An Algorithm for 2-Generator Fuchsian Groups

J. GILMAN & B. MASKIT*

Introduction

The purpose of this paper is to present a geometrically based algorithm for deciding whether or not two elements of PSL(2, **R**) generate a non-elementary discrete group. There is, however, an obvious difficulty with the word "algorithm," for that suggests a procedure that can, at least in principle, be programmed to run on a computer. The difficulty has to do with elliptic elements; there is no effective way to decide that an elliptic element does not have finite order. If we regard an algorithm as a procedure, involving computations in some field, that necessarily ends after finitely many steps, then we do indeed have such an algorithm, provided our field of computations includes all standard computations involving real numbers, including arithmetic operations, computation of the inverse cosine, and computations involving logarithms.

However, if we take the point of view that an algorithm is something that can, at least in principle, be programmed to run on a computer, then we can say that we have an algorithm to decide if two matrices in GL(2, **Z**), with positive determinant, generate a non-elementary free discrete subgroup of PSL(2, **R**).

The problem of finding criteria for discreteness of Fuchsian groups has been the source of considerable activity; our list of references includes only those that are specific to 2-generator groups (and not all of them), as opposed to more general criteria. Jørgensen's inequality [J] yields a necessary condition; sufficient conditions in some cases were given by Lyndon and Ullman [LU]; and necessary and sufficient conditions for the case of two parabolic generators were given by Beardon [B1]. Necessary and sufficient conditions, in the form of an algorithm, were given by Purzitsky ([P1], [P2], [P3]), Rosenberger ([R1], [R2], [R3], [R4], [R5]), Purzitsky and Rosenberger [PR], and by Kern-Isberner and Rosenberger [KR]. Their approach is primarily

Received January 31, 1989. Revision received October 30, 1990.

* The first author was supported in part by NSF grant #DMS 9001881; the second author was supported in part by NSF grant #DMS 9003361. Both authors wish to acknowledge the support of the Mathematical Sciences Research Institute at Berkeley; the original ideas for this paper were conceived there.

Michigan Math. J. 38 (1991).

algebraic; that is, they show that the problem can be solved in finitely many steps by "trace minimizing," but they do not give a geometric explanation of the meaning of their procedure.

Another, more geometric, approach was started by Matelski, but there are some difficulties with his procedures. In broad outline, we follow Matelski's procedure here, filling in the one difficult case, where both generators are hyperbolic, with non-intersecting axes. We also explore the relationship between the algebra and the geometry (see also [GM]).

We remark that there is one case that we do not treat here, and that is the case of two hyperbolic generators with intersecting axes. The algebraic treatment has been done by Rosenberger together with Purzitsky and Kern-Isberner (see [R5] and the references given there). A geometric treatment of this case can be found in [G].

The authors wish to thank A. Beardon for informative conversations. Also, the first author wishes to thank G. Rosenberger for helpful conversations and correspondence during the preparation of this and other related work.

The structure of our treatment is as follows. We assume that we are given two nontrivial 2×2 matrices, with either real or integer entries, and positive determinant. Call these matrices g and h. We give a step-by-step procedure to determine discreteness of the group $G=\langle g, h\rangle \subset \mathrm{PGL}(2, \mathbf{R})^+$, the subgroup of $\mathrm{PGL}(2, \mathbf{R})$ with positive determinant. We give geometric interpretations of the results of our procedure, along with the necessary proofs, at each step.

In the case that g and h are given with integral entries, then we regard the procedure as having ended as soon as we reach an elliptic element of G, for then G is either not free or not discrete. The programming necessary for each step is fairly simple, and is left to the interested reader; a fuller description of the computational algorithm can be found in [GM].

Our general procedure is that we regard parabolic elements as "simpler" than hyperbolic, and elliptic elements as "simpler" than parabolic. We start with two hyperbolic generators; this is Case I. We then go to one hyperbolic and one parabolic generator (Case II), then two parabolic generators (Case III). The last three cases are of interest only for the algorithm with real coefficients: one hyperbolic and one elliptic generator (Case IV); one parabolic and one elliptic generator (Case V); and two elliptic generators (Case VI).

In the cases that neither of the generators is elliptic (aside from the possibility, which we do not treat, of two hyperbolic generators with intersecting axes), the procedures are all very similar. We start with g and h, compute $T(g)$ and $T(h)$ (these are the traces if the determinant $=1$), and make necessary simple adjustments so that $0<T(g)<T(h)$. We show that $T(gh)<T(h)$, and that G is discrete if $T(gh)<-2$; we also know that gh is elliptic if $-2<T(gh)<2$. We replace the ordered pair of generators (g, h) by new generators (g, gh). The only difficulty in showing that the procedure ends after finitely many steps is in Case I, where both generators are hyperbolic.

Before starting our procedure, we make a remark about elementary groups. If g and h have integral entries and are both hyperbolic with the same fixed points, then it is not clear if there is an effective procedure to determine whether or not G is discrete. (In the real case, we need to assume we can take logarithms, and that we can tell if a number is rational.) If they both have integral entries, and are both parabolic with the same fixed point, then G is cyclic and discrete. We have already discussed the difficulties with elliptic elements.

0. First Computations

For any matrix

$$g=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with positive determinant, set $T(g)=(a+d)/(ad-bc)^{1/2}$; it is clear that $T(g)$ depends only on g as an element of $\mathrm{PGL}(2,\mathbf{R})^{+}$.

0-1. Compute $T(g)$ and $T(h)$. If $T(g)<0$, replace g by $-g$; similarly, if $T(h)<0$, replace h by $-h$.

0-2. If $T(g)>T(h)$, replace the pair (g,h) by (h,g). We now have that $0\le T(g)\le T(h)$.

0-3. If $T(g)>2$, go to Case I; if $T(g)=2$ and $T(h)>2$, go to Case II; if $T(g)=T(h)=2$, go to Case III; if $T(g)<2$ and $T(h)>2$, go to Case IV; if $T(g)<2$ and $T(h)=2$, go to Case V; if $T(h)<2$, go to Case VI.

Case I: Hyperbolic–Hyperbolic

I-0. If $T(g)>T(h)$, replace the pair of generators (g,h) with the pair of generators (h,g).

I-1. Find a_g and r_g, the attracting and repelling fixed points (resp.) of g; also find a_h and r_h, the attracting and repelling fixed points (resp.) of h. For the purpose of our program, we need only check that these are distinct, and then look at some inequalities involving the cross ratio.

I-2. If $a_g=a_h$ or $r_g=r_h$ or $a_g=r_h$ or $r_g=a_h$, then G is either not discrete or elementary. From here on we assume that these four points are distinct.

I-3. Compute the cross ratio

$$C=(r_g,a_g;r_h,a_h)=\frac{(r_g-r_h)(a_g-a_h)}{(r_g-a_h)(a_g-r_h)}.$$

I-4. If $C=0$ or ∞, then g and h have a common fixed point. It is well known that, in this case, G is not discrete.

I-5. If $C>1$, then replace h by h^{-1}; this replaces C by $1/C$, so from here on, in the case $C>0$, we can assume $0<C<1$.

I-6. Compute the Jørgensen number

$$\mu(g,h)=|T([g,h])-2|+|T^2(g)-4|.$$

If $\mu(g, h) < 1$, then G is not discrete [J].

I-7. Compute $T(gh)$; if $T(gh) < -2$, then G is free and discrete.

The statement above needs proof. We say that the hyperbolic elements g and h *bound* if the axes of g, h, and gh bound a common region (in the upper half-plane). If g, h, and gh are all hyperbolic, where the axes of g and h are disjoint, and they do not bound, then we say that they *separate;* that is, in this case, one of the three axes (and it can be any one of the three) separates the other two.

If g and h bound, then G is free and discrete. An easy proof of this fact is to consider the three common orthogonals to the axes of g, h, and gh. (One possible configuration of these hyperbolic lines, where, e.g., the axis of g is labelled as A_g, is given in Figure 1.) It is clear that g and h bound if and only

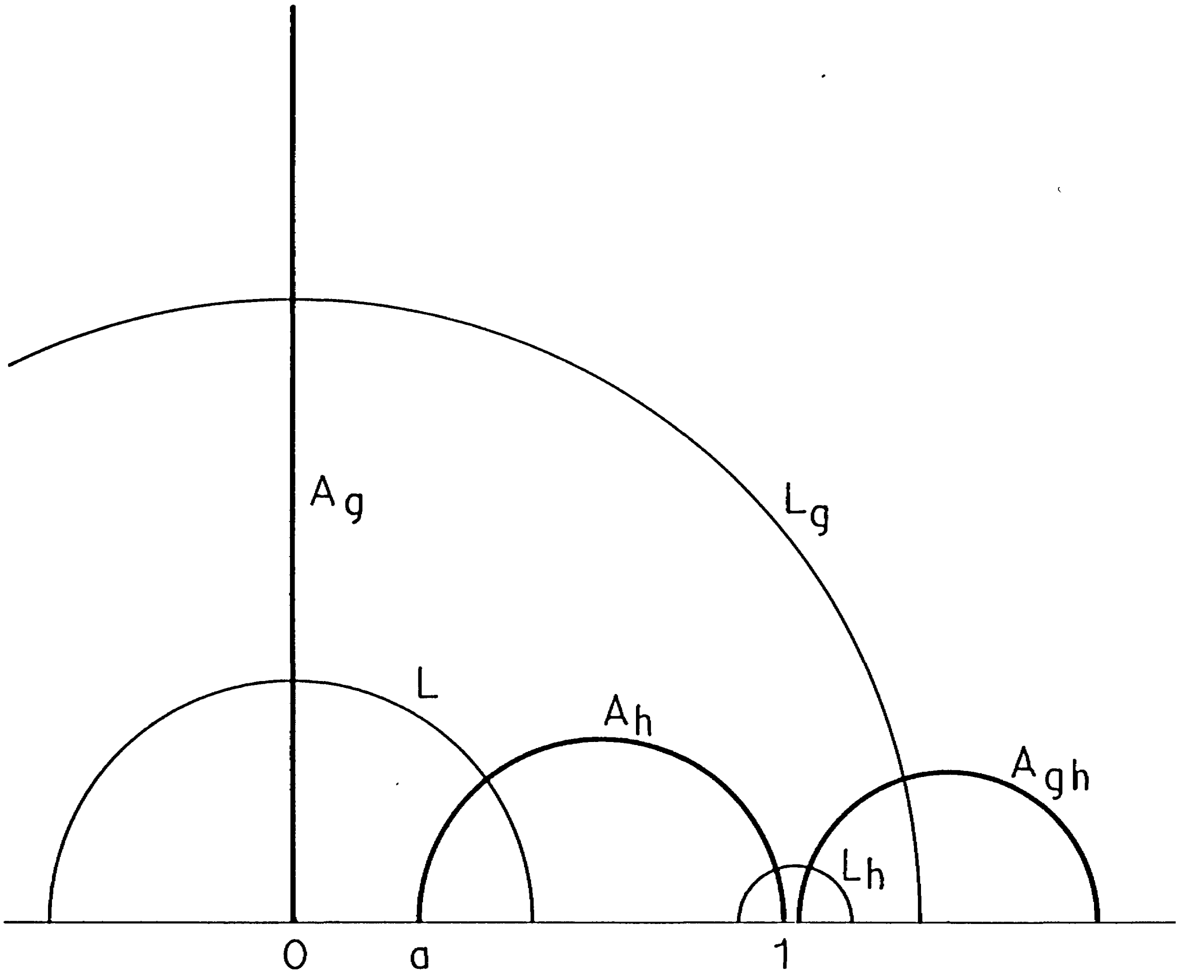

Figure 1

if these three hyperbolic lines also bound a common region. By Poincaré's polygon theorem (see [Ma1]), the group generated by the reflections in these three common orthogonals is discrete, and has no relations other than the fact that the three generators are involutions. The desired result now follows from the fact that G is the orientation-preserving half of this group.

THEOREM. *Let g and h be hyperbolic elements of* $\mathrm{SL}(2,\mathbf{R})$, *where g and h have no fixed points in common, the axes of g and h do not intersect, and gh is also hyperbolic.*

(a) *If $C(g,h)>1$ then $T(g)T(h)T(gh)>8$.*

(b) *If $C(g,h)<1$ then $T(g)T(h)T(gh)<-8$ if and only if g and h bound.*

Before going on to the proof of this theorem, we remark on its meaning. The condition $C(g,h)>1$ means that g and h are not oriented correctly with respect to one another; if we replace either g or h (but not both) by its inverse, then we go from $C(g,h)>1$ to $C(g,h)<1$. With this in mind, the theorem says the following. If the product of the traces is negative, then g and h are correctly oriented and their axes bound; if the product of the traces is positive, then either g and h are not correctly oriented, or their axes separate, or both.

Proof. We replace g and h by their corresponding matrices in $\mathrm{SL}(2,\mathbf{R})$; that is, we find g' and h' in $\mathrm{SL}(2,\mathbf{R})$ so that g and g', and h and h', represent the same element of $\mathrm{PGL}(2,\mathbf{R})$. Then $T(g')=T(g)$ and $T(h')=T(h)$; we then write g for g' and h for h'. With this notation, g is simultaneously a matrix, or an equivalence class of matrices, and a Möbius transformation; this should cause no confusion.

We normalize our transformations g and h as follows. Normalize so that the repelling fixed point r_g of g is at 0 and the attracting fixed point a_g is at ∞. Then either both fixed points of h are positive, or they are both negative. If necessary, conjugate by $z\to -z$, so that the fixed points of g are still at 0 and ∞, and the fixed points of h are both positive. Normalize further so that r_h is at 1. Notice that with this normalization $C(g,h)=a_h=a<1$ (this also shows that we need not consider the case $C(g,h)=1$, for that occurs precisely when h has only one fixed point). Since g has determinant 1, there is a number $R>1$ such that

$$g=\begin{pmatrix} R & 0 \\ 0 & R^{-1}\end{pmatrix},$$

and there is a number $K>1$ such that

$$h=\frac{1}{a-1}\begin{pmatrix} aK-K^{-1} & a(K^{-1}-K) \\ K-K^{-1} & aK^{-1}-K\end{pmatrix}.$$

We will make use of the following remark. If g and h are hyperbolic transformations of the hyperbolic plane, and if the axes of g and h intersect at a point, then so do the axes of g and gh (this is an easy computation once one normalizes the fixed points of g to be at 0 and ∞). It follows that if the axes of g and h intersect, then, for any Nielsen transformation α, the axes of $\alpha(g)$ and $\alpha(h)$ also intersect. Hence, if the axes of g and h do not intersect and gh is hyperbolic, then the axis of gh intersects neither the axis of g nor that of h.

With the normalization above, we have $T(g)=R+R^{-1}>2$ and $T(h)=K+K^{-1}>2$.

We prove part (a) by keeping the normalization as it is, with $0 < C(g, h) = a < 1$, and considering the transformation gh^{-1} instead of gh. It is easy to see that, with this normalization, the desired result follows from the statement $T(gh^{-1}) > 2$. Compute

$$\begin{aligned} T(gh^{-1}) &= \frac{R(aK^{-1}-K)+R^{-1}(aK-K^{-1})}{a-1} \\ &= \frac{a(RK^{-1}+R^{-1}K)-(RK+R^{-1}K^{-1})}{a-1} \\ &= RK^{-1}+R^{-1}K+\frac{RK+R^{-1}K^{-1}-RK^{-1}-R^{-1}K}{1-a}. \end{aligned}$$

Since $R^2K^2+1 > R^2+K^2$, we have $RK+R^{-1}K^{-1}-RK^{-1}-R^{-1}K > 0$. Thus $T(gh^{-1}) > RK^{-1}+R^{-1}K > 2$.

We now turn to part (b). We continue with our assumption that g and h are normalized as above, and that $0 < C(g, h) = a < 1$ (see Figure 2).

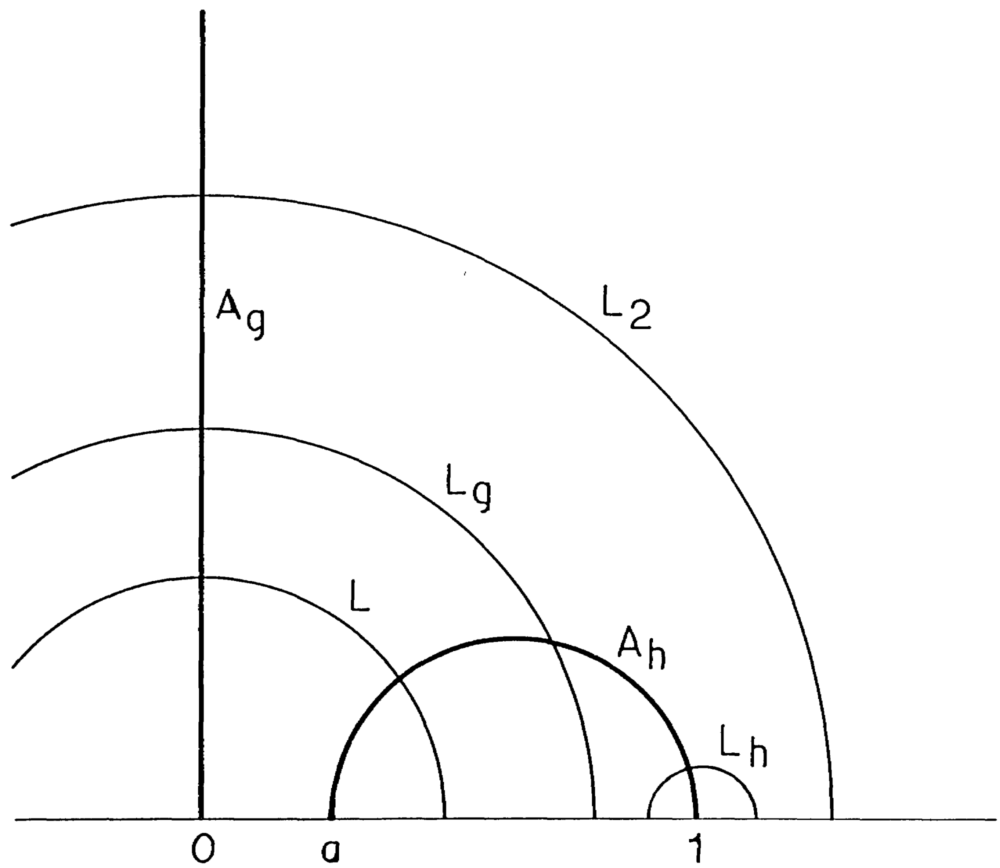

Figure 2

Since 0 is the repelling fixed point of g and a is the attracting fixed point of h, we have $0 < x < gh(x)$ for every point x with $0 < x < a$. Hence gh can have no fixed point in the interval $(0, a)$. Since the axes of g and h are disjoint, g and h separate if the fixed points of gh are negative, or if they lie between a and 1; hence g and h bound if and only if the average of the fixed points of gh is greater than 1. Write

$$gh=\frac{1}{a-1}\begin{pmatrix} R(aK-K^{-1}) & Ra(K^{-1}-K) \\ R^{-1}(K-K^{-1}) & R^{-1}(aK^{-1}-K) \end{pmatrix}=\frac{1}{a-1}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

The fixed points of gh satisfy

$$\gamma z^2+(\delta-\alpha)z-\beta=0.$$

Hence the average of the fixed points is greater than 1 if and only if

$$0<\alpha-\delta-2\gamma=\alpha+\delta-2(\gamma+\delta) \quad \text{(note that } \gamma>0\text{)},$$

or, equivalently,

$$\begin{aligned} 0>\frac{\alpha+\delta-2(\gamma+\delta)}{a-1}&=T(gh)-\frac{2(R^{-1}K-R^{-1}K^{-1}+aR^{-1}K^{-1}-R^{-1}K)}{a-1} \\ &=T(gh)-2R^{-1}K^{-1}. \end{aligned}$$

We have shown that g and h bound if and only if $T(gh)<2R^{-1}K^{-1}$. Since R and K are both greater than 1, $2R^{-1}K^{-1}<2$. Since either $T(gh)<-2$ or $T(gh)>2$, we can have $T(gh)<2R^{-1}K^{-1}$ if and only if $T(gh)<-2$. □

I-8. If $T(gh)=\pm 2$, then gh is parabolic; replace the generators (g,h) with the generators (g,gh), and go to Case II.

I-9. If $-2<T(gh)<2$, then gh is elliptic; replace the generators (g,h) with the generators (g,gh), and go to Case IV.

I-10. If $T(gh)>2$, then replace the generators (g,h) with the generators (gh,g) or $(-gh,g)$, so that $T(gh)>0$, and iterate the procedure above, starting with step I-0.

We need to show that this procedure terminates after finitely many steps. We start with an inequality.

PROPOSITION. *If g and h are hyperbolic with distinct fixed points and non-intersecting axes, where $T(h)\geq T(g)>2$ and $0<C(g,h)<1$, then $T(gh)<T(h)$.*

Proof. Normalize g and h as in step I-7, write g and h as above, and compute

$$T(gh)=\frac{aKR-RK^{-1}+aR^{-1}K^{-1}-R^{-1}K}{a-1}.$$

Then, using the fact that $T(h)=(aK-K^{-1}+aK^{-1}-K)/(a-1)$, we obtain

$$\begin{aligned} T(h)-T(gh)&=\frac{aK+aK^{-1}-K-K^{-1}-aKR+RK^{-1}-aK^{-1}R^{-1}+KR^{-1}}{a-1} \\ &=\frac{a(K-KR+K^{-1}-K^{-1}R^{-1})+K^{-1}R-K^{-1}+KR^{-1}-K}{a-1}= \end{aligned}$$

$$= \frac{aK(1-R)+aK^{-1}(1-R^{-1})+K(R^{-1}-1)+K^{-1}(R-1)}{a-1}$$

$$= \frac{a(K-K^{-1}R^{-1})(1-R)+(KR^{-1}-K^{-1})(1-R)}{a-1}$$

$$>0,$$

where we have used the fact that $K \geq R$ in the last inequality.

This proposition shows that, if we follow the outlined procedure, then $\max(T(g), T(h))$ keeps decreasing as long as we keep $T(g)$ and $T(h)$ positive. Our next goal is to bound the decrease from below. There are two cases to consider.

We keep g and h normalized as above; that is, g and h are as in step I-7, with $0 < a = C(g, h) < 1$, where a is the attracting fixed point of h and 1 is the repelling fixed point of h. Let L be the common orthogonal to the axes of g and h (see Figure 2 for one of the possible configurations). Then there are (hyperbolic) lines L_g and L_h, where L_g is orthogonal to the axis of g and L_h is orthogonal to the axis of h, so that the following holds. If we denote the reflection in L by r, the reflection in L_g by r_g, and the reflection in L_h by r_h, then $g = r_g r$ and $h = r r_h$. It follows that $gh = r_g r_h$. It is easy to see that L_g is the common orthogonal to the axes of g and gh, and that L_h is the common orthogonal to the axes of h and gh. By assumption, g and h do not bound, so L, L_g, and L_h do not bound a common region. Let ρ_g be the (hyperbolic) distance, measured along the axis of g, between the point of intersection with L and the point of intersection with L_g. Similarly, let ρ_h be the distance, measured along the axis of h, between the point of intersection with L and the point of intersection with L_h.

We denote the translation length of the element f by τ_f; then $2\rho_f = \tau_f$. It is well known that $\tau = \tau_f$ and $T = T_f$ are related by the formula

$$T = 2\cosh(\tau/2) = 2\cosh(\rho).$$

In particular, since $T(h) \geq T(g)$, $\rho_h \geq \rho_g$.

There are now several different cases to consider. Observe first that if L_g and L_h meet at a point z in the disc, then $gh = r_g r_h$ is elliptic with fixed point z. Similarly, if L_g and L_h meet at the circle at infinity, then gh is parabolic; we can assume that these cases do not occur. We have also assumed that the three lines L, L_g, and L_h do not bound a common region. The only other possibility is that L_g and L_h do not meet, even at the circle at infinity, and do not bound a common region. In this case, one of these lines separates the other two. In particular, either L_g crosses the axis of h, or L_h crosses the axis of g, or both.

If L_h lies between L and L_g, then the distance along the axis of g between L and L_h is less than ρ_g. The axis of h is the common orthogonal to L and L_h, so ρ_h (which is the distance between L and L_h measured along the axis of h) is less than the distance between these two lines measured along the axis of g. Since $\rho_g \leq \rho_h$, we cannot have that L_h lies between L and L_g.

We conclude that L_g lies between L and L_h; in particular L_g crosses the axis of h. There are now two cases to consider.

We first take up the case that L_h does not cross the axis of g; that is, one endpoint of L_h lies between a and 1, and the other is greater than 1 (see Figure 2). We draw a sequence of hyperbolic lines L_m, $m=2,3,\dots$, so that if r_m denotes reflection about L_m then $g^m = r_m r$. Since $\tau(g^m) = m\tau(g)$, the lines L_m, which are all orthogonal to the axis of g, are (hyperbolically) equally spaced along the axis of g, and the distance between successive lines is ρ_g. Also, since ∞ is the attracting fixed point of g, and hence of g^m, the lines L_m accumulate to ∞. Since L_h does not intersect the imaginary axis, either some L_m intersects L_h or there is a first m so that L_m, L_h, and L bound a common region ($m=2$ in Figure 2, and $m=3$ in Figure 3). In either case,

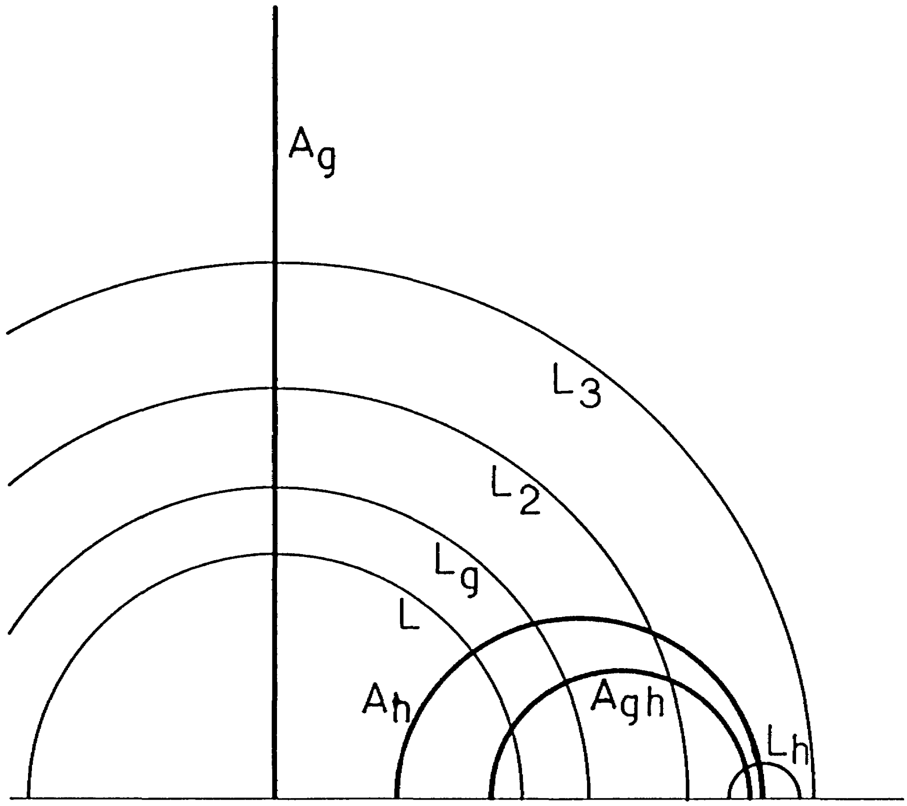

Figure 3

we can replace the generators (g,h) by the pair of generators $(g^m h, g)$; note that for these new generators, either their axes meet at the circle at infinity or they bound, since L_{m-1}, L_m, and L_h, the common orthogonals to their axes, bound a common region (in Figure 2, $m=2$ and $L_1 = L_g$). Note that these two axes cannot intersect, for the axes of any two hyperbolic generators of G do not intersect.

We return now to our algorithm. The instruction is to replace the pair of generators (g,h) with either (gh,g) or (g,gh), depending on whether $T(gh) < T(g)$ or $T(gh) > T(g)$. If $m=1$ in the argument above, then gh and

g bound, so we need not further consider this case. If $m>1$, we need to show that $T(gh)>T(g)$, for then the next step in our iteration will yield (g^2h, h) (or (g, g^2h)). We have shown that, for some m, either $g^m h$ is elliptic, or parabolic, or g and $g^m h$ bound. Hence once we show that $T(gh)>T(g)$ when $m>1$, then we will have shown that this part of our procedure necessarily ends after at most m steps.

Observe that both endpoints of the axis of gh must lie between a and 1, for L_h has one endpoint between a and 1 and the other endpoint is greater than 1. Since L_2 crosses the axis of h between L_g and L_h and does not intersect L_h, L_2 also crosses the axis of gh between L_g and L_h. Since ρ_g (the distance between L_g and L_2 measured along the axis of g) is less than the distance between L_g and L_2 measured along the axis of gh, which in turn is less than ρ_{gh} (see Figure 3), we conclude that if $m>1$ then $T(g)<T(gh)$.

We next take up the case that L_h crosses the axis of g. We renormalize so that 0 is the attracting fixed point of g and ∞ the repelling fixed point of g, so that the fixed points of h are both positive, and so that L, the common orthogonal to g and h, is the unit circle. We compute that the endpoints of L_g are symmetric with respect to the origin, and label them as $\pm y$. We label the endpoints of L_h as $w<0$ and $z>0$. We now have these points in the following order (see Figure 4):

$$-1<-y<w<0<z<y<1.$$

Using the canonical isomorphism between PGL(2, **R**) (including matrices with negative determinant) and the group of all isometries of the hyperbolic

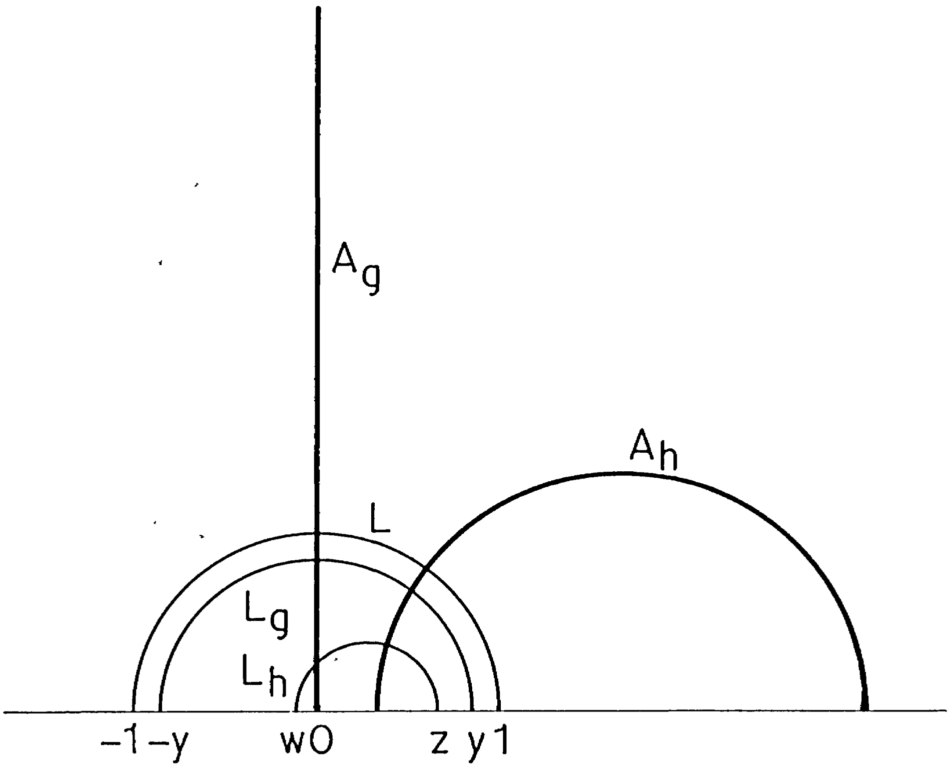

Figure 4

plane (including the orientation-reversing isometries; see [Ma2]), we write:

$$r=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad r_g=\begin{pmatrix}0 & y^2\\ 1 & 0\end{pmatrix},\quad r_h=\begin{pmatrix}-z-w & 2zw\\ -2 & z+w\end{pmatrix},$$

and compute:

$$g=\begin{pmatrix}y^2 & 0\\ 0 & 1\end{pmatrix}\quad\text{and}\quad h=\begin{pmatrix}-2 & z+w\\ -(z+w) & 2zw\end{pmatrix}.$$

We next use Jørgensen's inequality, and compute the Jørgensen number

$$\begin{aligned}\mu&=|T([g,h])-2|+|T^2(g)-4|\\ &=(z+w)^2(y-y^{-1})^2(z-w)^{-2}+(y-y^{-1})^2\\ &=2(z^2+w^2)(y-y^{-1})^2(z-w)^{-2}.\end{aligned}$$

If $\mu<1$ then, by Jørgensen's inequality, G is not discrete; we now assume that $\mu\geq 1$. Note that since $w<0$ and $z>0$, $(z^2+w^2)/(z-w)^2<1$. Hence Jørgensen's inequality [J] yields: $(y-y^{-1})^2\geq 1/2$. Since $y<1$, one easily translates this into the inequality $y\leq 2^{-1/2}$.

It is easy to see geometrically that $T(gh)<T(h)$. The axis of gh is the common orthogonal to L_g and L_h (see Figure 4). It does not matter where this line is located with respect to A_g and A_h (it could be in any one of three places). In any case, ρ_h is the distance between L_h and L measured along the common orthogonal A_h, and is clearly greater than the distance along A_h between L_g and L_h, which in turn is greater than ρ_{gh}, the distance between L_g and L_h measured along the common orthogonal.

Now repeat the same analysis as above for the generators g and $j=hg$. Only the last case, where L_g crosses the axis of j and L_j crosses the axis of g, is of concern, for in all the other cases, we either have that j is a simpler generator, or we conclude that G is discrete, or we conclude that G is not discrete, or we are in the case where we know the process ends after finitely many steps. Working with positive traces, an easy computation yields

$$T(gh)=\frac{2(y^2-zw)}{y(z-w)}\quad\text{and}\quad T(h)=\frac{2(1-zw)}{z-w}.$$

We bound the difference $T(h)-T(gh)$ from below as follows:

$$\begin{aligned}T(h)-T(gh)&=\frac{2(1-y)(zw+y)}{y(z-w)}\\ &\geq\frac{2(1-y)(y-y^2)}{y(z-w)}\\ &\geq\frac{(1-y)^2}{y}\\ &\geq\frac{(\sqrt{2}-1)^2}{\sqrt{2}}>0,\end{aligned}$$

where we have used the facts that $-y<w<0<z<y<1$ and $y\leq 1/\sqrt{2}$. Since $T(gh)<T(h)$, and the difference has a positive lower bound, we can remain in this last case for only finitely many steps. Hence, this part of our procedure ends after finitely many steps.

We remark that one could easily use the above inequality to bound the number of steps required in this case. However, bounds in the other cases are not easily obtained.

I-11. We turn now to the case that $C<0$, that is, when the axes of g and h intersect. An easy computation shows that this occurs if and only if $T([g,h])<2$. There are several proofs in the literature that $G=\langle g,h\rangle$ is discrete and free if and only if $T([g,h])\leq -2$ (see, e.g., [M] or [P1]).

If $|T([g,h])|<2$, then the commutator $[g,h]$ is necessarily elliptic, so G is either not free or not discrete. The question in this case as to when G is discrete is fully answered by the algorithm given by Theorem 4 of [P3], taken together with the corrections of [M]. This is also summarized in [R5].

Case II: Hyperbolic–Parabolic

II-0. If $T(g)>T(h)$, replace the generators (g,h) by the generators (h,g); after this step, h is hyperbolic and g is parabolic.

II-1. Compute the fixed points of g and h. If they have a common fixed point, then G is not discrete. From here on we assume these points are distinct.

II-2. Normalize so that the fixed point of g is at ∞; then $g(z)=z+\tau$. If $\tau<0$, replace g by g^{-1}; from here on, we can assume $\tau>0$.

II-3. If the attracting fixed point of h is larger than the repelling fixed point, replace h by h^{-1}. Now normalize further so that the fixed points of h are at ± 1. Note that -1 is the attracting fixed point and $+1$ is repelling.

Write

$$g=\begin{pmatrix}1 & \tau\\ 0 & 1\end{pmatrix} \quad\text{and}\quad h=\begin{pmatrix}a & b\\ b & a\end{pmatrix},$$

where $a^2-b^2=1$, $a>0$, and (since $+1$ is the repelling fixed point) $b<0$.

II-4. It is well known that, if the discrete group contains the parabolic element $z\to z+1$, then the radius of the isometric circle of any element of the group is at most 1 (this is sometimes known as the Shimizu–Leutbecher theorem; a proof can be found in [B2, p. 106]). Hence G is not discrete if $|b\tau|<1$.

II-5. Compute $T(gh)$; if $T(gh)<-2$, then G is free and discrete.

This statement needs proof. An easy computation shows that $T(gh)\leq -2$ if and only if $\tau/2\geq (a+1)/(-b)$.

Let L be the line orthogonal to the axis of h and ending at the fixed point of g. Let r denote the reflection in L. Then there are lines L_g and L_h, where L_h is orthogonal to the axis of h and L_g ends at the fixed point of g, so that (denoting reflection in L_g by r_g and reflection in L_h by r_h) we can write $g=r_g\circ r$ and $h=r\circ r_h$.

Observe that L_g is a Euclidean line segment parallel to the imaginary axis and that L_g ends at $\tau/2>0$. Also, L_h is orthogonal to the axis of h, which is the unit circle, and both its endpoints are positive; call these endpoints λ_1 and λ_2, where $\lambda_1<\lambda_2$. Notice that the (Euclidean) circle on which L_h lies is the isometric circle of h. Hence we can compute $\lambda_2=(a+1)/(-b)$. We have shown that $T(gh)\leq -2$ if and only if the three lines L, L_g, and L_h are disjoint and bound a common region. As we observed in the previous case, the fact that these three lines bound a common region implies that the group G is free and discrete.

II-6. If $-2<T(gh)<2$, then gh is elliptic. Replace the generators (g,h) by the generators (g,gh); these are parabolic and elliptic, respectively.

II-7. If $T(gh)=2$ then gh is parabolic. Replace the generators (g,h) by (g,gh), which are both parabolic.

II-8. If none of the above cases occur, then replace h by gh and return to step II-0. With g and h normalized as above, $T(gh)=2a+b\tau=T(h)+b\tau$. Since $b<0$ and $|b\tau|\geq 1$, we can remain in this case for only finitely many steps.

Case III: Parabolic–Parabolic

III-1. If $T(g)<0$, replace g by $-g$; also, if $T(h)<0$, replace h by $-h$.

III-2. Let x be the fixed point of g and let y be the fixed point of h. If g and h both have the same fixed point, then G is elementary. (In the case that g and h are given by integral matrices, then G is discrete and cyclic.)

III-3. Compute the cross ratio $C=(x,y;h(x),g(y))$. If $C>0$, replace h by h^{-1} (see step V-3).

III-4. If $-2<T(gh)<2$, then gh is elliptic; replace the generators (g,h) by the generators (g,gh) and go to case V.

III-5. If $T(gh)\leq -2$ then G is free and discrete. It never happens that $T(gh)\geq 2$.

In order to prove the above statements, normalize so that $g(z)=z+1$, and so that h has its fixed point at 0. Then $h(z)=z/(\tau z+1)$. One easily sees that $C<0$ if and only if $\tau<0$. Compute $T(gh)=2+\tau$. As $\tau<0$, $T(gh)<2$. Also, gh is elliptic if and only if $-2<T(gh)<2$. It remains to show that G is free and discrete if and only if $T(gh)\leq -2$.

Construct the line L with fixed points at 0 and ∞, and construct the line L_g with fixed points at $1/2$ and ∞. Then $g=r_g r$, where r is the reflection in L and r_g is the reflection in L_g. Similarly, construct the line L_h with endpoints at 0 and some point λ, so that $h=rr_h$. It is easy to see that the Euclidean circle containing L_h is the isometric circle for h, so λ is the point $2/(-\tau)$.

The lines L_g and L_h intersect if and only if $-2/\tau>1/2$, in which case gh is elliptic; otherwise, they bound a common region, in which case G is free and discrete.

We remark that the question of when two parabolic elements of $PSL(2,\mathbf{R})$ generate a discrete subgroup has been completely solved by Beardon [B1].

Case IV: Elliptic–Hyperbolic

IV-1. If g is elliptic and not of finite order, then G is not discrete. If g is of finite order then, for some positive α, g^α is a minimal rotation (i.e., $|\mathrm{tr}(g^\alpha)|$ is maximal); replace g by g^α.

IV-2. If g has order 2, then g and g^{-1} are indistinguishable. If the order of g is greater than 2, then we distinguish positive and negative rotations as follows. The transformation $z \to e^{it}z$ has positive rotation about the origin, for $0<t<\pi$, and has negative rotation about the origin for $\pi<t<2\pi$. For an arbitrary elliptic transformation, we define positive and negative rotation about a fixed point by conjugation. It is clear that if g has negative rotation about a fixed point x, then g^{-1} has positive rotation about x.

If g has negative rotation about its fixed point in the upper half-plane, then replace g by g^{-1}.

IV-3. Let x be the fixed point of g, and let A be the axis of h. If x does not lie on A, let L' be the line segment from x to A, orthogonal to A. L' divides A into two parts, the positive part having the attracting fixed point of h on it, and the negative part having the repelling fixed point of h on it. If x lies on A, then it divides A into the same two parts. If the order of g is greater than 2, then $g(L')$ lies in one of the two half-planes cut out by the full line L on which L' lies. If this half-plane does not also contain the negative half of A, replace h by h^{-1}. After this replacement, $g(L)$ and the negative half of A both lie in the same half-plane.

IV-4. If $T(g)<0$, replace g by $-g$; likewise, if $T(h)<0$, replace h by $-h$.

IV-5. If $T(gh)\le -2$ then G is discrete, and G is the free product $G=\langle g\rangle * \langle h\rangle$.

This assertion needs proof; we state it formally.

THEOREM. *If $T(gh)\le -2$ then G is discrete and $G=\langle g\rangle * \langle h\rangle$. Let L be the line through the fixed point of g orthogonal to the axis of h. Then there are unique lines L_g and L_h, so that $g=r_g r$ and $h=rr_h$, where r, r_g, and r_h are (resp.) the reflections in L, L_g, and L_h. If these three lines bound a convex polygon, where L_h does not intersect either L or L_g, then $T(gh)\le -2$.*

Proof. Normalize G so that x, the fixed point of g, is at i; so that L', the line segment between the fixed point of g and the axis of h, lies on the imaginary axis; and so that A, the axis of h, lies inside the (closed) unit disc. Let L be the full hyperbolic line defined by the imaginary axis. Then there are lines L_g and L_h, where L_g passes through the point i, so that $g=r_g r$ and $h=rr_h$, where r, r_g, and r_h denote (resp.) reflection in L, L_g, and L_h. Denote the endpoints of L_g by $b<0$ and $a>0$. Note that, since L_g passes through i, $ab=-1$. Similarly, let $c<d$ be the endpoints of L_h; note that, since A lies in the closed unit disc, $cd<1$ (see Figures 5 and 6).

It is easy to see that, since g is a minimal rotation, the angle between L' and L_g in the right half-plane is π/α, where α is the order of g. It follows that $|b|>a$.

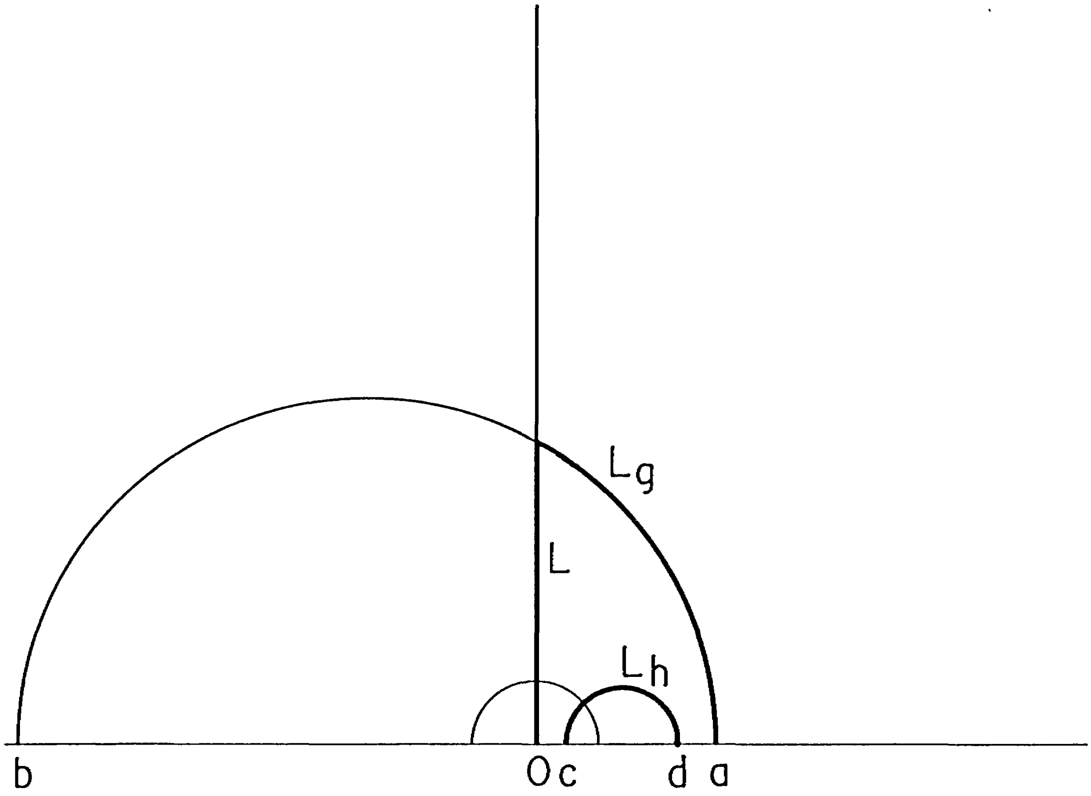

Figure 5

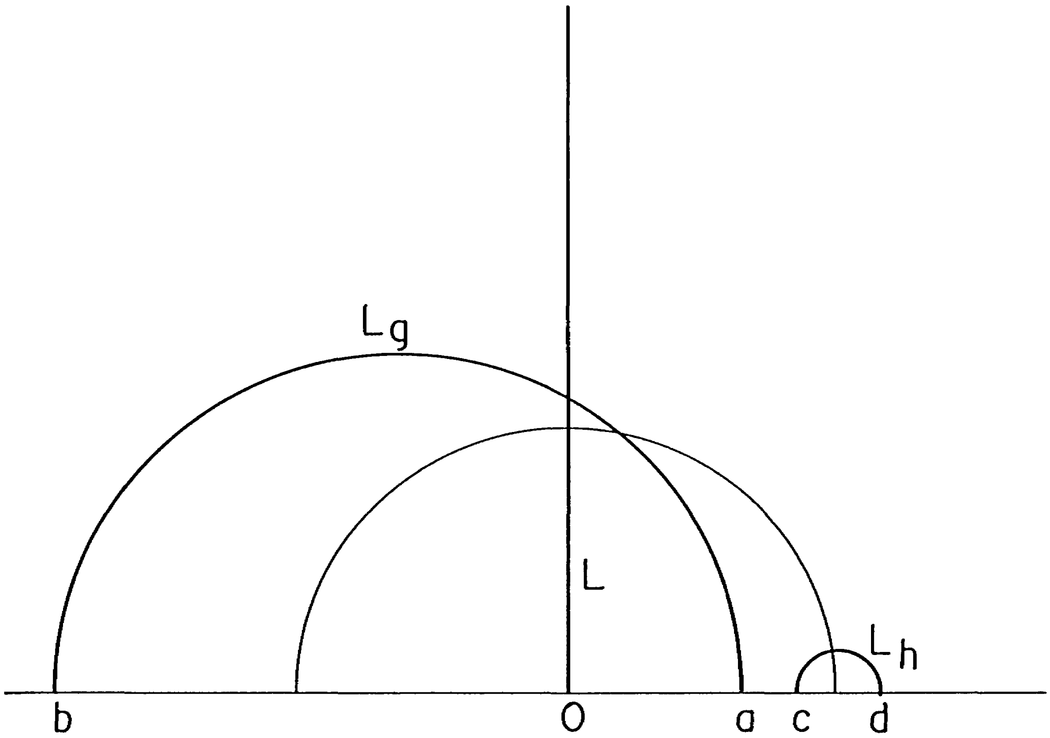

Figure 6

If the order of g is greater than 2, then the image of L' under g lies in the right half-plane. We have normalized so that the repelling fixed point of h lies in this same right half-plane; hence c and d are both positive. If the order of g equals 2, then L_g lies on the unit circle. If necessary in this case, we further normalize by conjugation by reflection in the imaginary axis; this leaves g unchanged and replaces h by h^{-1}. Hence, in this case as well, we can assume that c and d are both positive.

We now write

$$r_g = \begin{pmatrix} a+b & 2 \\ 2 & -(a+b) \end{pmatrix}, \quad r = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad r_h = \begin{pmatrix} c+d & -2cd \\ 2 & -(c+d) \end{pmatrix},$$

and compute (multiplying the matrices by -1 where necessary)

$$g = \begin{pmatrix} -(a+b) & 2 \\ -2 & -(a+b) \end{pmatrix} \quad \text{and} \quad h = \begin{pmatrix} c+d & -2cd \\ -2 & c+d \end{pmatrix}.$$

We fixed the sign of these matrices so that $T(g) = -2(a+b)/(a-b) \geq 0$ and $T(h) = 2(c+d)/(d-c) > 0$.

Using the fact that $ab = -1$, we now compute

$$T(gh) = \frac{-2(a+b)(c+d) - 4(1-cd)}{(a-b)(d-c)}.$$

We solve $T(gh) \leq -2$, and substitute $b = -1/a$ to get $ca^2 + a(1-cd) - d \geq 0$. Since $a > 0$, this is equivalent to $a \geq d$.

Observe that $a \geq d$ if and only if the entire hyperbolic line L_h lies in the sector between that part of L_g between a and i, and that part of the imaginary axis between 0 and i (the case $a \geq d$ is shown in Figure 5, where the three line segments needed for Poincaré's theorem are heavier than the others). Since the angle between these two line segments at i is π/α, where 2α is the order of g, Poincaré's theorem (see [Ma1]) tells us that the group generated by r, r_g, and r_h is discrete; further, the only relations are that the generators are involutions, and that $g^\alpha = 1$ (see Figure 6 for the case that $a < d$). □

IV-6. If $T(gh) = 2$ then replace h by gh, and go to Case V.

IV-7. If $-2 < T(gh) < 2$ then replace h by gh, and go to Case VI.

IV-8. If $T(gh) > 2$ then return to step IV-1 with the generators (g, gh).

It remains only to show that the iteration ends after finitely many steps. To this end, we consider the hyperbolic lines L_m, through the point i, so that $g^m = r_m r$, where r_m is reflection in L_m. Either there is a first such line which intersects L_h, in which case $g^m h$ is elliptic, or L_m intersects L_h at the circle at infinity, in which case $g^m h$ is parabolic, or there is a first such m so that L_m lies to the right of L_h, while L_{m-1} lies to the left of L_h.

In this last case, observe that the lines L_{m-1} and L_m meet at an angle of π/α at i; also note that L_h lies entirely inside the sector with this angle between them. It follows that $G = \langle g^m h, g\rangle$ is discrete and that $G = \langle g^m h\rangle * \langle g\rangle$.

In order to make use of these observations, it suffices to show that $T(gh) < T(h)$ (since g has finite order α, the process must then end after at most $\alpha - 1$ steps). We compute

$$\begin{aligned} T(h)-T(gh) &= \frac{2(c+d)}{d-c} - \frac{-2(a+b)(c+d)-4(1-cd)}{(a-b)(d-c)} \\ &= \frac{4a(c+d)+4(1-cd)}{(a-b)(d-c)} \\ &> 0. \end{aligned}$$

Case V: Elliptic–Parabolic

V-1. If g is elliptic and not of finite order, then G is not discrete. If g is of finite order then, for some positive α, g^α is a minimal rotation (i.e., $|\mathrm{tr}(g^\alpha)|$ is maximal); replace g by g^α.

V-2. If g has negative rotation about its fixed point in the upper half-plane, then replace g by g^{-1}.

V-3. Define the rotation of the parabolic element h (about its fixed point in the upper half-plane) as follows. If the fixed point of h is at ∞, then h has positive rotation if $h(0)>0$, and has negative rotation otherwise. If the fixed point is not at ∞, then use conjugation in $\mathrm{PSL}(2,\mathbf{R})$ to define positive and negative rotation. If h has negative rotation about its fixed point, replace h by h^{-1}.

V-4. If $T(g)<0$, replace g by $-g$; similarly, if $T(h)<0$, replace h by $-h$.

V-5. If gh is elliptic, replace h by gh and go to Case VI.

V-6. If gh is not elliptic, then G is discrete.

THEOREM. *If gh is not elliptic then $T(gh)\le -2$, G is discrete, and $G=\langle g\rangle * \langle h\rangle$.*

Proof. Normalize so that g has its fixed point (in the upper half-plane) at i, and so that h has its fixed point at 0. Then, as in Case IV, we can write

$$g=\begin{pmatrix} -(a+b) & 2 \\ -2 & -(a+b) \end{pmatrix} \quad \text{and} \quad h=\begin{pmatrix} 1 & 0 \\ \tau & 1 \end{pmatrix},$$

where $ab=-1$, $0<a\le 1$, and $\tau<0$.

Compute $T(gh)=[2\tau-2(a+b)]/(a-b)$, and observe that $T(gh)<T(g)<2$. Hence, either $T(gh)\le -2$ or gh is elliptic.

One observes that $T(gh)\le -2$ if and only if $-2/\tau\le a$. This last condition is easily seen to be equivalent to the following. We know that the imaginary axis intersects the isometric circle of g^{-1} at the point i and meets the isometric circle of h on the sphere at infinity. The isometric circle of h has its endpoints at 0 and $-2/\tau$, and the isometric circle of g^{-1} has its endpoints at a and $b=-1/a$. Hence $-2/\tau\le a$ if and only if these three hyperbolic lines (the imaginary axis, the isometric circle of g^{-1}, and the isometric circle of h) form a triangle to which we can apply Poincaré's polygon theorem, where the identifications of the sides are the three reflections. The result now follows. □

Case VI. Elliptic–Elliptic

VI-1. If g and h have the same fixed point in the upper half-plane, then G is either elementary or not discrete. From here on, we assume that g and h have distinct fixed points.

VI-2. If g is elliptic and not of finite order, then G is not discrete. If g is of finite order then, for some positive α, g^α is a minimal rotation (i.e., $|\mathrm{tr}(g^\alpha)|$ is maximal); replace g by g^α.

VI-3. If h is elliptic and not of finite order, then G is not discrete. If h is of finite order then, for some positive β, h^β is a minimal rotation (i.e., $|\mathrm{tr}(h^\beta)|$ is maximal); replace h by h^β.

VI-4. If g has negative rotation about its fixed point in the upper half-plane, then replace g by g^{-1}.

VI-5. If h has negative rotation about its fixed point in the upper half-plane, then replace h by h^{-1}.

VI-6. If gh is either hyperbolic or parabolic, then G is discrete and $G=\langle g\rangle*\langle h\rangle$.

To prove this statement, normalize so that the fixed point of g is at i, and so that the fixed point of h is at ti, $0<t<1$. Let L denote the hyperbolic line formed by the positive imaginary axis; let L_g and L_h be hyperbolic lines chosen so that $g=r_g r$ and $h=rr_h$, where r, r_g, and r_h denote (resp.) reflection in L, L_g, and L_h. Note that L_g is the isometric circle of g^{-1} and that L_h is the isometric circle of h.

Let L' be the segment of L between i and ti. We have chosen g and h so that the angle from L' to L_g, measured inside the right half-plane, is π/α, where α is the order of g. Similarly, the angle from L_h to L', also measured inside the right half-plane, is π/β, where β is the order of h.

If L_g and L_h do not meet in the right half-plane (except perhaps at the circle at infinity), then gh is either hyperbolic or parabolic. By Poincaré's polygon theorem, G is discrete and $G=\langle g\rangle*\langle h\rangle$.

If L_g and L_h do meet in the right half-plane, then gh is elliptic. From here on, we assume that this occurs.

VI-7. Compute the area of the triangle formed by L, L_g, and L_h; if it is less than $\pi/42$, G is not discrete.

VI-8. If gh is elliptic and not of finite order, then G is not discrete.

VI-9. If gh is elliptic and geometrically primitive (i.e., a rotation through an angle of $2\pi/n$, n an integer), then G is discrete.

The proof of this fact is again by Poincaré's polygon theorem; observe that the angle between L_g and L_h is half the rotation angle of gh.

VI-10. We now assume that gh is elliptic of finite order γ, but not geometrically primitive. Let T be the triangle formed by L, L_g, and L_h. Let x be the vertex between L and L_g, let y be the vertex between L and L_h, and let z be the vertex between L_g and L_h. Let T_1 be the triangle inside T, with vertices at x and z, where the angle at z is π/γ. Similarly, let T_2 be the triangle inside T, with vertices at y and z, where the angle at z is π/γ. It is clear that T_1 and

T_2 have disjoint interiors, so either the area of T_1 is less than half the area of T, or the area of T_2 is less than half the area of T. In the former case, replace h by gh; in the latter case, replace g by gh, and return to VI-1.

It is clear that this process terminates after finitely many steps.

We remark that instead of the iteration procedure of step VI-10, we could refer to Knapp's list of all possibilities for the reflections in the sides of T to generate a discrete group [K].

References

[B1] A. F. Beardon, *Fuchsian groups and nth roots of parabolic generators,* Holomorphic Functions and Moduli II, pp. 13–21, Math. Sci. Res. Inst. Publ. 11, Springer, New York, 1988.

[B2] ———, *The geometry of discrete groups,* Springer, New York, 1983.

[DJ] C. Doyle and D. James, *Discreteness criteria and high order generators for subgroups of* SL(2, **R**). Illinois J. Math. 25 (1981), 191–200.

[G] J. Gilman, *Two-generator discrete groups: the geometry of intersecting axes,* in preparation.

[GM] J. Gilman and B. Maskit, *Two generator discrete groups; algebra and geometry,* in preparation.

[J] T. Jørgensen, *On discrete groups of Möbius transformations,* Amer. J. Math. 98 (1976), 739–749.

[KR] G. Kern-Isberner and G. Rosenberger, *Über Diskretheitsbedingungen und die Diophantische Gleichung* $ax^2+by^2+cz^2=dxyz$. Arch. Math. (Basel) 34 (1980), 481–493.

[K] A. W. Knapp, *Doubly generated Fuchsian groups,* Michigan Math. J. 15 (1968), 289–304.

[LU] R. C. Lyndon and J. L. Ullman, *Pairs of real 2-by-2 matrices that generate free products,* Michigan Math. J. 15 (1968), 161–166.

[Ma1] B. Maskit, *On Poincaré's theorem for fundamental polygons,* Adv. in Math. 7 (1971), 219–230.

[Ma2] ———, *Parameters for Fuchsian groups I: Signature* (0, 4), to appear.

[M] J. P. Matelski, *The classification of discrete 2-generator subgroups of* PSL(2, **R**), Israel J. Math. 42 (1982), 309–317.

[P1] N. Purzitsky, *Two generator discrete free products,* Math. Z. 126 (1972), 209–223.

[P2] ———, *Real two-dimensional representation of two-generator free groups,* Math. Z. 127 (1972), 95–104.

[P3] ———, *All two-generator Fuchsian groups,* Math. Z. 147 (1976), 87–92.

[PR] N. Purzitsky and G. Rosenberger, *Two generator Fuchsian groups of genus one,* Math. Z. 128 (1972), 245–251. *Correction:* Math. Z. 132 (1973), 261–262.

[R1] G. Rosenberger, *Fuchssche Gruppen, die freies Produkt zweier zyklischer Gruppen sind, und die Gleichung* $x^2+y^2+z^2=xyz$, Math. Ann. 199 (1972), 213–228.

[R2] ———, *Von Untergruppen der Triangel Gruppen,* Illinois J. Math. 22 (1978), 404–413.

[R3] ———, *Eine Bemerkung zu einer Arbeit von T. Jørgensen,* Math. Z. 165 (1979), 261–265.

[R4] ———, *Some remarks on a paper of C. Doyle and D. James on subgroups of* SL(2, **R**), Illinois J. Math. 28 (1984), 348–351.

[R5] ———, *All generating pairs of all two-generator Fuchsian groups,* Arch. Math. (Basel) 46 (1986), 198–204.

[V] E. Van Vleck, *On the combination of non-loxodromic substitutions,* Trans. Amer. Math. Soc. 20 (1919), 299–312.

J. Gilman
Mathematics Department
Rutgers University
Newark, NJ 07102

B. Maskit
Mathematics Department
State University of New York
at Stony Brook
Stony Brook, NY 11794-3651