Ford circles, continued fractions, and rational approximation

How to cite:

© 2011 The Mathematical Association of America

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Accepted Manuscript

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.4169/amer.math.monthly.118.02.130

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
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 \( \alpha \) coincide with the rationals that are best approximations of \( \alpha \).

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 \( \alpha \) provided that, for each rational \( c/d \) such that \( d \leq b \), we have

\[ |b\alpha - a| \leq |d\alpha - c|, \]

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

\[
b_0 + \frac{1}{b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \ldots}}} ,
\]

where \( b_0 \) is an integer and the other coefficients \( b_i \) are positive integers. Either the sequence \( b_0, b_1, b_2, \ldots \) 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 \( A_0, A_1, A_2, \ldots \) and positive integers \( B_0, B_1, B_2, \ldots \) by the matrix recurrence relations

\[
\begin{pmatrix}
    A_n & A_{n-1} \\
    B_n & B_{n-1}
\end{pmatrix} = \begin{pmatrix}
    A_{n-1} & A_{n-2} \\
    B_{n-1} & B_{n-2}
\end{pmatrix} \begin{pmatrix}
    b_n & 1 \\
    1 & 0
\end{pmatrix}, \quad \begin{pmatrix}
    A_1 & A_0 \\
    B_1 & B_0
\end{pmatrix} = \begin{pmatrix}
    b_0b_1 + 1 & b_0 \\
    b_1 & 1
\end{pmatrix} ,
\]

(1.1)
for \( n \geq 2 \). 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 \( n \geq 1 \). Set \( z = \infty \) to see that

\[
\frac{A_0}{B_0} = b_0, \quad \frac{A_1}{B_1} = b_0 + \frac{1}{b_1}, \quad \frac{A_2}{B_2} = b_0 + \frac{1}{b_1 + \frac{1}{b_2}}, \ldots
\]

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 \( N \geq 1 \) then \( b_N \geq 2 \). This ensures that to each real number \( \alpha \) there corresponds a unique continued fraction with value \( \alpha \). The rationals \( A_n/B_n \) are known as the convergents of \( \alpha \), and they alternate from one side of \( \alpha \) to the other, in the sense that

\[
\frac{A_0}{B_0} < \frac{A_2}{B_2} < \frac{A_4}{B_4} < \cdots < \frac{A_5}{B_5} < \frac{A_3}{B_3} < \frac{A_1}{B_1}.
\] (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 \( \alpha \) if and only if it is a best approximation of \( \alpha \).

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

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 \( \text{rad}[C] \), 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

\[
|x - y|^2 \leq 4rs,
\] (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.

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.](image)

We define the continued fraction chain of a real number $\alpha$ 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 $\alpha$. Since $|A_nB_{n-1} - A_{n-1}B_n| = 1$ we see that any two consecutive circles in the continued fraction chain of $\alpha$ are tangent. Also, by (1.1) we have $B_0 = 1$, $B_1 = b_1$, and $B_n = b_nB_{n-1} + B_{n-2}$ for $n \geq 2$, 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 $\alpha$.

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

$$R_x(\alpha) = \frac{1}{2}|b\alpha - a|^2 = \frac{\alpha^2}{2} |\alpha - x|^2.$$  

(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 $\alpha$ that is tangent to $C_x$. Our main theorem is the following.

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

(i) $x$ is a convergent of $\alpha$;

(ii) $C_x$ is a member of the continued fraction chain of $\alpha$;
(iii) $x$ is a best approximation of $\alpha$;

(iv) if $z$ is a rational such that $\text{rad}[C_z] \geq \text{rad}[C_x]$ then $R_x(\alpha) \leq R_x(z)$, 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 $\text{rad}[C_z] \leq R_x(z)$, 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

$$\text{rad}[C_z] = \frac{1}{2d} \leq \frac{1}{2d}|ad - bc|^2 = R_x(z).$$

Equality holds if and only if $|ad - bc| = 1$; that is, if and only if $C_z$ and $C_x$ are tangent. \(\Box\)
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.](image)

**Proof.** Since $|z - x| < |y - x|$ we see from Lemma 2.1 (parts (i) and (ii)) that
\[
\text{rad}[C_z] \leq R_x(z) < R_x(y) = \text{rad}[C_y],
\]
and similarly $\text{rad}[C_z] < \text{rad}[C_x]$. \qed

**Lemma 2.3.** Let $C_x$ and $C_y$ be tangential Ford circles such that $\text{rad}[C_x] > \text{rad}[C_y]$, and suppose that a real number $\alpha$ 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.](image)

**Proof.** From Lemma 2.1 we have
\[
R_x(\alpha) < R_x(y) = \text{rad}[C_y].
\]
If $y$ lies between $z$ and $\alpha$ then $|y - z| < |\alpha - z|$, and by Lemma 2.1 we have
\[
\text{rad}[C_y] \leq R_y(y) < R_z(\alpha).
\]
Otherwise, $x$ must lie between $z$ and $\alpha$, so that $|x - z| < |\alpha - z|$, and again using Lemma 2.1 we have
\[
\text{rad}[C_y] < \text{rad}[C_x] \leq R_z(x) < R_z(\alpha).
\]
In summary, $R_z(\alpha) < \text{rad}[C_y] < R_z(\alpha)$, as required.

Next we prove that the Ford circle $C_x$ and the rational $\alpha$ 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 $\text{rad}[C_x] > \text{rad}[C_y]$, and suppose that a real number $\alpha$ lies strictly between $x$ and $y$. If $z$ is a rational such that $\text{rad}[C_z] \geq \text{rad}[C_x]$ then $R_z(\alpha) \leq R_z(\alpha)$, 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, $R_z(\alpha) < R_z(\alpha)$, 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 $C_{\alpha/b}$ is $1/(2\beta^2)$. For the remainder of the proof we denote the convergents of $\alpha$ by $A_0/B_0, A_1/B_1, A_2/B_2, \ldots$, and when $\alpha$ is rational the final convergent (which equals $\alpha$) 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 $\alpha$ is rational, $n < N - 1$. Note that since $x$ is not an integer, $n \geq 1$. Define $y = A_{n+1}/B_{n+1}$. Then $\text{rad}[C_x] > \text{rad}[C_y]$ and, by (1.2), $\alpha$ lies strictly between $x$ and $y$. Statement (iv) then follows from Corollary 2.4. When $\alpha$ 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 $B_N = b_BB_{N-1} + B_{N-2}$, with $b_B \geq 2$, and $|uB_N - vA_N| = 1$. When $\alpha$ 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 that then that $x$ is not a convergent of $\alpha$. Let $r_n$ denote the radius $1/(2B_n^2)$ of $C_{A_n/B_n}$. Then $r_1, r_2, \ldots$ is a strictly decreasing sequence which, if $\alpha$ is irrational, has limit $0$, and which, if $\alpha$ is rational, has terminal value $r_N = \text{rad}[C_n]$. In the latter case, if $\text{rad}[C_z] \leq \text{rad}[C_n]$ then statement (iv) fails because $R_\alpha(\alpha) = 0$ and $x \neq \alpha$. Thus we may assume that $\text{rad}[C_z] > \text{rad}[C_n]$.

Since $r_0 = 1/2$ (the greatest possible radius of a Ford circle) either (a) $r_0 = r_1$ and there is a unique integer $n \geq 1$ such that $r_n \geq \text{rad}[C_x] > r_{n+1}$, or (b) $r_0 > r_1$ and there is a unique integer $n \geq 0$ such that $r_n \geq \text{rad}[C_x] > r_{n+1}$. From (1.2), $\alpha$ 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, $R_{A_n/B_n}(\alpha) < R_z(\alpha)$. Let $x = A_n/B_n$. Then, because $\text{rad}(C_z) \geq \text{rad}(C_x)$, we see that statement (iv) fails. 

\[ \square \]
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


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