Tôhoku Math. Journ.
26 (1974), 203-218.

GENERATORS OF FUCHSIAN GROUPS

J. H. H. CHALK

(Received February 26, 1973)

1. Introduction. Let G denote a Fuchsian group of linear fractional transformations

$$g: z' = (\alpha z + \beta)/(\gamma z + \delta),\ \alpha\delta - \beta\gamma = 1$$

mapping the interior $\mathring{\mathscr{C}}$ of the disk $\mathscr{C}: |z| \leqq R$ in the complex z-plane onto itself. It is well-known, and readily verified, that this invariance assures that all isometric circles $|\gamma z + \delta| = 1$ of the group G intersect the fixed (principal) circle $C: |z| = R$ orthogonally. We propose to examine some geometrical extremal properties of a set of generators for G, defined in terms of these isometric circles. For this, the fundamental region[1] $\mathfrak{D} \subset \mathscr{C}$, as defined by Ford (cf. [4], Ch. III) for the action of G on $\mathscr{C}$, serves our purpose since (*loc. cit.*, Theorems 6, 10)

(i) its frontier $\partial\mathfrak{D}$ consists of arcs of isometric circles $C_i (1 \leqq i \leqq N)$ congruent in pairs under G; each such pair being of equal length, with corresponding points equidistant from 0,

(ii) the subset B of G whose elements provide the complete pairing of arcs of $\partial\mathfrak{D}$ under (i), is a generator set for G.

Since the interior $\mathring{\mathfrak{D}}$ of $\mathfrak{D}$ is given by

$$\mathring{\mathfrak{D}} = \mathscr{C} - \bigcup_{g \in G} \widetilde{\mathscr{C}}_g, \quad \text{where} \quad \widetilde{\mathscr{C}}_g: |\gamma z + \delta| \leqq 1\,,$$

we can view $\mathfrak{D}$ itself as a Riemann surface $\mathscr{F}$ when congruent points on $\partial\mathfrak{D}$ are identified. It is customary to call $\mathfrak{D}$ open or closed according as $\mathfrak{D}$ has frontier points on $\mathscr{C}$, or not. Then we can introduce the standard non-euclidean (N.E.) differentials

$$ds = \frac{2R\,|dz|}{R^2 - |z|^2}, \quad d\mu(z) = \frac{4R^2 r dr d\theta}{(R^2 - r^2)^2}, \quad (z = re^{i\theta})\,,$$

for N.E. elements of length and area. In particular, the formula

$$(1) \qquad \mu(\mathfrak{D}) = 4R^2 \iint_{z \in \mathfrak{D}} \frac{r dr d\theta}{(R^2 - r^2)^2}\,,$$

[1] For the sake of symmetry in $\mathfrak{D}$, we shall suppose that the point $z = +\infty$ is *not* a fixed point for any element $\neq$ of G; otherwise, we should proceed as in [4], Theorem 22. For our applications later, it is convenient not to fix on a value for R here.

gives the N.E. area $\mu(\mathfrak{D})$ of $\mathfrak{D}$. If Δ is any triangle bounded by three circles orthogonal to $\mathscr{C}$ and with vertex angles $\theta_1, \theta_2, \theta_3$ say, then

$$\mu(\Delta) = \pi - (\theta_1 + \theta_2 + \theta_3) .$$

A natural extension of this, by triangulation, is

$$(2) \qquad \mu(\mathfrak{D}) = N\pi - (\theta_1 + \theta_2 + \cdots + \theta_N) - 2\pi ,$$

when $\partial\mathfrak{D}$ consists of N arcs meeting at angles $\theta_1, \theta_2, \cdots, \theta_N$. From the theorem of Siegel-Tsuji (cf. [8], Theorem 1), we know that $\mu(\mathfrak{D}) = \infty$, unless $\mathfrak{D}$ is either closed, or open with at most a finite number n of parabolic vertices on C, (i.e. the Riemann surface $\mathscr{F}$ corresponding to $\mathfrak{D}$ is then either closed or obtained from a closed surface by the deletion of finitely many points) in which case $\mu(\mathfrak{D}) < \infty$. Indeed, if $\mu(\mathfrak{D}) < \infty$, it also provides an upper bound[2] to the number N in (2) of arcs bounding $\mathfrak{D}$; on writing

$$(3) \qquad \mu(\mathfrak{D}) = 2\sigma(\mathfrak{D}) = 2\sigma \text{ say}$$

we have

$$(4) \qquad N \leqq \frac{6}{\pi}\sigma + 6 - 2n .$$

In this article, we shall only consider the cases when

$$\sigma < \infty ,$$

so that the Siegel-Tsuji estimate can be re-cast in the shape

$$(5) \qquad \frac{\pi + \sigma}{N + 2n} \geqq \frac{\pi}{6} \quad \text{or} \quad \sin\left(\frac{\pi + \sigma}{N + 2n}\right) \geqq \frac{1}{2} .$$

Combining (2) and (4), we see that

$$(6) \qquad (\theta_1 + \theta_2 + \cdots + \theta_N) \leqq \frac{2\pi}{3}(N - n)$$

and we shall be concerned primarily with this form of the inequality for our refinement. An alternative source for (6) is given in Theorem 25 of Macbeath's Dundee Notes [7] where, on summing over the q ordinary cycles of $\mathfrak{D}$, one has

$$(7) \qquad \sum_{i=1}^{N} \theta_i = 2\pi \sum_{j=1}^{q} \frac{1}{k_j} = 2\pi \sum_{j=1}^{q} \nu_j / h_j ,$$

[2] In fact, it is a bound to the cardinality of the generator set B of G. However, a factor 2 can be gained if no mid-point of an arc of $\partial\mathfrak{D}_0$ is a fixed point of some elliptic element of G with period 2.

together with the relations

(i) $h_j = k_j\nu_j \geqq 3 \quad 1 \leqq j \leqq q$,

(ii) $\sum_{j=1}^{q} \nu_j = N - n$,

for the number ν_j of vertices of $\mathfrak{D}$ in a cycle and k_j is the order of the stabilizer of a vertex in that cycle, (applying the inequality in (i) to the right side of (7), gives (6) immediately). As it stands the constant $2\pi/3$ in (6) cannot be sharpened, as the example of the modular group, where the three vertices of $\mathfrak{D}$ have angles $(1/3)\pi$, $(1/3)\pi$, 0, shows. However, for large N and for special cases[3] it is of interest to seek more precise information about $\sum \theta_i$, or equivalently the relative magnitudes of N and σ in (4). Instances where such bounds can be usefully applied occur in the group $\mathscr{G}$ of units of a maximal order in a quaternion algebra. We first develop our refinement for the general Fuchsian group in the above setting and then apply it to the special types arising from an *indefinite* rational quaternion algebra $\mathscr{H}$. The refinement is given in Theorem 1 and is a natural generalization of a variational argument [10] I obtained for the estimation of "small" solutions of certain Pellian equations, but presented here with a new and more direct proof.

2. A Refinement of the Siegel-Tsuji Theorem.

NOTATION. Let $C_j = \{z \mid |\gamma_j z + \delta_j| = 1\}$, $1 \leqq j \leqq N$ denote the set of bounding isometric circles contributing to $\partial\mathfrak{D}$ and $\widetilde{C}_j$ the corresponding set of disks $|\gamma_j z + \delta_j| \leqq 1$. Then $\bigcup_{1\leqq j\leqq N} \widetilde{C}_j$ forms a covering of the principal circle $|z| = R$, and if $n = 0$, their interiors alone form an open covering. For $1 \leqq j \leqq N$, let O_j denote the centre $-\delta_j/\gamma_j$ of C_j,

$$\begin{aligned} r_j &= |\gamma_j|^{-1}, \text{ the radius of the } C_j\,, \\ \rho_j &= OO_j = |-\delta_j/\gamma_j|\,, \\ d_j &= O_jO_{j+1}\,. \end{aligned}$$

We may assume that $0 \leqq \arg(-\delta_1/\gamma_1) < \arg(-\delta_2/\gamma_2) < \cdots < \arg(-\delta_N/\gamma_N) < 2\pi \pmod{2\pi}$. Then, clearly

$$\rho_j^2 = R^2 + r_j^2\,,$$

since each circle C_j intersects the principal circle C orthogonally. We define

$$r = \max_{1\leqq j\leqq N} r_j\,, \tag{8}$$

and put

[3] Where, for example, the number of faces meeting at a vertex of the Dirichlet tessellation of G may be larger than 3.

$$\rho^2 = R^2 + r^2 \tag{9}$$

$$d = 2\rho \sin \frac{\pi}{N} . \tag{10}$$

We consider first the case $n = 0$ (when there are no parabolic cycles) and then extend the argument to $n > 0$.

THEOREM 1 ($n = 0$). $\sin((\sigma + \pi)/N) \geqq (1/2)(d/r)$, *or equivalently,*

$$\sum_{1 \leqq j \leqq N} \theta_j \leqq 2N \cos^{-1}\left(\frac{1}{2}\frac{d}{r}\right) .$$

REMARK. In some cases, for example certain Fuchsian groups arising from quaternion algebras, this inequality is stronger than the Siegel-Tsuji estimate (4), since $d \geqq r$. This may well be true generally.

PROOF. There are two stages in the proof:

(I) Replace each isometric disk $\tilde{C}_j$ by a new circular disk $\tilde{C}'_j$ orthogonal to the principal circle $|z| = R$ with radius $r = \max r_j$ and centre $-(\rho/\rho_j)\delta_j/\gamma_i$, so that

$$\tilde{C}'_j \supset \tilde{C}_j \qquad (1 \leqq j \leqq N) .$$

Then the set $\mathfrak{D}'$, where

$$\mathfrak{D}' = \mathscr{C} - \bigcup_{1 \leqq j \leqq N} \tilde{C}'_j \subset \mathfrak{D} = \mathscr{C} - \bigcup_{1 \leqq j \leqq N} \tilde{C}_j$$

is certainly contained in the Ford fundamental region $\mathfrak{D}$. Hence

$$2\sigma(\mathfrak{D}') = \mu(\mathfrak{D}') \leqq \mu(\mathfrak{D}) = 2\sigma(\mathfrak{D}) , \tag{11}$$

with strict inequality unless $r_1 = \cdots = r_N = r$. We observe too that the collection $\bigcup_{1 \leqq j \leqq N} \tilde{C}_j$ still forms a covering of $|z| = R$ and

$$\tilde{C}'_j \cap \tilde{C}'_{j+1} \neq \varnothing \qquad (1 \leqq j \leqq N) , \tag{12}$$

$$\tilde{C}'_j \not\subset \tilde{C}'_k \quad \text{for all} \quad j \neq k . \tag{13}$$

Let θ'_j $(1 \leqq j \leqq N)$ denote the corresponding angles for $\mathfrak{D}'$. Then

$$0 < \theta'_j < \pi \qquad (1 \leqq j \leqq N) , \tag{14}$$

by (12) and (13) and the distance d'_j between the centres of C'_j and C'_{j+1} may be expressed either, in the form

$$d'_j = 2r \cos \frac{1}{2}\theta'_j \tag{15}$$

or, in the form

(16) $$d'_j = 2\rho \sin \frac{1}{2}\alpha'_j ,$$

where $\alpha'_j = \arg(-\delta_j/\gamma_j) - \arg(-\delta_{j+1}/\gamma_{j+1})$ is the angle subtended by the centres at the origin O, $(1 \leqq j \leqq N)$. From (15) and (16), we deduce that

(17) $$\frac{1}{2}\theta'_j = \cos^{-1}\left[\frac{\rho}{r} \sin \frac{1}{2}\alpha'_j\right] ,$$

where

(18) $$\sum_{1 \leqq j \leqq N} \frac{1}{2}\alpha'_j = \pi .$$

(II) Application of Jensen's inequality to the function

$$f(x) = \cos^{-1}\left(\frac{\rho}{r} \sin \pi x\right) ,$$

where x takes the values

$$x = x_j = \frac{1}{2\pi}\alpha'_j \qquad (1 \leqq j \leqq N) .$$

Observe that

(i) $0 < f < (1/2)\pi$, by (14) and (17),

(ii) $\sum_{1 \leqq j \leqq N} x_j = 1$, by (18),

(iii) $f(x)$ is a strictly concave function of x over $(0, (1/\pi)\sin^{-1}(r/\rho))$.

It remains to verify (iii) and for this it is sufficient to show that $f''(x)$ is defined over $(0, (1/\pi)\sin^{-1}(r/\rho))$ and $f'' < 0$. By repeated differentiation of $\cos f$, we obtain successively

(19) $$f' \sin f = -\frac{\rho}{r}\pi \cos \pi x$$

$$f'' \sin f + f'^2 \cos f = \pi^2 \cos f ,$$

so that

(20) $$f'' = (\pi^2 - f'^2) \cot f .$$

Also, from (19), we have

$$\frac{\rho^2}{r^2} = \cos^2 f + \left(\frac{f'}{\pi}\right)^2 \sin^2 f ,$$

whence

$$\frac{1}{\pi^2}(f'^2 - \pi^2) \sin^2 f = \frac{\rho^2}{r^2} - 1 = \frac{R^2}{r^2} > 0 .$$

Thus, $f'' < 0$ and is certainly defined over $(0, (1/\pi)\sin^{-1}(r/\rho))$ by (i), (19) and (20). Thus, Jensen's inequality applies to $-f$ and we conclude that

$$-f\Big(\frac{1}{N}\sum_{1\leqq j\leqq N} x_j\Big) \leqq -\frac{1}{N}f(x_1) - \cdots - \frac{1}{N}f(x_N)\,,$$

or, more precisely,

$$\begin{aligned}\frac{1}{N}\sum_{1\leqq j\leqq N}\frac{1}{2}\theta'_j &= \frac{1}{N}\Big[\sum_{1\leqq j\leqq N}\cos^{-1}\Big(\frac{\rho}{r}\sin\pi x_j\Big)\Big]\\ &\leqq \cos^{-1}\Big[\frac{\rho}{r}\sin\Big(\frac{\pi}{N}\sum_{1\leqq j\leqq N} x_j\Big)\Big]\\ &= \cos^{-1}\Big[\frac{\rho}{r}\sin\Big(\frac{1}{2N}\sum_{1\leqq j\leqq N}\alpha'_j\Big)\Big]\\ &= \cos^{-1}\Big[\frac{\rho}{r}\sin\frac{\pi}{N}\Big]\\ &= \cos^{-1}\Big[\frac{1}{2}\frac{d}{r}\Big], \text{ by (18), (10)}\,;\end{aligned}$$

with strict inequality unless

$$\theta'_1 = \theta'_2 = \cdots = \theta'_N\,.$$

By (11), (2), we have

$$\frac{1}{N}\sum_{1\leqq j\leqq N}\theta_j \leqq \frac{1}{N}\sum_{1\leqq j\leqq N}\theta'_j \leqq 2\cos^{-1}\Big[\frac{1}{2}\frac{d}{r}\Big]\,,$$

with strict inequality, unless $\theta_j = \theta'_j$, $r_j = r$, $\rho_j = \rho$ for all j and

$$\theta_j = 2\cos^{-1}\Big[\frac{\rho}{r}\sin\frac{\pi}{N}\Big]\,.$$

We give the extension to the case when parabolic vertices exist as a corollary to Theorem 1.

COROLLARY. *If G possesses n parabolic vertices, then*

$$\sum_{1\leqq j\leqq N}\theta_j \leqq \sum_{k=1}^{n} 2(N^{(k)}-1)\cos^{-1}\Big(\frac{1}{2}d^{(k)}/r^{(k)}\Big)\,, \tag{21}$$

where $N^{(k)}$, $d^{(k)}$, $r^{(k)}$ have the previous meaning, but applied to the arc $\mathscr{C}_k$ of the principal circle $\mathscr{C}$ between the k^{th} pair of consecutive parabolic vertices in place of $\mathscr{C}$ itself.

A proof of this proceeds as for the Theorem itself, except for a modification in Stage I applied to each of the n arcs $\mathscr{C}_k$ of $\mathscr{C}$. If Δ_k denotes the infinite sector with angle ψ_k say, subtended by $\mathscr{C}_k$ from $z=0$, $N^{(k)}$ the number of isometric circles intersecting $\mathscr{C}_k$ and $r^{(k)}$ the radius of the largest such circle, we now arrange that each of these isometric circles is enclosed by a circle of radius $=r^{(k)}$ situated within Δ_k and orthogonal to $\mathscr{C}$. This is clearly possibly if we do *not* insist, as before, that the new centres maintain the same direction from $z=0$. Then stage II consists of applying Jensen's inequality to the sum of the angles of intersection of the $N^{(k)}$ equal circles ($k=1, 2, \cdots, n$); the k^{th} term on the right of (21) corresponding to the optimal case when the circles are situated symmetrically on the arc $\mathscr{C}_k$ with just two of them tangent to Δ_k at the end-points of $\mathscr{C}_k$ and with their centres at an equal distance $d^{(k)}$ apart. Thus,

$$d_k = 2[R^2 + r^{(k)2}]^{1/2} \sin \frac{1}{2}\alpha^{(k)} ,$$

where

$$\alpha_k = [N^{(k)} - 1]^{-1}\left[\psi^{(k)} - 2\tan^{-1}\frac{r^{(k)}}{R}\right] .$$

EXAMPLES. It is easy to provide examples where the theorem (or its Corollary) gives a sharper estimate than that of Siegel-Tsuji. Let G_D denote the group of elements of the form

$$z' = \frac{az + D\bar{c}}{cz + \bar{a}};\ a\bar{a} - Dc\bar{c} = 1 , \tag{22}$$

where $D \in Z^+$, $(a, c) \in Z^2[i]$. Since $i_0 \colon z' = -z$ is the only element of G_D, with ∞ as a fixed point (see footnote 1), we shall, by our convention, factor the subgroup $\langle i, i_0\rangle$ and consider $G'_D = G_D/\langle i, i_0\rangle$ instead. The case $D=1$ gives rise to a fundamental region $\mathfrak{D}$ having only parabolic vertices and does not test the inequalities (details, see Ford [4], Ch. 3, § 36, pp. 78–9). For $D=2$, when $N=8$, $n=4$,

$$\sum \theta_i = 2\pi , \qquad \left(\leqq \frac{8\pi}{3}, 2\pi\right) \tag{23}$$

and for $D=3$, when $N=12$, $n=0$,

$$\sum \theta_i = 6\pi , \qquad (\leqq 8\pi,\ 24\cos^{-1} 0.517 \cdots) , \tag{24}$$

where the estimates on the right are, respectively, from the Siegel-Tsuji Theorem and from Theorem 1 (or Corollary) and exhibit the improvements possible. In one further example we see that both inequalities are best possible. Take the modular group Γ and transform the real axis into the unit circle, by

$$z' = \frac{z+i}{z-i}\,.$$

Then Γ is mapped onto a group G of elements of the form (22) with $D = 1$, but a and c now belong to $(1/2)Z[i]$ and are subject to the conditions

$$\mathscr{R}(a+c) \in \mathbf{Z}\,, \qquad \mathscr{I}(a+c) \in \mathbf{Z}\,.$$

Then for $G' = G/\langle i, i_0\rangle$, we have $N = 4$, $n = 2$ and

$$\sum \theta_i = \frac{4\pi}{3}\,, \tag{25}$$

the constant $4\pi/3$ being that given by (6) and (21).

3. Indefinite rational quaternion algebras. Let $\mathscr{H}$ be an indefinite quaternion algebra over the rationals Q and let $\mathfrak{O}$ be an order of $\mathscr{H}$. Let

$$U = \{u \in \mathfrak{O} \mid u\mathfrak{O} = \mathfrak{O},\ n(u) = +1\}\,,$$

where $n(u)$ denotes the norm of u in $\mathscr{H}$. By the isomorphism between the matrix algebra $M_2(R)$ and $\mathscr{H} \otimes_Q \boldsymbol{R}$, the unit group U corresponds to a certain discrete subgroup G of $SL(2, R)$. It is therefore a Fuchsian group and is, in fact, of the first kind i.e., G is a properly discontinuous group and its quotient space H/G, where H is the upper half of the complex plane, has finite N.E. measure.

For our application, we shall utilize a canonical generation of $\mathscr{H}$ in the form

$$\mathscr{H} = [1, i, j, ij]\ ; \quad i^2 = -P\,, \quad j^2 = D\,,$$

where $D = q_1 q_2 \cdots q_{2n}$ is a product of $2r$ distinct rational primes and P is a prime $\equiv 3 \pmod 4$ satisfying

$$\left(\frac{-P}{q_s}\right) = -1 \qquad (s = 1, 2, \cdots, 2r)\,.$$

Albert [1] has shown that then the only maximal orders $\mathfrak{O}$ are generated by

$$\mathfrak{O} = [1, w, J, Jw]\ ; \quad w = \frac{1}{2}(1+i)\,, \quad J = P^{-1}(2\mu + j)i\,, \tag{26}$$

where μ is either one of two solutions to the congruence

$$4\mu^2 \equiv D \pmod P\,, \tag{27}$$

and is essentially unique since the other solution $-\mu$ gives the order $i^{-1}\mathfrak{O}i$. Thus, we suppose that the general element $v \in \mathfrak{O}$ is of the form

(28) $$v = \xi + J\eta\,, \qquad \begin{cases} \xi = x_0 + x_1 w \\ \eta = y_0 + y_1 w \end{cases}$$

where $(x_0, x_1, y_0, y_1) \in Z^4$ and ξ, η are elements of the complex quadratic field $Q(i)$. If we also put

(29) $$\zeta = \xi + 2P^{-1}\mu i\eta\,,$$

then the norm $N(v)$ takes a "diagonal" shape:

(30) $$N(v) = N(\xi + J\eta) = \zeta\zeta' - DP^{-1}\eta\eta'\,,$$

and if we multiply v by a unit $u = \alpha + J\beta$ say, with $N(u) = +1$ then, on introducing new variables ξ_1, η_1 taking values in $Q(i)$, defined by

$$uv = (\alpha + J\beta)(\xi + J\eta) \stackrel{\text{def}}{=} \xi_1 + J\eta_1$$

we obtain

(31) $$\begin{cases} \zeta_1 = \gamma\zeta + DP^{-1}\beta'\eta \\ \eta_1 = \beta\zeta + \gamma'\eta \end{cases}, \qquad \gamma\gamma' - DP^{-1}\beta\beta' = 1$$

where ζ_1 and γ correspond to ζ in (29), i.e.,

(32) $$\begin{pmatrix} \zeta_1 \\ \gamma \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \alpha \end{pmatrix} + \frac{2\mu}{P} i \begin{pmatrix} \eta_1 \\ \beta \end{pmatrix}.$$

The quaternion group G associated with (31) is then the group of elements of the form

(33) $$z_1 = \frac{\gamma z + DP^{-1}\beta'}{\beta z + \gamma'}\,,$$

where $\gamma\gamma' - DP^{-1}\beta\beta' = 1$. This is known (cf. [5]) to be a Fuchsian group of the first kind with principal circle $|z|^2 = DP^{-1}$ and the fundamental region[4] $\mathfrak{D}$ for G, (whose closure is the set of interior points of the principal circle which do not belong to any of the isometric disks $|\beta z + \gamma'| < 1$, $(\alpha, \beta) \in Z^2[w]$, $\gamma\gamma' - DP^{-1}\beta\beta' = 1$), has N.E. measure $\mu(\mathfrak{D})$, given by (cf. [3])

(34) $$\mu(\mathfrak{D}) = \frac{\pi}{3}\phi(D)\,.$$

A crude upper bound for the magnitude of the smallest unit $u = \alpha + J\beta$ with $\beta \neq 0$ is obtained by noting that the sum of the diameters of the bounding isometric circles exceeds the circumference of the principal circle (since their disks form an open covering of it!) and applying the Siegel-Tsuji estimate for the number N of such circles. Thus

[4] It is easy to verify that G has no fixed points at ∞ for $P>3$, since $P\equiv 3 \pmod 4$, and that there are no parabolic points for quaternion groups.

$$(35) \qquad \sum_{j=1}^{N} 2\,|\beta_j|^{-1} > 2\pi[DP^{-1}]^{1/2}\,,$$

which gives

$$(36) \qquad N\,|\beta_1|^{-1} > \pi[DP^{-1}]^{1/2}$$

or

$$(37) \qquad |\beta_1| < \frac{1}{\pi}[PD^{-1}]^{1/2}N\,,$$

if $|\beta_1|$ denotes the least value among the $|\beta_j|$. Then, from (37), (4) and (34),

$$(38) \qquad |\beta_1| < \frac{1}{\pi}[PD^{-1}]^{1/2}\Big(\frac{3}{\pi}\cdot\frac{\pi}{3}\phi(D) + 6\Big) = \frac{1}{\pi}\Big(\frac{\phi(D)}{D} + \frac{6}{D}\Big)[PD]^{1/2}\,.$$

A slightly better estimate can be obtained by observing that our fundamental region $\mathfrak{D}$ contains the disk

$$K\colon |z| < \rho - r$$

where, from our notation in section 2,

$$\rho - r = (R^2 + r^2)^{1/2} - r = [DP^{-1} + |\beta_1|^{-2}]^{1/2} - |\beta_1|^{-1}\,.$$

This disk has N.E. measure

$$\begin{aligned}\mu(K) &= 4R^2\int_K \frac{\bar r d\bar r d\theta}{(R^2 - \bar r^2)^2}\\ &= 2\pi\frac{R^2}{r(\rho - r)} - 4\pi\\ &= 2\pi[(R^2r^{-2} + 1)^{1/2} + 1] - 4\pi\\ &= 2\pi[(DP^{-1}|\beta_1|^2 + 1)^{1/2} - 1]\,.\end{aligned}$$

Then, from the inequality $\mu(K) \leqq \mu(\mathfrak{D})$ and (34), we have

$$2\pi[DP^{-1}|\beta_1|^2 + 1]^{1/2} \leqq \frac{\pi}{3}\phi(D) + 2\pi\,,$$

$$(39) \qquad [DP^{-1}]^{1/2}|\beta_1| < [DP^{-1}|\beta_1|^2 + 1]^{1/2} \leqq \frac{1}{6}\phi(D) + 1\,,$$

and

$$(40) \qquad |\beta_1| \leqq \frac{1}{6}\Big(\frac{\phi(D)}{D} + \frac{6}{D}\Big)[PD]^{1/2}\,,$$

which is an improvement on (38), by a factor $\pi/6 = 0.577\cdots < 1$, for large D.

Our main purpose in this section is to show that, while it is not known whether Theorem 1 includes that of Siegel-Tsuji, nevertheless, a suitable application of both, produces an estimate ((42) below) better than either (38) or (40).

THEOREM 2. *There is a unit u of the maximal order $\mathfrak{O}$ of $\mathscr{H}$ of the form $\alpha + J\beta$, where $J = P^{-1}(2\mu + j)i$, $\alpha \in Z[w]$, $\beta \in Z(w)$, $w = (1/2)(1 + i)$, where μ is a solution of $4\mu^2 \equiv D \pmod P$, with*

$$|\beta| < \frac{1}{6}[PD]^{1/2}\left(\frac{\phi(D)}{D} + \frac{6}{D}\right); \qquad \beta \neq 0 \tag{41}$$

and, for all sufficiently large D,

$$|\beta| \leqq \frac{1}{2}P^{1/2}D^{-1/2}\left[\sin\frac{\pi}{\phi(D) + 6}\right]^{-1} \sim \frac{1}{2\pi}[PD]^{1/2}\left(\frac{\phi(D)}{D} + \frac{6}{D}\right). \tag{42}$$

PROOF. From Theorem 1 we have

$$\sin\left[N^{-1}\left(\frac{1}{2}\mu(\mathfrak{D}) + \pi\right)\right] \geqq \frac{1}{2}\frac{d}{r}$$

where $r = |\beta_1|^{-1} = \max_{1 \leqq i \leqq N} |\beta_j|^{-1}$, for all N bounding isometric circles $|\beta_j z + \gamma_j| = 1$, with $\beta_j \neq 0$ and

$$d = 2[R^2 + r^2]^{1/2}\sin\frac{\pi}{N},$$

where $R^2 = DP^{-1}$. Thus

$$\sin\left[\frac{\pi}{6N}(\phi(D) + 6)\right] \geqq [DP^{-1}r^{-2} + 1]^{1/2}\sin\frac{\pi}{N},$$

$$[DP^{-1}]^{1/2}|\beta_1| < [DP^{-1}|\beta_1|^2 + 1]^{1/2} \leqq \frac{\sin\left[\frac{\pi}{N}\cdot\frac{1}{6}(\phi(D) + 6)\right]}{\sin\frac{\pi}{N}}. \tag{43}$$

Then, from (4) and (34), we know that

$$N \leqq \phi(D) + 6. \tag{44}$$

Since the expression on the right of (43) is difficult to estimate without further conditions on N, we proceed by contradiction, assuming henceforth that

$$|\beta_1| > \frac{3}{\pi}\cdot\frac{1}{6}\left(\frac{\phi(D)}{D} + \frac{6}{D}\right)(PD)^{1/2} \qquad \text{for } D \geqq D_0.$$

Then by (36) we can now suppose that

$$N > \pi[DP^{-1}]^{1/2}|\beta_1| > \frac{1}{2}(\phi(D) + 6). \tag{45}$$

For convenience in notation we put

$$(46)\qquad x = \frac{\pi}{N}\,, \qquad a = \frac{1}{6}(\phi(D) + 6) > 1$$

and consider the function

$$f(x) = \frac{\sin ax}{\sin x}\,.$$

Note that, by (44) and (45), we shall only need to consider values of x in the interval

$$(47)\qquad \frac{\pi}{6a} \leqq x < \frac{\pi}{3a}\,.$$

Firstly, we observe that

$$(48)\qquad f(x) = a\left[\frac{\sin ax}{ax}\Big/\frac{\sin x}{x}\right] < a = \frac{1}{6}(\phi(D) + 6)\,,$$

since $x^{-1}\sin x$ is strictly decreasing over $(0, \pi] \supset [\pi/6a, \pi/3]$; which gives (41), or (40), with strict inequality. Secondly, we note that

$$\frac{d^2}{dx^2}(\log f(x)) < 0\,,$$

for

$$\frac{d}{dx}(\log f) = a\cot(ax) - \cot x$$

and

$$\frac{d^2}{dx^2}(\log f) = -a^2\operatorname{cosec}^2(ax) + \operatorname{cosec}^2 x$$
$$= \operatorname{cosec}^2 ax\{f^2 - a^2\} < 0\,,$$

by (48). Since $(d^2/dx^2)(\log f)$ is defined over $[\pi/6a, \pi/3a)$, it follows that $(d/dx)(\log f)$ is strictly decreasing there. We now prove that there is a $D_0 > 0$ such that for all $D \geqq D_0$,

$$(49)\qquad \left[\frac{d}{dx}\log f\right]_{x=\pi/6a} < 0\,,$$

from which we conclude that $(d/dx)(\log f) < 0$ over $[\pi/6a, \pi/3a)$ and hence that

$$\log f(x) \leqq \log f\left(\frac{\pi}{6a}\right) \quad \text{for} \quad x \in \left[\frac{\pi}{6a}, \frac{\pi}{3a}\right]$$

or

$$\begin{aligned} f(x) &\leqq f\left(\frac{\pi}{6a}\right) \\ &= \frac{1}{2}\left[\sin\frac{\pi}{\phi(D)+6}\right]^{-1} \\ &\sim \frac{1}{2\pi}(\phi(D)+6), \quad \text{as} \quad D \to \infty . \end{aligned}$$

This is an improvement, for large D, on (41) by a factor asymptotic to $3/\pi < 1$. For (49), we have

$$\left[\frac{d}{dx}\log f\right]_{x=\pi/6a} = a\sqrt{3} - \cot\frac{\pi}{6a}$$

and

$$a\sqrt{3}\tan\frac{\pi}{6a} \to \frac{\pi}{2\sqrt{3}} < 1 \quad \text{as} \quad a = \frac{1}{6}(\phi(D)+6) \to \infty ,$$

as required.

4. In this section, we review briefly some open problems and some recent progress on generating sets for the general Fuchsian group Γ. Following Macbeath [7, p. 16, Def. 7], we normalize the group so that the principal circle is the real axis in the complex z-plane and view Γ as a discrete subgroup of $SL(2, R)$. If $\mathfrak{H}$ denotes the upper half z-plane, then Γ acts on $\mathfrak{H}$ by the operation

$$\mathfrak{H} \ni z \to g(z) = \frac{az+b}{cz+d} \in \mathfrak{H} \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma .$$

If we further suppose that Γ is a Fuchsian group of the first kind, when the quotient space $\mathfrak{H}/\Gamma$ has finite N.E. measure, Takeuchi [9], has given a characterization of those Γ which arises from indefinite quaternion algebras $\mathscr{H}$. Thus, if U is the unit group of some order of $\mathscr{H}$, where $\mathscr{H}$ is defined over Q, we know (cf. section 3) that $U \cong \Gamma_0$, a Fuchsian group of the first kind. Takeuchi calls a subgroup $\Gamma \subset \Gamma_0$ "derivable" from $\mathscr{H}$, when either $\Gamma = \Gamma_0$ or Γ is of finite index in Γ_0 and proves that the condition

$$\text{(50)} \qquad \operatorname{Tr}(\gamma) \in Z, \quad \text{for all} \quad \gamma \in \Gamma$$

is both necessary and sufficient for Γ to be derivable from $\mathscr{H}$.

Later, Kelly and I [6] showed that, subject only to a condition (51) on the spacings between values of the norms $\|\gamma\|$ of the elements $\gamma \in \Gamma$, where

$$||\gamma|| = [a^2 + b^2 + c^2 + d^2] \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

one can give effectively computable bounds, successively, to each member of a suitably chosen set of generators for Γ.

THEOREM. *Let $\Gamma \subset SL(2, R)$ be a Fuchsian group of first kind with compact quotient space. Suppose that there is some positive real constant $\lambda = \lambda(\Gamma)$ such that*

$$\lambda[||\gamma|| - 2] \equiv 0 \pmod 1, \quad \textit{for all} \quad \gamma \in \Gamma . \tag{51}$$

Then there is a set of generators $A_1, A_2, \cdots, A_N$ of Γ satisfying

(i) $||A_1|| \leqq ||A_2|| \leqq \cdots \leqq ||A_N||$

(ii) $||A_1|| \ll N$,

(iii) $||A_{j+1}|| \ll N ||A_j||^5$ $(j = 1, 2, \cdots, N-1)$.

For simplicity, we have omitted the actual constants in (ii) and (iii) and used the Vinogradow-notation "$\ll$", on the understanding that the implied constants depend only on λ. It follows that $\log\log ||A_j|| < 2N$, $(j = 1, 2, \cdots, N, N \geqq N_0(\lambda))$ and from the proof itself it can be inferred that $N < 3\pi^{-1}\mu + 6$ (cf. (4)), where μ is the N.E. measure of $\mathfrak{H}/\Gamma$. Although the restriction in (51) is somewhat artificial, it is certainly satisfied by the quaternion groups i.e., those Γ satisfying the Takeuchi condition (50)) and, indeed, by the arithmetic subgroups of $SL(2, R)$.

If S denotes any subset of Euclidean n-space R^n $(n \geqq 1)$, we shall say that S is δ-separated in R^n, if there is a constant $\delta > 0$ such that

$$x, y \in S \Rightarrow ||x - y||^{1/2} \geqq \delta ,$$

when $x \neq y$. Also, if we map $SL(2, R)$ into R^4, by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow (a, b, c, d) ,$$

(where $||\cdot||^{1/2}$ is the usual Euclidean distance function in R^4), the image of a subset L of $SL(2, R)$ will be denoted by L'. Then, from the above results[5], it would be natural to ask if every Fuchsian group Γ of the first kind is such that Γ' is δ-separated in R^4. In the arithmetic cases, $\Gamma' \subset \Lambda$, a lattice, and is certainly δ-separated in R^4. Also, if $||\Gamma||$ denotes the set

$$\{||\gamma|| \mid \gamma \in \Gamma\} ,$$

[5] For the general Fuchsian group $\Gamma \subset G = SL(2, R)$ of the first kind, it is known that $||\Gamma||$ is a *discrete* subset of $\boldsymbol{R}^1$ and I have recently verified that, if G/Γ has compact closure, then Γ' is δ-separated for some $\delta = \delta(\Gamma) > 0$.

one may ask whether the condition that $||\Gamma||$ be δ-separated in R^1 for some constant $\delta > 0$ assures that Γ is derivable from a quaternion algebra $\mathscr{H}$ over Q.

REFERENCES

[1] A. ALBERT, Integral domains of rational generalized quaternion algebras, Bull. Amer. Math. Soc., 40 (1934), 164–176.

[2] J. H. H. CHALK, An estimate for the fundamental solutions of a generalized Pell equation, Math. Ann. Bd 132 (1956), 263–276.

[3] M. EICHLER, Über die Einheiten der divisionsalgebren, Math. Ann. 114 (1937), 637–654, (cf. p. 649).

[4] L. R. FORD, Automorphic functions, (New York, Chelsea).

[5] R. HULL, On the units of indefinite quaternion algebras, Amer. J. Math. 61 (1939), 365–374.

[6] B. G. A. KELLY AND J. H. H. CHALK, Generating sets for Fuchsian groups, Proc. Royal Edinburgh Soc., (to appear)

[7] A. M. MACBEATH, Fuchsian Groups, (Lecture Notes, Queen's College, Dundee, 1961).

[8] C. L. SIEGEL, Some remarks on discontinuous groups, Ann. of Math., 46 (1945), 708–718; (cf. Theorem 5).

[9] K. TAKEUCHI, On some discrete subgroups of $SL(2, R)$, J. Fac. Sci., Univ. Tokyo, Sec. I (1969), 97–100.

[10] M. TSUJI, Theory of Fuchsian groups, Japanese J. of Math., 21 (1951), 1–27; (cf. Theorem 1 and footnote).

Addendum: On a recent visit to Toronto, Professor A. M. Macbeath constructed an example of a Fuchsian group G with arbitrarily large genus, for which the inequality of Siegel-Tsuji (cf. (6) with $n = 0$) and that of Theorem 1 are both exact. As this provides additional significance to the non-invariant quality of the theorem, it is perhaps worthwhile to indicate briefly his construction and proof.

Let F be a regular N.E. polygon in $\mathring{\mathscr{C}}$ with $N = 6(2m + 1)$ sides ($m \geqq 1$) and $\theta_i = 2\pi/3$ ($i = 1, 2, \cdots, N$). The sides of F are labelled $x_2, \cdots, x_{2m+1}, y_0, \cdots, y_{2m}, a_1, a_2, \cdots, a_{2m+2}$ together with their inverses, the order around the polygon F being determined according to the surface symbol

$$\prod_{v=0}^{2m} y_{2m-v} \prod_{v=0}^{2m-1} x_{2m+1-v} \prod_{v=1}^{m} (a_{2v-1} y_{2v-1}^{-1} a_{2v} x_{2v}^{-1}) a_{2m+1} a_{2m+2} y_0^{-1} a_1^{-1} \cdot$$
$$\prod_{v=1}^{m} a_{2v}^{-1} y_{2v}^{-1} a_{2v+1}^{-1} x_{2v+1}^{-1}) a_{2m+2}^{-1} \cdot$$

This pairing of the sides of F assures that the cycles consist of exactly 3 vertices whose angles sum to 2π. Hence, by Poincaré's theorem, there exists a Fuchsian group G with F as a fundamental region and preserving the pairing. From the symmetry of the tessellation of $\mathring{\mathscr{C}}$

by F and its images under G, it also follows that F is a Dirichlet region for G with centre $z = 0$; the sides being arcs of isometric circles of G. Thus the equality sign is essential in both (6), when $n = 0$, and in Theorem 1.

UNIVERSITY OF TORONTO
TORONTO, CANADA