Ford circles, continued fractions, and rational approximation

Ian Short

Abstract

We give an elementary geometric proof using Ford circles of a well-known theorem of Lagrange, which says that the convergents of the continued fraction expansion of a real number α coincide with the rationals that are best approximations of α.

1 Introduction

This paper is about a geometric view of the relationship between continued fractions and approximation of real numbers by rationals. Whenever we speak of a rational u/v we mean that u and v are coprime integers and v is positive. A rational a/b is a *best approximation* of a real α provided that, for each rational c/d such that $d \leqslant b$, we have

$$|b\alpha - a| \leqslant |d\alpha - c|\,,$$

with equality if and only if $c/d = a/b$. This definition can be found in [3, Section 6], where the phrase *best approximation of the second kind* is used; Khinchin also defines *best approximation of the first kind*, but that concept does not concern us here.

A *continued fraction* is an expression of the form

$$b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \cfrac{1}{b_3 + \cdots}}}\,,$$

where b_0 is an integer and the other coefficients b_i are positive integers. Either the sequence $b_0, b_1, b_2, \dots$ is infinite, in which case the continued fraction is said to be *infinite*, or there is a final member b_N of this sequence, in which case the continued fraction is said to be *finite*.

We define integers $A_0, A_1, A_2, \dots$ and positive integers $B_0, B_1, B_2, \dots$ by the matrix recurrence relations

$$\begin{pmatrix} A_n & A_{n-1} \\ B_n & B_{n-1} \end{pmatrix} = \begin{pmatrix} A_{n-1} & A_{n-2} \\ B_{n-1} & B_{n-2} \end{pmatrix} \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} A_1 & A_0 \\ B_1 & B_0 \end{pmatrix} = \begin{pmatrix} b_0 b_1 + 1 & b_0 \\ b_1 & 1 \end{pmatrix}, \tag{1.1}$$

for $n \geqslant 2$. Taking determinants in these equations we see that $|A_n B_{n-1} - A_{n-1}B_n| = 1$, which means that A_n and B_n are coprime. Using the well-known homomorphism from the group of two-by-two invertible complex matrices to the group of Möbius transformations, we can reinterpret (1.1) in terms of Möbius transformations: we find that if, for $n = 0, 1, 2, \ldots$, $t_n(z) = b_n + 1/z$ and $T_n = t_0 \circ t_1 \circ \cdots \circ t_n$, then $T_n(z) = (A_n z + A_{n-1})/(B_n z + B_{n-1})$ for $n \geqslant 1$. Set $z = \infty$ to see that

$$\frac{A_0}{B_0} = b_0\,, \qquad \frac{A_1}{B_1} = b_0 + \frac{1}{b_1}, \qquad \frac{A_2}{B_2} = b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2}}, \ldots .$$

The *value* of a finite continued fraction is the final term A_N/B_N (which is a rational number) and the *value* of an infinite continued fraction is the limit of the sequence A_n/B_n (which is an irrational number). Since the finite continued fraction with coefficients $b_0, b_1, \ldots, b_{N-1}, 1$ has the same value as the finite continued fraction with coefficients $b_0, b_1, \ldots, b_{N-2}, b_{N-1}+1$, we can assume that if $N \geqslant 1$ then $b_N \geqslant 2$. This ensures that to each real number α there corresponds a unique continued fraction with value α. The rationals A_n/B_n are known as the *convergents* of α, and they alternate from one side of α to the other, in the sense that

$$\frac{A_0}{B_0} < \frac{A_2}{B_2} < \frac{A_4}{B_4} < \cdots < \alpha < \cdots < \frac{A_5}{B_5} < \frac{A_3}{B_3} < \frac{A_1}{B_1}. \tag{1.2}$$

All these facts about continued fractions can be found in [3], as can the next theorem of Lagrange [3, Theorems 16 and 17].

Theorem 1.1. *A rational x that is not an integer is a convergent of a real number α if and only if it is a best approximation of α.*

One can easily check that Theorem 1.1 also holds when x is an integer, unless $x + \frac{1}{2} \leqslant \alpha < x+1$, in which case x is a convergent, but not a best approximation, of α.

Classical proofs of Theorem 1.1, such as that given in [3], are algebraic. Irwin proves Theorem 1.1 using plane lattices in [2]. Our aim is to give an illuminating geometric proof based on the theory of Ford circles. Ford circles, developed by Ford in [1], are objects most naturally associated with hyperbolic geometry, and our proof has undertones of hyperbolic geometry. We now give a brief description of Ford circles and their relationship to continued fractions (full details can be found in [1]).

A circle C in the complex plane that is tangent to the real axis at a point x, and that otherwise lies in the upper half-plane, is called a *horocircle*. We denote the radius of C by $\mathrm{rad}[C]$, and describe the point x as the *base point* of C. Two horocircles with radii r and s and distinct base points x and y intersect if and only if

$$|x-y|^2 \leqslant 4rs, \tag{1.3}$$

with equality if and only if the horocircles are tangent. This can be proven by applying Pythagoras's theorem to the triangle with vertices $x + is$, $x + ir$, and $y + is$ (the latter two vertices are the centers of the horocircles).

The *Ford circle* C_x of a rational number $x = a/b$ is the horocircle with radius $1/(2b^2)$ and base point x. It follows from (1.3) that two Ford circles C_x and C_y, where $x = a/b$ and $y = c/d$, are tangent if and only if $|ad - bc| = 1$, and if they are not tangent then they are wholly external to one another (Ford circles do not overlap). Some Ford circles are shown in Figure 1.1.

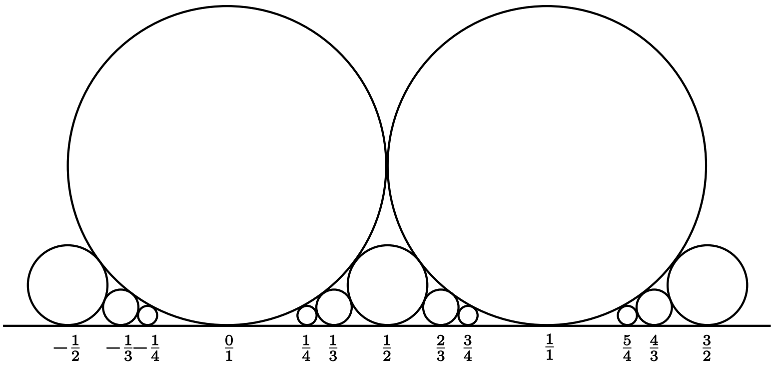

Figure 1.1: Ford circles.

We define the *continued fraction chain* of a real number α to be the sequence of Ford circles $C_{A_0/B_0}, C_{A_1/B_1}, C_{A_2/B_2}, \ldots$, where A_n/B_n are the convergents of α. Since $|A_n B_{n-1} - A_{n-1} B_n| = 1$ we see that any two consecutive circles in the continued fraction chain of α are tangent. Also, by (1.1) we have $B_0 = 1$, $B_1 = b_1$, and $B_n = b_n B_{n-1} + B_{n-2}$ for $n \geqslant 2$, which means that the sequence $B_1, B_2, \ldots$ of positive integers is increasing, and hence the sequence of radii $1/(2B_1^2), 1/(2B_2^2), \ldots$ is decreasing. The first few Ford circles from a continued fraction chain are shown in Figure 1.2 (in black). Observe that, because of (1.2), the members of the continued fraction chain alternate from the left to the right side of α.

Given a rational $x = a/b$ and a real α, we define

$$R_x(\alpha) = \tfrac{1}{2}|b\alpha - a|^2 = \tfrac{b^2}{2}\left|\alpha - x\right|^2 . \tag{1.4}$$

Note that $R_x(x) = 0$. When $\alpha \neq x$, we see from (1.3) (with equality) that $R_x(\alpha)$ is the radius of the unique horocircle with base point α that is tangent to C_x. Our main theorem is the following.

Theorem 1.2. *Let α be a real number. Given a rational x that is not an integer, the following are equivalent:*

(i) *x is a convergent of α;*

(ii) *C_x is a member of the continued fraction chain of α;*

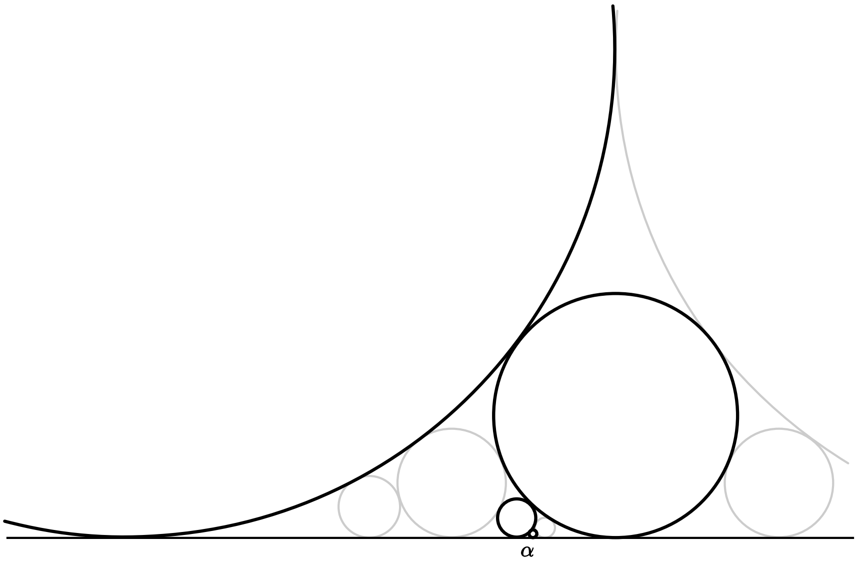

Figure 1.2: A continued fraction chain.

(iii) *x is a best approximation of α;*

(iv) *if z is a rational such that $\mathrm{rad}[C_z] \geqslant \mathrm{rad}[C_x]$ then $R_x(\alpha) \leqslant R_z(\alpha)$, with equality if and only if $z = x$.*

Statement (ii) of Theorem 1.2 is merely a geometric reformulation of statement (i), and statement (iv) is merely a geometric reformulation of statement (iii). The equivalence of statements (i) and (iii) yields Theorem 1.1.

2 Proof of Theorem 1.2

We begin with two basic properties of the function R_x.

Lemma 2.1. *Given a rational $x = a/b$,*

(i) *if $|\alpha - x| < |\beta - x|$ then $R_x(\alpha) < R_x(\beta)$;*

(ii) *if z is a rational distinct from x then $\mathrm{rad}[C_z] \leqslant R_x(z)$, with equality if and only if C_z and C_x are tangent.*

Proof. Part (i) follows immediately from (1.4). For (ii), let $z = c/d$; then

$$\mathrm{rad}[C_z] = \tfrac{1}{2d^2} \leqslant \tfrac{1}{2d^2}|ad - bc|^2 = R_x(z).$$

Equality holds if and only if $|ad - bc| = 1$; that is, if and only if C_z and C_x are tangent. □

Next we need two elementary lemmas about basic properties of Ford circles.

Lemma 2.2. *Let C_x and C_y be tangential Ford circles. If a rational z lies strictly between x and y then C_z has smaller radius than both C_x and C_y.*

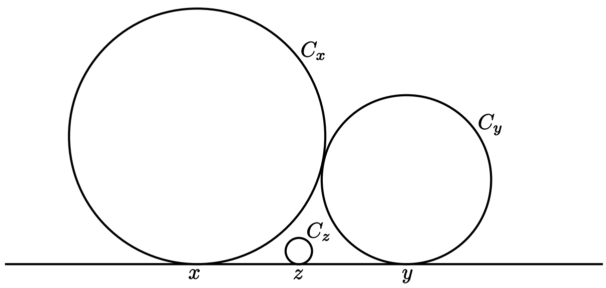

Figure 2.1: The geometry of Lemma 2.2.

Proof. Since $|z-x| < |y-x|$ we see from Lemma 2.1 (parts (i) and (ii)) that

$$\mathrm{rad}[C_z] \leqslant R_x(z) < R_x(y) = \mathrm{rad}[C_y],$$

and similarly $\mathrm{rad}[C_z] < \mathrm{rad}[C_x]$. □

Lemma 2.3. *Let C_x and C_y be tangential Ford circles such that* $\mathrm{rad}[C_x] > \mathrm{rad}[C_y]$, *and suppose that a real number α lies strictly between x and y, and a rational z lies strictly outside the interval bounded by x and y. Then $R_x(\alpha) < R_z(\alpha)$.*

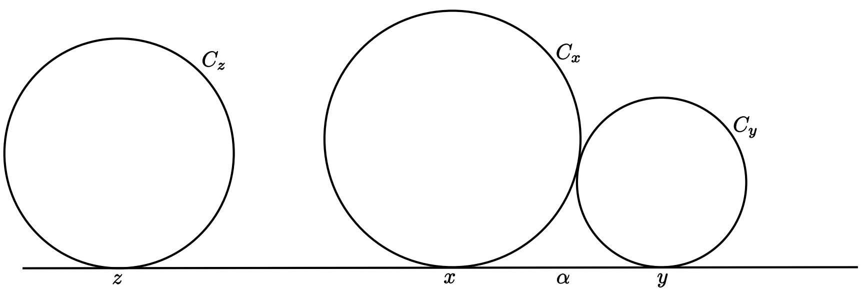

Figure 2.2: The geometry of Lemma 2.3.

Proof. From Lemma 2.1 we have

$$R_x(\alpha) < R_x(y) = \mathrm{rad}[C_y].$$

If y lies between z and α then $|y-z| < |\alpha - z|$, and by Lemma 2.1 we have

$$\mathrm{rad}[C_y] \leqslant R_z(y) < R_z(\alpha).$$

Otherwise, x must lie between z and α, so that $|x - z| < |\alpha - z|$, and again using Lemma 2.1 we have

$$\mathrm{rad}[C_y] < \mathrm{rad}[C_x] \leqslant R_z(x) < R_z(\alpha).$$

In summary, $R_x(\alpha) < \mathrm{rad}[C_y] < R_z(\alpha)$, as required. □

Next we prove that the Ford circle C_x and the rational α described in Lemma 2.3 satisfy statement (iv) of Theorem 1.2.

Corollary 2.4. *Let C_x and C_y be tangential Ford circles such that $\mathrm{rad}[C_x] > \mathrm{rad}[C_y]$, and suppose that a real number α lies strictly between x and y. If z is a rational such that $\mathrm{rad}[C_z] \geqslant \mathrm{rad}[C_x]$ then $R_x(\alpha) \leqslant R_z(\alpha)$, with equality if and only if $z = x$.*

Proof. By Lemma 2.2, if $z \neq x$ then z lies outside the closed interval bounded by x and y, and then by Lemma 2.3, $R_x(\alpha) < R_z(\alpha)$, as required. □

We can now prove Theorem 1.2. Statements (i) and (ii) of Theorem 1.2 are equivalent by the definition of a continued fraction chain. Statements (iii) and (iv) can be seen to be equivalent using (1.4), and the fact that the radius of $C_{a/b}$ is $1/(2b^2)$. For the remainder of the proof we denote the convergents of α by $A_0/B_0, A_1/B_1, A_2/B_2, \dots$, and when α is rational the final convergent (which equals α) is A_N/B_N. We prove that (i) and (iv) are equivalent.

First we prove that (i) implies (iv). Suppose that $x = A_n/B_n$, and for the moment suppose also that when α is rational, $n < N - 1$. Note that since x is not an integer, $n \geqslant 1$. Define $y = A_{n+1}/B_{n+1}$. Then $\mathrm{rad}[C_x] > \mathrm{rad}[C_y]$ and, by (1.2), α lies strictly between x and y. Statement (iv) then follows from Corollary 2.4. When α is rational and $x = A_{N-1}/B_{N-1}$ we may apply Corollary 2.4 in the same way but with $y = u/v$, where $u = A_N - A_{N-1}$ and $v = B_N - B_{N-1}$. One must check that u and v are coprime, v is positive, and the hypotheses of Corollary 2.4 are satisfied; these facts follow from the equations $B_N = b_N B_{N-1} + B_{N-2}$, with $b_N \geqslant 2$, and $|uB_N - vA_N| = 1$. When α is rational and $x = \alpha$, again statement (iv) holds, because $R_\alpha(\alpha) = 0$.

Finally we prove that if (i) *fails* then (iv) also fails. Suppose then that x is *not* a convergent of α. Let r_n denote the radius $1/(2B_n^2)$ of C_{A_n/B_n}. Then $r_1, r_2, \dots$ is a strictly decreasing sequence which, if α is irrational, has limit 0, and which, if α is rational, has terminal value $r_N = \mathrm{rad}[C_\alpha]$. In the latter case, if $\mathrm{rad}[C_x] \leqslant \mathrm{rad}[C_\alpha]$ then statement (iv) fails because $R_\alpha(\alpha) = 0$ and $x \neq \alpha$. Thus we may assume that $\mathrm{rad}[C_x] > \mathrm{rad}[C_\alpha]$.

Since $r_0 = \frac{1}{2}$ (the greatest possible radius of a Ford circle) either (a) $r_0 = r_1$ and there is a unique integer $n \geqslant 1$ such that $r_n \geqslant \mathrm{rad}[C_x] > r_{n+1}$, or (b) $r_0 > r_1$ and there is a unique integer $n \geqslant 0$ such that $r_n \geqslant \mathrm{rad}[C_x] > r_{n+1}$. From (1.2), α lies strictly between A_n/B_n and A_{n+1}/B_{n+1}. By Lemma 2.2, x lies outside the closed interval bounded by A_n/B_n and A_{n+1}/B_{n+1}, and by Lemma 2.3, $R_{A_n/B_n}(\alpha) < R_x(\alpha)$. Let $z = A_n/B_n$. Then, because $\mathrm{rad}(C_z) \geqslant \mathrm{rad}(C_x)$, we see that statement (iv) fails. □

3 Concluding remarks

Let j denote the point $(0, 0, 1)$ in three-dimensional Euclidean space. The *Ford sphere* S_x of $x = a/b$, where a and b are coprime Gaussian integers, is the sphere with centre $x + j/(2|b|^2)$ and radius $1/(2|b|^2)$. Ford spheres share many properties with Ford circles, and they can be used in the study of Gaussian integer continued fraction expansions of complex numbers. A brief account can be found at the end of [1]. It would be of interest to investigate whether the techniques of this paper can be applied to Ford spheres and Gaussian integer continued fractions to give results on approximation of complex numbers by quotients of Gaussian integers.

Acknowledgments. I thank the referees, whose suggestions have greatly improved this paper.

References

[1] L. R. Ford, Fractions, *Amer. Math. Monthly* **45** (1938) 586–601.

[2] M. C. Irwin, Geometry of continued fractions, *Amer. Math. Monthly* **96** (1989) 696–703.

[3] A. Ya. Khinchin, Continued Fractions (trans. of 3rd, 1961 Russian ed.), Dover, Mineola, NY, 1997; reprint of University of Chicago Press, Chicago, IL, 1964.

Ian Short received his B.A. in 2000 and his Ph.D. in 2005, both from the University of Cambridge. He is currently a lecturer in analysis in The Open University.
Mathematics and Statistics Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
i.short@open.ac.uk