ARTICLES

Continued Fractions Without Tears

Ping-pong using Farey sequences is proposed as an alternative to the traditional fraction chain.

IAN RICHARDS
University of Minnesota
Minneapolis, MN 55455

The traditional presentation of continued fractions is via an infinite sequence of quotients within quotients. For example:

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cdots}}}.$$

This "python descending a staircase" format has the advantage of historical priority, and it suggests important generalizations. However, there are approaches that are more conceptual, and I would like to outline one of them. (For the sake of comparison, we will untangle the infinite fraction in its traditional form and tie it together with our theory at the end of the paper.) If by the end of this paper you feel that this is all pretty trivial, then I have succeeded; if you think it's tough, then this approach isn't your cup of tea. Our goal will be to discover and prove the essential facts about continued fractions and to develop a computer algorithm for them.

The spirit of this presentation is geometrical, but it is geometry in an unusual setting: that of one-dimensional space. Remember the traditional drill sergeant's complaint: "Can't you tell your left hand from your right hand?" Our approach will be, essentially, to keep careful track of which points lie right and which ones lie left.

The Farey process

Now let's take a look at the theory. A pair of nonnegative fractions,

$$\frac{a}{b} < \frac{c}{d},$$

is called a **Farey pair** if $bc - ad = 1$. This means, of course, that the difference between the fractions is $1/bd$. The **mediant** of these two fractions is defined to be $(a+c)/(b+d)$. For example, $2/3 < 3/4$ is a Farey pair, with mediant $5/7$. A trivial calculation shows that the mediant always lies between a/b and c/d, so we have:

$$\frac{a}{b} < \underset{\text{(mediant)}}{\frac{a+c}{b+d}} < \frac{c}{d}.$$

For convenience, we call the interval $[a/b, c/d]$ a **Farey interval** if a/b and c/d are a Farey pair.

These ideas have a curious history. Farey discovered some properties of the mediant but was unable to prove them; the proof was supplied by Cauchy, who named the theory after its supposed discoverer. However, both were unaware that Haros had proved the same theorems several years before. A poignant comment on the capriciousness of fame was made by Hardy and Wright [**2**]: "Farey has a notice of twenty lines in the [British] 'Dictionary of National Biography,' where he is described as a geologist. As a geologist he is forgotten, and his biographer does not mention the one thing in his life which survives."

LEMMA 1. *Let* $[a/b, c/d]$ *be a Farey interval and consider the mediant* $(a+c)/(b+d)$ *(see* FIGURE 1). *Then*:

(i) *the two subintervals formed by inserting the mediant are also Farey intervals*;

(ii) *among all the fractions* x/y *lying strictly between* a/b *and* c/d, *the mediant is the one (and only one) with the smallest denominator.*

As we shall see, part (ii) of Lemma 1 is the key to the whole theory.

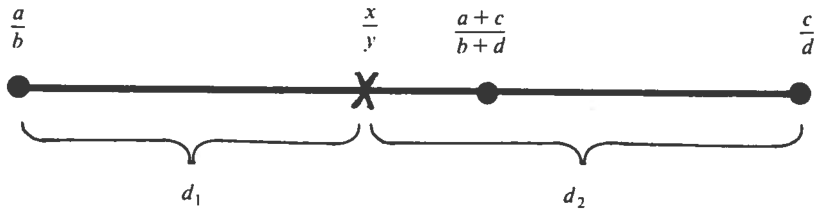


FIGURE 1. **The geometric setup in Lemma 1.**

Proof. Part (i) is easy. For part (ii), take any $x/y \neq (a+c)/(b+d)$ in the open interval $(a/b, c/d)$. We first assume that $y<(b+d)$ and derive a contradiction. The distance $d_1 = (x/y) - (a/b)$ (see FIGURE 1) is equal to $(bx-ay)/by$. The numerator $(bx-ay)$ is a positive integer and thus $\geqslant 1$; hence

$$d_1 \geqslant \frac{1}{by}.$$

Similarly

$$d_2 = \frac{c}{d} - \frac{x}{y} \geqslant \frac{1}{dy},$$

and so

$$d_1 + d_2 \geqslant \frac{1}{by} + \frac{1}{dy} = \frac{b+d}{y} \cdot \frac{1}{bd}.$$

But since $a/b, c/d$ is a Farey pair, the distance $d_1 + d_2 = 1/bd$, and the assumption that $y<(b+d)$ leads to a contradiction.

Now we come to the case $y = b+d$. We can handle this without computation, using the following device. From part (i), the mediant $(a+c)/(b+d)$ partitions the original interval into two Farey subintervals. We apply what we already know to these subintervals. The mediants for these subintervals must have denominators larger than $(b+d)$. So by what we have already proved, there is no x/y *inside* either of these subintervals with $y \leqslant (b+d)$. (Yes, $\leqslant$.)

The procedure used in the last paragraph of the proof above provides an introduction to a technique which is called the **slow continued fraction algorithm**. This is a method for finding the "best" rational approximations to an irrational number α. (For convenience, we will assume that $0<\alpha<1$.) However, it turns out to be easier to forget about α for a moment, and study the technique in the absence of its object. In that case, the technique is called the **Farey process**. Later it will become clear how we apply the Farey process to zero in on α.

0/1								1/1
0/1				1/2				1/1
0/1		1/3		1/2		2/3		1/1
0/1	1/4	1/3	2/5	1/2	3/5	2/3	3/4	1/1

TABLE 1. **The Farey process.**

1/3								2/5
1/3				3/8				2/5
1/3		4/11		3/8		5/13		2/5
1/3	5/14	4/11	7/19	3/8	8/21	5/13	7/18	2/5

TABLE 2. **A continuation of Table 1, starting with the Farey pair 1/3 and 2/5.**

Here is the Farey process. Start with a Farey pair a/b and c/d. Take their mediant; this creates two subintervals and two new Farey pairs. Form all possible mediants again; this gives four subintervals.... The process is illustrated in TABLES 1 and 2. (For a curious property of this algorithm, see the paper of Shrader-Frechette in this *Magazine* [3].) Now it is clear how we zero in on an irrational number α. We simply, at each stage, keep the interval that contains α and discard the rest. This is the slow continued fraction algorithm. (The "fast" or standard algorithm is a refinement of the slow one, and we will describe it presently.)

If α were a rational number, $\alpha = p/q$, then the situation could be different, for p/q might appear as one of the division points in the Farey process. In fact that always happens! This, essentially, is the theorem that Farey discovered but couldn't prove. (Actually Farey looked at it differently, which is one reason why he couldn't prove it.) Farey's formulation—which has many uses, but is the wrong approach here—can be found in any standard textbook on number theory, e.g., [2], p. 23.

From now on, we will assume that the Farey process begins with the numbers 0/1 and 1/1.

THEOREM 1. *Every rational number p/q in lowest terms, with $0<p/q<1$, appears at some stage of the Farey process.*

EXAMPLE. The reader might want to continue the process in TABLE 2 until he finds the fraction 37/100, which lies between 7/19 and 3/8.

Proof of Theorem 1. As expected, we use Lemma 1 above. After that, the proof almost writes itself. Suppose that a given fraction p/q between 0 and 1 never shows up in the Farey progression. Then at every stage, p/q remains squeezed between two adjacent fractions in the Farey process. By part (i) of Lemma 1, these fractions will always be a Farey pair. But the denominators of these fractions increase without bound (on one side of p/q, at least). Eventually the sum of the denominators exceeds q, and this violates Lemma 1, part (ii).

We seem to have strayed from our objective, which is the good approximation of irrational numbers by rationals. Now we come back to it. First we define what we mean by a "good" approximation.

DEFINITION 1. Let α be an irrational number with $0<\alpha<1$. Then a fraction p/q is called a **best left** (respectively, **best right**) approximation to α if:

(i) $p/q<\alpha$ (respectively, $p/q>\alpha$);
(ii) there is no fraction x/y between p/q and α with a denominator $y \leq q$.

Thus we put the left and right approximations to α into separate categories which do not compete against each other (like the American League and National League in baseball, before the World Series). After that, we give preference to fractions with small denominators. The small denominators, of course, are the whole point. (It doesn't take all this fuss to prove that the rational numbers are dense in the reals!) What we seek are classy approximations, like the famous estimates 22/7 and 355/113 for π. Thus 355/113 gives π correctly to six decimal places, although the denominator in this fraction is scarcely over a hundred. [The next term in the continued fraction expansion for π cannot be found on a hand calculator, because calculators round off after about ten digits, and the next term is more accurate than that.] The following theorem shows how to discover these good approximations.

THEOREM 2. *Take any irrational number α, with $0<\alpha<1$. The slow continued fraction algorithm (= the Farey process, zeroed in on α) gives a sequence of best left and right approximations to α. Every best left/right approximation arises in this way.*

Proof. We use Lemma 1 and Theorem 1. Thus consider any Farey pair a/b and c/d. Lemma 1 tells us that all fractions lying between a/b and c/d have denominators larger than either b or d. Hence these fractions are not in competition with a/b or c/d, and automatically, a/b and c/d furnish best left/right approximations to all irrational numbers α lying between them. This proves the first part.

Now from Theorem 1, which tells us that all fractions occur in the Farey process, we will prove the second part. Take any fraction p/q between 0 and 1 and consider the first time it appears in the Farey process. Then p/q must be the mediant of its two neighbors: call them a/b and c/d. Thus we have the three adjacent terms (see also TABLE 2):

$$\frac{a}{b}<\frac{p}{q}<\frac{c}{d}$$

with $p=a+c, q=b+d$. Now there are two possibilities. If α lies between a/b and c/d (and hence in an interval also bounded by p/q), then p/q is a term in the slow continued fraction algorithm for α. If not (so that α lies left of a/b or right of c/d), then one of the fractions $a/b, c/d$ beats p/q on all counts, by being closer to α and having a smaller denominator; and thus p/q is not a "best" approximation to α.

The fast continued fraction algorithm

We turn now to the fast or standard continued fraction algorithm. This involves a selection of certain exceptionally good approximations furnished by the slow algorithm. It is important to note that the fast algorithm finds nothing new; in fact, it finds less, but by doing so goes faster. Here is a description of it.

Recall that the slow algorithm involves a series of shrinking Farey intervals zeroing in on an irrational number α. At each stage in this shrinking process, one end of the Farey interval moves in closer to α, and the other end stays put. Now it may happen that the same end, say the left one, moves several times in a row before the right-hand end moves again. Suppose that the left endpoint moves altogether s times. In the slow continued fraction algorithm we would keep all of the resulting points, since all of them furnish best left approximations to α. In the fast algorithm, we retain only the last (or sth) point.

To understand this better, it may be helpful to consult FIGURE 2. The left-hand endpoint $a/b=a_0/b_0$ moves successively to $a_1/b_1, a_2/b_2, \ldots, a_s/b_s$ and then stops. How are these a_k/b_k computed? Just ask, what is the Farey process? Each a_{k+1}/b_{k+1} is simply the mediant of a_k/b_k and c/d, and the algorithm successively computes these mediants until the mediant a_{s+1}/b_{s+1} is to the right of α. We stop the rightward migration of the left endpoints of the Farey pairs with the fraction a_s/b_s, before we cross the point α. This rightward migration is then followed by a similar leftward migration (of the right-hand endpoint) and so back and forth ad infinitum. It is conventional to index these migrations by n, the rightward moving ones being even, the others odd. I have suppressed the variable n to avoid double superscripts, but the reader should understand it is implicit. It is important to note that c/d was the fraction retained from the leftward migration immediately preceding the situation shown in FIGURE 2.

FIGURE 2. A stage in the slow continued fraction algorithm, in which the left endpoint a/b moves rightward s times. The next point a_{s+1}/b_{s+1} is on the wrong side of α. For each k, a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d. The fast algorithm retains the fraction a_s/b_s and (when $s>1$) discards all of the a_k/b_k with $1 \leqslant k < s$.

EXAMPLE. Referring to TABLE 2, let $a/b = 1/3$ and $c/d = 2/5$, so that c/d was the last term in a previous migration. Suppose that α lies between $5/13$ and $7/18$. Then the left endpoint $a/b = 1/3$ moves over two times, to $3/8$ and $5/13$, giving $s = 2$. The next rightward movement would bring the left endpoint to $7/18$, which is on the other side of α.

We have explained what the fast algorithm is. Now we ask: is it a good method? Have our choices of which points to keep, which discard, been good ones? We will define a precise sense in which the approximations which we have kept are better than those we threw away.

DEFINITION 2. Let α be any number with $0<\alpha<1$, and let p/q be a fraction. We define the **ultra-distance** from p/q to α to be $q|(p/q)-\alpha|$, i.e., q times the ordinary distance. We call p/q an **ultra-close approximation** to α if, among all fractions x/y with denominators $y \leqslant q$, p/q has the least ultra-distance to α. (Here we make no distinction between left and right approximations.)

Thus each fraction p/q is handicapped by having its distance to α multiplied by q. It is worth noting that between two fractions a/b and c/d, the ultra-distance is not commutative, because the denominators b and d are probably different. However, for our proof it is much more important (crucial, in fact) to look at the good side. The ultra-distance from p/q to α has nothing to do with the number theoretic structure of α. If α moves closer to p/q, then the ultra-distance from p/q to α decreases.

This may seem very artificial, and it would be nice to have an intrinsic idea of what the ultra-distance means. Luckily there is a very natural one. In what we have done so far it is really the denominators of fractions which are interesting. Suppose we consider together all of the fractions $0/q, 1/q, 2/q, \ldots$ with a fixed denominator q. A little thought shows that any number α can be approximated by one of these fractions to within a distance of $1/2q$. This approximation is guaranteed, so to speak. Now the ultra-distance is just the ratio

$$\frac{|(p/q)-\alpha|}{(1/q)}$$

Thus (ignoring the $1/2$) the ultra-distance tells us how much better the fraction p/q does than what we could automatically expect.

A comparison of Definitions 1 and 2 shows that "ultra-close" implies "best." (Just as with modern advertising, "best" isn't really very good.) Hence, by Theorem 2, the slow continued fraction algorithm furnishes the only possible candidates for "ultra-close" status. We aim to prove that the fast algorithm makes the correct choices from this list.

LEMMA 2. *Let a/b and c/d be a Farey pair with mediant $(a+c)/(b+d)$. Then the ultra-distances from either a/b or c/d to the mediant are the same.*

In terms of the ordinary distance on the number line, the mediant is not equidistant from a/b and c/d, but in terms of the ultra-distance, it is.

Proof. The ultra-distance from a/b to $(a+c)/(b+d)$ is

$$b\cdot\left(\frac{a+c}{b+d}-\frac{a}{b}\right),$$

and a simple calculation (using $bc-ad=1$) reduces this to $1/(b+d)$. By symmetry, the same holds for c/d.

THEOREM 3. *Take any irrational number* α, $0<\alpha<1$. *The fast continued fraction algorithm gives precisely the set of all ultra-close approximations to* α.

Proof. Since "ultra-close" approximations are also "best," it follows from Theorem 2 that we need consider only the terms produced by the slow continued fraction algorithm. Recall that the slow algorithm gave us a sequence of terms $a_1/b_1,\dots,a_s/b_s$ moving inward towards α, and that we chose a_s/b_s and discarded the rest (see FIGURE 2). The question is: did we choose correctly; i.e., is a_s/b_s ultra-close to α? (Here we must not jump to conclusions. It is not a priori clear that the fractions which are closer to α are better, because the ultra-distance also involves the denominators, and the denominators in the sequence $a_1/b_1,\dots,a_s/b_s$ are increasing.)

We should also remember that the slow algorithm involves a *sequence* of leftward and rightward migrations, and FIGURE 2 shows only a single stage in this process. By induction, we can assume that our theorem holds for all of the previous stages. In particular, we can assume that c/d (which was the term retained from the previous migration) is an ultra-close approximation to α. So now the question reduces to: which of the a_k/b_k (if any) are better than c/d?

Take any a_k/b_k. Recall that by definition of the Farey process, a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d. Hence by Lemma 2, a_k/b_k and c/d have the same ultra-distance to a_{k+1}/b_{k+1}. Thus the contest is decided by whether α lies to the left or right of a_{k+1}/b_{k+1}. If left, a_k/b_k wins. If right, c/d wins. But (cf. FIGURE 2), α lies to the left of a_{k+1}/b_{k+1} $(1\leqslant k\leqslant s)$ if and only if $k=s$.

So the proof is, after all, just a matter of knowing your left hand from your right hand.

How to program it

The Farey process we have described is a concrete algorithm for approximation of an irrational number, and is suitable for programming on a computer or programmable calculator. Here, for the sake of brevity, we will give the reasoning as a chain of assertions, which the reader is invited to prove. We will conclude with an actual program. (We are not here concerned with the minutiae of programming languages, and common sense dictates that we continue to follow the notations of this paper.)

We assume that the fractions a/b and c/d in FIGURE 2 have already been found. Of course, we view this as a set of four integers a,b,c,d, not two real numbers. Our objective is to compute the integers a_s and b_s. To achieve this, we will first show how to compute a_k and b_k for any k, and then see how to find the stopping index s.

1. The integers a_k and b_k are given by $a_k=a+kc, b_k=b+kd$. (Hint: a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d.)

2. The function $f(x)=(a+xc)/(b+xd)$ is strictly increasing for $x>0$. If we define the real number γ by the condition $f(\gamma)=\alpha$, then $\gamma=(\alpha b-a)/(c-\alpha d)$. (Note: this depends on the assumption that $a/b<c/d$; if that inequality were reversed, then we should replace the word "increasing" by "decreasing." We mention in passing that $(\alpha b-a)$ is the ultra-distance from a/b to α, and similarly for $(c-\alpha d)$.)

3. The stopping index s is the greatest integer $\leqslant\gamma$. (Hint: use the fact that $f(x)$ is increasing.) This data gives us the basis for a workable program. To make a "do loop" out of it, we compute γ, then s, then a_s and b_s. Then we make the replacements $a=c, b=d, c=a_s, d=b_s$ and start over.

4. There is a way to speed things up. Let γ and s be as above, and let γ' be the next value of γ. Then $\gamma'=1/(\gamma-s)$.

5. The best starting point for this program is $a/b=0/1, c/d=1/0$ which we think of as "infinity." Then after step 2, $\gamma=\alpha$. (By starting with $c/d=1/0$ instead of $1/1$, we remove the restriction that $\alpha<1$.)

Now let's write the program. The variables are $\gamma, s, a, b, c, d, a_s, b_s$, of which only γ is not a positive integer. The number to be approximated is α. The fractions a_s/b_s are the approximations.

PROGRAM.

Start with $a=0, b=1, c=1, d=0, \gamma=\alpha$.

Main Loop:

$s=\text{int}(\gamma)$

$a_s=a+sc \qquad (1)$

$b_s=b+sd \qquad (2)$

Print out a_s, b_s (and perhaps s)

$a=c$

$b=d$

$c=a_s$

$d=b_s$

$\gamma=1/(\gamma-s) \qquad (3)$

Repeat the main loop.

This is an infinite loop, of course; you can just stop the program by hand or else fix it. The variable s occurs in the standard presentations of continued fractions, where it is usually written a_n. If you take the number $e\cong 2.718\ldots$ you will find that the s values follow an interesting pattern; you will also find that this pattern breaks down eventually (why?). The s values for the square roots of various integers are quite interesting. Try also the golden mean $(1+\sqrt{5})/2$; here the b_s values form a well-known sequence, and the s values are also interesting. Nothing is known about higher level algebraic irrationals. A famous unsolved problem is to prove that the sequence of s values is unbounded for any algebraic number of degree greater than two.

If we apply our program to the number π, we obtain the following data:

a_s	b_s	s
3	1	3
22	7	7
333	106	15
355	113	1
103,993	33,102	292
⋮	⋮	⋮

There is little point in carrying the calculations much further, since the approximation $a_s/b_s=103{,}993/33{,}102$ is already good to nine decimal places, and machine round-off error would soon render the results meaningless. (For example, the same program on my hand calculator produced a 293 in place of the 292 at the bottom of the s column.) Looking at some of the other fractions

a_s/b_s, we find the familiar approximations 22/7 and 355/113 mentioned earlier. Finally we observe that the numbers in the s column are just those which appeared in the "python" at the beginning of the article. This is no accident, as we now show.

Relation to the traditional format

Following our notations, the continued fraction expansion of an irrational number $\alpha > 0$ can be expressed in closed form by the equation:

$$\alpha = s_0 + \cfrac{1}{s_1 + \cfrac{1}{s_2 + \cfrac{\ddots}{\ddots + \cfrac{1}{s_{n-1} + \cfrac{1}{\gamma_n}}}}}. \tag{4}$$

Here the s_i are integers, and γ_n is an irrational number > 1 chosen so that equality holds. Then s_n is the greatest integer $\leqslant \gamma_n$. *If in* (4) γ_n *is replaced by* s_n, *the fraction chain becomes a rational number* p_n/q_n. *These* p_n/q_n *are the terms in the (fast) continued fraction algorithm for* α; *they are left approximations* a/b *if* n *is even, and right approximations* c/d *otherwise*. The variable n, which we suppressed in our discussion of the fast continued fraction algorithm, indexes the rightward and leftward migrations in our geometric presentation.

To prove these statements, we will use the computer algorithm as our pivot. We have already seen that the geometric theory leads to this algorithm; now we deduce the same algorithm from the fraction chain (4), thus proving the equivalence of the two theories.

First, it is easy to prove by induction that, given α, a unique fraction chain (4) exists for every n. For, assuming this holds for n, we immediately deduce the recursion

$$\gamma_{n+1} = 1/(\gamma_n - s_n).$$

We notice that this matches formula (3) in the computer algorithm. Now our objective is to recover formulas (1) and (2). This requires another induction.

We think of the fraction chain (4) as a function of γ_n, with the s_i held fixed and γ_n varying. In order to prove the assertion in italics, we will show that the function (4) has the form:

$$\frac{p_{n-2} + \gamma_n p_{n-1}}{q_{n-2} + \gamma_n q_{n-1}}. \tag{5}$$

(To match the notation of the computer algorithm we would write $a/b = p_{n-2}/q_{n-2}$ and $c/d = p_{n-1}/q_{n-1}$.) Now to make the induction from n to $n+1$, we recall that p_n/q_n is the result of replacing γ_n by s_n in (4), and that $\gamma_{n+1} = 1/(\gamma_n - s_n)$. Thus:

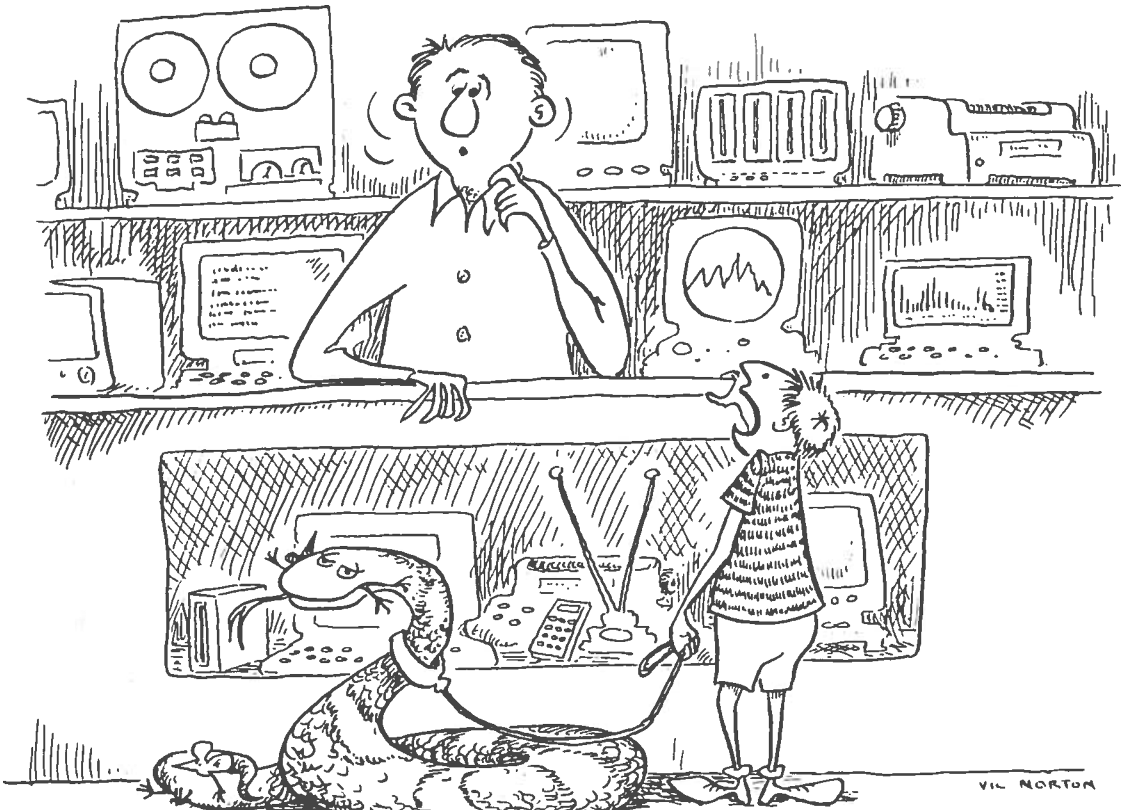
$$\frac{p_{n-1} + \gamma_{n+1} p_n}{q_{n-1} + \gamma_{n+1} q_n} = \frac{p_{n-1} + \gamma_{n+1}(p_{n-2} + s_n p_{n-1})}{q_{n-1} + \gamma_{n+1}(q_{n-2} + s_n q_{n-1})},$$

and using $1/\gamma_{n+1} = \gamma_n - s_n$ yields

$$\frac{(\gamma_n - s_n) p_{n-1} + (p_{n-2} + s_n p_{n-1})}{(\gamma_n - s_n) q_{n-1} + (q_{n-2} + s_n q_{n-1})} = \frac{p_{n-2} + \gamma_n p_{n-1}}{q_{n-2} + \gamma_n q_{n-1}}.$$

This completes the induction and establishes formulas (1) and (2).

Thus our previous theory, in its geometric setting, is equivalent to that determined by the endless fraction.



I WANT TO TRADE MY PYTHON FOR AN ELECTRONIC **FAREY** PING-PONG GAME.

Suggestions for further reading

An interesting geometrical approach to continued fractions is given in the book by Harold Stark [**4**]. Roughly speaking, Stark's approach is two dimensional (based on the slopes of lines), whereas our approach is one dimensional. For the classical "infinite fraction chain" viewpoint, see almost any book on elementary number theory. My favorite is Hardy and Wright [**2**]. An elementary introduction to fractions (Egyptian fractions, Farey fractions, continued fractions, decimal fractions) can be found in [**1**].

References

[1] A. Beck, M. Bleicher, D. Crowe, Excursions into Mathematics, Worth, NY, 1969, pp. 414–475.
[2] G. H. Hardy and E. M. Wright, The Theory of Numbers, 4th ed., Oxford Univ. Press, 1960.
[3] M. Shrader-Frechette, Modified Farey sequences and continued fractions, this MAGAZINE, 54 (1981) 60–63.
[4] H. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.