Möbius Maps and Periodic Continued Fractions

A. F. BEARDON
University of Cambridge
Cambridge CB3 0WB UK

Most people meet continued fractions for the first time in number theory, but for well over 100 years, people have studied complex continued fractions of the form

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \ddots}}, \tag{1}$$

where the a_i and b_j are complex numbers (possibly zero). The arguments used in number theory are inadequate for discussing these continued fractions, and a broader view is necessary. One such view is based on the use of Möbius maps, that is, on complex functions of the form

$$g(z) = \frac{az+b}{cz+d}, \tag{2}$$

where a, b, c and d are complex numbers with $ad - bc \neq 0$. To see why Möbius maps are relevant here, let $s_1(z) = (az+1)/z = a + 1/z$ and $s_2(z) = (bz+1)/z = b + 1/z$; then

$$s(z) = s_1\big(s_2(z)\big) = a + \cfrac{1}{b + \cfrac{1}{z}}$$

is a finite continued fraction and also a Möbius map, namely

$$s(z) = \frac{(ab+1)z + a}{bz+1}.$$

We will revisit Möbius maps after introducing some terminology about continued fractions that will allow us to state a theorem by Galois.

This paper is primarily about using Möbius transformations to analyze continued fractions of the form

$$[b_0, b_1, b_2, \ldots] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \ddots}} = \lim_{n\to\infty} [b_0, b_1, \ldots, b_n], \tag{3}$$

where the b_j are integers, with $b_1, b_2, \ldots$ positive and where

$$[b_0, b_1, \ldots, b_n] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{\cdots + \cfrac{1}{b_n}}}. \tag{4}$$

Math. Mag. **88** (2015) 272–277. doi:10.4169/math.mag.88.4.272. © Mathematical Association of America
MSC: Primary 40A15, Secondary 30B70; 11A55.

The limit in (3) always exists, and $[b_0, b_1, \ldots, b_n]$ in (4) is calculated according to the usual rules of arithmetic.

A continued fraction $[b_0, b_1, b_2, \ldots]$ is *periodic* with period k if $b_n = b_{n+k}$ for all n and *eventually periodic* if $b_n = b_{n+k}$ for all sufficiently large n. We prefer this terminology (borrowed from dynamical systems) to the classical terminology for continued fractions which would be *purely periodic* and *periodic*, respectively. If $[b_0, b_1, b_2, \ldots]$ is periodic with period k, then we write $[\overline{b_0, \ldots, b_{k-1}}]$ for $[b_0, b_1, b_2, \ldots]$. Although in general b_0 can be any integer, if $[b_0, b_1, b_2, \ldots]$ is periodic with period k, then $b_0 = b_k \geqslant 1$. In particular, the value of a periodic continued fraction exceeds 1.

A real number x is a *quadratic irrational* if and only if it is irrational and also the zero of a quadratic polynomial P with integer coefficients. The polynomial P here is unique up to a scalar multiple, and the second zero of P is the *algebraic conjugate* x^* of x. Equivalently, a quadratic irrational is an irrational number that can be written as either $q + \sqrt{r}$ or $q - \sqrt{r}$, where q and r are rational; then $q + \sqrt{r}$ and $q - \sqrt{r}$ are the *algebraic conjugates* of each other.

It is well known that each irrational number has a unique infinite continued fraction expansion of the form (3). Suppose now that x is irrational and that $x = [b_0, b_1, b_2, \ldots]$. The significance of quadratic irrationals is that $[b_0, b_1, b_2, \ldots]$ *is eventually periodic if and only if x is a quadratic irrational* and that $[b_0, b_1, b_2, \ldots]$ *is periodic if and only if x is a quadratic irrational whose algebraic conjugate x^* lies in* $(-1, 0)$. Euler proved that a periodic continued fraction is a quadratic irrational, and the converse was proved by Lagrange; for proofs of this see, for example, [**3**, Theorems 176 and 177], [**7**, p. 119], and [**6**, pp. 40–41]. The last result here is due to Galois.

Galois studied periodic continued fractions and proved the following result (see, for example, [**6**, p. 46]).

Galois' theorem. If $x = [\overline{b_0, \ldots, b_{k-1}}]$ then $[\overline{b_{k-1}, \ldots, b_0}] = -1/x^*$.

Note that if the continued fraction expansion of x is periodic, then $x > 1$. Similarly, $-1/x^* > 1$ so that $x^* \in (-1, 0)$.

The purpose of this paper is to highlight the application of Möbius transformations to continued fractions by using them to prove Galois' theorem (which is usually proved by the manipulation of solutions of recurrence relations). The use of Möbius transformations gives additional insight by showing the significance of the map $z \mapsto -1/z$ and also provides a proof when the b_i are *real* numbers greater than 1 (and not restricted to integers). We end this section with a simple example in which these ideas are completely transparent and which are used to prove Galois' theorem in a special case.

Example. Let a and b be positive integers, and let $\alpha = [\overline{a, b}]$. Simple substitution yields

$$\alpha = a + 1/(b + 1/[\overline{a, b}]) = a + 1/(b + 1/\alpha)$$

so that α is a fixed point of s, where $s(z) = a + 1/(b + 1/z)$. Now the fixed points of s are the solutions of $bz^2 - abz - a = 0$, so the roots of this equation are α and its algebraic conjugate α^*. Since $\alpha\alpha^* = -a/b < 0$, we see that $\alpha > 0 > \alpha^*$. Now let $\beta = [\overline{b, a}]$. Then, similarly, β is a fixed point of the map $z \mapsto b + 1/(a + 1/z)$, and β and β^* are the roots of the equation $az^2 - abz - b = 0$. Note that $\beta > 0 > \beta^*$. By considering the transformation $w = -1/z$, we find that $\{\alpha, \alpha^*\} = \{-1/\beta, -1/\beta^*\}$ so that $\beta = -1/\alpha^*$, and this is Galois' result in this case.

The broader view

We begin by recalling the theory of Möbius maps. Let $\mathbb{C}$ be the complex plane. We adjoin a "new" point, which we label ∞, to $\mathbb{C}$ to form the *extended complex plane* $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. A map g with domain $\mathbb{C}_\infty$ is a *Möbius map* if it can be written in the form (2) where a, b, c, and d are complex numbers with $ad - bc \neq 0$. If $c \neq 0$, we interpret (2) to say that $g(\infty) = a/c$ and $g(-d/c) = \infty$. If $c = 0$, then $g(\infty) = \infty$. Then each Möbius map g is a bijection of $\mathbb{C}_\infty$ onto itself, and the inverse of a Möbius map is a Möbius map. Moreover, the set of Möbius maps is a group under composition.

The Möbius map g in (2) preserves the extended real axis $\mathbb{R} \cup \{\infty\}$, which we denote by $\mathbb{R}_\infty$, if and only if a, b, c, and d are real numbers. Further, as

$$\mathrm{Im}[g(z)] = \frac{(ad - bc)\mathrm{Im}[z]}{|cz + d|^2},$$

we see that g preserves the upper half-plane $\mathbb{H} = \{x + iy : y > 0\}$ if and only if $ad - bc > 0$. Note that if

$$h(z) = \frac{az + b}{cz + d},$$

where a, b, c, and d are real and $ad - bc = -1$, then h can be written as

$$h(z) = \frac{iaz + ib}{icz + id}, \quad (ia)(id) - (ib)(ic) = 1,$$

and h interchanges $\mathbb{H}$ with the lower half-plane $\mathbb{H}^-$ (defined by $y < 0$). This is the case, for example, when $h(z) = 1/z$. On the other hand, $z \mapsto -1/z$ preserves both $\mathbb{H}$ and $\mathbb{H}^-$.

The remarks on metric spaces and hyperbolic geometry that follow are not necessary for the proofs in the paper, but they do explain the main ideas and motivation that lie behind this work. It is easy to make $\mathbb{C}_\infty$ into a metric space, for we can project it (using stereographic projection) onto the unit sphere $\mathbb{S}$ in $\mathbb{R}^3$ and then transfer the Euclidean metric from $\mathbb{S}$ to a metric χ on $\mathbb{C}_\infty$. Now that $(\mathbb{C}_\infty, \chi)$ is a metric space, we can show that each Möbius map g is a homeomorphism of $\mathbb{C}_\infty$ onto itself. Better still, with an appropriate definition (which we omit here), the group of Mobius maps (under composition) is the group of all conformal mappings of $(\mathbb{C}_\infty, \chi)$ onto itself. With this, we have set up the equipment to deal with continued fractions of the form (1).

Observe that the boundary of $\mathbb{H}$ is $\mathbb{R}_\infty$ and that $\mathbb{H}$, with the metric $|dz|/y$, is one of the standard models of the hyperbolic plane. Moreover, the (conformal) isometries of the hyperbolic plane are precisely the Möbius maps that leave $\mathbb{H}$ invariant. Indeed, the hyperbolic metric on $\mathbb{H}$ is a "natural" metric from the point of view of complex analysis since it is (up to a scalar multiple) the only metric for which the conformal self-maps of $\mathbb{H}$ are isometries. Although we shall not use hyperbolic geometry in this article, we would be remiss if we did not mention the fact that Möbius maps (when suitably defined) exist in Euclidean spaces of all dimensions and form the conformal isometry groups of hyperbolic space of the appropriate dimension. Sadly, the usual introduction of Möbius maps acting on $\mathbb{C}_\infty$ (where they are not isometries) completely misses this point.

The modular group

The *modular group* Γ is the group of Möbius maps of the form (2) where now a, b, c, and d are *integers* and $ad - bc = 1$. For example, the map $s(z) = a + 1/(b + 1/z)$ in the example is in the modular group. The modular group acts on the hyperbolic plane $\mathbb{H}$ as a group of isometries of the hyperbolic metric, and its action on the boundary $\mathbb{R} \cup \{\infty\}$ of $\mathbb{H}$ is intimately connected to the theory of continued fractions. Indeed, a modern view of continued fractions is to see them from the geometric perspective of the action of the modular group Γ on the boundary of $\mathbb{H}$. To study continued fractions by using only real methods is, in effect, ignoring the fact that the real axis is the boundary of $\mathbb{H}$ and that the essential mathematics here is to be found in the action of Γ on $\mathbb{H}$. To see how these ideas are used to study continued fractions see, for example, [**1, 2, 5, 8**].

Of particular interest to us are the *loxodromic isometries* of $\mathbb{H}$. These are the Möbius maps that preserve $\mathbb{H}$, and have two distinct real fixed points (for example, $z \mapsto 2z$ that fixes 0 and ∞). With these, we can see quadratic irrationals in a different light for it is a (nontrivial) fact that *a real x is a quadratic irrational if and only if it is a fixed point of some loxodromic element g of the modular group* Γ, and then *its algebraic conjugate x^* is the second fixed point of g*. This fact makes it completely transparent why, for example, quadratic irrationals have a role to play in continued fraction theory whereas cubic (and higher order) irrationals do not. Indeed, in [**4**], Khinchin proves that the real number x has an eventually periodic continued fraction expansion if and only if it is a quadratic irrational and then writes (on page 50) that "no proofs analogous to this are known for continued fractions representing algebraic irrational numbers of higher degrees." The reason for this is now apparent for the fundamental result here is that a number is a quadratic irrational if and only if it is fixed by a loxodromic element in the modular group.

There is one more fact that is crucial to our argument: If g is a loxodromic map with fixed points u and v, then one of u and v, let us say u, is an *attracting fixed point*, and then v is a *repelling fixed point*. This means that if g^n is the n-iterate of a loxodromic map g (that is the composition obtained by applying g exactly n times), then $g^n(z) \to u$ if $z \neq v$. For example, ∞ and 0 are the attracting and repelling, respectively, fixed points of $z \mapsto \lambda z$ when $|\lambda| > 1$. Informally, a fixed point w of a general map f is attracting or repelling according as $|f'(w)| < 1$ or $|f'(w)| > 1$.

Finally, the *extended modular group* Γ^* (which we shall need later) is the group of all maps $z \mapsto (az+b)/(cz+d)$, where a, b, c, and d are integers but now $|ad - bc| = 1$. It is easy to see that Γ^* is a group, with Γ a subgroup of index two in Γ^*. Note that if a is an integer, then the map $z \mapsto a + 1/z$ is in Γ^*.

The basic lemma

Our proof of Galois' theorem is based on the following simple lemma that, since the b_j need not be integers, generalizes the number-theoretic notion of an algebraic conjugate in geometric terms. In this lemma, $s_1 s_2(z) = s_1\big(s_2(z)\big)$, and so on.

Lemma. A finite composition, say $S = s_1 \cdots s_k$, of k maps of the form $s : z \mapsto b + 1/z$, where $b \geqslant 1$, has an attracting fixed point, say ζ, in $(1, +\infty)$, and a repelling fixed point, say $\tilde{\zeta}$, in $(-1, 0)$.

Proof. Let $S = s_1 \cdots s_k$, where $s_j(z) = b_j + 1/z$ and $b_j \geqslant 1$. As each s_j maps $[1, +\infty)$ onto a bounded subinterval of $(1, +\infty)$, the composition S does the same, so we conclude (from the usual elementary fixed point theorem) that S has a fixed point, say ζ, in $(1, +\infty)$. Since $|s_j'(x)| < 1$ on $(1, +\infty)$, the chain rule implies that $|S'(\zeta)| < 1$ so that ζ is an attracting fixed point of S. Now let $\tilde{S} = s_k \cdots s_1$. Then, for exactly the same reason, $\tilde{S}$ has an attracting fixed point, say $\tilde{\zeta}$, in $(1, +\infty)$. Let

$\sigma(z) = -1/z$. Then $\sigma s_j = s_j^{-1}\sigma$ so that $\sigma S = \tilde{S}^{-1}\sigma$. It follows easily that S also fixes $\sigma(\tilde{\zeta})$ and that this is the repelling fixed point of S in $(-1, 0)$.

If the b_j in the lemma are positive integers, then S is in the extended modular group, so the quadratic equation for the fixed points of S has integer coefficients and then ζ and $\tilde{\zeta}$ are algebraic conjugates in the usual sense. As we noted in our discussion of attracting and repelling fixed points, we see that $S^n(z) \to \zeta$ as $n \to \infty$, providing that $z \neq \tilde{\zeta}$. Finally, we recall that we can compute the composition of Möbius maps by multiplying their associated matrices. Thus, the matrix for S is the product of the matrices for the s_j, namely

$$\begin{pmatrix} b_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_k & 1 \\ 1 & 0 \end{pmatrix}.$$

We do not have to compute this matrix, but we do note that it has determinant $(-1)^k$. We conclude that S is in Γ if k is even and in Γ^* (but not in Γ) if k is odd.

The proof of Galois' theorem

Now consider the collection of maps $s_j(z) = b_j + 1/z$, $j = 0, 1, 2, \ldots$, where, for each j, $b_j \geqslant 1$ (note that the b_j may, but need not, be positive integers). Then, by definition,

$$[b_0, b_1, b_2, \ldots] = \lim_{n\to\infty} s_0 \cdots s_n(\infty).$$

Suppose now that $b_0, b_1, b_2, \ldots$ is periodic with period k. Let $S = s_0 \cdots s_{k-1}$, and let ζ and $\tilde{\zeta}$ be the attracting and repelling fixed points of S. Thus, $\zeta > 1$ and $\tilde{\zeta} \in (-1, 0)$. Next, let

$$K = \{\infty, s_0(\infty), s_0 s_1(\infty), \ldots, s_0 s_1 \cdots s_{k-2}(\infty)\}.$$

Then, by the lemma, $K \subset [1, +\infty)$, and since $\tilde{\zeta} < 0$, we see that $S^n(z) \to \zeta$ for each z in K. This is the same as saying that $s_0 \cdots s_n(\infty) \to \zeta$ as $n \to \infty$; thus,

$$[\overline{b_0, \ldots, b_{k-1}}] = [b_0, b_1, \ldots] = \zeta.$$

The argument in the proof of the lemma, combined with the discussion here, shows that

$$[\overline{b_{k-1}, \ldots, b_0}] = \tilde{\zeta}.$$

Since S fixes $-1/\tilde{\zeta}$ (see the proof of the lemma) we see that $-1/\tilde{\zeta} = \zeta^*$ (the algebraic conjugate of ζ), and hence,

$$[\overline{b_{k-1}, \ldots, b_0}] = -1/\zeta^*,$$

which is Galois' result. In fact, this proves more since our argument does not require the b_j to be integers.

REFERENCES

1. A. F. Beardon, Continued fractions, discrete groups and complex dynamics, *Comput. Methods Funct. Theory* **1** no. 2 (2001) 535–594.

2. W. B. Jones, W. J. Thron, Continued fractions: Analytic theory and applications, in *Encyclopedia of Mathematic & Its Applications*. Addison-Wesley, Reading, MA, 1980.
3. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*. Fifth edition. Oxford Science Pub., Clarendon Press, Oxford, 1979.
4. A. Ya Khinchin, *Continued Fractions*. Third edition. Univ. of Chicago Press, Chicago, IL, 1964.
5. L. Lorentzen, H. Waadeland, *Continued Fractions with Applications*. Stud. Comput. Math., Vol. 3, North-Holland, Amsterdam, 1992.
6. A. M. Rockett, P. Szüsz, *Continued Fractions*. World Scientific Press, Hackensack, NJ, 1992.
7. H. E. Rose, *A Course in Number Theory*. Oxford Science Pub., Clarendon Press, Oxford, 1988.
8. C. M. Series, The modular surface and continued fractions, *J. London Math. Soc.* **2** no. 31 (1985) 69–80.

Summary. We describe some of the relationships that exist between quadratic irrationals, continued fractions, Möbius maps, and hyperbolic geometry, and we illustrate these by giving a simple geometric proof of Galois' result on dual continued fractions.

ALAN F. BEARDON (MR Author ID: 33080) obtained his Ph.D. from Imperial College, London in 1964, and taught at the University of Maryland, and the University of Kent at Canterbury before moving to the University of Cambridge in 1968 where he stayed until retirement. His main mathematical interests are in hyperbolic geometry and geometric function theory; other interests include walking, wildlife photography, and woodwork.