason outputs the terms 7 5 1 when it should have produced 6 9. As soon as permitted, it will simply recant with 0 -1 -5 and continue with the correction -1 9. The sequence 7 5 1 0 -1 -5 -1 9 is equivalent to 6 9 because b 0 c is the same as b+c. In order to undo these computations, z violates the condition (Outside -1 1) when it is 0 -1 -5 ... . This condition is obeyed by nearly all convergent continued fractions after their first term, and its violation will very probably cause further retractions among the functions dependent upon z.

This computation reversal trick is also handy for mechanizing and denoting imprecise quantities. Instead of 2.997930 ±.000003, we have 2 1 481 0 2, meaning between 2 1 481 and 2 1 483. Similarly, 137 26 0 1 replaces 137.0373 ±.0006.

Successive approximations methods benefit considerably from not requesting terms until needed. Consider Newton's method for algebraic roots. We expect successive approximations to have about twice as many correct terms each time. Since the production of these terms cannot be aided by reading incorrect terms, the additional correct terms must be produced before the bad ones of the previous approximation are used. But this means that there is no need to read in the bad ones at all. By feeding back the output terms in place of the approximation, we get the correct answer directly! (69% of the credit for this goes to Schroeppel.)

The basic eight variable form exemplified above by z(x,y) is not the only form preserved by continued fraction term transactions. We need only four variables and a single interval check to compute z(x) = (ax+b)/(cx+d), the homographic function of one argument. On the other hand, z(w,x,y) (linear in all three arguments) requires sixteen variables and a twelve way interval check. Each of these forms can be solved for x in terms of z etc. to get a function of the same form. This is not true of

$$z(x) = (ax^2+bx+c)/(dx^2+ex+f),$$

for example, even though this form is also preserved. This form is not guaranteed monotone, thus theoretically invalidating the interval check algorithm, but it hardly ever errs. Even if it did, it would quickly correct itself anyway. This form is not only more economical than z(x,x), it is essential for the success of the Newton's method feedback trick, which must know when two variables are really the same one.

By choosing the eight coefficients a through h properly, it should be possible to rewrite arithmetic expressions as compositions of considerably fewer of these forms than one for each +, -, *, and /. The reader is invited to investigate the problem of trying to find minimal representations. Depending on the metric for minimality, the question can be complicated by allowing higher powers of x and y. If the highest powers of x, y, z, ... in an invariant form are i, j, k, ..., then the number of integer variables required for the coefficients (mostly because of all of the cross terms) is $2(i+1)(j+1)(k+1)\ldots$ .

It is awkward in this system to evaluate transcendental functions of irrational arguments. The problem is that you may need any number of continued fraction (or series) terms which, instead of being numbers, are symbolic functions of x, some infinite continued fraction. My suggestion is to represent each symbolic term of the function by a subroutine which is a function of x and the next term, with this next term really a dummy until actually called upon for output, whereupon it replaces itself with a full fledged term subroutine which in turn refers to x and a new dummy.

Sad to say, the integer variables in these algorithms do not usually shrink on outputs as much as they grow on inputs. Fortunately, the operations for input and output only require (besides addition) multiplication by terms which are almost invariably small. (I have not seen a term exceed 20776 except in specially constructed numbers.) It is fairly safe, then, to declare any function which has gotten (Outside $-2^{35}$ $2^{35}$) to be infinite, thus terminating its continued fraction. Better still, note that the term 20776 is equivalent to the terms 20000 0 700 0 70 0 6, i.e., a very large term can be transmitted piecewise. Although this is just thinly disguised multiprecision multiplication, that first piece of the term will probably satisfy its recipient for quite some time.

In some special cases, the integer variables will become periodic rather than large, especially when all but one of the arguments to a function have terminated. Then, we have the form $z(x) = (ax+b)/(cx+d)$, known as a homographic function. If ad-bc is $\pm 1$, then a, b, c, d will eventually become 1, 0, 0, 1, whereupon z will output the terms of x unmodified. Periodicity will also occur when x is a Hurwitz number, i.e., when the terms of x are the values of one or more polynomials evaluated on consecutive integers and then interleaved. Coth 1/69, $\sqrt{105}$, and e are Hurwitz numbers whose polynomials are linear or constant. Hurwitzness is preserved by homographic functions. If one can show that $\pi$ is not a Hurwitz number, one confirms the long standing conjectures that $e*\pi$, $e+\pi$, $e/\pi$, etc. are all irrational.

If z, x, and y are all regular, then it generally won't be possible to reduce z by finding a GCD of a through h which is > 1. However, it has been determined empirically that much reduction is often possible in other cases. This reduction is almost always by a divisor of an input or output term numerator (or 10 if output is decimal digits) and can be facilitated by keeping cetain of the integer variables around modulo these quantities.

ITEM 101C (Gosper):
Problem: Given an interval, find in it the rational number with smallest numerator and denominator.
Solution: Express the endpoints as continued fractions. Find the first term where they differ and add 1 to the lesser term, unless it's last. Discard the terms to the right. What's left is the continued fraction for the "smallest" rational in the interval. (If one fraction terminates but matches the other as far as it goes, append an infinity and proceed as above.)