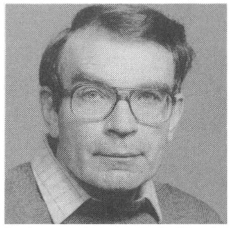
Geometry of Continued Fractions

M. C. IRWIN



MICHAEL CHARLES IRWIN wrote his Ph.D. thesis (Cambridge University, U.K., 1962) on "Embeddings of Polyhedral Manifolds," and his early papers were concerned with piecewise linear topology, but from about 1970 onwards his research was mainly on dynamical systems. His most recent publications included work on invariant paths for Anosov diffeomorphisms. He was the author of the book *Smooth Dynamical Systems* published by Academic Press in 1980. He taught at the University of Liverpool, U.K., from 1961 until his death on March 11, 1988.

The representation of real numbers by continued fractions dates back to Bombelli in the 16th century. In the 17th century Huygens used them in constructing a model of the solar system; he had to approximate the ratios of periods of planets by ratios of numbers of teeth on corresponding gear wheels, keeping the latter within reasonable bounds. The nice thing about the continued fraction process is that, being completely intrinsic, it brings out very strongly the personality of each individual number α. The digits in the decimal expansion of α are much less revealing since they relate to the arbitrary choice of 10 as basis. On the other hand, the great defect of continued fractions is that it is virtually impossible to use them for even the simplest algebraic computation involving two or more numbers.

There are several books devoted entirely to the subject of continued fractions (e.g., **[1]**, **[2]**, **[3]**, **[5]**), and many books on number theory give an elementary introduction to the subject. The proofs are not difficult, but they are usually algebraic, and I find that when I read them I have a tendency to lose sight of where they are leading. The following geometrical treatment of the easiest results may be of some help to readers who, like myself, need a picture of what is going on. Harold M. Stark has already given such a treatment in his book **[4]**, but I think that the version below produces the main results with greater economy.

The idea of continued fractions comes from the observation that (i) any number between $1/3$ and $1/2$ (say) can be written as $1/(2$ plus a remainder), and (ii) the remainder, being between $1/(n+1)$ and $1/n$ for some n, is susceptible to the same treatment as the original number. More formally, we can express any real number α as the sum of an integer $[\alpha]$, the *integral part* of α, and a number $\{\alpha\}$ with $0 \leq \{\alpha\} < 1$, the *fractional part* of α. If we define inductively $\alpha_0 = \alpha$ and $\alpha_n = 1/\{\alpha_{n-1}\}$ for $n \geq 1$, we obtain a sequence of integers $a_n = [\alpha_n]$, with $a_n \geq 1$ for $n \geq 1$. Of course the process terminates if $\{\alpha_n\} = 0$, and, in this case, α is rational and has the value

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_n}}}} = \frac{p_n}{q_n}, \tag{1}$$

say, where p_n/q_n is in its lowest terms. If $\{\alpha_n\} \neq 0$, then the rational number p_n/q_n of (1) is intuitively some sort of approximation to α, since we get it by ignoring the remainder in $1/(a_n$ plus a remainder). We write

$$p_n/q_n = [a_0; a_1, a_2, \ldots, a_n].$$

It is called the *nth-order convergent* or *approximant* of α. If α is irrational, then we find that $p_n/q_n \to \alpha$ as $n \to \infty$, and we write

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} = [a_0; a_1, a_2, a_3, \ldots, a_n, \ldots].$$

Example 1.

$$\frac{2}{3} = [0;1,2] = \cfrac{1}{1 + \cfrac{1}{2}},$$

and

$$\frac{2}{3} = [0;1,1,1] = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1}}},$$

the continued fraction process for 2/3 giving the former expression. In general, we have the trivial equality

$$[a_0; a_1, \ldots, a_{n-1}, a_n] = [a_0; a_1, \ldots, a_{n-1}, a_n - 1, 1].$$

Example 2. The golden ratio $(1 + \sqrt{5})/2$ has continued fraction expansion $[1; 1, 1, \ldots]$, since both the ratio and the continued fraction are positive numbers satisfying the quadratic equation derived from $x = 1 + 1/x$.

We begin by examining in what sense the convergents of α approximate α. We say that a fraction p/q $(q > 0)$ is a *best approximation* to α if, for all fractions p'/q' with $0 < q' \leq q$, $|q\alpha - p| < |q'\alpha - p'|$ unless $q = q'$, $p = p'$. At first sight, it would be more natural to use the inequality $|\alpha - p/q| < |\alpha - p'/q'|$, which, written in the form

$$\frac{1}{q}|q\alpha - p| < \frac{1}{q'}|q'\alpha - p'|,$$

is clearly weaker. This would allow more fractions to be called best approximations, but the relation with continued fractions is more complicated (see [2]). Convergents of a number α are characterized as best approximations to α in the following way:

THEOREM 1. (i) *If* $\{\alpha\} < 1/2$, *then the best approximations to* α *are precisely the convergents* p_n/q_n *for* $n \geq 0$.

(ii) *If* $\{\alpha\} > 1/2$, *then the best approximations to* α *are precisely the convergents* p_n/q_n *for* $n \geq 1$.

Note (i). If $\{\alpha\} > 1/2$ then trivially the integral best approximation to α is $[\alpha] + 1$. Since $1 < 1/\{\alpha\} = \alpha_1 < 2$, $a_1 = 1$, and we have two integral convergents of α, $p_0/q_0 = a_0 = [\alpha]$ and $p_1/q_1 = a_0 + 1/1 = [\alpha] + 1$, of which only the latter is a best approximation.

(ii) If $\{\alpha\} = 1/2$, then α has no integral best approximation, since $[\alpha]$ and $[\alpha] + 1$ are equally good. Thus the first convergent $p_0/q_0 = [\alpha]$ is not strictly speaking a best approximation. Of course $p_1/q_1 = \alpha$ is a best approximation.

Proof of Theorem 1. The outline of the proof is as follows. We associate geometrically best approximations to α with points in the plane, which we call *best approximation points*. We describe a geometrical construction which, when applied repeatedly, picks out in succession all best approximation points. Finally we compute the coordinates of the best approximation points to show that the best approximations to α are the given convergents. We shall also find that several further properties of convergents of α emerge immediately from this proof (see Corollaries 3 to 5 below).

Notations and definitions. Following Stark, we use the same notation $P = (b, a)$ for points in the plane and for vectors. The *lattice* generated by two vectors P and Q is the set of all points $mP + nQ$ where m and n are integers. The (closed) *positive* (P, Q)*-cone with vertex* U is the set of points $U + aP + bQ$ for all numbers $a \geq 0$ and $b \geq 0$. The *distance from a point* A *to a line* l *in the direction* P (not parallel to l) is the length of the line segment joining A to the unique point $A + aP$ on l. The ratio

$$(\text{distance from } A \text{ to } l)/(\text{distance from } A' \text{ to } l)$$

is independent of the direction P in which it is measured, so we are free to order distances from points to l using any direction that suits us. See FIGURE 1 for an illustration of the above ideas.

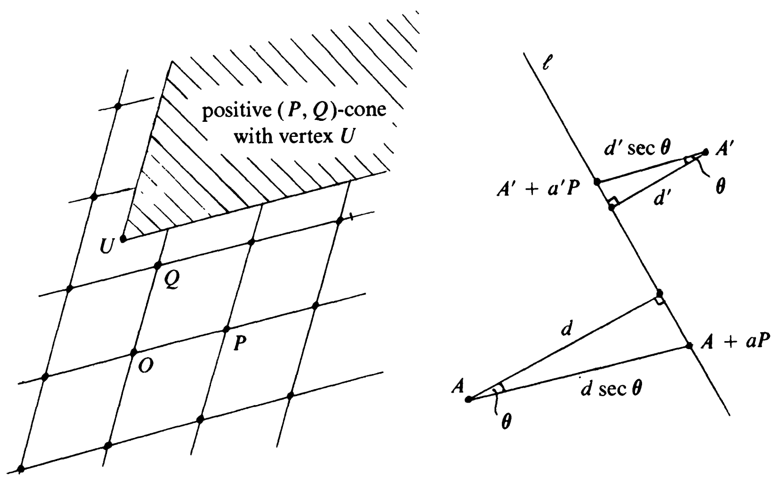


FIG. 1.

Geometrical interpretation of best approximation. Let l be the line $y = \alpha x$. The geometrical interpretation of p/q being a best approximation to α is that, amongst all points (q', p') of the integer lattice Z^2 (the lattice generated by $(1,0)$ and $(0,1)$) with x-coordinate satisfying $0 < q' \leq q$, (q, p) is uniquely the nearest to l. Here we use the fact that $|q'\alpha - p'|$ is the vertical distance from (q', p') to l (i.e., in the direction $(0,1)$). We call (q, p) a *best approximation point* for l.

The outpoint construction. Consider the construction illustrated in FIG. 2: We are given a real number α, and two integer lattice points A, B. We wish to construct the point C in FIG. 2, which we will call the *outpoint* of A, B and will denote by $C = C(A, B; \alpha)$.

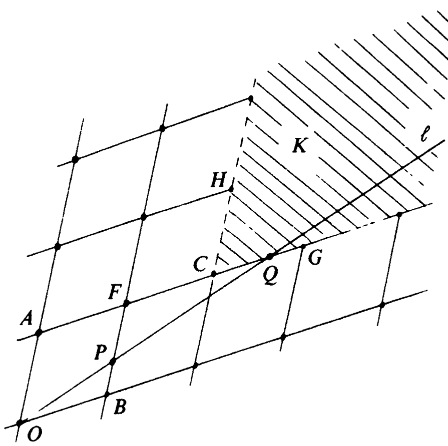


FIG. 2.

The outpoint construction is the following. Let l be the line $y = \alpha\chi$. Suppose l meets parallelogram $OBFA$ in the origin O, and in a point P that is interior to the side BF. Let l meet the extension of AF in the point Q.

Then, $P = \theta A + B$ for some $\theta, 0 < \theta < 1$, and clearly $Q = A + (1/\theta)B$. Hence Q lies on the lattice edge between $C = A + [1/\theta]B$ and $G = B + C$. This completes the construction of the outpoint $C = (A, B; \alpha)$. □

To what extent is C a good lattice approximation to a point of l? Apart from 0 and $-C$, the reflection of C in 0, C is nearer to l than is any lattice point outside (i) the positive (A, B)-cone K with vertex at C and (ii) the reflection of K in 0. This is because the distance from any such lattice point to l in the B-direction is at least a positive integral multiple of $|B|$, whereas the distance from C to l in this direction is $\varphi|B|$, and $\varphi < 1$. Moreover we can get away with a half-open cone for K, since any point on the A-edge of K other than C has distance greater than $|A|$ from l in the A-direction. We also comment that, outside the two cones, the next nearest lattice points to l after $0, C$ and $-C$ are B and $-B$, since these have distance $\theta|A|$ and $|B|$ in the A- and B-direction respectively.

Repeating the outpoint construction. We are going to iterate the outpoint construction, and in doing so, it is important to observe two facts. First, the lattice generated by A and B is precisely the same as the lattice generated by B and C (because $C = A + [1/\theta]B$, $A = C - [1/\theta]B$ and $[1/\theta]$ is an integer). Second, at any vertex, the positive (A, B)-cone, open along the A-edge, contains the positive (B, C)-cone, open along the B-edge (because $C = A + [1/\theta]B$ and $[1/\theta] \geq 1$). Begin with a real number α and two integer lattice points $V_{-1} = (0, 1)$ and $V_0 = (1, a_0)$, where $a_0 = [\alpha]$. Then for each $n = 1, 2, 3, \ldots$ let $V_n = C(V_{n-2}, V_{n-1}; \alpha)$. □

FIG. 3 illustrates the construction of V_1, V_2, V_3 when $\alpha = \sqrt{5} - 1 = [0; 1, 1, 1, \ldots]$.

Trivially, if $\{\alpha\} < 1/2$, $p_0/q_0 = [\alpha]$ is a best approximation for α. For $n \geq 1$, using the observations about lattices and positive cones, we can say that if (q', p') is a point of Z^2 with $q' > 0$ and if l is no further from (q', p') than from V_n, then (q', p') is in the positive (V_{-1}, V_0)-cone with vertex V_n, open along the V_{-1}-edge. Thus, by the geometrical interpretation, p_n/q_n is a best approximation to α. Also, by the next-nearest property of B and $-B$, l is closer to V_{n-1} than to any point of Z^2 with x-coordinate between q_{n-1} and q_n. Thus there is no best approximation p/q with $q_{n-1} < q < q_n$. We deduce that there are no best approximations other than p_n/q_n, $n \geq 1$, and p_0/q_0 for $\{\alpha\} < 1/2$.

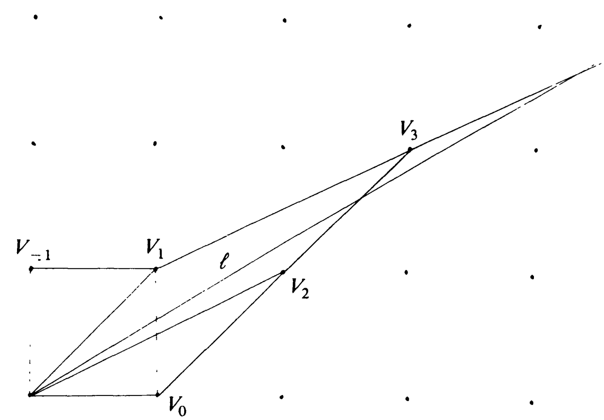


FIG. 3.

Best approximations as convergents. It remains to show that our p_n/q_n satisfy (1). Let l' be the line $y = (p_n/q_n)x$. First note the important fact that, for $0 \leq r \leq n$, the inductive sequence of constructions yields precisely the same points V_r and integers a_r for the line l' as for the line l (but, of course, unless $l = l'$ we get a new sequence of numbers θ_r' replacing θ_r). This is because, for $0 \leq r < n$, l' intersects the interior of the line segment joining V_r to $V_r + V_{r-1}$. For certainly l does so. Thus if l' does not, then one of the lattice points V_r or $V_r + V_{r-1}$ separates l and l' in the half-plane $x > 0$. Since either point has x-coordinate $\leq q_n$, it is nearer to l than V_n is, and this contradicts the fact that p_n/q_n is a best approximation to α.

Let d_r' denote the vertical distance from V_r to l'. Then $d_r'/d_{r-1}' = \theta_r'$ (This is because the distance ratio CQ/OB equals φ in the basic geometrical construction).

Hence,

$$d'_{r-2}/d'_{r-1} = 1/\theta'_{r-1} = a_r + \theta'_r = a_r + d'_r/d'_{r-1}$$

for $1 \le r \le n$, and so

$$d'_{r-1}/d'_{r-2} = 1/(a_r + d'_r/d'_{r-1}).$$

Now $d'_0/d'_{-1} = \{p_n/q_n\}$, and $d'_n = 0$. By substituting successively for the ratios $d'_1/d'_0, \ldots, d'_{n-1}/d'_{n-2}$, we obtain $\{p_n/q_n\} = [0; a_1, \ldots, a_n]$, and hence $p_n/q_n = [a_0; a_1, \ldots, a_n]$. This completes the proof of Theorem 1.

We can easily deduce several basic properties of continued fractions from the above proof.

COROLLARY 2.

$$q_n = a_n q_{n-1} + q_{n-2} \quad \text{for } n \ge 1.$$
$$p_n = a_n p_{n-1} + p_{n-2}$$

Proof. Immediate from (2). In practice, one uses these relations to compute q_n and p_n inductively. The pair q_n, p_n so obtained has H.C.F. 1, by the definition of best approximation. In the case of irrational α, we deduce that the sequence q_n, $n \ge 0$, tends to ∞ at least as fast as the Fibonacci sequence $1, 1, 2, 3, 5, 8, \ldots$ (since $q_n \ge q_{n-1} + q_{n-2}$). A cruder estimate is $q_n \ge 2^{n/2}$ for $n \ge 2$ (since $q_n \ge 0$, $q_n \ge q_{n-1}$ and thus $q_n \ge 2q_{n-2}$).

COROLLARY 3. *If $\alpha \neq p_n/q_n$, then α is between p_{n-1}/q_{n-1} and p_n/q_n. The sequence p_n/q_n increases for even n and decreases for odd n.*

Proof. Immediate from the geometrical construction.

COROLLARY 4. $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$ *for* $n \ge 0$

Proof. The left-hand side is the oriented area of the parallelogram with vertices $0, V_n, V_n + V_{n-1}, V_{n-1}$. The result follows from the value $+1$ when $n = 0$, and the observation that, in the basic geometrical construction, the parallelograms $OBFA$ and $OCGB$ have the same base and height but opposite orientations.

Corollaries 2 and 3 give that, for α irrational,

$$|p_n/q_n - \alpha| < |p_n/q_n - p_{n-1}/q_{n-1}| = 1/q_n q_{n-1},$$

which shows that $p_n/q_n \to \alpha$ as $n \to \infty$, as we asserted in the introduction.

COROLLARY 5. $p_{n-2}q_n - p_n q_{n-2} = (-1)^{n-1} a_n$.

Proof. This follows from Corollary 4 with $n - 1$ replacing n, and the observation that, in the basic geometrical construction, the parallelograms $OBFA$ and $OCHA$ (where $H = A + C$) have the same base OA and their heights are in the ration OB/AC, which is $1/[1/\theta]$.

Two numbers α and β are said to be (*rationally*) *equivalent* if, for some integers a, b, c, d with $ad - bc = \pm 1$,

$$\frac{a\alpha + b}{c\alpha + d} = \beta.$$

Geometrically, the second condition can be interpreted as a linear map f with matrix

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

taking $(1, \alpha)$ to a multiple of $(1, \beta)$, or, equivalently, taking the line $y = \alpha x$ onto the line $y = \beta x$. The determinant condition $ad - bc = \pm 1$ is precisely the condition for f to take the integer lattice Z^2 onto itself. It is also the condition for f to induce a homeomorphism of the two-dimensional torus (which is obtained from the plane by identifying all points that differ by vectors in Z^2), so the idea is important, for example, when studying dynamical systems on the torus. Any rational number $\alpha = p/q$ (in its lowest terms) is equivalent to 0, since $mp + nq = 1$ for some integers m and n, and thus

$$\begin{pmatrix} p & -q \\ n & m \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence any two rational numbers are equivalent. We finish this article with a well-known characterization of equivalence for irrational numbers. We shall say that the two continued fraction expansions $[a_0, a_1, a_2, \dots]$ and $[b_0; b_1, b_2, \dots]$ are *eventually identical* if for some $M \geq 0$ and $N \geq 0$, $a_{M+i} = b_{N+i}$ for all $i \geq 1$.

THEOREM 6. *Two irrational numbers are equivalent if and only if their continued fraction expansions are eventually identical.*

Proof. Let α and β have continued fraction expansions as in the above definition of eventual identity. We prove that α is equivalent to γ, where $\gamma = [0; a_{M+1}, a_{M+2}, \dots]$. The similar result for β then gives α equivalent to β. Let l be the line $y = \alpha x$ and let V_n be the associated best approximation points, as constructed in Theorem 1. Let f be the linear map that takes V_{M-1} and V_M to $(0,1)$ and $(1,0)$ respectively. By Corollary 4, f has determinant ± 1. Let $l' = f(l)$ be the line $y = \gamma x$ and let $W_i = f(V_{M+i})$ for $i \geq -1$. Since f is linear, it takes parallelograms to parallelograms, interior intersections to interior intersections, and linear sums $A + \theta B$ to $f(A) + \theta f(B)$. Thus it preserves the basic geometrical construction of Theorem 1. Hence it is clear that W_i, $i = 1, 2, \dots$ are the best approximation points for l', and hence that γ has continued fraction expansion $[0; a_{M+1}, a_{M+2}, \dots]$ as required.

Conversely, suppose that $\alpha = [a_0; a_1, a_2, \dots]$ is equivalent to $\beta = [b_0; b_1, b_2, \dots]$. By the preceding paragraph, we can assume that $a_0 = 0$. Let f be a linear map taking Z^2 onto itself and the line l, $y = \alpha x$, onto the line m, $y = \beta x$. Again let V_n be the best approximation points for l, and, this time, let $W_n = f(V_n)$. We can assume that the line $W_{-1}W_0$ intersects m in some point R lying in the half-plane $x > 0$ (otherwise, consider $-f$). By our comment of the preceding paragraph, W_n is the sequence of points obtained by applying the inductive geometrical construction to the line m, beginning with the pair W_{-1} and W_0, and a_n is the corresponding sequence of integers obtained. Now let d_n denote the distance from W_n to m in the direction of the line $W_{-1}W_0$. As in Theorem 1,

$$d_{n-2} = d_n + a_n d_{n-1},$$

whence $d_{n-1} \geq d_n$ and so $d_{n-2} \geq 2d_n$. Thus $d_n \to 0$ as $n \to \infty$. Now the line through W_n parallel to $W_{-1}W_0$ intersects m on the opposite side of R to 0, so such intersections have a positive minimum distance from the half-plane $x \leq 0$. Thus, for large enough n, W_n has a positive x-coordinate. This implies that, for large enough n, the positive (W_{n-2}, W_{n-1})-cone at W_n lies to the right of W_n. Hence, by the lattice approximation property of the basic geometrical construction, we recognise such W_n as successive best approximation points of m. We deduce that the corresponding a_n are successive terms in the continued fraction expansion of β, and this completes the proof of the theorem.

REFERENCES

1. W. B. Jones and W. J. Thron, Continued Fractions, Encyclopaedia of Mathematics and its Applications 11, Addison-Wesley, 1980.
2. A. Y. Khinchin, Continued Fractions, University of Chicago Press, 1964.
3. O. Perron, Die Lehre von den Kettenbrücke, Band I, Teubner, Stuttgart, 1954.
4. H. M. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.
5. H. S. Wall, Analytic Theory of Continued Fractions, Van Nostrand, New York, 1948.