Drawing Limit Sets of Kleinian Groups Using Finite State Automata

Greg McShane, John R. Parker and Ian Redfern

CONTENTS

The theory of automatic groups (groups that can be treated by means of finite-state automata), and in particular the techniques of automatic coset enumeration developed by Redfern, are applied to the problem of computing limit sets of a large class of Kleinian groups. We discuss in detail the case of groups in the Maskit embedding of the Teichmüller space of the punctured torus.

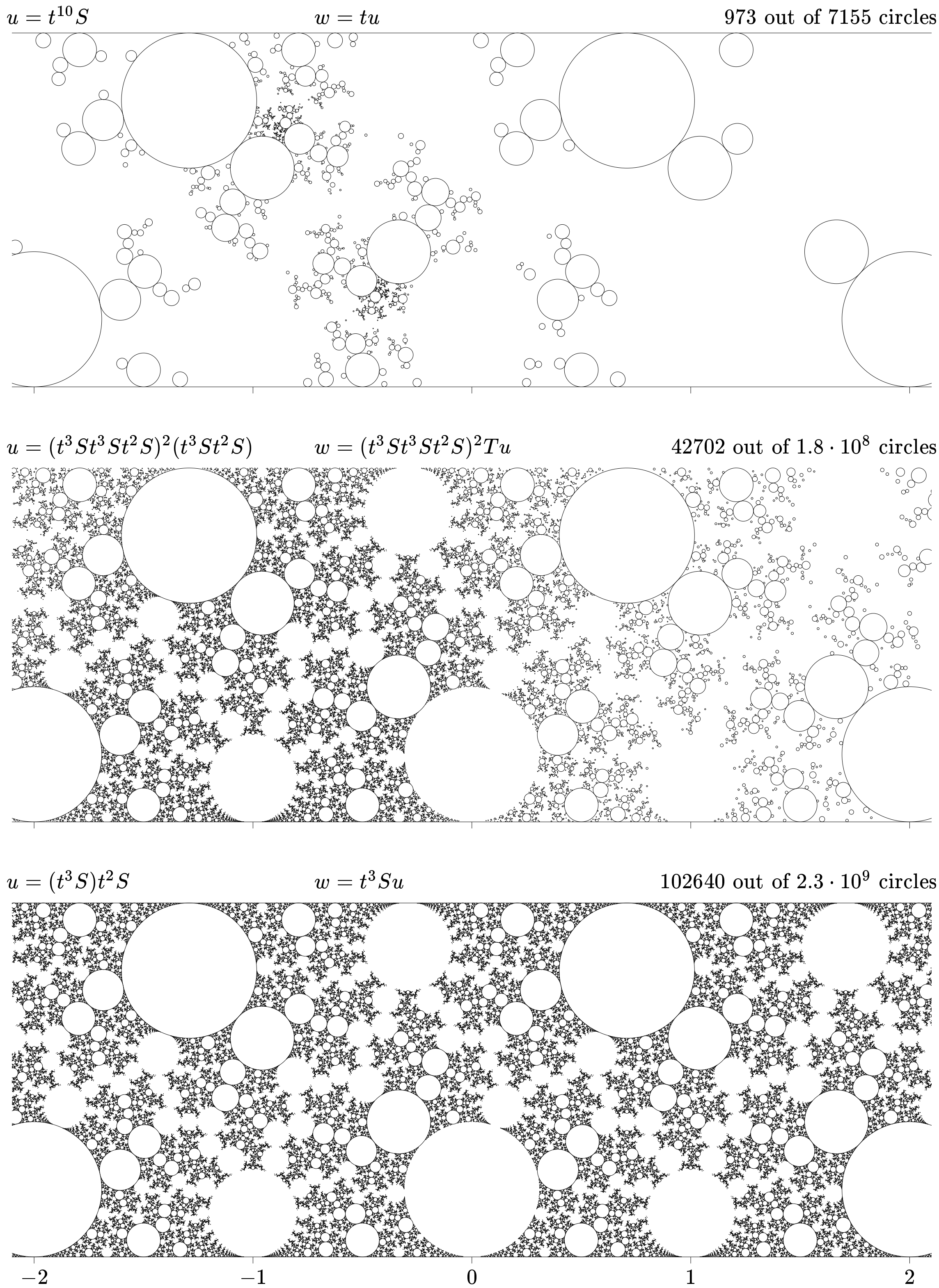
1. INTRODUCTION

Computer drawn pictures of limit sets of groups in the Maskit embedding of the Teichmüller space of the punctured torus led David Wright to make various conjectures about these groups and their parameter space [Wright 1987]. Many of these conjectures were proved in a paper [Keen and Series 1993] that began the theory of pleating coordinates for Teichmüller space.

In extending this theory to Teichmüller spaces other than the punctured torus, Keen, Parker and Series needed pictures of limit sets for other groups [Keen et al.]. Producing such pictures is an ideal application for the techniques developed by Redfern [1993] for enumerating coset systems using finite-state automata. There are several automata that can be used for this, each one depending on a choice of generators for the group. Making the right choice of generators is vital if detailed pictures are to be drawn efficiently. This is illustrated by Figure 1, which represents five attempts to plot the same limit set, with various degrees of success. The only difference among the first three plots (on page 154) is the choice of generating set used to derive the automaton.

McShane and Parker were supported by S.E.R.C. research fellowships. Redfern was supported by an S.E.R.C. research studentship.

1058-6458/96 $0.50 per page



$u = t^{10}S$
$w = tu$
973 out of 7155 circles
$u = (t^3St^3St^2S)^2(t^3St^2S)$
$w = (t^3St^3St^2S)^2Tu$
42702 out of $1.8 \cdot 10^8$ circles
$u = (t^3S)t^2S$
$w = t^3Su$
102640 out of $2.3 \cdot 10^9$ circles
-2
-1
0
1
2

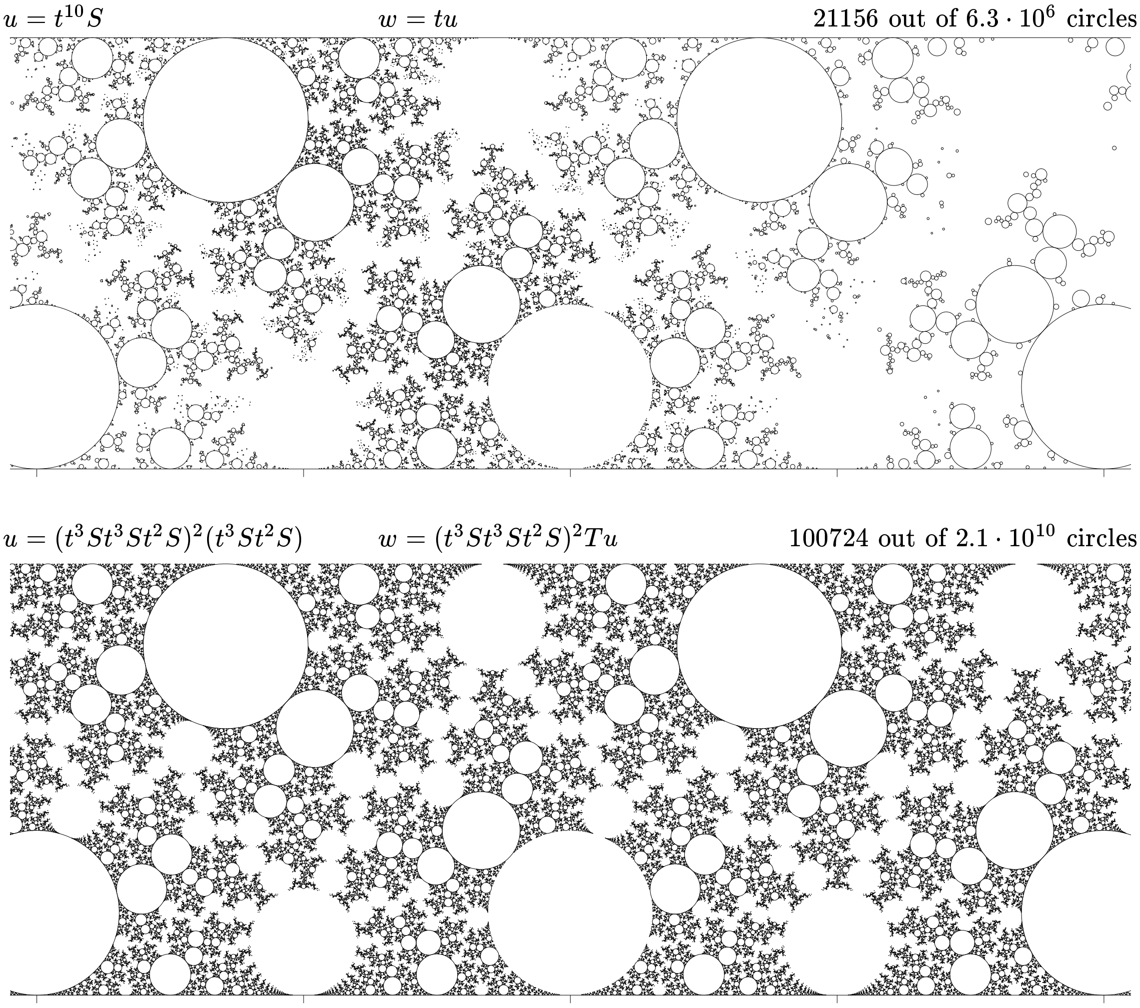


FIGURE 1. Five attempts to draw the limit set of the Kleinian group G_μ generated by the fractional linear transformations (2.1), with $\mu = 0.7082909515 + 1.617996654i$. The three plots on the facing page are obtained with cut-off radius $r_{\text{cut}} = 10^{-7}$ and visibility radius $r_{\text{vis}} = .0005$ (roughly speaking, we examine circles of radius greater than r_{cut}, and plot those of radius greater than r_{vis}; see Section 5). Each plot on this page has $r_{\text{cut}} = 10^{-10}$, and otherwise the same parameters as the plot to its left on the facing page. The decrease in r_{cut} yields an improvement in quality, but at a significant cost in running time. (The notation "N out of M circles" indicates that M circles of radius greater than r_{cut} were examined, and of those N had radius greater than r_{vis}. The running time is proportional to M.)

A more efficient way to get higher quality is to choose a better generating set $\{u, w\}$ (Section 6) for the finite-state automaton. Thus the plot immediately above and the one immediately to the left are of equal quality, but one took ten times longer than the other to run. The generating set for each example is given in terms of the transformations S and T of (2.1), with $t = T^{-1}$. See also the discussion on page 164.

All our limit sets will be invariant under the translation $S : z \mapsto z + 2$ and bounded by the "circles" $\operatorname{Im} z = 0$ and $\operatorname{Im} z = \operatorname{Im} \mu$, the extended real axis and its image under T.

The methods we describe can be used for any Kleinian group whose limit set is a circle packing or, more generally, the closure of the orbit of a circle under the group. Examples of such groups are the points in the Maskit embedding of the Teichmüller space of any Riemann surface of finite type [Maskit 1974], and the tetrahedron groups considered in [Bullett and Mantica 1992]. In practice the methods have been used for the Maskit embedding of the punctured torus, the twice punctured torus [Keen et al.], the four-punctured sphere, and the closed surface of genus two, as well as for a tetrahedron group. But full details have only been worked out in generality for the punctured torus, and it is on this case that we focus here.

The goal of this paper, then, is to explain how finite-state automata can be used to draw pictures of limit sets of groups in the Maskit embedding for the punctured torus, and what factors influence the quality of the output. We have implemented these ideas in a C program, which we used to generate all the plots of limit sets shown here. Using the program on a desktop Sun IPC, a useful plot can be generated in a minute, and higher-quality plots in about an hour, corresponding to several million circles being examined and several tens of thousand being plotted. On one of the University of Warwick's fast SparcServer 2000's, this time is reduced to a matter of minutes.

We begin with an outline of the Maskit embedding in Section 1. Sections 2 and 3 give background on the automata, much of which can be found in [Epstein et al. 1992] or [Redfern 1993]. Sections 4 and 5 are the real heart of the paper, and explain how to use the automata to draw limit sets in an efficient manner. Sections 6 and 7 are concerned with measuring this efficiency.

2. THE MASKIT EMBEDDING FOR THE PUNCTURED TORUS

Recall that Kleinian groups are discrete subgroups of $\mathrm{PSL}(2,\mathbb{C})$. A Kleinian group G acts on the Riemann sphere $\hat{\mathbb{C}}$ by Möbius transformations. The set of limit points of G-orbits on $\hat{\mathbb{C}}$ is called the *limit set* $\Lambda(G)$ of G, and is the smallest closed G-invariant subset of $\hat{\mathbb{C}}$. The complement of $\Lambda(G)$ is called the *ordinary set* of G and is denoted $\Omega(G)$. The action of G on $\Omega(G)$ is properly discontinuous and, when G is finitely generated, the quotient $\Omega(G)/G$ is a finite union of Riemann surfaces of finite type (*Ahlfors' finiteness theorem*).

For details on the material in the remainder of this section, see [Keen and Series 1993].

For $\mu \in \mathbb{C}$ in the upper half-plane ($\operatorname{Im}\mu > 0$), let G_μ be the group generated by

$$S = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad T = T_\mu = \begin{pmatrix} i\mu & i \\ i & 0 \end{pmatrix}. \tag{2.1}$$

Let $\mathcal{M}_{1,1}$ be the set of μ such that G_μ is a Kleinian group and $\Omega(G_\mu)/G_\mu$ is topologically the union of the punctured torus $\Sigma_{1,1}$ with the thrice punctured sphere $\Sigma_{0,3}$. The punctured torus is covered by a single connected component $\Omega_0(G_\mu)$ of $\Omega(G_\mu)$, invariant under the group, and the thrice punctured sphere is covered by infinitely many disks.

$\mathcal{M}_{1,1}$ gives a parametrization of the Teichmüller space of the punctured torus, that is, each complex structure on $\Sigma_{1,1}$ up to isotopy arises as the structure on the punctured torus component of $\Omega(G_\mu)/G_\mu$, for exactly one $\mu \in \mathcal{M}_{1,1}$. We call $\mathcal{M}_{1,1}$ (or its image under the correspondence $\mu \mapsto G_\mu$) the *Maskit embedding* of the Teichmüller space of the punctured torus.

Now fix $\mu \in \mathcal{M}_{1,1}$. One of the components of $\Omega(G_\mu)$ that cover the thrice punctured sphere is the lower half-plane $\mathbb{H}^* = \{z \in \mathbb{C} : \operatorname{Im} z < 0\}$, and $\Lambda(G_\mu)$ is the closure of all the G_μ-images of the extended real axis $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. The stabiliser of $\hat{\mathbb{R}}$ is the subgroup

$$H = \left\langle S = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},\ T^{-1}ST = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \right\rangle,$$

so precomposing by an element of H does not affect the image circle.

To obtain a good approximation to the limit set of G_μ we must draw all of these circles that are larger than the smallest dot our drawing method can produce. For efficiency's sake, we want to draw each circle only once and in the simplest way possible. We choose a set of generators for G_μ—not necessarily S and T (but the group itself remains the same). Every element of G_μ has an expression as a string in these generators. Our goal is to find a unique shortest string representing each coset of the stabilizer H of $\hat{\mathbb{R}}$ in G_μ; these strings, when applied to $\hat{\mathbb{R}}$, will give us the desired circles. We shall use a finite-state automaton to do this enumeration, as explained in the next three sections.

The construction applies to any Kleinian group G whose limit set is the closure of all G-orbits of a circle γ with stabiliser H. In the case of a group in the Maskit embedding of the punctured torus, $G = G_\mu$ is a free group on two generators and $\gamma = \hat{\mathbb{R}}$, as just discussed.

Before moving on, we stress a technical point of great practical importance. Traditionally, in hyperbolic geometry groups act on the left, so the cosets appear on the right; whereas in combinatorial group theory groups act on the right and so the cosets appear on the left. Using left and right actions inconsistently leads to disastrous results—circles are plotted not once but hundreds of times, and so on. We were deceived more than once because of this. So we now establish the convention that, throughout this paper, *groups act on the left.* In particular, strings should be read from right to left, and we will say "suffix" where a combinatorial group theorist would say "prefix".

3. REWRITE RULES

Let G be a group generated by the elements $g_1, \dots, g_n$. We assume this set of generators is closed under inversion, and we write G_i for g_i^{-1} (and likewise for other letters). Let w be a string over the alphabet $A = \{g_1, \dots, g_n\}$, that is, an element of the free monoid on $g_1, \dots, g_n$. The image of w in G, denoted $\bar{w}$, is defined in the obvious way: formally, it is characterized inductively by the conditions $\bar{\varepsilon} = 1$ (where ε is the empty string) and $\overline{vw} = \bar{v}\bar{w}$. We denote by $|w|$ the length of w, and by $w(t)$ the suffix of w of length t. We can think of w as a path from 1 to $\bar{w}$ in the Cayley graph of G. Shortest strings correspond to geodesic paths.

In the language of combinatorial group theory, our problem is: Given a group G with presentation $\langle g_1, \dots, g_n \mid R_1, \dots, R_m \rangle$ and a subgroup $H = \langle h_1, \dots, h_k \rangle$, find a finite-state automaton whose language contains, for each coset wH, precisely one string in $g_1, \dots, g_n$ that represents wH and is moreover as short as possible with this property. (For the definition of an automaton and its language, see Section 3.) If there is such an automaton, the coset system G/H is *automatically enumerable.*

To characterize a unique shortest string, we pick an order $g_1 < \cdots < g_n$ for the alphabet A, and use lexicographic ordering (from right to left!) among strings of the same length. Together with the ordering by length, this gives the so-called *ShortLex ordering* on the set of strings over A, making this monoid into a well-ordered set.

We will describe the coset system in terms of rewrite rules. A *rewrite rule* $l \Rightarrow r$ is a pair of equivalent strings (that is, $\bar{l} = \bar{r}$) such that $l > r$; any occurrence of l as a substring of a string can be replaced by r without changing the image in the group, and vice versa.

We always have the rewrite rules $g_j G_j \Rightarrow \varepsilon$ and $G_j g_j \Rightarrow \varepsilon$ for each generator, where ε is the empty string. If the group is not free, we also have the rule $R_j \Rightarrow \varepsilon$ for each relator.

We also need rules to describe the coset system: we introduce them by adding the new letter H to our alphabet. For each chosen generator h_j of H we have the rules $h_j H \Rightarrow H$ and $H_j H \Rightarrow H$.

We can get from any string of the form wH to its least equivalent by repeatedly substituting one side of a rewrite rule by the other. This is called *reducing* a string to an *irreducible* equivalent string. Unfortunately, to do this we may need to use rules backwards, that is, to make the string longer before it becomes shorter again. A *complete* system of rules is one in which rules need only be applied in the forward direction to reduce a string to its least equivalent. Since applying rules in the forward direction makes strings lesser under the well-ordering, having a complete system of rules means that any string can be reduced to its smallest equivalent in a finite number of steps.

A complete system of rules always exist (the trivial instance being the set of all possible reduction rules), but unless we can generate it in some manageable way this reduction algorithm is not of much use. Fortunately, we have the following result:

Proposition 3.1. *Let G be a free group of rank n, with generators $g_1, \dots, g_n, g_{n+1} = G_1, \dots, g_{2n} = G_n$. If H is a finitely generated subgroup, G/H has a finite complete rewrite rule system.*

Proof. Let H be generated by $h_1, \dots, h_k$ (also closed under inversion). We first show that H is *quasiconvex.* This means that if v is a shortest string in the generators of G whose image is in H, the image of any suffix of v is in a ρ-neighborhood of H in the Cayley graph of G, for some fixed ρ. In other words, walking along a geodesic toward a point of H never takes us far from H.

To show this, let w be a shortest string in the h_i with $\bar{w} = \bar{v}$, and let u be obtained from w by replacing each h_i with its expression in the generators of G. Because G is free, we can get v from u by cancellation of inverses (rewrite rules $g_j G_j \Rightarrow \varepsilon$ and $G_j g_j \Rightarrow \varepsilon$), so given a suffix of v there is a suffix of u with the same image in G. This image is therefore within distance ρ from H, where ρ is half the maximum of the lengths of the expressions of the h_i in terms of the g_i.

Let E be the set of rewrite rules whose left-hand side has length at most $\rho + 1$ (plus cancellation of inverses). This set can be constructed because we can solve the word problem in G. We shall show that any reducible string uH can be reduced using E, and thus that E is complete.

We can assume that u is reduced in G. Suppose $uH = vH$, with $v < u$ likewise reduced in G. Set $u = rp$ and $v = rq$, where r is the largest (possibly trivial) common prefix of u and v. Then $\bar{q}^{-1}\bar{p} \in H$, so $pH \Rightarrow qH$ is a rewrite rule that applies to reduce u (we have $q < p$ because deleting a common prefix preserves order). If $|p| \le \rho + 1$, this rule is in E and we are done. Otherwise, note that $q^{-1}p$ is irreducible in G, since G is free and we have removed the largest common prefix. Thus $q^{-1}p$ gives a geodesic to a point in H. Consider the suffix $p(\rho + 1)$. By the quasiconvexity of H, there is $h \in H$ and some string x of length at most ρ such that $\overline{p(\rho+1)} = \bar{x}h$. We deduce the rule $p(\rho + 1)H \Rightarrow xH$, which is again in E. □

Remark. This proof is a special case of a more general existence argument that shows that quasiconvex subgroups of hyperbolic groups are coset automatic [Epstein et al. 1992].

In practice, E can be much smaller than in the proof just given. Still assuming that G is a free group, it is enough to take one rule for each h_i, as follows: write h_i in the form $\bar{R}l$ without cancellation, with r, the "string inverse" to R, being less than l, and l being as short as possible (about half the length of h_i). Then add the rule $lH \Rightarrow rH$. So if $h_i = \overline{abc}$, the rule is $BAH \Rightarrow cH$.

The condition that the generating set be inverse-closed and free is not necessary to guarantee that G/H is coset automatic; it is needed for finite completeness, however.

4. MAKING AN AUTOMATON

Given the finite complete set of rewrite rules just constructed, we want to make a finite-state automaton that has as its language the set of irreducible strings. This will allow us to enumerate the unique ShortLex string representing each coset.

A *deterministic finite-state automaton* over an alphabet A is a finite directed graph whose edges, or *arrows*, are labelled with letters from A, so that at each node, or *state*, has exactly one outgoing arrow for each letter in A. Some nodes are special: there are zero or more *accept states*, and exactly one *start state.*

A string is *accepted* by such an automaton if, by starting at the start state, and following the path described by the string, we finally arrive at an accept state. The set of accepted strings is the *language* of the automaton.

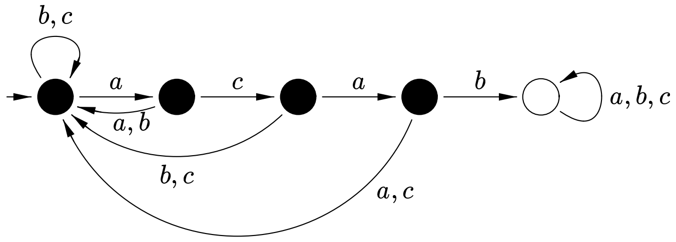


FIGURE 2. An automaton that rejects exactly those strings over the alphabet $\{a, b, c\}$ that contain the substring *acab*. The start state is at the left, and the accept states are shown in black. Arrows with same source and destination have been combined.

Now suppose we have a set of k rules of the form $l_j \Rightarrow r_j$ and want to construct an automaton to decide whether a string is reducible or not. For each l_j, it is easy to construct an automaton M_j that rejects exactly those strings that have l_j as a substring: see Figure 2. Then we take the *cartesian product* M of the automata M_j, as follows: the states of M are k-tuples of states of the M_i, and the arrows act independently on each entry of the k-tuple. The accept states of M are k-tuples of accept states of the M_j. Clearly, a string is rejected by M if and only if it is rejected by one of the M_j.

Remark. This construction is a particular case of a well-known procedure due to Kleene, Rabin, and Scott to find automata for so-called regular languages. See, for example, [Epstein et al. 1992, Ch. 1].

An important feature of the language of our automaton is that it is *suffix-closed*, that is, if a string is accepted by the automaton, so is any suffix of it (remember that strings are read right-to-left). If a string is rejected, any extension of it is also rejected. We can in effect remove the reject states from the automaton and the arrows that lead to it, getting what is called a *partial automaton*. A string is accepted if we can follow the path it prescribes without getting stuck at a state lacking an outgoing arrow for the next letter.

The discussion above has focused on the case when G is a free group. An automaton to recognize least strings in the ShortLex order can also be constructed when G is the fundamental group of any closed surface of finite type. In this case the coset system does not have a finite complete rewrite rule set, so alternative methods have to be used, which involve generating a large sample of the rule set and deducing the automaton behind it. This algorithm is contained in [Redfern 1993] and is too long to describe here. Background can be found in [Epstein et al. 1992, Chs. 2, 5 and 6], for example.

5. USING THE AUTOMATON

Having an automaton, we can list all the strings it accepts, that is, all possible paths from the start state: conceptually, this amounts to untying all the loops in the automaton to produce a (generally infinite) tree, as in Figure 3. (We're thinking in terms of the partial automaton: see two paragraphs above.)

Of the two basic methods for traversing a tree, depth-first and breadth-first, we choose the first because it is easy to program and requires very little storage space. Essentially all that we have to keep track of is where we are in the tree, and we store that information in the form of a "current string". Beginning at the start state with the empty string, we follow the arrow with the least label (a in Figure 3), and append that label to our current string. From the new state we again follow the arrow with the least label, and write down that label to the left of the first in the current string (now aa in the figure). We keep going until we hit a state from which no arrows issue: at that point we must backtrack. We return to the previous state and take the next greater arrow out of it, replacing the last label written by the label of the new arrow (so we get the string ab in the figure). If we run out of arrows going out of a state, we must backtrack another level, and so on.

There is one problem with this: if there are loops in the automaton—and there must be, unless the accepted language is finite—we start going ever deeper, writing ever longer strings, along the branch corresponding to the first loop we encounter. Later branches, such as the branch starting with b in Figure 3, are never reached. Obviously, we need a cut-off criterion. One solution is to stop going along a branch when we reach a certain depth. But our purpose is really to draw the circles associated with the strings we're generating—those circles that fall within the page and are bigger than some *visibility radius* r_{vis}. If we go down to the same depth on every branch, we typically generate enormous numbers of uninteresting circles for the sake of a few interesting ones. What we need are geometric criteria for cut-off.

We can't use an out-of-bounds cut-off because, no matter how far out, a circle might still map within bounds if we were to persevere along that branch.

What we can do is pick a *cut-off radius* r_{cut} that is much smaller than the visibility radius, and stop following a branch when the circle gets smaller than the cut-off radius, on the assumption that no circle further down along that branch will be visible. This assumption is false, but in Section 6 we show how to overcome the problems it introduces. We also do impose a depth cut-off, but make it very large.

To summarize, then, we use a depth-first search with two cut-off criteria: depth and circle radius. The cut-off radius is significantly smaller than the radius of the smallest circle we are interested in. As already indicated, this method is economical in terms of storage space; there is no need to store a list of circles for the next pass, as there would be with a breadth-first search. Overhead is minimal, and virtually all the time is spent in multiplying matrices and trying matrices on points to determine the image circle. Since there are infinitely many circles in $\Lambda(G_\mu)$ and our procedure only plots a finite number, it will miss almost all circles—in fact it will usually omit some visible circles, as just explained. It will nevertheless give an acceptable picture provided the presentation for the group is chosen correctly, as explained in the next section.

6. GEOMETRICAL CHOICE OF GENERATORS

We stress that the true picture, that is, the limit set $\Lambda(G_\mu)$, does not change when we alter the presentation of G_μ. But a change in presentation does affects the order in which coset representatives are enumerated, and therefore the part of the circle distribution that the depth-first search shows us. We now turn to the question of how to change the presentation advantageously.

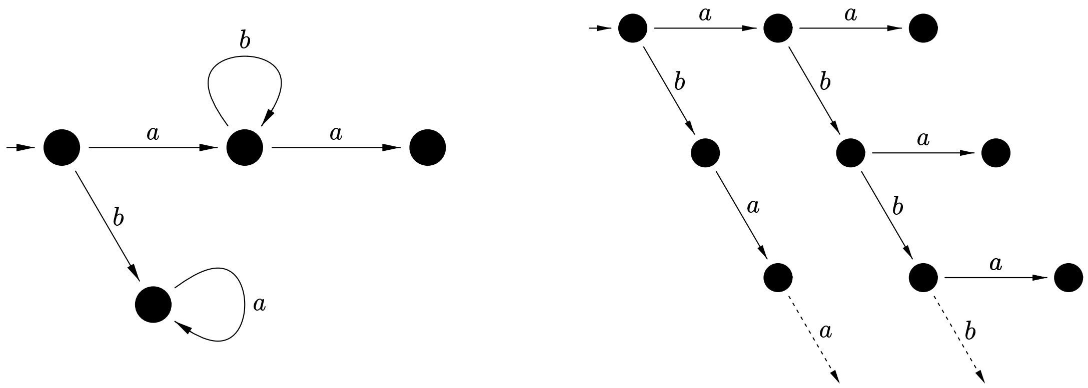


FIGURE 3. The tree of accepted strings implied by an automaton. A depth-first search yields the strings ε, a, aa, ab, aba, abb, $abba$, ... A cut-off criterion must be imposed if all branches are to be visited.

We are free to choose any set of generators for G_μ, in any order (and the order generally matters). To take advantage of the especially simple situation covered by Proposition 3.1, however, we will work with generator sets of four elements—two pairs of mutually inverse transformations. Within these restrictions, the idea is to try to ensure that elements of the group that have undesirable properties are not represented by short strings in the resulting language of least representatives in the ShortLex order.

Since the limit set is invariant under the translation $S : z \mapsto z+2$ of (2.1), we can concentrate on a fundamental domain for S, say the strip $-1 < \operatorname{Re} z < 1$. We want circles that meet this strip to have low depth (correspond to short strings) in our enumeration. Similarly, we want circles with large radii to have low depth.

(Reaching the circles of interest at low depth is important not only to save time but also to reduce the number of arithmetic operations required. Each time we generate a new node in the tree we multiply the matrix held in the parent node by some generator; since we are using floating point arithmetic, this introduces a rounding error. The cumulative error over hundreds of multiplications, say, could result in visible inaccuracies.)

If S can be expressed as a short string in the chosen generators, this expression is likely to occur as a suffix of many strings in the (truncation of the) language of least representatives, and circles lying outside the fundamental domain of interest will appear frequently. Thus one strategy is to avoid generating sets in terms of which S has a short expression.

Another goal is to try to avoid *clumping*, a phenomenon illustrated by the following example. (In the next section we will see how to measure it quantitatively.) Let A and B be a pair of Möbius transformations with A loxodromic with a long translational component. Let L_n be the set of strings in A and B having length at most n. If n is small and z is a point of the complex plane not fixed point by A, the translates $\{\gamma z \mid \gamma \in L_n\}$ will tend to cluster round the attracting fixed point of A. To see this, consider the proportion of strings in L_n that have a run of (say) $\frac{1}{2}n$ or more A's as a prefix. There are around $2^{n/2}$ such strings, out of a total of $2^{n+1}-1$. For large n this ratio rapidly approaches zero, but for n small it is significant. Thus, for small n, a significant proportion of the γz for $\gamma \in L_n$ will be found in a neighbourhood of the attracting fixed point of A (see the first plot on page 155).

We conclude that to reduce clumping we should not choose loxodromic generators with long translational parts. To see how to do this in practice, we must explore in more detail the structure of the groups G_μ.

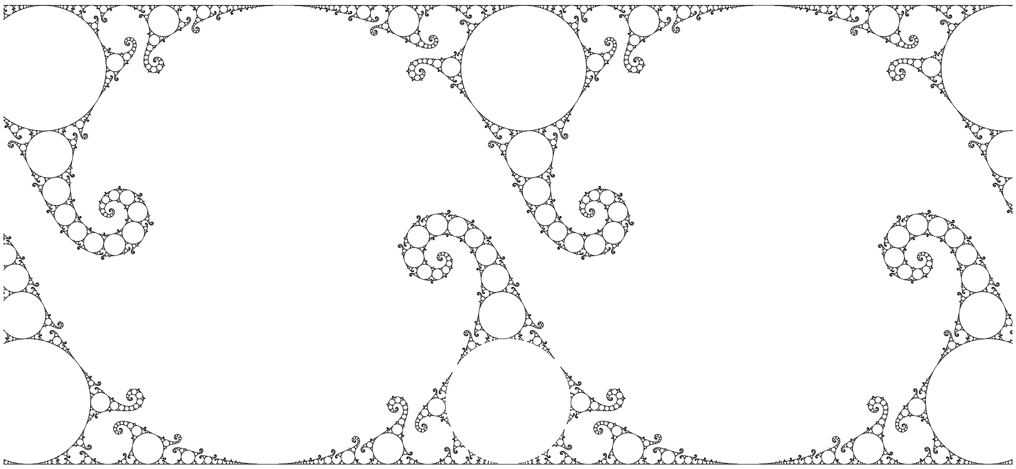
Every element of G_μ represents a homotopy class of closed curves on $\Omega(G_\mu)/G_\mu$. It is well known that the homotopy classes of *simple* closed curves on $\Sigma_{1,1}$ are naturally parametrised by $\mathbb{Q} \cup \{\infty\}$. The parametrisation, associating to each p/q (in lowest terms, with $q \geq 0$) an element $W_{p/q} \in G_\mu$, is easy to describe. First, $W_\infty = W_{1/0} = S^{-1}$, $W_0 = W_{0/1} = T$ and $W_1 = W_{1/1} = S^{-1}T$. Then, given any rational number in lowest terms, write it as $(p+r)/(q+s)$, where p, q, r, s are integers with $ps - qr = 1$; the rationals p/q and r/s are called *Farey neighbours* and $(p+r)/(q+s)$ is their *Farey sum.* We then have

$$W_{(p+r)/(q+s)} = W_{p/q} W_{r/s}.$$

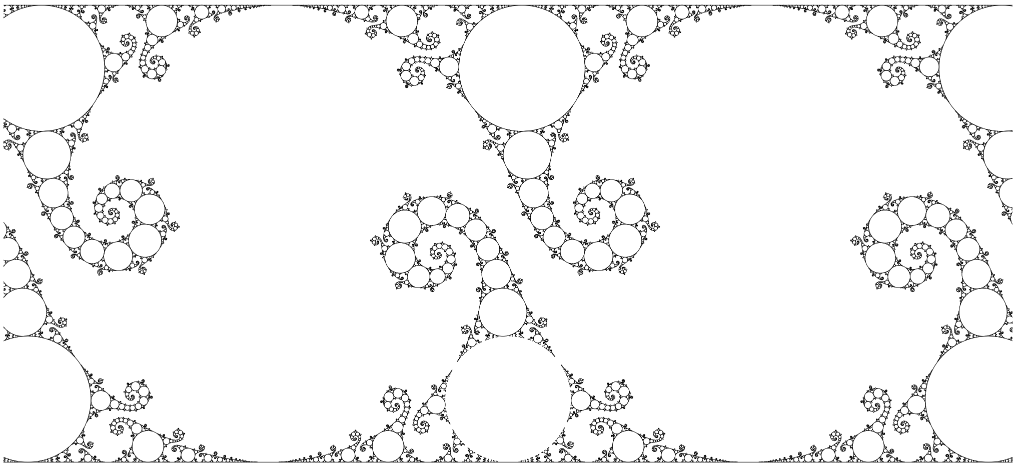
For example, $W_{1/n} = S^{-1}T^n$ for all positive integers n.

Generically, $W_{p/q}$ is a loxodromic transformation, but by varying μ we can make its trace tend to two, and so, in the limit, become parabolic (the exception is $p/q = 1/0$, since $S = W_{1/0}$ is already parabolic). Geometrically, this corresponds to making the length of the geodesic in the given homotopy class tend to zero—in the limit, the two sides of the neck around the geodesic become two punctured disks. This process is called *pinching* [Maskit 1970], and the group produced in this way, which lies on the boundary of the Maskit embedding $\mathcal{M}_{1,1}$, is called a *cusp group.* The value of

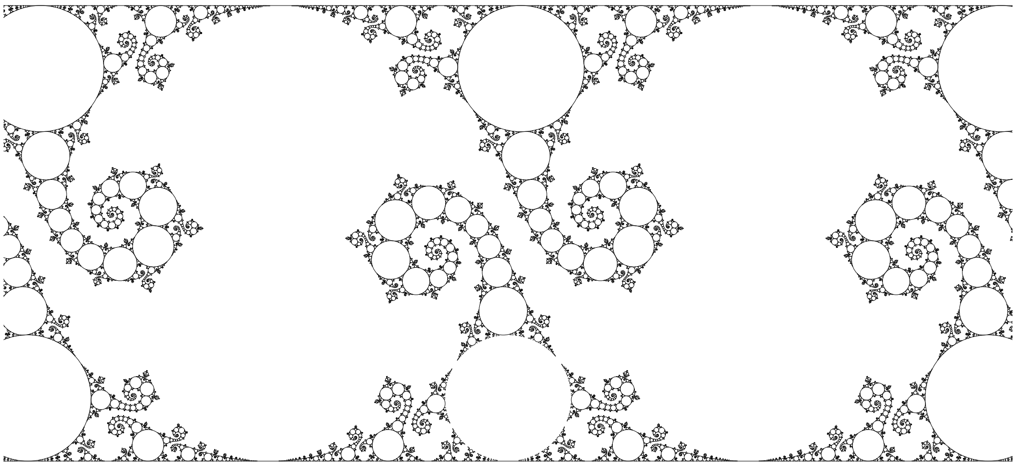
$\mu = 0.05744 + 1.904i$ 21689 out of $2.4 \cdot 10^7$ circles



$\mu = 0.06469 + 1.912i$ 14707 out of $1.0 \cdot 10^7$ circles



$\mu = 0.0533 + 1.9i$ 29018 out of $4.7 \cdot 10^7$ circles



$\mu = \mu_{1/9} = 0.0487530129 + 1.896407251i$ 48598 out of $1.3 \cdot 10^8$ circles

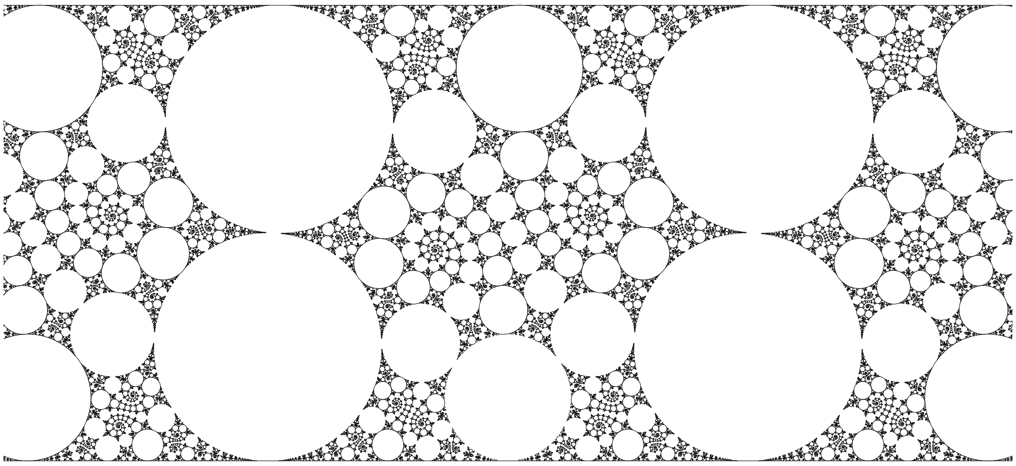


FIGURE 4. Sequence of limit sets as G_μ approaches the cusp group for $p/q = 1/9$ along the pleating ray [Keen and Series 1993]. Notice the complementary family of circles in the case of the cusp group (last panel). The following parameters are common to all plots: generators $u = t^8 S$, $v = tu$; radii $r_{\text{vis}} = 0.0005$, $r_{\text{cut}} = 10^{-7}$. Note the steady increase in the number of circles examined (and plotted) as we approach the cusp group.

μ for which $W_{p/q}$ is pinched is denoted $\mu_{p/q}$. See Figure 4.

For a cusp group, the ordinary set $\Omega(G_\mu)$ no longer has a group-invariant simply connected component covering $\Sigma_{1,1}$; this component becomes a infinite collection of discs on which the group acts transitively. Let Δ_0 be one of these disks, and δ_0 its boundary [Keen and Series 1993]. The images of δ_0 under the group are part the limit set: $\Lambda(G_\mu)$ has become a circle packing. Because of this, better pictures of the limit set are obtained by drawing not only those circles that are images of $\hat{\mathbb{R}}$, but also those that are images of δ_0.

The condition $\operatorname{tr}(W_{p/q}) = 2$ gives a polynomial equation in μ of degree q. The particular root $\mu_{p/q}$ that leads to pinching of the curve associated with $W_{p/q}$ is characterised as follows [Keen and Series 1993]. Take the locus of μ values for which $\operatorname{tr}(W_{p/q})$ is real, and remove the set of points where $\operatorname{tr}(W_{p/q})$ equals 2. There is a unique component of this set that is asymptotic to the line $\operatorname{Re}\mu = p/q$. The finite endpoint of this component is $\mu_{p/q}$. In practice it is often easier to search for the correct root by trial and error: taking the root with the largest imaginary part usually suffices.

Remark. Cusp groups are dense in the boundary of $\mathcal{M}_{1,1}$ (compare [McMullen 1991]), and it is conjectured that $\partial\mathcal{M}_{1,1}$ is a Jordan curve.

Let's return to the choice of generators. It is clear from the preceding discussion that, when we are at or close to the cusp $\mu_{p/q}$, the element $W_{p/q}$ is a loxodromic with small translation length, or a parabolic. We therefore choose $W = W_{p/q}$ as one of the generators, and must choose a complementary generator, hopefully with the same good properties.

At the cusp there is always a complementary generator U whose trace is $-i\mu_{p'/q}$ (see Proposition 7.1 below). If $-1 < p'/q < 1$, this is close to two, and U has a small translation length. For other values of p'/q, we compose U with powers of $W_{p/q}$ to get one with short translation length.

We found that these choices are not good enough when q is large. Because the generators have rather small isometric circles in this case, the detail is still concentrated around the fixed points, although the attraction is slower than when we use a loxodromic generator with large translational part. This can be avoided by using $W_{r/s}$ and a complementary generator, where s is around 10 and $|ps - qr|$ is as small as possible. For example, r/s can be a truncation of the continued fraction expansion of p/q.

The three first plots in Figure 1 (page 154) illustrate this phenomenon. The value of μ for that figure is equals $\mu_{13/34}$. The middle picture is obtained using $W_{13/34}$ and a complementary generator; it has good definition around δ_0 (the circle tangent to the real axis at -1), but quality drops noticeably for $\operatorname{Re} z > 0$. The much better picture at the bottom is obtained using the automaton corresponding to $W_{3/8}$ and a complementary generator. The top figure, using $W_{1/11}$, is characterized by clumping around two small regions. The remaining two plots in Figure 1 show how quality may be improved by decreasing the cut-off radius, but at a high cost in running time.

A similar situation arises in Figure 5, which displays the limit set of G_μ for $\mu = 10/109$. At the top, we use $W_{10/109}$ and a complementary generator, and the result is poor. In the middle, we decrease the visibility radius; all that happens here is that the area with most detail becomes very black and the detail is obscured, while the areas where there were few circles before remains unchanged. By contrast, in the bottom plot we change the generators; this gives a much better picture (but the running time increases substantially).

7. THE DUALITY OF CUSPS

The next section will describe a method for assessing how good an approximation to the limit set we obtain with the algorithm above. The method applies equally well to groups in the interior of the Maskit embedding and to cusp groups on its boundary. For the latter we need some auxiliary results, which we develop in this section.

As we have seen, for a cusp group $G = G_{\mu_{p/q}}$ the set $\Omega(G)$ consists of two families of discs, G acting transitively on each. The quotient of each family is a thrice punctured sphere, one puncture on each corresponding to $K = [T^{-1}, S^{-1}]$ and the other pair corresponding to S on one and $W_{p/q}$ on the other. We now show that there is an involution on the set of cusp groups that swaps these two families of discs.

Proposition 7.1. *For every coprime $p, q \in \mathbb{Z}$ there is a Möbius involution P and an integer p' coprime to q such that P conjugates $G = G_{\mu_{p/q}}$ to $G' = G_{\mu_{p'/q}}$, and sends each family of discs in $\Omega(G)$ to the other family in $\Omega(G')$.*

Proof. Let $T = T_{\mu_{p/q}}$. The group element $W_{p/q}$ for G is parabolic, and there is $U \in G$ such that $G = \langle U, W_{p/q}\rangle$ and that

$$K = [T^{-1}, S^{-1}] = [W_{p/q}^{-1}, U^{-1}]$$

(see [Keen and Series 1993]; in fact we may take $U = W_{r/s}^{-1}$, where $ps - qr = -1$). Because $G = \langle U, W_{p/q}\rangle$ and $W_{p/q}$ is parabolic, we can find $P \in \mathrm{PSL}(2, \mathbb{C})$ so that $PW_{p/q}P^{-1} = S$ and $PUP^{-1} = T_{\mu'}$ for some $\mu' \in \mathbb{C}$. In other words, conjugation by P renormalises G so that $W_{p/q}$ becomes the distinguished parabolic generator. Furthermore,

$$\begin{aligned} PKP^{-1} &= PW_{p/q}^{-1}U^{-1}W_{p/q}UP^{-1} \\ &= S^{-1}T_{\mu'}^{-1}ST_{\mu'} = K^{-1}. \end{aligned}$$

Since P conjugates the parabolic element K to its inverse, it has order two.

Now write S in terms of $W_{p/q}$ and U. Replacing $W_{p/q}$ by S and U by $T' = T_{\mu'}$ in this expression gives PSP^{-1}, which is parabolic. Thus $G' = \langle S, T'\rangle$ is a cusp group, that is, $\mu' = \mu_{p'/q'}$ and $PSP^{-1} = W_{p'/q'}$ for $G_{\mu'}$. We now show that $q = q'$. Observe that q is the intersection number of the curves represented by S and $W_{p/q}$, and this is the same as the intersection number q' of the curves represented by $PSP^{-1} = W_{p'/q'}$ and $PW_{p/q}P^{-1} = S$. It follows also that p' and q are coprime.

Next, P must fix -1 and send ∞ to the fixed point of $W_{p/q}$. This means that P sends $\hat{\mathbb{R}}$ to the circle through that fixed point and tangent to $\mathbb{R}$ at -1. This is the basic circle δ_0 for the other family [Keen and Series 1993]. The boundaries of the second family of discs in $\Omega(G_{p/q})$ are given by

$$\bigcup_{g\in G} g(\delta_0) = \bigcup_{g\in G} g(P(\hat{\mathbb{R}})) = P\Big(\bigcup_{g\in G'} g(\hat{\mathbb{R}})\Big).$$

These are the circles bounding the first family of discs in $\Omega(G')$. As P swaps the two families of circles, it also swaps the two families of discs. □

Remark. From this proof we see that

$$W_{p/q}(\mu_{p/q}) = P^{-1}SP = PSP^{-1} = W_{p'/q}(\mu_{p'/q}).$$

We can explicitly obtain p' from p and q. First note that p' is only determined up to congruence mod q, since adding an integer to p/q does not change $G_{p/q}$, it merely changes the generators. Now consider the curves represented by S and $W_{p/q}$ on a punctured torus; call them $\gamma_{1/0}$ and $\gamma_{p/q}$. On another punctured torus consider the curves $\gamma_{p'/q}$ and $\gamma_{1/0}$ represented by $W_{p'/q}$ and S. There is an (orientation-reversing) element of the mapping class group of the punctured torus sending $\gamma_{1/0}$ to $\gamma_{p'/q}$ and $\gamma_{p/q}$ to $\gamma_{1/0}$. Composing with the orientation-reversing element of the mapping class group sending γ_x to γ_{-x} for all $x \in \mathbb{Q}\cup\{\infty\}$, we get an element of $\mathrm{PSL}(2,\mathbb{Z})$, the orientation-preserving mapping class group of the punctured torus, sending $\gamma_{1/0}$ to $\gamma_{p'/q}$ and $\gamma_{-p/q}$ to $\gamma_{1/0}$. Such an element must be of the form $\left(\begin{smallmatrix} p' & * \\ q & p \end{smallmatrix}\right)$. This implies $pp' = 1$ modulo q, and determines p'.

Example 7.2. If $p = 1$, we can choose $p' = 1$. Thus $PG_{1/q}P^{-1} = G_{1/q}$. We now show this explicitly. We have already seen that $W_{1/q} = W = S^{-1}T^q$. We may take $U = S^{-1}TS$ as the other generator. Now $W^{-1}U^q = S$ and $W^{-1}UW = T$. As above, we can find a map P satisfying $PWP^{-1} = S$ and $PUP^{-1} = T'$. Then $S^{-1}T'^q = PW^{-1}U^qP^{-1} = PSP^{-1}$, which is parabolic. Since there is a unique value of μ for which the group is a cusp group with $S^{-1}T'^q$ parabolic, we see that $\mu' = \mu_{1/q}$ and $PUP^{-1} = T$, as required.

Example 7.3. If $p = 2$ and $q = 7$, we get $p' = 4$. This situation is illustrated in Figures 6 and 7. Although the pictures of the limit sets for $G_{2/7}$ and $G_{4/7}$ look quite different (in particular, $\mu_{2/7}$ and $\mu_{4/7}$ have different imaginary parts), they are related by a Möbius transformation that swaps the two families of circles (the family arising from δ_0 is in each case not drawn explicitly, but only implied by the accumulation of circles of the other family). The circle δ_0 and the fixed point of $W_{p/q}$ are the same in both cases.

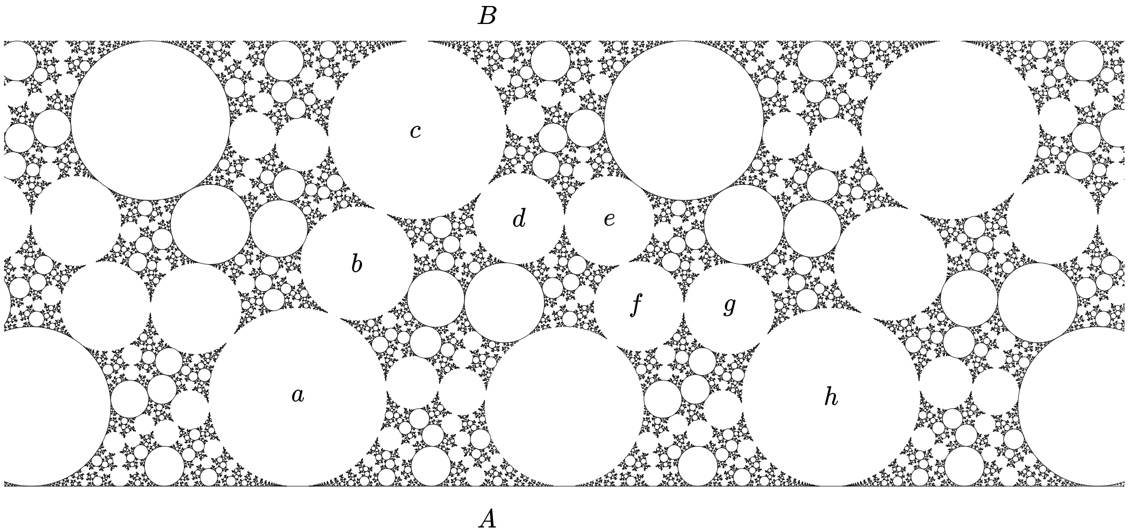


FIGURE 6. The limit set of G_μ for $\mu = \mu_{2/7} \approx 0.448059 + 1.6708i$. This cusp group is conjugate to that of Figure 7; corresponding disks are marked with the same letters.

8. CALCULATING THE CIRCLE PACKING EXPONENT

We now state a performance test for the algorithm of this paper. Let G_μ be a group in the Maskit embedding. Our primary performance criterion is how successfully we are able to enumerate large circles, and to do this we first estimate the circle packing exponent of $\Lambda(G_\mu)$ (defined below). We begin by showing that for any $\mu \in \mathcal{M}_{1,1}$ the circle packing exponent of $\Lambda(G_\mu)$ equals the Hausdorff dimension of $\Lambda(G_\mu)$ (see [Parker] for a more general version of this result).

Recall that we defined $\Omega_0(G_\mu)$, for $\mu \in \mathcal{M}_{1,1}$, as the unique G_μ-invariant component of $\Omega(G_\mu)$. The only other $\langle S\rangle$-invariant components of $\Omega(G_\mu)$ are the lower half-plane $\mathbb{H}^*$ and its image under T, the half-plane $\operatorname{Im} z > \operatorname{Im}\mu$.

Let $\Omega_1(G_\mu) = \Omega_0(G_\mu) \cup \mathbb{H}^* \cup T(\mathbb{H}^*)$ be the union of the three $\langle S\rangle$-invariant components, and choose a fundamental domain E for $\langle S\rangle$ such that $\partial E \subset \Lambda(G_\mu) \cup \Omega_1(G_\mu)$. Also, define

$$K = (\hat{\mathbb{C}} - \Omega_0(G_\mu)) \cap \bar{E};$$

this is a compact subset of $\mathbb{C}$. The intersection $\Omega(G_\mu) \cap E$ is a collection of disjoint open discs, one from each $\langle S\rangle$ orbit of components of $\Omega(G_\mu) - \Omega_1(G_\mu)$. Each of these discs is the image of $\mathbb{H}^*$ under an element of G_μ. Removing these discs gives a circle packing Π of K.

Let r denote the radius of an arbitrary circle in Π. The circle packing exponent of Π is defined as [Boyd 1973; Bullett and Mantica 1992]

$$e = \sup\Big(t : \sum_{\text{circles in } \Pi} r^t = \infty\Big) = \inf\Big(t : \sum_{\text{circles in } \Pi} r^t < \infty\Big).$$

It is clear that $e \le 2$, because for $t = 2$ the sum is $\operatorname{Area}(K)/\pi$.

The radius of the image of $\hat{\mathbb{R}}$ under the map $g(z) = (az + b)/(cz + d)$, where $ad - bc = 1$, is $|c\bar{d} - d\bar{c}|^{-1}$. In order to count each circle in Π exactly once we need to sum over the cosets $\langle S\rangle\backslash G_\mu/H$ excluding the coset of the identity I and the coset of T (since these give the lines $\hat{\mathbb{R}}$ and $T(\hat{\mathbb{R}})$). We index these cosets by considering the points $g^{-1}(\infty)$ in a given fundamental domain D for the action of H on $\Omega(H) = \hat{\mathbb{C}} - \hat{\mathbb{R}}$ [Parker]. Thus $\sum_{\text{circles in } \Pi} r^t$ equals

$$\sum_{\langle S\rangle\backslash G_\mu/H - \{I,T\}} |c\bar{d} - d\bar{c}|^{-t} = \sum_{g^{-1}(\infty)\in D} |c\bar{d} - d\bar{c}|^{-t}.$$

The fundamental domain we choose is

$$D = \{z : -1 < \operatorname{Re} z \le 1,\ |2z-1| > 1,\ |2z+1| \ge 1\}.$$

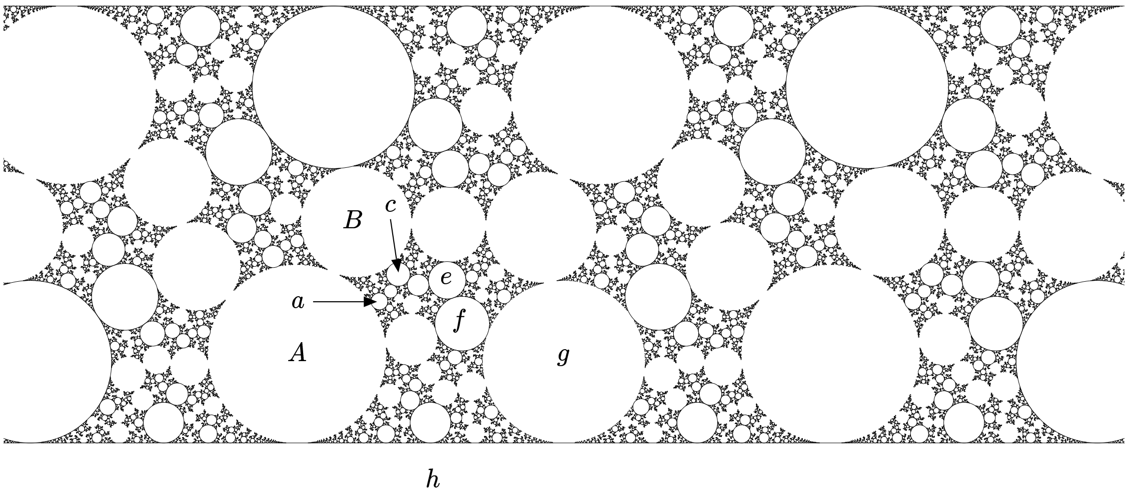


FIGURE 7. The limit set of the cusp group G_μ for $\mu = \mu_{4/7} \approx 1.13661 + 1.63996i$. Compare Figure 6, and see also Example 7.3.

Furthermore, each point $g^{-1}(\infty)$ must be in $\Lambda(G_\mu)$ and so does not lie in $\mathbb{H}^*$ nor in any of the discs Δ_0, $S(\Delta_0)$ or $T^{-1}(\mathbb{H}^*)$ (recall that Δ_0 is the disc in $\Omega(G_\mu)$ with boundary δ_0). These discs are tangent to $\mathbb{R}$ at the points -1, $+1$ and 0 respectively. Thus there exists a positive constant ε so that $\operatorname{Im} g^{-1}(\infty) = \operatorname{Im}(-d/c) \geq \varepsilon$ for $g^{-1}(\infty) \in D$. So, for $t \leq 2$, we have

$$\begin{aligned}\sum_{g^{-1}(\infty)\in D} |c\bar{d} - d\bar{c}|^{-t} &\leq (2\varepsilon)^{-t} \sum_{g^{-1}(\infty)\in D} |c|^{-2t} \\ &\leq (2\varepsilon)^{-2} \sum_{g^{-1}(\infty)\in E'} |c|^{-2t},\end{aligned}$$

where E' is the strip $\{z : -1 < \operatorname{Re} z \leq 1\}$, the last inequality being a consequence of $D \subset E'$. Now E' is a fundamental domain for $\langle S\rangle$, so the points $g^{-1}(\infty)$ in E' index the cosets $\langle S\rangle\backslash G_\mu/\langle S\rangle - \{I\}$. Thus we obtain

$$\sum_{g^{-1}(\infty)\in E'} |c|^{-2t} = \sum_{\langle S\rangle\backslash G_\mu/\langle S\rangle-\{I\}} |c|^{-2t}.$$

It follows that

$$\begin{aligned}e &\leq \sup\Bigl(t : \sum_{\langle S\rangle\backslash G_\mu/\langle S\rangle-\{I\}} |c|^{-2t} = \infty\Bigr) \\ &= \inf\Bigl(t : \sum_{\langle S\rangle\backslash G_\mu/\langle S\rangle-\{I\}} |c|^{-2t} < \infty\Bigr),\end{aligned}$$

which is just the exponent of convergence of the Poincaré series [Beardon 1968]. Since G_μ is geometrically finite, this is the Hausdorff dimension $d = d(\mu)$ of Π [Nicholls 1989, Theorem 9.3.6]. In other words, $e \leq d$. But it is known that $e \geq d$ for all sphere packings of compact subsets of $\mathbb{R}^n$ [Larman 1966]. It follows that $e = d$.

This proof also works for cusp groups $\mu = \mu_{p/q}$ in the boundary of $\mathcal{M}_{1,1}$, because they are also geometrically finite [Keen et al. 1993]. For a cusp group the component $\Omega_0(G_{\mu_{p/q}})$ has itself degenerated to a circle packing. By Proposition 7.1, there is an involution $P \in \mathrm{PSL}(2,\mathbb{C})$ and p' such that the second family of circles in $\Lambda(G_{\mu_{p/q}})$ become the first family of circles in $\Lambda(G_{\mu_{p'/q}})$ and vice versa. Because the Hausdorff dimension is a Möbius invariant, we have

$$e(\mu_{p/q}) = d(\mu_{p/q}) = d(\mu_{p'/q}) = e(\mu_{p'/q}).$$

So we only need to consider one family of circles and it does not matter which we choose.

To estimate the circle packing exponent we use the following procedure [Bullett and Mantica 1992]. We plot the number of circles of radius at least x against x in double logarithmic scale; the slope of the resulting graph is the circle packing exponent. More precisely, we start with, say, a thousand intervals along the x-axis, which we think of as numbered empty buckets. When we draw a circle of radius r we add one to the count in bucket number $-\lfloor 100 \log_{10} r\rfloor$. We then plot the log of the cumulative sum of the counts on the y-axis against bucket number on the x-axis. The resulting graph is of course discontinuous, but for radii less than .01, say, the jumps are small enough that they don't matter. If, in that region, the log-log graph is linear with slope a, the number of circles found that have diameter greater than r is proportional to r^{-a}.

If the algorithm for plotting the limit set were 100% accurate we would get a linear graph everywhere (apart from the early discontinutities), with slope equal to the circle packing exponent; this follows from the definition of the exponent by a simple reasoning. In practice, the graph starts approximately linear and then tails off. The point where it begins to tail off marks where the algorithm breaks down. The algorithm can be regarded as a good way of enumerating circles larger than this radius.

Figure 8 shows these graphs for the first three plots of Figure 1 (page 154). The true graph would be a line with slope equal to the circle packing exponent. The solid line, corresponding to the bottom plot on page 154, is approximately linear for $r < 10^{-5}$ (which is less than the visibility radius). Notice, however, we don't know that the slope is in fact the true circle packing exponent.

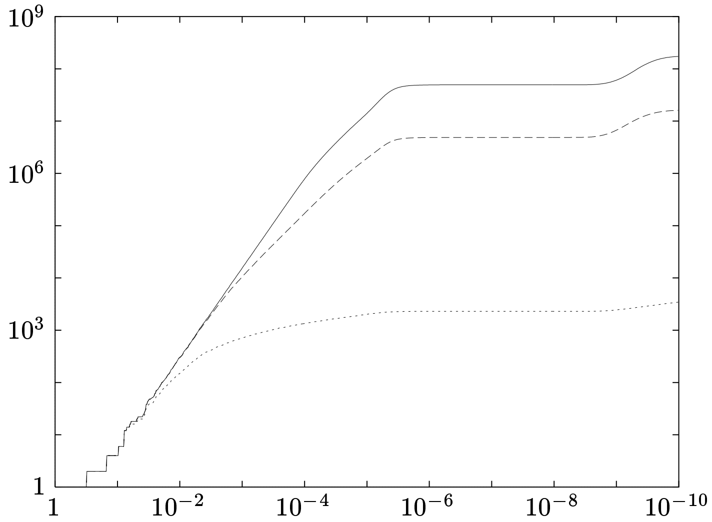


FIGURE 8. Cumulative number of circles against radius for the plots of page 154. The poor plot at the top of page 154 gives a graph that tails off very quickly (dotted line). The middle plot gives the graph in the middle, and the best plot, at the bottom of page 154, gives the solid line, which remains roughly linear until r is down to about 10^{-5} (solid line).

There is one case where theoretical estimates of the circle packing exponent exist, namely, the group G_μ for $\mu = 2i$. The limit set of this group is the circle packing of Apollonius, shown in Figure 9, for which the circle packing exponent is known to be $1.300197 < e < 1.314534$ [Boyd 1973]. With parameters as in Figure 9, for values of the radius between 10^{-2} and 10^{-4}, we get a slope of 1.300393. We conclude that, at least for this group, the algorithm described in this paper gives an efficient method of enumerating large circles first, and hence generates good pictures of the limit set.

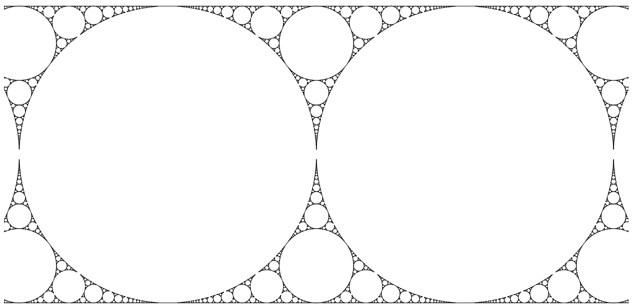


FIGURE 9. The Apollonius circle packing, $\Lambda(G_{2i})$. Generators: $W = T$ and $U = s$. Radii: $r_{\text{vis}} = .0005$ and $r_{\text{cut}} = 10^{-7}$. There were 6447 circles plotted out of $2.1 \cdot 10^8$.

REFERENCES

[Beardon 1968] A. F. Beardon, "The exponent of convergence of the Poincaré series", *Proc. London Math. Soc.* (3) **18** (1968), 461–483.

[Boyd 1973] D. W. Boyd, "Improved bounds for the disk packing constant", *Aeq. Math.* **9** (1973), 99–106.

[Bullett and Mantica 1992] S. Bullett and G. Mantica, "Group theory of hyperbolic circle packings", *Nonlinearity* **5** (1992), 1085–1109.

[Epstein et al. 1992] D. B. A. Epstein et al., *Word Processing in Groups*, Jones and Bartlett, Boston, 1992.

[Keen et al. 1993] L. Keen, B. Maskit and C. Series, "Geometric finiteness and uniqueness for Kleinian groups with circle packing limit sets", *J. reine angew. Math.* **436** (1993), 209–219.

[Keen et al.] L. Keen, J. R. Parker and C. Series, "Pleating coordinates for the Teichmüller space of the twice punctured torus" (in preparation).

[Keen and Series 1993] L. Keen and C. Series, "Pleating coordinates for the Maskit embedding of the Teichmüller space of punctured tori", *Topology* **32** (1993), 719–749.

[Larman 1966] D. G. Larman, "On the exponent of convergence of a packing of spheres", *Mathematika* **13** (1966), 57–59.

[Nicholls 1989] P. Nicholls, *The Ergodic Theory of Discrete Groups*, London Math Soc. Lecture Notes **143**, Cambridge University Press, New York and Cambridge, 1989.

[Maskit 1970] B. Maskit, "On boundaries of Teichmüller spaces and on Kleinian groups II", *Annals of Math.* **91** (1970), 607–639.

[Maskit 1974] B. Maskit, "Moduli of Marked Riemann surfaces," *Bull. Amer. Math. Soc.* **80** (1974), 773–777.

[McMullen 1991] C. T. McMullen, "Cusps are dense", *Annals of Math.* **133** (1991), 217–247.

[Parker] J. R. Parker, "Kleinian circle packings", to appear in *Topology*.

[Redfern 1993] I. D. Redfern, "Automatic coset systems", Ph.D. Thesis, University of Warwick, 1993.

[Wright 1987] D. J. Wright, "The shape of the boundary of Maskit's embedding of the Teichmüller space of once punctured tori", privately circulated preprint.

Greg McShane, Mathematics Institute, University of Warwick, Coventry CV4 7AL, England (mcshane@maths.warwick.ac.uk)

John R. Parker, Mathematics Institute, University of Warwick, Coventry CV4 7AL, England (jrp@maths.warwick.ac.uk)

Ian Redfern, Mathematics Institute, University of Warwick, Coventry CV4 7AL, England (redferni@dodo.logica.co.uk)

Received April 1, 1994; accepted in revised form June 27, 1994.