Fractional Linear Transformations

Mohamed Bouchouata, Emile Hantrais-Smith,
Yuvraj Kumar, Reyansh Sharma

August 2023

Abstract

This paper is a write-up for a group exploration of Fractional Linear Transformations at PROMYS Europe 2023. We are mainly interested in the cycles generated by such transformations on the set $\mathbb{P}_p$, but also discuss other properties and criteria, and look at infinite fields. In this paper, we show possible cycle lengths and prove conjectures concerning them. Our methods involve a correspondence of fractional linear transformations with matrices, using linear algebra to prove results.

Contents

1 Introduction

1.1 Overview of this paper

Throughout this paper, we will be looking at the correspondence between matrices and FLTs, using linear algebra to compute and prove results. In particular, we'll be studying ideas about possible cycle lengths such as whether or not we can have one cycle of length $|\mathbb{P}_p| = p + 1$, understanding the criteria for fixed points of f, characterizing cycles of particular functions, compositions and inverses of FLTs, computing the number of possible FLTs on $\mathbb{P}_p$, the relationship between FLTs and continued fractions when considering them over $\mathbb{R}$, and other interesting conjectures.

2 Definitions and properties

2.1 Fractional linear transformations

We start by defining the principal object that we will explore throughout this paper:
Definition A fractional linear transformation (FLT), is a function of the form

$$f(x) = \frac{ax + b}{cx + d}$$

with the condition that $ad - bc \neq 0$.

2.2 Domain and range $\mathbb{P}_p$

If we want to define an FLT f on the field $\mathbb{Z}_p$, the image of x is not defined if $cx + d = 0$. To avoid this, we adjoin a new element ∞, defining $\frac{e}{0} = \infty$ for $e \neq 0$ and $f(\infty) = \frac{a}{c}$. Let $\mathbb{P}_p$ denote the union $\mathbb{Z}_p \cup \{\infty\}$.
The function f defined on $\mathbb{P}_p$ is now well-defined, as we never encounter $ax + b = 0, cx + d = 0$ simultaneously, since this would imply $ad - bc = 0$.

2.3 Composition and inverse

Consider the two FLTs $f_i(x) = \dfrac{a_i x + b_i}{c_i x + d_i}$ for $i = 1, 2$. We start by computing the composition of f_1 and f_2

$$(f_2 \circ f_1)(x) = \frac{(a_2a_1 + b_2c_1)x + (a_2b_1 + b_2d_1)}{(c_2a_1 + d_2c_1)x + (c_2b_1 + d_2d_1)}.$$

We can also compute the inverse of f to get

$$f^{-1}(x) = \frac{1}{ad - bc}\frac{dx - b}{-cx + a}$$

(well-defined as $ad - bc \neq 0$).
Notice that composing and inverting FLTs resembles multiplying and inverting matrices with the same coefficient. We will later see how we can formalise this notion and use it to our advantage...

2.4 Cycles and cycle lengths

The function f is a permutation of the finite set $\mathbb{P}_p$. Applying f repeatedly to x generates a subset $\mathbb{P}_p$: the cycle generated by x. It is a finite set, and we say that the cycle generated by x has length l if this set has size l. Define the *length* L_f of an FLT f to be the ordered tuple of its cycle lengths:

$$L_f = (l_1, l_2, ..., l_k)$$

Where $l_1 \leq l_2 \leq ... \leq l_k$ are the lengths of the cycles of f.

3 Matrix and vector representations

3.1 Vector representation

We notice that there exists a bijection between element $\mathbb{P}_p$ and the set of lines of $(\mathbb{Z}_p^2)^\times$ that pass through the origin, as we can assign to each line of $(\mathbb{Z}_p^2)^\times$ a slope in $\mathbb{P}_p$.
We can therefore map the vectors of $(\mathbb{Z}_p^2)^\times$ to elements of $\mathbb{P}_p$ with the map $\rho : (a,b) \mapsto \frac{a}{b}$.
Notice that $\rho(a,b) = \rho(a',b') \iff (a',b') = \lambda(a,b)$ for some $\lambda \in \mathbb{Z}_p^\times$.
In the following parts of the paper, we will say that the vector (a,b) represents $\frac{a}{b}$ and we may represent $\frac{a}{b}$ by any element of $\rho^{-1}(\frac{a}{b}) = \{\lambda(a,b) \mid \lambda \in \mathbb{Z}_p^\times\}$.

3.2 Matrix representation

In this section we will formalise the relation between FLTs and matrices.
Consider the map

$$\begin{aligned} \phi : \mathrm{GL}_2(\mathbb{Z}_p) &\longrightarrow \mathbb{FLT} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\longmapsto \frac{ax+b}{cx+d} \end{aligned}$$

And in general we will use the notation M_f to denote any matrix with $\phi(M_f) = f$

Proposition $f(\rho(v)) = \rho(M_f v)$

Proof. Write $M_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so $f = \phi(M_f) = \frac{ax+b}{cx+d}$, and $v = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, so $x = \rho(v) = \frac{\alpha}{\beta}$. Then in the case where $\beta \neq 0$, we have

$$f(\rho(v)) = f\left(\frac{\alpha}{\beta}\right) = \frac{a(\frac{\alpha}{\beta})+b}{c(\frac{\alpha}{\beta})+d} = \frac{a\alpha+b\beta}{c\alpha+d\beta} = \rho\left(\begin{pmatrix} a\alpha+b\beta \\ c\alpha+d\beta \end{pmatrix}\right) = \rho(M_f v)$$

In the case $\beta = 0$, $\alpha \neq 0$, so $\rho(v) = \infty$, meaning

$$\rho(M_f v) = \rho\left(\begin{pmatrix} a\alpha \\ c\alpha \end{pmatrix}\right) = \frac{a}{c} = f(\infty) = f(\rho(v)))$$

□

We check that two matrices that are mapped to the same FLT are scalar multiples of each other:

Proposition For $M, M' \in \mathrm{GL}_2(\mathbb{Z}_p)$, $\phi(M) = \phi(M') \iff M = \lambda M'$ for some $\lambda \in \mathbb{Z}_p^\times$

Proof. Write $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $M' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$ and $\phi(M) = f$, $\phi(M') = g$

($\Rightarrow$) : If $\phi(M) = \phi(M')$ and M', then we have that $\forall \alpha, \beta \in \mathbb{P}_p$

$$\rho\left(M\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) = \rho\left(M'\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right)$$

Therefore

$$\rho\left(\begin{pmatrix} a\alpha + b\beta \\ c\alpha + d\beta \end{pmatrix}\right) = \rho\left(\begin{pmatrix} a'\alpha + b'\beta \\ c'\alpha + d'\beta \end{pmatrix}\right)$$

Taking $(\alpha, \beta) = (1, 0), (0, 1)$ yields

$$\rho\left(\begin{pmatrix} a \\ c \end{pmatrix}\right) = \rho\left(\begin{pmatrix} a' \\ c' \end{pmatrix}\right) \quad \text{and} \quad \rho\left(\begin{pmatrix} b \\ d \end{pmatrix}\right) = \rho\left(\begin{pmatrix} b' \\ d' \end{pmatrix}\right)$$

So $a = \mu a'$, $c = \mu c'$ and $b = \nu b'$, $d = \nu d'$ for some $\mu, \nu \in \mathbb{Z}_p^\times$
Similarly, by taking $(\alpha, \beta) = (1, 1)$, we get

$$\rho\left(\begin{pmatrix} a + b \\ c + d \end{pmatrix}\right) = \rho\left(\begin{pmatrix} a' + b' \\ c' + d' \end{pmatrix}\right)$$

And so $a + b = \omega(a' + b'), c + d = \omega(c' + d')$ for some $\omega \in \mathbb{Z}_p^\times$.
Combining these equations gives us

$$\begin{cases} a' + \nu b' = \omega(a' + b') \\ \mu c' + \nu d' = \omega(a' + d') \end{cases}$$

So we have

$$M' \begin{pmatrix} \mu \\ \nu \end{pmatrix} = M' \begin{pmatrix} \omega \\ \omega \end{pmatrix}$$

And since $M' \in \mathrm{GL}_2(\mathbb{Z}_p)$, it is invertible, giving

$$\begin{pmatrix} \mu \\ \nu \end{pmatrix} = \begin{pmatrix} \omega \\ \omega \end{pmatrix}$$

So $\mu = \nu$, and therefore $M = \mu M'$

($\Leftarrow$) : If $M = \lambda M'$, then for any $v \in (\mathbb{Z}_p^2)^\times$, $Mv = (\lambda M')v = M'(\lambda v) \implies \rho(Mv) = \rho(M'(\lambda v)) \implies f(\rho(v)) = g(\rho(\lambda v)) = g(\rho(v))$, therefore $f(x) = g(x)\ \forall x \in \mathbb{P}_p$, and so $f = g$.

□

3.3 Fixed points and eigenvectors

We recall that λ is an eigenvalue of a matrix M if and only if there exists a non-zero vector v such that $Mv = \lambda v$. This relation is equivalent to $(M - \lambda I)v = 0$, which has a non-zero solution if and only if $\det(M - \lambda I) = 0$. We hence define the characteristic polynomial $c_M(x) = \det(xI - M) = x^2 - \mathrm{Tr}Mx + \det M$, whose roots are the eigenvalues of M.

Proposition $x = \rho(v)$ is a fixed point of $f \iff v$ is an eigenvector of M_f over $\mathbb{Z}_p$

Proof. ($\Leftarrow$): Given an eigenvector v of M_f whose eigenvalue $\lambda \in U_p$, we have that $M_f v = \lambda v$ so $\rho(M_f v) = \rho(\lambda v) = \rho(v)$, so $f(\rho(v)) = \rho(v)$, meaning $\rho(v)$ is a fixed point of f.

($\Rightarrow$): Given a fixed point $x \in \mathbb{P}_p$, since ρ is a surjection, there exists $v \in (\mathbb{Z}_p^2)^\times$ such that $\rho(v) = x$. So $f(x) = x = \rho(v), f(\rho(v)) = \rho(Mv) = \rho(v)$, which implies $Mv = \lambda v$ for some $\lambda \in \mathbb{Z}_p$, therefore v is an eigenvector.

□

Proposition For FLTs f and g, if $M_f \sim M_g$, then $L_f = L_g$

Proof. $M_f \sim M_g$ implies that there exists an invertible matrix P such that $M_f = P^{-1}M_gP$.
For any $k \in \mathbb{N}$, an eigenvector v of M_f^k and its corresponding eigenvalue λ will have

$$\lambda v = M_f^k v = P^{-1}M_g^k Pv$$

And multiplying by P gives us

$$P\lambda v = \lambda(Pv) = M_g^k(Pv)$$

Therefore there is a bijection between the eigenvectors of M_f^k and the eigenvectors of M_g^k. Hence their corresponding FLTs generate the same cycle lengths, so $L_f = L_g$. □

4 How many unique FLTs are there in $\mathbb{P}_p$?

4.1 Counting

4.1.1 Definitions and motivations

Due to the correspondence between FLTs and matrices, to determine the number of unique FLTs in $\mathbb{P}_p$ we should look at the number of invertible 2x2 matrices modulo p. Our matrices are invertible since $ad - bc \neq 0$.

Lemma 3.1 *Given a prime p,* $|GL_2(\mathbb{Z}_p)| = p(p-1)^2(p+1)$

Proof.

We start by counting all elements of $M_2(\mathbb{Z}_p)$. We have freedom for each of the 4 entries to be any of the p elements of $\mathbb{Z}_p$ and thus there are p^4 distinct elements of $M_2(\mathbb{Z}_p)$, so $|M_2(\mathbb{Z}_p)| = p^4$.

Now we wish to exclude non-invertible elements M of $M_2(\mathbb{Z}_p)$, i.e. M for which $\det(M) = 0$, taking $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, this is equivalent to $ad = bc$. So, we split into the cases of different numbers of entries that are 0.

Case 1. *All four entries are* 0.
There is trivially only 1 such matrix.

Case 2. *Exactly three entries are* 0.
By the pigeon-hole principle, both ad and bc are equal to zero and so we only have freedom on the non-zero variable of which there are $p-1$ possibilites meaning there are $\binom{4}{1}\cdot(p-1) = 4(p-1)$ total possiblities in this case.

Case 3. *Exactly two entries are* 0.
As we want $ad - bc = 0$, we need that exactly one of a or d are zero and exactly one of b and c are zero. Therefore there are $(p-1)^2$ possiblities for the other two values and so there are $\binom{2}{1}\cdot\binom{2}{1}\cdot(p-1)^2 = 4(p-1)^2$ total possiblities in this case.

Case 4. *Exactly one entry is* 0.
If either side of $ad = bc$ is zero then the other side must also be zero. Since there are no zero divisors in $\mathbb{Z}_p$, we would have more than one variable equal to zero which is a contradiction to our case. Therefore, there are 0 total possibilities in this case.

Case 5. *Exactly zero entries are* 0.
Suppose a, b, c are given, then $ad = bc \iff d = bca^{-1}$ where a^{-1} must exist in $\mathbb{Z}_p$ since $(a, p) = 1 \implies a$ is invertible. This means that we have freedom of choice on three entries (given that they are non-zero) and so there $(p-1)^3$ total possibilities in this case.

So we have

$$\begin{aligned}&(p-1)^3+0+4(p-1)^2+4(p-1)+1\\&=(p-1)^3+4(p-1)^2+4(p-1)+1\end{aligned}$$

non-invertible matrices.
By the inclusion-exclusion principle, we have that

$$\begin{aligned}|\mathrm{GL}_2(\mathbb{Z}_p)| &= p^4-((p-1)^3+4(p-1)^2+4(p-1)+1)\\&=p(p-1)^2(p+1)\end{aligned}$$

□

Now, since for $M, M' \in \mathrm{GL}_2(\mathbb{Z}_p), \phi(M)=\phi(M') \iff M=\lambda M'$ for some $\lambda \in \mathbb{Z}_p^\times$, we have that each $f \in \mathbb{FLT}$ is mapped to by exactly $(p-1)$ elements of $\mathrm{GL}_2(\mathbb{Z}_p)$, meaning

$$|\mathbb{FLT}| = \frac{|\mathrm{GL}_2(\mathbb{Z}_p)|}{p-1} = (p-1)p(p+1)$$

Alternative proof 1:

Proof.

Consider the FLT f. We distingish between two cases:

Case 1: $a=0$:
Since $a=0$ and $ad-bc\neq 0$, $b\neq 0$ and we can fix b to be 1. We now have p possible choices for d. To find the number of choices for c, we notice that $ad-bx=0$ has a unique solution for $b\neq 0$. Hence we have $p-1$ choices for c, and $(p-1)p$ FLTs.

Case 2:
$a\neq 0$:
We can fix a to be 1. We then choose b and c with p choices for each. In the same way as in the previous case, $ax-bc=0$ has a unique solution as $a\neq 0$. Hence we have $p-1$ choices for d. This yields $(p-1)p^2$ FLTs.

Therefore, we get $(p-1)p+(p-1)p^2=(p-1)p(p+1)$ FLTs in total.

□

Alternative proof 2:

Lemma: *For FLTs f and g over $\mathbb{P}_p$ we have that*

$$f(0) = g(0) \quad f(1) = g(1) \quad f(\infty) = g(\infty)$$
$$\iff f(x) = g(x)$$

Proof.

$$f(x) = \frac{ax+b}{cx+d}$$
$$g(x) = \frac{a'x+b'}{c'x+d'}$$

We consider the following system of equations.

$$f(\infty) = g(\infty) \Leftrightarrow \frac{a}{c} = \frac{a'}{c'}$$

$$f(0) = g(0) \Leftrightarrow \frac{b}{d} = \frac{b'}{d'}$$
$$f(1) = g(1) \Leftrightarrow \frac{a+b}{c+d} = \frac{a'+b'}{c'+d'}$$

This gives us that

$$a = \mu a' \quad c = \mu c'$$
$$b = \lambda b' \quad d = \lambda b'$$
$$\Rightarrow a+b = \omega(a'+b') \quad , \quad c+d = \omega(c'+d')$$

For some $\mu, \lambda, \omega \in U_p$

Substituting into the 3rd pair of equations gives

$$\mu a' + \lambda b' = \omega(a'+b') \quad \mu c' + \lambda d' = \omega(c'+d')$$

Dividing one by the other leaves

$$\frac{\mu a' + \lambda b'}{\mu c' + \lambda d'} = \frac{\omega a' + \omega b'}{\omega c' + \omega d'} \quad \frac{(\frac{\mu}{\lambda})a' + b'}{(\frac{\mu}{\lambda})c' + d'} = \frac{a'+b'}{c'+d'}$$

Or equivalently

$$g\Big(\frac{\mu}{\lambda}\Big) = g\Big(1\Big)$$

And since all FLTs are invertible, we have $\frac{\mu}{\lambda} = 1$, so $\mu = \lambda$, so we have:

$$f(x) = \frac{ax+b}{cx+d} = \frac{\mu(a'+b')}{\mu(c'+d')} = \frac{a'x+b'}{c'x+d'} = g(x)$$

□

Using this property, we can uniquely identify a FLT by the values it maps 0, 1 ∞ to. And given the mappings of 0, 1 and ∞, it is trivially true that there exists an FLT which gives these mappings. Therefore it follows that given a prime p, there is a bijection between distinct FLTs and mappings of 0, 1, and ∞ to distinct elements of $\mathbb{P}_p$. So there are exactly $(p-1)p(p+1)$ distinct FLTs.

5 Properties of cycles generated over $\mathbb{P}_p$

Lemma 5.1 *If a FLT f has more than 2 fixed points, then $f = Id$.*

Proof.

Suppose that $f \neq \text{Id}$. we have that f has a fixed point at x if and only if $f(x) = \frac{ax+b}{cx+d} = x$, meaning $cx^2 + (d-a)x - b = 0$. This is a non-zero polynomial of degree ≤ 2 in $\mathbb{Z}_p[x]$, so has at most 2 roots in $\mathbb{Z}_p$, meaning f has at most 2 fixed points.

Notice that we must consider the special case where $x = \infty$ is a fixed point, which happens if and only if $c = 0$ and $a \neq 0$, in which case the polynomial $cx^2+(d-a)x-b$ is in fact linear, so can yield at most 1 solution, so in this case we still get that f has at most 2 fixed points.

For the case where f is the identity, every point in the domain is a fixed point, so f has more than 2 fixed points.

□

Definition *(Discriminant) Given a FLT $f(x) = \frac{ax+b}{cx+d}$, we define Δ_f to be the discriminant of the quadratic equivalent to $f(x) = x$, so*

$$\Delta_f = (a-d)^2 + 4bc$$

Corollary 5.1 Given a FLT $f \neq \text{Id}$
f has 2 fixed points $\iff$ Δ_f is a QR mod p
f has 1 fixed point $\iff$ $\Delta_f \equiv 0 \bmod p$
f has no fixed points $\iff$ Δ_f is QNR mod p

Definition *(Order) The order of a FLT, $ord(f)$, is defined as:*

$$ord(f) := \min\{n \in \mathbb{N} : f^n = \text{Id}\}$$

This function is well-defined since given a prime p, the set of all FLTs is finite and forms a group under composition.

Lemma 5.2 *$\forall a \in \mathbb{P}_p$, there exists $k \in \mathbb{N}$ such that $f^k(a) = a$*

Proof.

Consider the sequence $S = (f^n(a)))_{n \in \mathbb{N} \cup \{0\}}$. Since $\mathbb{P}_p$ is finite and S is an infinite sequence whose elements are all in $\mathbb{P}_p$, we must have that at least 1 element of S appears twice. Say that the element $f^m(a)$ appears twice, then $f^{m-1}(a)$ must also appear twice, since f^{-1} is well defined. And we can repeat this inductively to get that $f^0(a) = a$ appears twice. Therefore there is some $k \in \mathbb{N}$ such that $f^k(a) = a$.

□

Lemma 5.3 If $f^k(a) = a$ then $cyc(a) = \{a, f(a), f^2(a), ..., f^{k-1}(a)\}$

Proof.

For any $t \in \mathbb{N} \cup \{0\}$ we have that $f^{t+k}(a) = f^t(f^k(a)) = f^t(a)$. Therefore $t \equiv s \bmod t$ implies $f^s(a) = f^t(a)$.
So all iterates $f^r(a)$ with $r \geq t$ are equal to some element in the set

$$\{a, f(a), f^2(a), ..., f^{k-1}(a)\}$$

meaning $cyc(a) = \{a, f(a), f^2(a), ..., f^{k-1}(a)\}$.

□

Lemma 5.4 *If $a \in cyc(b)$, $cyc(a) = cyc(b)$*

Proof.

From the definition of a cycle, $a = f^k(b)$ for some $k \in \mathbb{N} \cup \{0\}$. Then (from the definition of a cycle) we have that $f^t(a) = f^{t+k}(b) \in cyc(b)$ for all $t \in \mathbb{N} \cup \{0\}$ so $cyc(a) \subseteq cyc(b)$.

Claim: $b \in cyc(a)$
FTSOC, suppose $b \notin cyc(a)$, then for all $r \in \mathbb{N}$ with $r \geq k$, $f^r(b) = f^{r-k}(a) \neq b$. The same holds for all r such that $0 < r < k$ since $f^r(b) = b$ implies (by Lemma 4.3) that $cyc(b) = \{b, f(b), f^2(b), ..., f^{r-1}(b)\}$, but clearly a is not in this set. So there is no $r \in \mathbb{N}$ such that $f^r(b) = b$, which contradicts Lemma 4.2.

Therefore $b \in cyc(a)$, so $cyc(b) \subseteq cyc(a)$, so $cyc(a) = cyc(b)$.

□

Corollary 5.2 *The sum of the cycle lengths generated by f over $\mathbb{P}_p$ is $|\mathbb{P}_p| = (p+1)$.* This follows from Lemma 4.4 since we get that the cycles generated by f over $\mathbb{P}_p$ partition $\mathbb{P}_p$, so the sum of their lengths $= |\mathbb{P}_p| = (p + 1)$

Lemma 5.5 *ℓ is the smallest natural number such that $f^\ell(a) = a \iff |cyc(a)| = \ell$*

Proof.

($\Rightarrow$)
By Lemma 5.3 $cyc(a) = \{a, f(a), f^2(a), ..., f^{\ell-1}(a)\}$. Furthermore, each of these elements must be distinct, since if we had $f^i(b) = f^j(b)$ for some $0 \leq i < j < \ell$ then we would have $f^{j-i}(b) = b$, which contradicts the minimality of ℓ. Therefore $|cyc(a)| = \ell$.

($\Leftarrow$)
Notice that by Lemma 5.2 we have that there is $k \in \mathbb{N}$ such that $f^k(a) = a$, so there must be a minimal value in $\mathbb{N}$ with this property, say k, then the same reasoning as above shows that $cyc(a) = \{a, f(a), f^2(a), ..., f^{\ell-1}(a)\}$ and that $k = |cyc(a)| = \ell$, so $k = \ell$.

□

Lemma 5.6 $\forall x \in \mathbb{P}_p$, $|cyc(x)| \leq ord(f)$

Proof.

Take $n = |cyc(x)|$, then $\forall x \in cyc(x)$, By Lemma 4.5 we have that

$$x, f(x), f^2(x), f^3(x), ..., f^{n-1}(x)$$

are distinct. FTSOC, suppose that $n > ord(f)$. Then $x \neq f^{ord(f)}(x)$, which contradicts the definition of $ord(f)$.

□

Lemma 5.7 *If* $|cyc(a)| = \ell$ *then* f^ℓ *fixes all* $b \in cyc(a)$

Proof.

By Lemma 5.4, for all $b \in cyc(a)$, $cyc(b) = cyc(a)$, so by Lemma 4.5 f^ℓ fixes b.

□

Theorem 5.1 *A FLT has at most 2 distinct cycle lengths*

Proof.

Take $\alpha, \beta, \gamma \in \mathbb{P}_p$ with

$$a = |cyc(\alpha)| \quad b = |cyc(\beta)| \quad c = |cyc(\gamma)|$$

And FTSOC, suppose a, b, c are distinct. WLOG we can assume $a < b < c$. We consider the following 2 cases:

Case 1. $a = 1$
In this case we get $b > a = 1$, so $b \geq 2$. Since α is in a cycle of length 1, f^k fixes α for all $k \in \mathbb{N}$. And by Lemma 4.7 we have that f^b fixes all $x \in cyc(\beta)$, therefore f^b fixes ≥ 3 points, so by Lemma 4.1 $f^b = \text{Id}$. But by Lemma 4.6, we get

$$c = |cyc(\gamma)| \leq ord(f) \leq b < c$$

Contradiction.

Case 2. $a \neq 1$
In this case we get $b > a \geq 2$, so $b \geq 3$, which leads to the same contradiction.

□

Theorem 5.2 *If a FLT has no fixed points, it has 1 distinct cycle length*

Proof.

FTSOC suppose we have a FLT f with no fixed points and 2 distinct cycle lengths a and b. WLOG we can assume $a < b$, and since f has no fixed points, there are no cycles of length 1, so we have $1 < a < b$.

We consider the following 2 cases:
Case 1. $a \geq 3$
By Lemma 5.7, f^a fixes every element in $cyc(a)$, meaning f^a has ≥ 3 fixed points, so by Lemma 5.1 $f^a = \text{Id}$, so $ord(f) = a < b$, which contradicts Lemma 5.6.

Case 2. $a = 2$
In this case we must have that f^2 has 2 fixed points by Lemma 5.7 and is not the identity (else we reach the same contradiction as in case 1).
By Corollary 4.1, if we take $f(x) = \frac{ax+b}{cx+d}$, then Δ_f is a QNR mod p. But notice that $\Delta_{f^2} = (a+d)^2((d-a)^2 + 4bc) = (a+d)^2\Delta_f$.
Since f^2 has fixed points, we have that $\Delta_{f^2} = 0$ or Δ_{f^2} is a QR mod p. But Δ_{f^2} cannot be a QR, since Δ_f is a QNR, meaning we must have $\Delta_{f^2} = 0$, but since $\Delta_f \neq 0$, $(a+d)^2 = 0$ so $a + d = 0$, leaving

$$f^2(x) = \frac{(a^2+bc)x + b(a+d)}{(a+d)x + (bc+d^2)} = \frac{(a^2+bc)x}{a^2+bc} = x$$

Contradiction.

□

Remarks

Combining the results of Theorem 5.1 and 5.2, we can describe the general structure of a FLT's cycles over $\mathbb{P}_p$ given its number of fixed points:

Case 1. *No fixed points:*
d cycles of length $\frac{p+1}{d}$

Case 2. *1 fixed point:*
1 cycle of length 1 and 1 cycle of length p

Case 3. *2 fixed points:*
2 cycles of length 1 and d cycles of length $\frac{p-1}{d}$

So we have that given an odd prime p, the set of possible cycle lengths is a subset of the set of positive divisors of $(p-1), p$ and $(p+1)$.

In the next section we prove that in fact these two sets are equal...

6 Possible cycle lengths

In this section, we will prove the main results of this paper:

Theorem *(Classification of FLTs) Let $\lambda_{1,2}$ be the eigenvalues of M_f (possibly in $\mathbb{Z}_p[\sqrt{\delta}]$) and d the order of $\frac{\lambda_1}{\lambda_2}$. Then the possible lengths L_f are*

- $L_f = (1, \ldots, 1) \iff M_f = I$
- $L_f = (1, 1, d, \ldots, d) \iff M_f$ *has two eigenvalues in* $\mathbb{Z}_p$
- $L_f = (1, p) \iff M_f$ *has one eigenvalue and* $M_f \neq I$.
- $L_f = (d, \ldots, d) \iff M_f$ *has two eigenvalues in* $\mathbb{Z}_p[\sqrt{\delta}]$

We first need a lemma:
Lemma *If $M \in GL_2(\mathbb{Z}_p)$ has three eigenvectors v_1, v_2, v_3, pairwise independent, then $M = I$.*

Proof. In $\mathbb{Z}_p^2$, M has either two eigenspaces $E_{\lambda_1}, E_{\lambda_2}$ of dimension 1 or one eigenspace E_{λ_1} of dimension at most 2. By the Pigeonhole principle, one eigenspace E_{λ_i} contains two idependent vectors v_i, v_j. As v_i, v_j are independent, they span the set $\mathbb{Z}_p^2$. Hence $E_{\lambda_i} = \mathbb{Z}_p^2$. That is, for all $v \in \mathbb{Z}_p^2, Mv = \lambda_i v$. This equivalent to $(M - \lambda_i I)v = 0$. This yieds $M - \lambda_i I = 0$, and thus $M = \lambda_i I = I$. □

We now prove the theorem:

Proof. ($\Rightarrow$) This direction of the proof is pretty straightforward, by comparing the number of the 1-cycles with the number of eigenvalues.
($\Leftarrow$) We consider each of the cases:

- If $M_f = I$, then all vectors are eigenvectors of M_f, and $L_f = (1, ..., 1)$.
- If M_f has two eigenvalues, then M_f has two fixed points, so $l_1 = l_2 = 1$. Now, the next cycle, of length l_3, provides a new eigenvalue to $M_f^{l_3}$. $M_f^{l_3}$ has now three eigenvectors. By the previous lemma, $M_f^{l_3} = I$, and $L_f = (1, 1, l_3, ..., l_3)$. To show that $l_3 = d$, write $D = \mathrm{diag}(\lambda_1, \lambda_2)$ so that $M = PDP^{-1}$ for some $P \in GL_2$, we notice that $M_f^k = I$ if and only if $\lambda_1^k = \lambda_2^k$, as $M_f^k = PD^kP^{-1} = I$ implies $I = D^k = \mathrm{diag}(\lambda_1^k, \lambda_2^k)$.
- If M_f has no eigenvalues in $\mathbb{Z}_p^\times$ then the eigenvalues of M_f satisfy $\lambda_2 = \overline{\lambda_1}$. We have that $M_f^{l_1}$ has a eigenvector, so $\lambda_1^{l_1} = \overline{\lambda_1}^{l_1} = \lambda_2^{l_1}$. This implies that $M_f^{l_1} = I$ as we can write $M_f^{l_1} = PD^{l_1}P^{-1} = I$ for P invertible matrix. This implies that $I = D^{l_1} = \mathrm{diag}(\lambda_1^{l_1}, \lambda_2^{l_1})$. By minimality of l_1, $l_1 = d$. We get that $L_f = (d, ..., d)$.
- If M_f has one eigenvalue, then $l_1 = 1$ and $l_2 \leq 2$. This implies that $M_f^{l_2}$ has at least three eigenvectors. By the previous lemma, $M_f^{l_2} = I$ and $L_f = (1, l_2, ..., l_2)$. Now we have $l_2 \mid (p+1) - 1 = p$ and $l_2 \neq 1$, so $l_2 = p$. Hence $L_f = (1, p)$.

□

Theorem *The set of possible cycles lengths is the set of divisors of* $(p-1)$*,* p *and* $(p+1)$*.*

Proof. We distinguish between the three following cases:

- If $d \mid p-1$, then let g be a generator of U_p. The function represented by the matrix $\text{diag}(g^{\frac{p-1}{d}}, 1)$ had cycle lengths $(1, 1, d, ..., d)$.
- If $d \mid p$, then we just need to consider $d = p$. We can check that the function $x+1$ gives a 1-cycle ∞ and a p-cycle $0 \mapsto 1 \mapsto ... \mapsto p-1$.
- If $d \mid p+1$ then we only need to find a function f that gives a $p+1$-cycle, as then $f^{\frac{p+1}{d}}$ gives d-cycles. We will prove this in the next section...

□

7 Cycle of length $p+1$

We will show that for all odd primes p, we can find a FLT that generates one $(p+1)$-cycle.
Let f a FLT such that M_f has no eigenvalues in $\mathbb{Z}_p$. First, notice that as there are no 1-cycles, the characteristic polynomial c_f of the matrix M_f has no eigenvalues in $\mathbb{Z}_p$.

Proposition: *Let δ be a QNR mod p. Then the ring $\mathbb{Z}_p[\sqrt{\delta}]$ is a field.*

Proof.

We can check that the inverse of $a+b\sqrt{\delta}$ is $\dfrac{a-b\sqrt{\delta}}{a^2-b^2\delta}$, which is well-defined since $a^2-b^2\delta \neq 0$, since δ is a QNR. □

Corollary: *$\mathbb{Z}_p[\sqrt{\delta}]^\times$ is cyclic.*

Proposition: *Let g be a generator of $\mathbb{Z}_p[\sqrt{\delta}]^\times$. Then $p+1 | ord(\frac{g}{\overline{g}})$.*

Proof.

Notice that the order of g is p^2-1, as it is the number of elements in $\mathbb{Z}_p[\sqrt{\delta}]^\times$. Let d be the order of $\dfrac{g}{\overline{g}}$. Then $g^d = \overline{g}^d = \overline{g^d}$, so $g^d \in \mathbb{U}_p$. We claim that g^d is a primitive root of $\mathbb{U}_p$.
Indeed suppose that there exists $x \in \mathbb{U}_p$ such that $g^{dk} \neq x$ for all $k \in \mathbb{N}$. Then we can write $x = g^e$ for some e such that d does not divide e. Then there exist q, r such that $r = e - dq$ and $0 < r < d$. But then $g^r = g^{e-dq} \in \mathbb{Z}_p$ and $\left(\dfrac{g}{\overline{g}}\right)^r = 1$. Contradiction.
Now as g^d is a primitive root, its order is $p-1$. In particular, $(g^d)^{p-1} = g^{d(p-1)} = 1$, and therefore $p^2-1 \mid d(p-1)$, or $p+1 \mid d$.

□

Now consider the matrix

$$T = \begin{pmatrix} g+\overline{g} & -g\overline{g} \\ 1 & 0 \end{pmatrix}$$

This matrix has eigenvalues, g and $\overline{g}$. By the Theorem of Classification of FLTs,

$$L_T = \left(\text{ord}\left(\frac{g}{\overline{g}}\right)\right) = (p+1).$$

8 FLTs that generate the same cycle lengths for all primes

In this section we will study a few examples of FLTs over $\mathbb{P}_p$ that only generate cycles of length 1 and ℓ regardless of the prime p.

Example 1: *The FLT* $f(x) = \dfrac{x-1}{x+1}$ *only generates cycles of lengths 1 and 4*

The corresponding matrix

$$M_f = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

has characteristic polynomial $(1-\lambda)^2 + 1 = 0$, so the eigenvalues of M_f, are $\lambda_1 = 1 + i = \sqrt{2}e^{\frac{\pi i}{4}}$ and $\lambda_2 = 1 - i = \sqrt{2}e^{-\frac{\pi i}{4}}$

and we have that

$$\lambda_1^2 = i, \lambda_1^3 = -1 + i, \lambda_1^4 = -1$$
$$\lambda_2^2 = -i, \lambda_2^3 = -1 - i, \lambda_2^4 = -1$$

Therefore, $\lambda_1^4 = \lambda_2^4 \in \mathbb{Z}_p$ for all primes p, so M_f^4 has one eigenvalue, and this eigenvalue is in $\mathbb{Z}_p$ for all p prime. So we must always have $ord(f) = 4$, so all other cycles must be 4 cycles.

A further observation is that $\lambda_1, \lambda_2 \in \mathbb{Z}_p$ if and only if -1 is a QR mod p, or equivalently, $p \equiv 1 \pmod 4$, so we get $\frac{p+1}{4}$ cycles of length 4 for $p \equiv 3 \pmod 4$ and get 2 1-cycles and $\frac{p-1}{4}$ 4 cycles.

Example 2: $f(x) = \dfrac{x-3}{x+1}$ *only generates cycles of lengths 1 and 3*

A similar analysis of the corresponding matrix $\begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}$, yields eigenvalues $\lambda_1 = 2e^{\frac{\pi i}{3}}, \lambda_2 = 2e^{-\frac{\pi i}{3}}$, and that the smallest power we must raise the eigenvalues to for them to be equal is 3. And these eigenvalues are in $\mathbb{Z}_p$ for all primes p so we always have $ord(f) = 3$.

Constructions: *Here we construct FLTs that only generates cycles of lengths 1 and* ℓ

It is easy to see from the reasoning in the above examples that if the eigenvalues λ_1, λ_2 of M_f are such that λ_1 is a multiple of a ℓ^{th} primitive root of unity in $\mathbb{C}$ such that $\lambda_1 \in \mathbb{Z}_p[\sqrt{\delta}]$ or $\lambda_1 \in \mathbb{Z}_p$ for all primes p and $\lambda_2 = \overline{\lambda_1}$. Then we have that f will only generate cycles of lengths 1 and ℓ.

The case $\ell = 2$ is simple since we can take $\lambda_1 = i$ and $\lambda_2 = -i$, giving the characteristic polynomial $\lambda^2 + 1 = 0$ so $\text{Tr}(M_f) = 0$ and $\det(M_f) = 1$. An example of a suitable corresponding matrix is $M_f = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$, so we get that $f(x) = \dfrac{x-2}{x-1}$ only generates cycles of lengths 1 and 2.

The case $\ell = 5$ is difficult given that $e^{\frac{\pi i}{5}} = \frac{1}{4}(1+\sqrt{5}) + \sqrt{\frac{5}{8} - \frac{\sqrt{5}}{8}}i$ which has no multiple which is in $\mathbb{Z}_p[\sqrt{\delta}]$ or $\mathbb{Z}_p$ in general.

However, for $\ell = 6$, we have that $e^{\frac{\pi i}{6}} = \frac{\sqrt{3}}{2} + \frac{1}{2}i$, so we can choose $\lambda_1 = 2\sqrt{3}e^{\frac{\pi i}{6}} = 3 + \sqrt{3}i$ and $\lambda_2 = 2\sqrt{3}e^{-\frac{\pi i}{6}} = 3 - \sqrt{3}i$. This gives the characteristic polynomial $\lambda^2 - 6\lambda + 12 = 0$, so $\mathrm{Tr}(M_f) = 6$ and $\det(M_f) = 12$, so a suitable matrix is

$$M_f = \begin{pmatrix} 3 & 1 \\ -3 & 3 \end{pmatrix} \quad \text{so} \quad f(x) = \frac{3x+1}{-3x+3}$$

So we have that this f generates only cycles of lengths 1 and 6 for all primes p, with cycles of length 1 if and only if 3 is a QR mod p, or equivalently $p \equiv \pm 1 \pmod{12}$.

9 Real Numbers

In this section we'll be looking at FLTs over the field of real numbers.

9.1 Fixed Point iteration on FLTs

The fixed point iteration of some function f is given by

$$f(x_n) = x_{n+1}$$

for some initial value $x_0 \in \mathbb{R}$. Notice that this is analogous to cycles in finite fields. It came to the author in a dream that for continuous functions f, the sequence $x_0, x_1, ..., x_n$ converges to a fixed point given that $|f'(x_0)| < 1$.

By Lemma 5.21 a non-identity FLT can have at most 2 fixed points.

Lemma 8.1 *For any initial value $x_0 \in \mathbb{R}$, the fixed point interation of a FLT f converges to at most one fixed point.*

Proof.

Let $f(x)$ be a FLT.
The lemma is equivalent to proving that if there exist two fixed points, $|f'(x)| < 1$ at exactly one of the fixed points.

We further claim that the gradient at one fixed point is the reciprocal of the gradient at the other fixed point. This is enough to prove the lemma.

The asymptotes of any FLT are the lines $x = \frac{-d}{a}$ and $y = \frac{a}{c}$. Translating doesn't affect the gradient, and the translates of the fixed points are the fixed points of the translated function so without loss of generality, translate our FLT so that the asymptotes become $x = 0$ and $y = 0$ forcing $a = 0$ and $d = 0$. This means that $f(x)$ translates to $g(x) = \frac{b}{cx}$ which is the same as $g(x) = \frac{k}{x}$ where $k = \frac{b}{c}$. We also assume that k is positive without losing generality. If you take some point (x_0, y_0) on g, reflecting this across the line $y = -x$ maps it to $(-y_0, -x_0)$ which is also on $g(x)$ since $y = \frac{k}{x} \iff x = \frac{k}{y}$. Let (x_f, x_f) be a fixed point on g then reflecting across $y = -x$ gives us the second fixed point $(-x_f, -x_f)$. This means that g has a line of symmetry with gradient -1 meaning that any FLT has a line of symmetry with gradient -1 or 1 by reversing the translation and considering the generalisations made. Since a fixed point reflected under such a line of symmetry maps to the other fixed point, and also because the x and y coordinates swap, it must be true that the gradients of the fixed points are inverses.

□

9.2 Diophantine Approximation

Lemma 8.2 *Given a non-square natural number d, and $m \in \mathbb{R}_{>0} : m > \sqrt{d}$, we claim that for*

$$f(x) = \frac{mx + d}{x + m}$$

$$\forall x \in \mathbb{R}_{>0} : \lim_{n\to\infty} f^n(x) = \sqrt{d}$$

Proof.

$$f(x) = \frac{mx + d}{x + m} = m + \frac{d - m^2}{x + m}$$

$$\Rightarrow f'(x) = \frac{m^2 - d}{(x + m)^2}$$

$$\frac{m^2 - d}{(x + m)^2} = 1 - \frac{x^2 + 2mx + d^2}{x^2 + 2mx + m^2}$$

Note that $f'(x)$ is monotone decreasing in the positive reals, hence the maximum value of $f'(x)$ is precisely $f'(0)$ and $f'(0) = 1 - \frac{d^2}{m^2}$.
Hence for all $x \in \mathbb{R}_{>0}$ we have that $f'(x) < 1 - \frac{d^2}{m^2} < 1$. Hence

$$\frac{|f(x) - f(f(x))|}{|f(x) - x|} < 1 - \frac{d^2}{m^2} < 1$$

Then, define the sequence $(x_n)_0^\infty := f^n(x_0)$. Let $\epsilon = 1 - \frac{d^2}{m^2}$. Then it follows that

$$|x_n - x_{n+1}| < \epsilon \cdot |x_n - x_{n-1}|$$

$$\Rightarrow \epsilon \cdot |x_n - x_{n-1}| < \epsilon^2 \cdot |x_{n-1} - x_{n-2}|$$

Then by induction on n it follows that

$$|x_n - x_{n+1}| < \epsilon^n \cdot |x_1 - x_0|$$

$$\Rightarrow \lim_{n\to\infty} |x_n - x_{n+1}| = 0$$

Therefore the sequence (x_n) converges to some limit L in the positive reals.

$$\lim_{n\to\infty} f^n(x) = L \Rightarrow f(L) = L$$

$$\Rightarrow f(L) = \frac{mL + d}{L + m} = L$$

$$\Rightarrow L^2 = d \Rightarrow L = \sqrt{d}$$

As required. □

9.3 Numerical Examples

We consider the above proof to explore the structure of the convergents generated by the sequence $(a_n)_0^\infty$ compared the the continued fraction convergents.

$$\sqrt{2} = [1; \overline{2}]$$

		1	2	2	2	2	2	2
0	1	1	3	7	17	41	99	239
1	0	1	2	5	12	29	70	169

$$f_2(x) = \frac{2x+2}{x+2}$$

n	1	2	3	4	5	6	7
x	4	7	24	41	140	239	816
y	3	5	17	29	99	169	577

Note that there are common terms between the sequence (a_n) and the continued fraction convergents of $\sqrt{2}$. Namely

$$\frac{x_n}{y_n} = \frac{P_k}{Q_k}$$

for $(n,k) = (2,3),(4,5),(5,6),(6,7)$ for $n,k \leq 7$.

n	1	2	3	4	5	6	7
$\lvert\sqrt{2} - \frac{x_n}{y_n}\rvert$	0.0801	0.0142	0.0024	0.0004	0.000072	0.0000239	0.00000212
$\frac{1}{{y_n}^2}$	0.1111	0.0400	0.0034	0.0018	0.000102	0.0000350	0.00000300

Hence, we note that for $f_2(x)$ it appears that for all $n \in \mathbb{N}$

$$\mid \sqrt{2} - \frac{x_n}{y_n} \mid < \frac{1}{y_n^2}$$

$$\sqrt{3} = [1; \overline{1,2}]$$

		1	1	2	1	2	1	2	1
0	1	1	2	5	7	19	26	71	97
1	0	1	1	3	4	11	15	41	56

$$f_3(x) = \frac{3x+3}{x+3}$$

n	1	2	3	4	5	6	7
x	3	5	12	19	45	71	168
y	2	3	7	11	26	41	97

Note that for $(n,k) = (2,3),(4,5),(6,7)$ for $n,k \leq 7$.

$$\mid \sqrt{2} - \frac{x_n}{y_n} \mid < \frac{1}{y_n^2}$$

n	1	2	3	4	5	6	7
$\lvert\sqrt{3} - \frac{x_n}{y_n}\rvert$	0.232	0.0654	0.0177	0.00477	0.00072	0.0000239	0.00000212
$\frac{1}{{y_n}^2}$	0.250	0.1111	0.0204	0.00826	0.00148	0.0005940	0.00010000

Hence, we note that for $f_3(x)$ it appears that for all $n \in \mathbb{N}$

$$\mid \sqrt{3} - \frac{x_n}{y_n} \mid < \frac{1}{y_n^2}$$

Conjecture 8.1: For any non-square $d \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that for the FLT

$$f(x) = \frac{mx + d}{x + m}$$

and the corresponding sequence, for some z in the positive reals

$$(a_n)_0^\infty = \frac{x_n}{y_n} = f^n(z)$$

has the property that, for $\frac{P_n}{Q_n}$ being the nth continued fraction convergent to $\sqrt{d}$, there exist infinitely many solutions in $(n, k) \in \mathbb{N}$ such that

$$\frac{x_n}{y_n} = \frac{P_k}{Q_k}$$

Conjecture 8.2: For any non-square $d \in \mathbb{N}$, there exists $m \in \mathbb{N}$ such that for the FLT

$$f(x) = \frac{mx + d}{x + m}$$

and the corresponding sequence, for some z in the positive reals

$$(a_n)_0^\infty = \frac{x_n}{y_n} = f^n(z)$$

has the property that for all $n \in \mathbb{N}$

$$\mid \sqrt{d} - \frac{x_n}{y_n} \mid < \frac{1}{y_n^2}$$

Observation: Given $\alpha \in \mathbb{R}$ with continued fraction $[a_1, \overline{a_2, a_3, ..., a_{\ell+1}}]$ periodic with period ℓ, the FLT

$$f(x) = \frac{P_{\ell+1}x + (P_{\ell+1}(a_2 - a_1) + P_\ell)}{Q_{\ell+1}x + (Q_{\ell+1}(a_2 - a_1) + Q_\ell)}$$

has the property that it has a fixed point at α and for all $n \in \mathbb{Z}$,

$$f^n(a_1) = \frac{P_{n\ell+1}}{Q_{n\ell+1}}$$

10 Acknowledgements

We would like to express our gratitude to PROMYS Europe, Clay Mathematics Institute, The Mathematical Institute of the University of Oxford and Wadham College for giving us the opportunity to work on this exploration project. We especially want to thank our supervisor Diego Vurgait.