

GROUPS OF THE SECOND KIND WITHIN THE MODULAR GROUP

BY

W. W. STOTHERS

The modular group Γ can be studied using (hyperbolic) geometry, group theory or graph theory. The division of the subgroups into those of the first and of the second kind has always been described in geometrical terms. We provide other criteria. We begin by discussing the various views of the whole group.

Geometrically, Γ is $LF(2, \mathbf{Z})$, the group of integral bilinear transformations of the Poincaré plane

$$\mathcal{H} = \{z \in \mathbf{C}: \text{Im}(z) \geq 0\} \cup \{\infty\}.$$

Is generated by $x = \pm(0 \ -1 \mid 1 \ 0)$ and $y = \pm(0 \ -1 \mid 1 \ 1)$. In fact,

$$\Gamma = \langle x, y: x^2 = y^3 = I \rangle, \quad (1)$$

so that each element of Γ can be written uniquely as a word in x , y and y^2 . The element $u = \pm(1 \ 1 \mid 0 \ 1)$ is very significant. The elements x , y and u generate, respectively, the Γ -stabilizers of i , $\phi = \frac{1}{2}(-1 + i\sqrt{3})$ and ∞ .

Let G be a subgroup of index n in Γ ; we allow $n = \infty$. The Γ -orbits of i , ϕ and ∞ each split into G -orbits. The G -stabilizers of points in the same G -orbit are clearly G -conjugate. Let r (resp. s , h_0) be the number of G -orbits in Γi (resp. $\Gamma \phi$, $\Gamma \infty$) which have non-trivial G -stabilizers. Let h_∞ be the number of other G -orbits in $\Gamma \infty$. These are the geometrical definitions of the parameters which appear in [6].

As a group, Γ is the free product of cyclic groups of order two and three, see (1). By a careful application of the ideas behind the Kurosh subgroup theorem, the subgroup G has a presentation

$$\langle x(1), \dots, x(r), y(1), \dots, y(s), u(1), \dots, u(h_0), a(1), \dots, a(t_1):$$

$$x(1)^2 = \dots = x(r)^2 = y(1)^3 = \dots = y(s)^3 = w = I \rangle, \quad (2)$$

where w is a known word in the generators, and is trivial when $n = \infty$. In (2), each $x(j)$ (resp. $y(j)$) is a Γ -conjugate of x (resp. y), and each $u(j)$ a Γ -conjugate of $u^{m(j)}$ for some non-zero integer $m(j)$. The parameter t_1 is equal to twice the genus of \mathcal{H}/G when n is finite. The existence of the presentation (2) is proved group-theoretically in [2], [3]. For n finite, it is proved geometrically in [4], for n infinite it is proved graph-theoretically in [6]. The parameters r , s , and h_0

Received July 31, 1979.

appear in (2), though it is not yet clear that they are well defined. The parameter h_∞ is less natural in this approach. It can be taken as the number of equivalence classes of cosets Gv having $G^v \cap \langle u \rangle = \{I\}$, where $Gv_1 \sim Gv_2$ if and only if $Gv_1 = Gv_2 u^m$ for an integer m . Here, as later, we write G^v for the subgroup vGv^{-1} when $G \leq \Gamma$ and $v \in \Gamma$. We can give similar definitions for r, s , and h_0 , thus proving that they are well defined.

Using graph theory, we study the subgroups of Γ by their coset diagrams, see [6], [7]. Coset diagrams are edge-coloured, (partially) directed pseudographs. That for the subgroup G has vertex set $\{Gv: v \in \Gamma\}$ with directed red (resp. blue) edges to indicate the effect on the cosets of post-multiplication by x (resp. y). As Γ is generated by x and y , these edges are sufficient to determine the effect of post-multiplication by any element of Γ . In particular, if $Gvx = Gw$ and $Gwy = Gz$, then, as $u = xy$, $Gvu = Gz$. It is often convenient to add green edges to indicate the effect of u . By the preceding remark, a green edge is equivalent to a red edge then the following blue edge. As $x^2 = I$, the red edges form loops and digons. For simplicity, we replace each (directed) digon by a single undirected red edge between the relevant vertices. As $y^3 = I$, the blue edges form loops and triangles. The usual parameters can be determined from the coset diagram with added green edges: r (resp. s) is the number of red (resp. blue) loops, h_0 (resp. h_∞) is the number of finite (resp. infinite) green polygons. The index n is, of course, equal to the number of vertices.

DEFINITION 1. If $v = \pm(a \ b | c \ d) \in LF(2, \mathbf{R})$ has $c(v) = |c| = 0$, then the isometric circle of v is $\mathcal{I}(v) = \{z \in \mathbf{C}: |cz + d| \leq 1\}$.

By direct calculation, $c(u^m v u^{-m}) = c(v)$ and

$$\mathcal{I}(u^m v u^{-m}) = \{z + m: z \in \mathcal{I}(v)\}. \tag{3}$$

We note that $\mathcal{I}(v)$ has center $-d/c (\in \mathbf{R})$ and has radius $c(v)^{-1}$. For $v \in \Gamma$, $c(v)$ is an integer and so, when it is defined, $\mathcal{I}(v)$ has radius at most one.

Suppose that $G \leq LF(2, \mathbf{R})$ is such that $c(v) \neq 0$ whenever $v \in G - \{I\}$. Then, by Section 20 of [1], G has fundamental domain

$$\mathcal{F}(G) = \mathcal{H} - \bigcup \{\mathcal{I}(v): v \in G - \{I\}\}. \tag{4}$$

From (3) and the preceding remark, $\mathcal{F}(u^m G u^{-m})$ is defined whenever $\mathcal{F}(G)$ is defined, and is equal to the latter shifted by m units to the right.

Every discrete subgroup G of $LF(2, \mathbf{R})$ has a fundamental domain $\mathcal{F}^*(G)$ defined using isometric circles (though not, in general, by (4)), see p. 77 of [1]. We may as well assume that $\mathcal{F}^*(G) = \mathcal{F}(G)$ whenever the latter is defined. Much as above, a suitable domain for G^v can be obtained by applying v to one for G .

DEFINITION 2. A discrete subgroup G of $LF(2, \mathbf{R})$ is of the second kind if and only if $\mathcal{F}^*(G)$ contains an open subset in $\mathbf{R} \cup \{\infty\}$. A maximal such set is called a free side.

Each element of $LF(2, \mathbf{R})$ is an open mapping preserving $\mathbf{R} \cup \{\infty\}$, so that G^v is of the second kind if and only if G is.

We observe that $v \in LF(2, \mathbf{R})$ fixes ∞ if and only if $c(v) = 0$. As $\langle u \rangle$ is the Γ -stabilizer of ∞ , $v \in \Gamma$ has $c(v) = 0$ if and only if $v = u^m$ for some integer m . Thus, for $v \in \Gamma$, $\mathcal{F}(G)$ is defined if and only if $\langle u \rangle \cap G = \{I\}$.

DEFINITION 3. Let $G = LF(2, \mathbf{R})$. A point $z \in \mathcal{H}$ is an ordinary point if it is not fixed by an element of $G - \{I\}$.

Then $\mathcal{F}(G)$ is defined if and only if ∞ is an ordinary point of G . Finally, each point of a free side of a fundamental domain is an interior point (in the topology induced on \mathcal{H}) and so is an ordinary point of the subgroup.

We shall see that the subgroups of the second kind cannot be characterised by the values of the parameters defined above. It is known that G is of the first kind if either n is finite, p. 109 of [4], or $n = \infty, h_0 = 0, h_\infty = 1$ [8]. Corollary 4 below shows that a subgroup of the second kind must have $h_\infty = \infty$. This includes the above results, the former since a subgroup of finite index contains a power of each Γ -conjugate of u , so that $h_\infty = 0$ for such a subgroup, e.g., by the group characterisation above.

THEOREM 1. *A subgroup G of Γ is of the second kind if and only if, for some $\alpha \in \mathbf{R}$, there is an element $v \in \Gamma$ such that $\mathcal{F}(G^v)$ is defined and $(-\infty, \alpha) \subseteq \mathcal{F}(G^v)$.*

Proof. Suppose that G is of the second kind. Then $\mathcal{F}^*(G)$ includes an open interval I of R . There is a rational point, *dc* say, in I . We may as well assume that c and d are coprime, so that there are integers a and b with $ad + bc = 1$. Then $v = \pm \begin{pmatrix} a & b \\ -c & d \end{pmatrix} \in \Gamma$. As d/c is an ordinary point of G , $\infty = v(d/c)$ is an ordinary point of G^v , see p. 87 of [4]. Then, by an earlier remark, $\mathcal{F}(G^v)$ is defined. As ∞ is on a free side of $F(G^v)$, there is an $\alpha \in \mathbf{R}$, with $(-\infty, \alpha) \subseteq \mathcal{F}(G^v)$.

Conversely, suppose that $(-\infty, \alpha) \subseteq \mathcal{F}(G^v)$ for some $\alpha \in \mathbf{R}, v \in \Gamma$. As $(-\infty, \alpha)$ is open, G^v and so also G is of the second kind.

We note that the proof works, with minor modifications, for any Δ of the first kind. In such a group, the stabilizer of ∞ is of the form $\langle \pm \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \rangle$ with $k > 0$, and $\Delta\infty$ is dense in \mathbf{R} .

From (1), each element of Γ is a word in x, y, y^2 . We split $\Gamma - \{I\}$ into four disjoint classes:

$$C_1 = \{x \cdots x\} \cup \{x\},$$

$$C_2 = \{y^a \cdots x : a \in \{1, 2\}\},$$

$$C_3 = \{x \cdots y^a : a \in \{1, 2\}\},$$

$$C_4 = \{y^a \cdots y^b : a, b \in \{1, 2\}\} \cup \{y, y^2\}.$$

We note that, if $w \in C_2$, then $w^{-1} \in C_3$, and conversely. Also, if $w = \pm(a \ b \ c \ d)$ has $c \neq 0$, then the signs of the centres of $\mathcal{F}(w)$ and $\mathcal{F}(w^{-1})$ are determined by those of a, c and d . We show that these signs are determined by the class C_j to which w belongs.

Notation. Let \mathcal{P} (resp. \mathcal{Q}) denote the set of non-negative (resp. non-positive) integers. Write $(X_1 \ X_2 \ X_3 \ X_4)$ for the set of $(a_1 \ a_2 \ a_3 \ a_4)$ with $a_j \in X_j, j = 1, 2, 3, 4$, and each $X_j = \mathcal{P}$ or \mathcal{Q} . Finally, write $w \in (X_1 \ X_2 \ X_3 \ X_4)$ if $w = \pm(a \ b \mid c \ d)$ with $(a \ b \ c \ d)$ or $(-a \ -b \ -c \ -d) \in (X_1 \ X_2 \ X_3 \ X_4)$. We refer to $(X_1 \ X_2 \ X_3 \ X_4)$ as a *sign pattern*.

LEMMA. $C_1 \subseteq (\mathcal{Q} \ \mathcal{P} \ \mathcal{Q} \ \mathcal{P}), C_2 \subseteq (\mathcal{P} \ \mathcal{Q} \ \mathcal{Q} \ \mathcal{P}), C_3 \subseteq (\mathcal{P} \ \mathcal{P} \ \mathcal{P} \ \mathcal{P}),$
 $C_4 \subseteq (\mathcal{Q} \ \mathcal{Q} \ \mathcal{P} \ \mathcal{P}).$

Proof. We outline a proof by induction on the word length.

It is easy to verify the result for words of length one or two, i.e., $x, y, y^2, xy, xy^2, yx, y^2x$.

The induction step is accomplished by considering the pre-multiplication of elements of C_2, C_4 by x , and of C_1, C_3 by y and y^2 . For example, if $w = \pm(a \ b \mid c \ d) \in C_2$, then, replacing each entry by its negative if necessary, we have $a, d \in \mathcal{P}$ and $b, c \in \mathcal{Q}$. Then $xw = \pm(c \ d \mid -a \ -b) \in C_1$, as required.

THEOREM 2. *A subgroup G of Γ is of the second kind if and only if there is an element $v \in \Gamma$ with $G^v \subseteq \{I\} \cup C_1$.*

Proof. Suppose that $G^v \subseteq \{I\} \cup C_1$. If $G^v \cap \langle u \rangle \neq \{I\}$, then there is a positive integer m with u^m and hence $u^{-m} = \pm(1 \ -m \mid 0 \ 1) \in G^v$. But $u^{-m} \in C_2$, contradicting our hypothesis. Hence $w \in G^v - \{I\}$ has $c(w) \neq 0$, so that $\mathcal{F}(G^v)$ is defined. By the definition of $C_1, w \in G^v - \{I\}$ has $\mathcal{I}(w)$ having non-negative center and radius at most one. Then $\mathcal{I}(w)$ lies entirely to the right of the line $\text{Re}(z) = -1$. Now, by (4), $(-\infty, -1) \subseteq \mathcal{F}(G^v)$. By Theorem 1, G is of the second kind.

Now suppose that G is of the second kind. By Theorem 1, there are $\alpha \in \mathbf{R}, v' \in \Gamma$ with $(-\infty, \alpha) \subseteq \mathcal{F}(G^{v'})$. We can choose an integer m with $\alpha + m < 1$. Then, by the remarks after (4), $(-\infty, 1) \subseteq \mathcal{F}(G^{v'})$ with $v = u^m v'$. Let $w = \pm(a \ b \ c \ d) \in G^v - \{I\}$. Then the center of $\mathcal{I}(w)$ lies to the right of $\text{Re}(z) = 1$, see (4), so that without loss of generality, $d > -c \geq 1$. By the lemma, $w \in C_1$ or C_2 . Now $w^{-1} = \pm(d \ -b \mid -c \ a) \in G^v$ so that, if $w \in C_2$, then $a \geq 0$ and $\mathcal{I}(w^{-1})$ has center $-(a/-c) = a/c \leq 0$ as $c < 0$. This contradicts the fact that the centers of isometric circles for elements of G^v are positive. Hence $w \in C_1$, as required.

In [6] we defined X , the D_1 -diagram (i.e. the coset diagram without green edges) for the subgroup $\langle x \rangle$ of Γ . This diagram is central to the graph-theoretic characterisation, so we shall describe it in some detail. We construct X in stages. We begin with a directed blue triangle $T(0, 1)$, with a red loop at a

vertex P . P will be the distinguished vertex of the final diagram. At the k th stage, $k > 0$, we have a diagram consisting of $2^k - 1$ directed blue triangles, 2^{k-1} of which have two vertices with no red edges attached. We add 2^k new directed blue triangles, $T(k, 1), \dots, T(k, 2^k)$. Each is added by drawing a red edge from one of its vertices, $P(k, j)$, to one of the vertices of the original (i.e. k th stage) triangles with only red edge attached. The new structure has $2^{k+1} - 1$ triangles, the original $2^k - 1$ now have red edges at all vertices, the 2^k new ones have a red edge at only one vertex. We note that, at each stage, we get a connected diagram. Further, for $K > 0$, the edges at the vertices of the $T(k, j)$, $k < K$, remain unaltered after the $(K + 1)$ st stage is complete. By then, each such vertex is on one directed blue triangle and is involved in one red edge. Ultimately, the process yields an infinite pseudograph X which is the D_1 -diagram of some subgroup of Γ . To establish the properties we require, we follow the procedure of Section 2 of [6]. We delete from X the red loop at P ($= P(0, 1)$), and, from each blue triangle, the blue edge not meeting $P(k, j)$, $k \geq 0$, $0 \leq j \leq 2^k$. This produces a pseudograph X' . We note that X' is connected.

LEMMA. X' is a tree.

Proof. We must show that any cycle at P in X' must involve an edge immediately followed by the same edge taken in the opposite direction. We refer to such a pair as an excursion. It then follows that each cycle is essentially trivial, i.e., reduces to the null cycle once we remove all excursions.

Suppose that \mathcal{C} is a cycle at P in X' . Since \mathcal{C} is finite, there is a greatest integer k such that \mathcal{C} visits a vertex of triangle $T(k, j)$ for some j . First suppose that $k = 0$. Then \mathcal{C} is a cycle in the sub-pseudograph consisting of two edges of $T(0, 1)$. As this is a tree, \mathcal{C} is essentially trivial. Now suppose that k is positive. Suppose, further, that the first triangle $T(k, -)$ visited by \mathcal{C} is $T(k, j)$, and that \mathcal{C} enters at the vertex V . Since all triangles visited earlier by \mathcal{C} are $T(K, -)$ with $K < k$, the definition of X shows that $V = P(k, j)$ —the other vertices are linked to triangles of the form $T(k + 1, -)$. Similarly, \mathcal{C} must leave $T(k, j)$ for the first time (and each subsequent time) at the vertex $P(k, j)$. Hence \mathcal{C} involves at least one subcycle at $P(k, j)$ entirely within $T(k, j) \cap X'$. Since this consists of just two edges at $P(k, j)$, the subcycle is essentially trivial. It is therefore null or involves an excursion. In the latter case, we are done. In the former case, since the only edge at $P(k, j)$ other than those of $T(k, j)$ is the red edge from some $T(k - 1, j')$, \mathcal{C} must contain an excursion involving this edge. Hence, in either case, \mathcal{C} has an excursion.

Since X' is connected, it must now be a spanning tree for X . In the language of [6], it is a T_1 for X . By Proposition 1.2 of [6], the subgroup G corresponding to X is generated by x which corresponds to the red loop at P . By Theorem 1.3 of [6], G has $n = \infty$, $r = 1$, $s = 0$. By Lemma 2.5, and formula (5) of that paper, it has $h_0 = t_1 = 0$. Finally, by Lemma 4.2, $h_\infty = \infty$.

As we have seen, X has a red loop. It follows that it can be joined (in the sense defined in [6]) to any other diagram having a red loop. In this context, the null pseudograph is allowed as one of the diagrams to be joined.

THEOREM 3. *A subgroup G of Γ is of the second kind if and only if the coset diagram for G can be obtained by joining X to another diagram with a red loop.*

Proof. We show that the condition on diagrams is equivalent to that in the statement of Theorem 2.

For convenience, we discuss D_1 -diagrams with all loops and one edge of each blue triangle deleted, like X' in the above lemma.

Suppose that the diagram D' for G is obtained by joining X' to a diagram E' in which a red loop has been deleted from vertex Q . After the remarks preceding Theorem 1.10 of [7], we may replace G by a conjugate if necessary and assume that P is the special vertex of D' . As in [6], the elements of G correspond to routes (cycles without excursions) at P in D' . The only edge joining a vertex of X' to one of E is the red edge PQ introduced in the process of joining. This edge corresponds to x . It is clear that any cycle at P can be viewed as a sequence of subcycles, alternately within X' and of the form $PQ \cdot \mathcal{C} \cdot QP$, with \mathcal{C} within E' . If the cycle is a route, then those of the former type are trivial, since a route and hence any subcycle has no excursions. It follows that a non-trivial route must consist of a single cycle of the latter type. The corresponding word has the form $x \cdots x$. Thus, $G - \{I\} \subset C_1$.

The converse is similar.

COROLLARY 4. *A subgroup of the second kind in Γ has $h_\infty = \infty$.*

Proof. This follows from the theorem, the fact that $\langle x \rangle$ has $h_\infty = \infty$ (proved above), and Lemma 3.2 of [6], which shows that the h_∞ of the join of X and E is equal to the sum of those of X and E .

COROLLARY 5. *A finitely generated subgroup of Γ is of the first kind if and only if it is of finite index.*

Proof. A subgroup of finite index has a finite diagram. Since X is infinite, any diagram produced by joining using X is also infinite. Hence a group of finite index is of the first kind.

Suppose now that we have a finitely generated subgroup of infinite index. Since it is finitely generated, the coset diagram has finitely many loops and finitely many non-trivial cycles other than blue triangles, see Theorem 1.3, Lemma 2.4 of [6]. If we cut (see [6]) a red edge outside the finite region containing the basic cycles and loops, then the diagram must split into two. Also, one component has one red loop, introduced by cutting, and no other loops or non-trivial cycles. This is the diagram X . Hence the original diagram can be made by a join involving X , so the subgroup is of the second kind.

Previous proofs of this result have used the hyperbolic area of the fundamental domains; see p. 51 of [5].

COROLLARY 6. *A non-trivial normal subgroup of Γ is of the first kind.*

Proof. From Theorem 1.10 of [7], a coset diagram without a distinguished vertex corresponds to a conjugacy class of subgroups; each choice of distinguished vertex determines a member of the conjugacy class. As the diagram for a subgroup is unique, any choice of vertex on the diagram of a normal subgroup produces the same picture.

Suppose that G is a normal subgroup of the second kind. Let D be the diagram for G . Then, by Theorem 3, D is the join of X and a diagram E , with P on X linked to Q on E say. If we consider the vertices P then Q as distinguished vertices for D , then, by the above symmetry, we see that $X \cong E$ as coset diagrams. Then, much as in the proof of the theorem, we see that G has no non-trivial generators, so that $G = \{I\}$.

Previous proofs have been geometrical; see Exercise 1 on p. 21 of [5], noting that a "horocyclic group" is a group of the first kind.

The condition in Corollary 4 is not sufficient: there are subgroups of the first kind with $h_\infty = \infty$.

Example. We begin by describing another of the coset diagrams from [6], the diagram C of Section 4. The blue edges of C form directed triangles $P_n Q_n R_n$, $n \geq 0$. For each n , there is a red edge from P_n to Q_{n+1} . All other vertices have red loops. Then $P_0 Q_0 R_0$ has two red loops, at Q_0 and R_0 , and, for $n > 0$, $P_n Q_n R_n$ has one red loop, at R_n . The corresponding subgroup clearly has $r = \infty$, $s = 0$. By an analysis similar to that for X , we find that $h_0 = t_1 = 0$. A sketch showing the associated green edges demonstrates that $h_\infty = 1$.

We add to C countably many more copies of C , labeled $C(0)$, $C(1)$, ... to form a diagram D . We join Q_0 on $C(n)$ to R_n on C . If we consider the joins being made successively and apply Lemma 3.2 of [6] each time, then we see that G , the subgroup corresponding to D , has $h_\infty = \infty$. Each blue triangle of D belongs to a $C(n)$, in which case there is a red loop at the vertex R_n , or to C , in which case it is adjacent to a triangle with a red loop (on the adjacent $C(n)$). Now consider X . The triangle $T(2, 1)$ has no red loop and is adjacent to triangles $T(k, j)$ with $k = 1$ or 3 , none of which have red loops. As joining cannot create red loops, our diagram D cannot be the join of X and another diagram. Thus, by Theorem 3, the subgroup G is of the first kind.

REFERENCES

1. L. R. FORD, *Automorphic functions*, McGraw-Hill, New York, 1929.
2. A. HOARE, M. HOWARD, A. KARRASS, and D. SOLITAR, *Subgroups of finite index in Fuchsian groups*, Math. Zeitschr., vol. 120 (1971), pp. 289–298.
3. ———, *Subgroups of infinite index in Fuchsian groups*, Math. Zeitschr., vol. 125 (1972), pp. 59–69.

4. J. LEHNER, *Discontinuous groups and automorphic functions*, Amer. Math. Soc., Providence, R.I., 1964.
5. ———. *A short course in automorphic functions*, Holt Rinehart and Winston, New York, 1966.
6. W. W. STOTHERS, *Subgroups of infinite index in the modular group*, Glasgow Math. J., vol. **19** (1978), pp. 33–43.
7. ———. *Diagrams associated with subgroups of Fuchsian groups*, Glasgow Math. J., vol. **20** (1979), pp. 103–114.
8. C. TRETAKOFF, *Non-parabolic subgroups of the modular group*, Glasgow Math. J., vol. **16** (1975), pp. 91–102.

UNIVERSITY OF GLASGOW
GLASGOW, SCOTLAND