PAIRS OF REAL 2-BY-2 MATRICES THAT GENERATE FREE PRODUCTS

R. C. Lyndon and J. L. Ullman

1. INTRODUCTION

This note arose from the observation that certain results of Newman [4] follow very simply from a theorem of Macbeath [3]. A result of Brenner [1] follows by a similar argument.

We are concerned with the question when two real unimodular 2-by-2 matrices A and B generate a free group $\mathfrak{G}$, or a group $\mathfrak{G}$ that is the free product of the cyclic group $\mathfrak{A}$ generated by A and the cyclic group $\mathfrak{B}$ generated by B. We obtain several sufficient conditions for $\mathfrak{G}$ to be a free product. The conditions are stated in terms of the arrangement of the fixed points of A, B, AB, and $ABA^{-1}B^{-1}$ under the action of these matrices as linear fractional transformations on the extended real axis.

2. A LEMMA ON PERMUTATION GROUPS

We could obtain our results by applying Macbeath's theorem directly to the action of our matrices as linear fractional transformations acting on the open upper half of the complex plane. In the cases that we treat, this would also enable us to show that $\mathfrak{G}$ operates discontinuously on the upper half-plane and is therefore discrete. But we have preferred to recast Macbeath's theorem in the form of a lemma that enables us to confine attention to the action of $\mathfrak{G}$ on the extended real axis.

LEMMA. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be groups of permutations of a set Ω, and let $\mathfrak{G}$ be the group generated by $\mathfrak{A}$ and $\mathfrak{B}$ together. Suppose that Ω contains two disjoint nonempty sets Γ and Δ such that each nontrivial element of $\mathfrak{A}$ maps Γ into Δ and each nontrivial element of $\mathfrak{B}$ maps Δ into Γ. Then either $\mathfrak{G}$ is the free product of its subgroups $\mathfrak{A}$ and $\mathfrak{B}$, or else both $\mathfrak{A}$ and $\mathfrak{B}$ have order 2 and $\mathfrak{G}$ is a dihedral group.*

Proof. Suppose that $\mathfrak{A}$ has order greater than 2, hence more than one nontrivial element. Since the images ΓA of Γ under the nontrivial elements A of $\mathfrak{A}$ are disjoint nonempty subsets of Δ, it follows that each such ΓA is properly contained in Δ, that is, $\Gamma A < \Delta$. Let $W = A_1 B_1 \cdots A_n B_n$, where $n \geq 1$, and where $1 \neq A_i \in \mathfrak{A}$ and $1 \neq B_i \in \mathfrak{B}$ for all i. Then $\Gamma A_1 < \Delta$, whence $\Gamma A_1 B_1 < \Delta B_1 \leq \Gamma$; by a continuation of this argument, $\Gamma W < \Gamma$. Thus $W \neq 1$. This shows that $\mathfrak{G}$ is the free product of $\mathfrak{A}$ and $\mathfrak{B}$.

The same conclusion holds if $\mathfrak{B}$ has order greater than 2, and also (trivially) if either $\mathfrak{A}$ or $\mathfrak{B}$ has order 1. The case remains where $\mathfrak{A}$ is generated by an element A, and $\mathfrak{B}$ by B, with $A^2 = B^2 = 1$. Any further relations between A and B can be reduced to the form $(AB)^n = 1$ $(n > 0)$, and indeed to at most one such relation. If such a relation holds, $\mathfrak{G}$ is a dihedral group of order 2n. If none holds, $\mathfrak{G}$ is the infinite dihedral group, a free product of $\mathfrak{A}$ and $\mathfrak{B}$.

Received October 2, 1967.
This work was sponsored by the National Science Foundation.

This lemma has a converse. Suppose that $\mathfrak{G}$ is the free product of its subgroups $\mathfrak{A}$ and $\mathfrak{B}$, and that $\mathfrak{G}$ acts regularly on a set Ω (for example, on $\Omega = \mathfrak{G}$ under the regular representation). Then Ω contains sets Γ and Δ, as in the hypothesis of the lemma. Choosing any $p \in \Omega$, we let Γ be the set of all elements of the form pW, where W is a nontrivial element of $\mathfrak{G}$ with normal form ending in a factor from $\mathfrak{B}$, and we let Δ be the set of all elements pW, where W is a nontrivial element ending in a factor from $\mathfrak{A}$. If, moreover, $\mathfrak{G}$ is a countable group of real unimodular matrices acting on a real projective space E, then some point p in E is not a fixed point for any nontrivial element of $\mathfrak{G}$. Therefore $\mathfrak{G}$ acts regularly on the orbit $pG = \Omega$, and we may choose Γ and Δ as described. These sets Γ and Δ may have a very complicated geometric structure. In this note we have limited attention to the case where $\mathfrak{G}$ is the free product of two cyclic groups $\mathfrak{A}$ and $\mathfrak{B}$ of real unimodular 2-by-2 matrices, and where, under the action of $\mathfrak{G}$ on the extended real axis $\mathbb{R}^*$, the sets Γ and Δ can be chosen so that each is an open interval or the union of two open intervals.

COROLLARY (Newman). *Let* $A = \begin{pmatrix} -a & b \\ -c & d \end{pmatrix}$ *and* $B = \begin{pmatrix} -e & -f \\ g & h \end{pmatrix}$, *where*

$$a, \cdots, h \geq 0, \qquad \det A = \det B = 1, \qquad |\mathrm{Tr}\, A|, |\mathrm{Tr}\, B| \geq 2.$$

Then A *and* B *generate a free group.*

Proof. Let $\mathfrak{A}$ be the group generated by A, $\mathfrak{B}$ that generated by B, and $\mathfrak{G}$ that generated by A and B together. Let $\mathfrak{G}$ act as a group of linear fractional transformations on $\mathbb{R}^*$. Let $\Gamma = (\infty, 0)$ and $\Delta = (0, \infty)$. Evidently, $z < 0$ implies that $zA > 0$, while $z > 0$ implies that $zB < 0$; that is, $\Gamma A \leq \Delta$ and $\Delta B \leq \Gamma$. Since $|\mathrm{Tr}\, A| \geq 2$, A must have two real fixed points (possibly coincident), and they must lie in $[0, \infty]$. The points $0, 0A, 0A^2, \cdots$ must converge monotonically to one of the fixed points of A, hence all lie in $[0, \infty]$. Therefore all of $\Gamma A, \Gamma A^2, \cdots$ are contained in Δ. Since $\Gamma \cap \Gamma A = \emptyset$ implies that $\Gamma \cap \Gamma A^{-1} = \emptyset$, we see that $\Gamma A^{-1} \leq \Delta$, and a similar argument shows that all negative powers of A map Γ into Δ. Similarly, $\Delta B^h \leq \Gamma$ for all $h \neq 0$.

3. REAL, NONALTERNATING FIXED POINTS

THEOREM. *Let* A, B, *and* C = AB *be real unimodular* 2*-by-*2 *matrices, all with real fixed points. Suppose that the fixed points of each of these matrices, under action on* $\mathbb{R}^*$, *lie in an interval of* $\mathbb{R}^*$ *containing no fixed point of the other two. Then* A *and* B *generate a free group.*

Proof. It is convenient to change notation by replacing B by B^{-1}. The hypothesis now asserts that $C = AB^{-1}$ has real fixed points, that is, points p such that $pA = pB$. If A has distinct fixed points, one is a source a^- and one a sink a^+; if A is parabolic, we take $a^- = a^+$ to be its sole fixed point. We employ a similar notation b^- and b^+ for the fixed points of B. The points a^-, a^+, b^-, and b^+ divide $\mathbb{R}^*$ into four open intervals (some possibly empty). The hypothesis requires that both fixed points (possibly coincident) of C lie in the same one of these intervals, and not in one bounded by two fixed points of A or by two fixed points of B.

If $p \in (a^+, b^+)$, then $pA \in (a^+, p)$ and $pB \in (p, b^+)$. Since $(a^+, p) \cap (p, b^+) = \emptyset$, it is impossible that $pA = pB$. Thus no fixed point of C lies in an interval (a^+, b^+), or, similarly, in an interval (b^+, a^+).

Suppose that (a^-, b^+) is one of the four intervals. Then the fixed points occur in cyclic order a^-, b^+, b^-, a^+. Evidently, both A and B map $[a^-, b^+]$ into $[a^-, a^+]$. Now $a^- A = a^-$ precedes $a^- B$, in the natural order of elements of $[a^-, a^+]$. Also, $b^+ B = b^+$ precedes $b^+ A$. Thus the number of solutions p (counting multiplicity) of the equation $pA = pB$ in the interval (a^-, b^+) must be odd. Thus C cannot have both its fixed points in this interval, and hence it has no fixed point in (a^-, b^+). A similar argument applies to (b^-, a^+), (b^+, a^-), and (a^+, b^-).

It follows that C has both its fixed points in an interval (a^-, b^-) or both in an interval (b^-, a^-). By symmetry, we may suppose that C has a fixed point p in (a^-, b^-). Let $q = pA = pB$. The fixed points of A and B must occur in the cyclic order a^-, b^-, b^+, a^+. Since $pA \in (p, a^+)$ and $pB \in (b^+, p)$, we find that

$$q = pA = pB \in (p, a^+) \cap (b^+, p) = (b^+, a^+).$$

Let $\Gamma = (p, q)$ and $\Delta = (q, p)$. We have the cyclic order a^-, p, b^-, b^+, q, a^+. Since $pA = q$, it follows that $\Gamma A^h \leq (q, a^+) < \Delta$ for all $h > 0$. Also, $qA^{-1} = p$ implies that $\Gamma A^h \leq (a^{-1}, p) < \Delta$ for all $h < 0$. Similarly, $\Delta B^k < \Gamma$ for all $k \neq 0$. The lemma applies, and the theorem is proved.

The following result, for the case where $m = n$, was obtained by Brenner.

COROLLARY. *If* m *and* n *are real and* $|mn| \geq 4$, *then* $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$ *generate a free group.*

Proof. Replacing A by $A^{-1} = \begin{pmatrix} 1 & -m \\ 0 & 1 \end{pmatrix}$, if necessary, we may suppose that $mn \leq -4$. A has a single fixed point ∞, and B a single fixed point 0. The matrix $C = AB = \begin{pmatrix} 1+mn & m \\ n & 1 \end{pmatrix}$ has trace $2 + mn \leq -2$; hence C has real fixed points. These are the roots of the equation $zC = z$, that is, of the equation $nz^2 - mnz - m = 0$, and since $-m/n > 0$, they have the same sign. The fixed points of C thus lie in an interval containing neither 0 nor ∞, and the preceding theorem applies.

However, this result (without the restriction that m and n be real) can be obtained from the following stronger result of Brenner. We note that Brenner's result was improved by Chang, Jennings, and Ree [2], and that the present authors have strengthened it further, in a paper that will appear later. The corollaries below are due to Chang, Jennings, and Ree.

PROPOSITION. *If* $m \in \mathbb{C}$ *and* $|m| \geq 2$, *then* $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 0 \\ m & 1 \end{pmatrix}$ *freely generate a free group.*

Proof. Let $C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then $C^2 = 1$ and $CAC = B^{-1}$. Let $\mathfrak{G}$ be the group generated by A and C, and let $\mathfrak{A}$ and $\mathfrak{C}$ be the cyclic subgroups with generators A and C. The lemma applies, if we let Γ be the interior of the unit circle and Δ its exterior. Thus $\mathfrak{G}$ is the free product of $\mathfrak{A}$ and $\mathfrak{C}$. It follows that A and B freely generate a free subgroup of $\mathfrak{G}$.

COROLLARY. *If* $m, n \in \mathbb{C}$ *with* $|mn| \geq 4$, *then* $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$ *generate a free group.*

Proof. For $0 \neq d \in \mathbb{C}$, conjugation by $D = \begin{pmatrix} d & 0 \\ 0 & 1/d \end{pmatrix}$ replaces m and n by $m' = d^2 m$ and $n' = d^{-2} n$, leaving $|mn|$ unchanged. Choosing d so that $m' = n'$, we see that $|m'^2| \geq 4$ and hence $|m'| \geq 2$.

COROLLARY. *Let* A *and* B *be* 2*-by-*2 *complex unimodular matrices. If* A *and* B *are parabolic and* $|\mathrm{Tr}\ AB - 2| \geq 4$, *then* A *and* B *generate a free group.*

Proof. The fixed points a of A and b of B are distinct, since otherwise $|\mathrm{Tr}\ AB| = 2$. After conjugation we may suppose that $a = \infty$ and $b = 0$, whence $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$ for some m and n in $\mathbb{C}$. Now $\mathrm{Tr}\ AB = 2 + mn$, and it follows that $|mn| \geq 4$.

4. REAL, ALTERNATING FIXED POINTS

THEOREM. *Let* A *and* B *be two hyperbolic linear fractional transformations with distinct fixed points such that every interval containing both fixed points of one transformation contains also a fixed point of the other. Suppose that the commutator* $C = ABA^{-1}B^{-1}$ *has a real fixed point. Then* A *and* B *generate a free group.*

Proof. The fixed points a^- and a^+ of A, together with the fixed points b^- and b^+ of B divide $\mathbb{R}^*$ into four quarters, each with one endpoint $a^{\pm}$ and one $b^{\pm}$. This, as well as the further hypothesis that pAB = pBA for some real p, is clearly preserved under interchange of A and B. Moreover, if pAB = pBA = q, then $qA^{-1}B^{-1} = qB^{-1}A^{-1} = p$, whence the hypothesis remains valid under simultaneous replacement of A by A^{-1} and B by B^{-1}.

We shall now show that pAB = pBA implies that p lies in one of the two quarters with endpoint b^-. It will then follow that such p must lie in a quarter with endpoint a^-. By symmetry, we may then suppose that a quarter (a^-, b^-) contains p. By the last sentence in the preceding paragraph, we conclude also that $q \in (a^+, b^+)$.

To begin, we conjugate by a transformation carrying the fixed points of A into 0 and ∞. Then A will be given by $zA = kz$, for some $k > 0$ $(k \neq 1)$. Now B will be given by some matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with real entries and $ad - bc = 1$. We are free to take $c \geq 0$. The fixed points of B are the roots of $cx^2 + (d - a)z - b = 0$, and since they separate the fixed points 0 and ∞ of A, they are of opposite sign, whence $bc > 0$ and $b, c > 0$. The relation $ad = bc + 1 > 0$ implies that $ad > 0$. Since A^{-1} satisfies the same hypotheses as A, while B^{-1} is given by the matrix $\begin{pmatrix} -d & b \\ c & -a \end{pmatrix}$ with $c \geq 0$, and since $d \leq 0$ implies $-a \geq 0$, we may suppose $d > 0$. Thus $a, b, c, d > 0$.

Since $0B = b/d > 0$ and $\infty B = a/c > 0$, we see that $(0, \infty)B \subseteq (0, \infty)$, whence $b^- < 0 < b^+$. Now, if zAB = zBA, then

$$\frac{akz + b}{ckz + d} = k\,\frac{az + b}{cz + d},$$

$$(cz + d)(akz + b) = (ckz + d)(akz + bk),$$

$$ackz^2 + (adk + bc)z + bd = ack^2 z^2 + (adk + bck^2)z + bdk,$$

$$ac(k^2 - k)z^2 + bc(k^2 - 1)z + bd(k - 1) = 0.$$

Since $k \neq 1$, this is equivalent, upon division by $k - 1$, to the relation

$$ackz^2 + bc(k + 1)z + bd = 0.$$

Since all the coefficients of this equation are positive, the roots are negative; therefore $pAB = pBA$ implies that p lies in one of the quarters with endpoint b^-.

The argument given earlier now shows that we have points

$$b^-,\ p,\ a^-,\ b^+,\ q = pAB = pBA,\ a^+$$

in this cyclic order on $\mathbb{R}^*$. It follows further that we have the points

$$b^-,\ p,\ a^-,\ pB,\ b^+,\ q,\ a^+,\ pA,$$

in this cyclic order. Now let

$$\Gamma = (pB,\ q = pBA) \cup (pA,\ p) \qquad \text{and} \qquad \Delta = (p,\ pB) \cup (q = pAB,\ pA).$$

It follows immediately that the hypotheses of the lemma hold.

5. TWO ELLIPTIC ELEMENTS

If $\mathfrak{A}$ is generated by an elliptic transformation A of infinite order, then, for any real p, the set of images of p under $\mathfrak{A}$ is dense in $\mathbb{R}^*$, and we have no hope of applying the lemma with Γ containing any open interval. Therefore we confine our attention to elliptic transformations A of finite order m. We call A *minimal* if, viewed as a rotation of the hyperbolic upper half-plane, it has the least positive angular displacement of any element of $\mathfrak{A}$. In real terms, this means that $|\mathrm{Tr}\ A|$ is maximal for all nontrivial elements of $\mathfrak{A}$, or, more explicitly, that for each real p the points $p, pA, \cdots, pA^{m-1}$ occur on $\mathbb{R}^*$ in that cyclic order. Clearly, every nontrivial cyclic subgroup of elliptic transformations has a (unique) minimal generator.

THEOREM. *Let* A *and* B *be minimal elliptic transformations. If* $AB \neq 1$ *and* AB *has real fixed points, then the group* $\mathfrak{G}$ *generated by* A *and* B *is the free product of the subgroup* $\mathfrak{A}$ *generated by* A *together with the subgroup* $\mathfrak{B}$ *generated by* B.

Proof. If p is a fixed point of AB, then $pA = q$ with $qB = p$. Let $\Gamma = (p, q)$ and $\Delta = (q, p)$. Since the images of p under $\mathfrak{A}$ occur in the cyclic order $p, pA, \cdots, pA^{m-1}$, it is clear that $\Gamma A^h \leq \Delta$ whenever $A^h \neq 1$. Similarly, whenever $B^k \neq 1$, $\Delta B^k \leq \Gamma$. The lemma applies, and, even if A and B are involutions, the hypotheses ensure that AB has infinite order, whence $\mathfrak{G}$ is the free product of $\mathfrak{A}$ and $\mathfrak{B}$.

6. THE MIXED CASE

THEOREM. *Let* A *be a minimal elliptic transformation, and* B *a transformation with real fixed points. Suppose that* AB *has a real fixed point* p *such that the fixed points of* B *lie in the interval* (p, pA). *Then the group* $\mathfrak{G}$ *generated by* A *and* B *is the free product of its subgroup* $\mathfrak{A}$ *generated by* A *together with the subgroup* $\mathfrak{B}$ *generated by* B.

Proof. As before, $pA = q$ and $qB = p$, with $p, pA, \cdots, pA^{m-1}$ occurring in that cyclic order. Again we take $\Gamma = (p, q)$ and $\Delta = (q, p)$. It follows as before that $A^h \neq 1$ implies that $\Gamma A^h \leq \Delta$. From the assumption that the fixed points of B lie in (p, q), and the fact that $qB = p$, it follows that $\Delta B^k \leq \Gamma$ for all $k \neq 0$. The lemma applies.

REFERENCES

1. J. L. Brenner, *Quelques groupes libres de matrices.* C. R. Acad. Sci. Paris 241 (1955), 1689-1691.
2. B. Chang, S. A. Jennings, and R. Ree, *On certain pairs of matrices which generate free groups.* Canad. J. Math. 10 (1958), 279-284.
3. A. M. Macbeath, *Packings, free products and residually finite groups.* Proc. Cambridge Philos. Soc. 59 (1963), 555-558.
4. M. Newman, *Pairs of matrices generating discrete free groups and free products.* Michigan Math. J. 15 (1968),

The University of Michigan
Ann Arbor, Michigan 48104