Proceedings of the Edinburgh Mathematical Society (1983) **26**, 89–96 ©

THE SOLUTION OF LENGTH THREE EQUATIONS OVER GROUPS

by JAMES HOWIE*

(Received 22nd September 1981)

Let G be a group, and let $r=r(t)$ be an element of the free product $G*\langle t\rangle$ of G with the infinite cyclic group generated by t. We say that the equation $r(t)=1$ *has a solution in* G if the identity map on G extends to a homomorphism from $G*\langle t\rangle$ to G with r in its kernel. We say that $r(t)=1$ *has a solution over* G if G can be embedded in a group H such that $r(t)=1$ has a solution in H. This property is equivalent to the canonical map from G to $\langle G,t\,|\,r\rangle$ (the quotient of $G*\langle t\rangle$ by the normal closure of r) being injective.

In general it is not possible to find a solution to an arbitrary equation $r(t)=1$ over an arbitrary group G. It is necessary to place some sort of restriction on the group, on the equation, or possibly on both. One possible restriction on the equation is that the exponent sum of t in r be non-zero. Under this hypothesis, but with no restriction on the group G, it is an open problem whether a solution over G always exists.

It is known that a solution exists if G is either locally residually finite [**6**] or locally indicable [**1**, **3**, **9**], and other known results give solutions under restrictions on r. Levin [**4**] showed that a solution exists if t occurs in r only with positive exponent. Thus the simplest remaining case (up to conjugacy and inversion) is when $r(t)$ has the form $atbtct^{-1}$ $(a,b,c\in G)$. Lyndon [**5**, Corollary 5.3] has solved this case under certain restrictions on the "coefficients" a,b,c. These restrictions are based on small cancellation theory, and concern the relations which can hold in G between a, b and c.

It is the purpose of this note to remove the restrictions from Lyndon's result, and show that any equation of the form

$$atbtct^{-1}=1$$

over any group G has a solution over G. Combined with Levin's theorem, this solves the problem whenever t occurs at most 3 times in $r(t)$.

Note that the above equation can be transformed to one of the form

$$a't^2c't^{-1}=1$$

by applying the automorphism $g\mapsto g$ $(g\in G)$, $t\mapsto tb^{-1}$ of $G*\langle t\rangle$, so we are reduced to the case $b=1$. Also, if the equation has a solution over the subgroup G_0 of G generated by

*Research supported by a William Gordon Seggie Brown Fellowship.

the coefficients a, b and c (say in a group $H \supset G_0$), then it has a solution over G (in the group $G *_{G_0} H$). Hence we may assume $G = G_0$, and so in particular G is a 2-generator group.

The method of proof is a variant of Lyndon's Dehn diagrams [**5**, **8**] essentially the dual of that developed by Short [**9**] (see also Rourke [**7**]). The strategy is to infer from some diagram sufficiently many relations between the 2 generators of G to deduce that G belongs to a class of groups for which the solution to the problem is known. Specifically, we deduce that G is at worst residually finite, and then the result follows from a well-known theorem of Gerstenhaber and Rothaus [**2**, **6**].

I am grateful to the referee for a number of useful comments.

1. Relative diagrams

Let

$$r(t) = g_1 \cdot t^{\varepsilon(1)} \cdot g_2 \cdot t^{\varepsilon(2)} \dots \cdot g_n \cdot t^{\varepsilon(n)} \in G * \langle t \rangle \qquad (g_i \in G,\ \varepsilon(i) = \mp 1).$$

The elements $g_1, \dots, g_n$ are called the *coefficients* of r, the $\varepsilon(i)$ are the *exponents* and their sum $\varepsilon(1) + \cdots + \varepsilon(n)$ the *exponent-sum.* The integer n is the *t-length* of r.

A *relative diagram* for the equation $r(t) = 1$ over G is a triple (D, v_0, ϕ), where D is a cellular subdivision of the 2-sphere S^2, with oriented 1-skeleton $D^{(1)}$; v_0 is a vertex (0-cell) of D; and ϕ is a "labelling function" which associates to each edge (1-cell) of D the element t, and to each corner of each face (2-cell) of D an element of G; such that the following conditions are satisfied.

(i) Reading the labels around any face in the clockwise direction from a suitable starting point gives either r or r^{-1} in cyclically reduced form. (Here an edge is to be read as t or t^{-1} depending on its orientation).

(ii) The product of the labels, read anti-clockwise around any vertex $v \neq v_0$ of D (the *vertex-label* of v), is equal to 1 in G.

Remarks. (1) It follows from (i) that the label of each corner is one of the coefficients or its inverse. Hence by (ii) the vertex-labels of vertices other than v_0 yield relations between the coefficients which hold in G.

(2) The vertex-label of v_0 is also defined (up to conjugacy) and is an element of the intersection of G with the normal closure in $G * \langle t \rangle$ of r. It is therefore a necessary condition for the existence of a solution over G to the equation $r(t) = 1$, that in any relative diagram for $r(t) = 1$, the vertex label of v_0 is equal to 1 in G. That this condition is also sufficient is the crux of our method, and we state it in the form of a Lemma. This can be proved using standard Dehn diagram methods [**5**, **8**]. Alternatively, the Lemma can be regarded as the dual of [**9**, Proposition 2.17], and can be proved using transversality. We give an outline of the latter argument.

Lemma 1. *If the equation $r(t) = 1$ has no solution over G, then there exists a relative diagram (D, v_0, ϕ) for $r(t) = 1$ such that the vertex label of v_0 is nontrivial.*

Proof. Let (L, K) be a geometric realisation of the relative presentation $\langle G, t | r \rangle$ [**3**].

That is K is a connected CW-complex with $\pi_1(K)=G$, and $L=K\cup e^1\cup_r e^2$. Then the inclusion-induced map $\pi_1(K)\to\pi_1(L)$ is not injective, so there exists a map $f:(D^2,S^1)\to(L,K)$ whose restriction to S^1 is essential in K.

Let $\Gamma\subset L$ be a tamely embedded graph with 2 vertices, one in the interior of each cell of $L\backslash K$, and an edge joining the two for each occurrence of t in r. Then Γ has a regular neighbourhood in $L\backslash K$, so f is homotopic rel S^1 to a map f_0 which is transverse to Γ. Then $\Delta=f_0^{-1}(\Gamma)$ is a graph in $\operatorname{Int} D^2$, $f_0(D^2\backslash\Delta)\subset L\backslash\Gamma$, which is homotopy equivalent to K, and the restriction of f_0 to S^1 is essential in $L\backslash\Gamma$.

An elementary argument enables us to find a connected component Δ_1 of Δ, and a map $f_1: D^2\to L$ with $f_1^{-1}(\Gamma)=\Delta_1$, and f_1 restricted to S^1 essential in $L\backslash\Gamma$. Then Δ_1 is the 1-skeleton of a cellular subdivision of $S^2=D^2/S^1$, and the dual subdivision D gives rise to a relative diagram (D,v_0,ϕ) in the obvious way. Here v_0 is the vertex of D corresponding to the unique non-simply connected component of $D^2\backslash\Delta_1$, and has vertex label in the conjugacy class $[f_1|S^1]$.

2. The result

Theorem 2. *Let G be any group, and let $r\in G*\langle t\rangle$ be an element of t-length 3. Then the equation $r(t)=1$ has a solution over G.*

Proof. As remarked in the introduction, we may assume $r(t)$ has the form $atbtct^{-1}$, that $b=1$ in G, and that G is generated by a and c. We will indeed make these assumptions, but will retain the symbol b for convenience as a label.

Suppose that the equation has no solution over G. Then by Lemma 1 there is a relative diagram (D,v_0,ϕ) for $r(t)=1$ such that the vertex-label of v_0 is non-trivial. Let us assume that D is chosen with the smallest possible number of faces. In particular, the vertex labels are all cyclically reduced words in the symbols a, b, c—otherwise two faces may be "cancelled" in the diagram (see e.g. fig. 1).

Note that a corner labelled $a^{\mp 1}$ separates two edges oriented away from that corner; while one labelled $c^{\mp 1}$ separates edges oriented towards it; and one labelled $b^{\mp 1}$ separates an edge oriented towards and an edge oriented away. It follows that all vertex labels have the form a^m, c^m, or $a^{m(1)}b^{-1}c^{n(1)}b\ldots a^{m(k)}b^{-1}c^{n(k)}b$ (up to conjugacy). In particular, the number of occurrences of b is an even integer, no larger than half the index of the vertex.

By Euler's formula, at least one vertex other than v_0 has index 5 or less. If b appears in the label of such a vertex, that label has the form $a^m b^{-1}c^n b$ with $|m|+|n|\leqq 3$. It follows that G is cyclic, and the equation $r=1$ has a solution in G, which is a contradiction. Hence we may assume that any vertex (except possibly v_0) of index $m\leqq 5$ is either a source (label $a^{\mp m}$) or a sink (label $c^{\mp m}$).

If some vertex has index 1, then again G is cyclic, so we may assume no vertex (other than v_0) has index 1. Similarly, we may assume that the vertex labels a^m, a^n ($|m|$, $|n|\leqq 5$) cannot both occur unless $m=n$ or $\{m,n\}=\{2,4\}$. We may also assume that no two vertices of index 5 or less are adjacent (unless one of them is v_0), for then b appears in at least one of the vertex labels.

Form a new subdivision $\bar{D}$ of S^2 from D as follows. Remove any vertex (other than

v_0) of index $m \leqq 5$, together with all incident edges and faces, and replace the m triangular faces removed in this way by a single m-gon. Call a face of $\bar{D}$ a *new* face if it arises from the removal of a vertex of D, and say it is of type a or c depending on the vertex label of the vertex which was removed being $a^{\mp m}$ or $c^{\mp m}$. The corners of the new faces inherit labels from the labelling on D — $(cb)^{\mp 1}$ for a new face of type a, and $(ba)^{\mp 1}$ for a new face of type c. The faces of $\bar{D}$ which are not new are called *old* faces. All old faces are triangular.

Note that no two new faces of the same type can be adjacent in $\bar{D}$, for otherwise cancellation occurs in D. Also, if two edges of an old face meet new faces of the same type (say a) then the corresponding portion of the label of their common vertex reads $(cb)a^{\mp 1}(cb)^{-1}$. In particular, if the sequence of faces around a vertex includes the sequence new, old, new, old, new, then these 3 new faces cannot all be of the same type. Finally, if the vertex label of some vertex includes the sequence $b \cdot a^n \cdot b^{-1}$ (resp. $b^{-1}c^n b$), where n is a multiple of the order of a (resp. c) in G, then D may be altered to give a relative diagram with fewer faces (fig. 2). By the assumption of minimality therefore, the

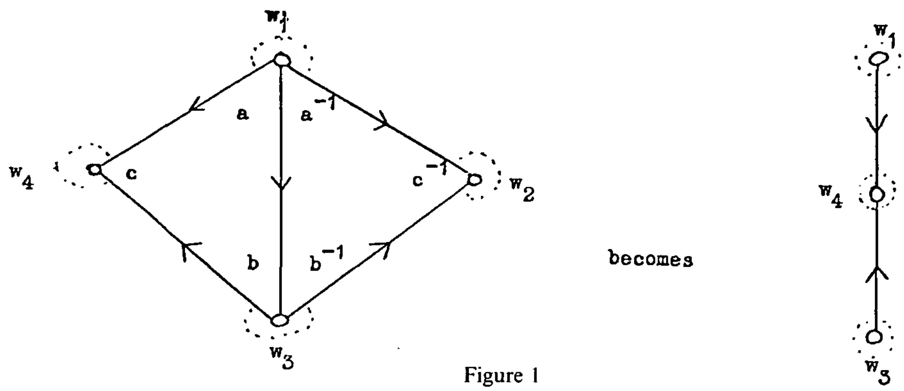

Figure 1

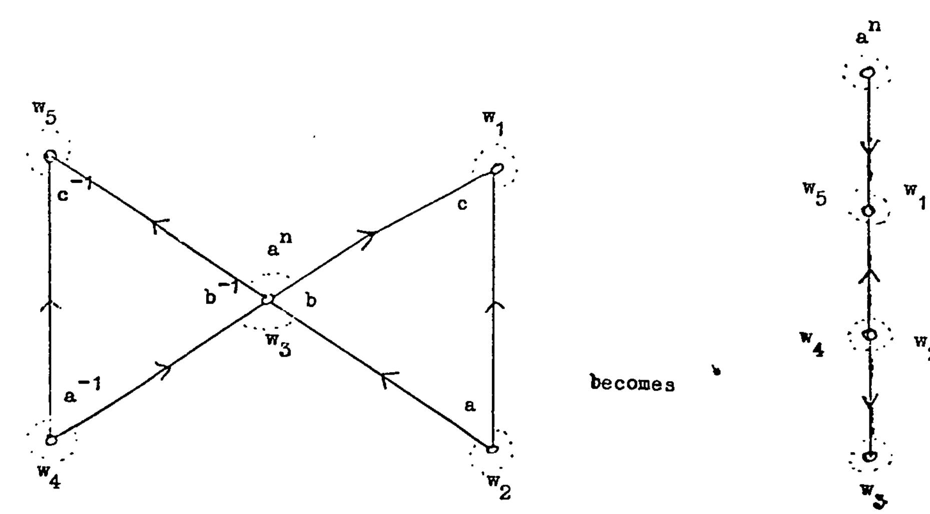

Figure 2

The symbols w_i represent words in $\{a, b, c\}$ which are (parts of) vertex labels.

vertex labels are cyclically reduced words in the free product $\langle a\rangle * \langle b^{-1}cb\rangle$ of the cyclic subgroups of G generated by a and $b^{-1}cb$.

Now associate to each corner of each m-gon of $\bar{D}$ the angle $(m-2)\pi/m$. The sum of all these angles is $2\pi(V-2)$ where V is the number of vertices in $\bar{D}$. It follows that for some vertex v other than v_0, the sum of the angles around v is strictly less than 2π. The argument proceeds by examining the various possible combinations of faces around v.

(1) 2-gons of type a and of type c both occur. Then $a^2=c^2=1$ in G, so G is either infinite dihedral or finite.

From now on, assume that at most one type of 2-gon can occur. In particular, no two are adjacent, and the index of v in $\bar{D}$ is at most 10.

(2) 2-gons (of type a, say) occur, and possibly also 4-gons of the same type, but no other new faces. Then at most 5 old faces occur, and at most 3 new faces. The index of v in $\bar{D}$ is at most 8, and in D at most 11, so b occurs at most 4 times in the vertex label of v. Hence G satisfies the relation $a^2=1$ together with one of:

$$ac^n=1 \qquad (|n|\leqq 8) \quad \text{or} \quad ac^m ac^n=1 \qquad (|m|+|n|\leqq 5).$$

It follows that G is finite, except possibly in one of the cases

$$a^2=1=acac^{\mp 1} \quad \text{or} \quad a^2=1=ac^2ac^{\mp 2}.$$

(3) 2-gons (of type a, say) and new 3-gons (of type c) occur. The index of v in $\bar{D}$ is at most 10, and in D at most 20. Then G satisfies $a^2=c^3=1$, together with a relation of the form

$$(ac)^n ac^{\mp 1}=1 \qquad (0\leqq n<3) \quad \text{or} \quad (ac)^2(ac^{-1})^2=1$$

$$\text{or} \quad (acac^{-1})^2=1 \quad \text{or} \quad (ac)^5=1$$

(The last relation is the only possibility arising from a vertex of index 20). In all cases, G is finite.

(4) 2-gons (of type a) and 5-gons (of type c) occur. There are at most 3 2-gons, at most 3 5-gons, and at most (6—2k) old faces, where k is the number of 5-gons. The index of v in D is at most 12, so G satisfies $a^2=c^5=1$, together with one of the following:

$$ac^m=1 \qquad (5\nmid m) \quad \text{or} \quad ac^m ac^n=1 \qquad (|m|+|n|\leqq 6;\ m,n\neq 0)$$

$$\text{or} \quad (ac)^2 ac^{\mp 1}=1.$$

In all cases, G is finite.

(5) Only 3-gons occur, of which there are at most 5, and at most 4 are new. If both types a and c occur, then $a^3=c^3=1$, and one of

$$ac^{\mp 1}=1 \quad \text{or} \quad ac^{\mp 1}a^{\mp 1}c^{\mp 1}=1 \text{ in } G, \text{ so } G \text{ is finite.}$$

Otherwise, at most 2 new faces occur (say of type a), and G satisfies either

$$a^m c^n = 1 \qquad (|m| + |n| \leqq 3) \quad \text{or} \quad a^3 = 1 \quad \text{and} \quad a^m c^n = 1 \qquad (|m| + |n| \leqq 5).$$

Again G is finite.

(6) 5-gons (of type a) occur, and no 2-gons. There are either 2 5-gons and at most 2 3-gons; or 1 5-gon, 1 4-gon and at most 2 old faces; or 1 5-gon and at most 4 3-gons, at most 2 of which can be new. In all cases, v has index at most 8 in D. If the index is less than 8, then G satisfies

$$a^5 = 1 = a^m c^n \qquad (|m| + |n| \leqq 5),$$

so is finite cyclic. If the index is 8, there are 2 new 3-gons, so G satisfies $a^5 = c^3 = 1$ along with one of

$$ac^{\mp 1} a^{\mp 1} c^{\mp 1} = 1.$$

Again G is finite.

(7) 4-gons (of type a), but no 2-gons or 5-gons occur. There are either 2 4-gons and at most 2 3-gons, or 1 4-gon and at most 4 3-gons, of which at most 2 can be new. The index of v in D is at most 8. If the index is 8, then there are 2 new 3-gons, so G satisfies $a^4 = c^3 = 1$, along with one of $ac^{\mp 1} a^{\mp 1} c^{\mp 1} = 1$. In all cases, G is finite. If the index is 7, then there is 1 new 3-gon, so G satisfies $a^4 = c^3 = 1$ along with one of $a^m c^n = 1$ $(|m| + |n| = 5)$. Hence G is finite cyclic. If the index is less than 7, then G satisfies $a^4 = 1$ and $a^m c^n = 1$ with $|m| + |n| \leqq 4$ $(m, n \neq 0)$. Again, G is finite, except possibly in the cases

most 2 3-gons, or 1 4-gon $a^4 = 1 = a^{\mp 2} c^{\mp 2}$.

(8) 4-gons (of type a) and 2-gons (of type c) occur. There are at most 3 4-gons, at most 3 2-gons, and at most 4, 2 or 1 old faces, depending on whether there are 1, 2, or 3 4-gons. In any case, the index of v in D is at most 13. Hence G satisfies $a^4 = c^2 = 1$, along with one of

$$a^n c = 1 \qquad (4 \nmid n) \quad \text{or} \quad aca^n c = 1 \qquad (4 \nmid n) \quad \text{or} \quad (ac)^2 a^n c = 1 \qquad (4 \nmid n)$$

$$\text{or} \quad aca^{-1} ca^2 c = 1 \quad \text{or} \quad (a^2 c)^2 = 1.$$

In all cases, except possibly the last, G is finite.

A further 7 cases occur when a and c are interchanged in cases (2)–(8) above, but the symmetry between a and c allows us to treat these additional cases in a similar manner. We have thus covered all possible combinations of faces around v, and discovered that G is finite except in a few exceptional cases, when it is a homomorphic image of one of the following groups:

$$\langle a,c \mid a^2=c^2=1\rangle \quad \text{(case 1)}$$

$$\langle a,c \mid a^2=(ac)^2=1\rangle \quad \text{(case 2)}$$

$$\langle a,c \mid a^2=(ac^2)^2=1\rangle \quad \text{(case 2)}$$

$$\langle a,c \mid a^2=[a,c]=1\rangle \quad \text{(case 2)}$$

$$\langle a,c \mid a^2=[a,c^2]=1\rangle \quad \text{(case 2)}$$

$$\langle a,c \mid a^4=a^2c^2=1\rangle \quad \text{(case 7)}$$

$$\langle a,c \mid a^4=c^2=(a^2c)^2=1\rangle \quad \text{(case 8)}$$

Now each of the groups listed above (and so also any homomorphic image of one of these groups) has a free abelian subgroup of finite index, and so in particular is residually finite.

We have deduced that G is residually finite, so we may apply the theorem of Gerstenhaber and Rothaus [**2**, **6**] to show that any equation with non-zero exponent sum has a solution over G, contradicting the hypothesis that $r(t)=1$ has no solution. This completes the proof of Theorem 2.

3. Remarks

The proof of Theorem 2 is somewhat unsatisfactory, as it involves much tedious checking of cases. It also hinges very strongly on the fact that G is in this case essentially a 2-generator group, so a few easily obtainable relations suffice to show G is residually finite. There is some hope that a similar approach would work for other equations of small t-length, but it seems unlikely that this type of argument would lead to a general solution of the problem.

REFERENCES

1. S. D. Brodskii, Equations over groups and groups with one defining relator (Russian), *Uspekhi Mat. Nauk* **35–4** (1980), 183.

2. M. Gerstenhaber and O. S. Rothaus, The solution of sets of equations in groups, *Proc. Nat. Acad. Sci. U.S.A.* **48** (1962), 1531–1533.

3. J. Howie, On pairs of 2-complexes and systems of equations over groups, *J. reine angew. Math.* **324** (1981), 165–174.

4. F. Levin, Solutions of equations over groups, *Bull. American Math. Soc.* **68** (1962), 603–604.

5. R. C. Lyndon, On Dehn's algorithm, *Math. Ann.* **166** (1966), 208–228.

6. O. S. Rothaus, On the nontriviality of some group extensions given by generators and relators, *Ann. Math.* **106** (1977), 559–612.

7. C. P. Rourke, Presentations and the trivial group, *Topology of low-dimensional manifolds, Proc. 2nd Sussex Conf.* (Lecture Notes in Mathematics 722, 1979), 134–143.

8. P. E. Schupp, On Dehn's algorithm and the conjugacy problem, *Math. Ann.* **178** (1968), 119–130.

9. H. B. Short, *Topological methods in group theory: the adjunction problem* (Ph.D. Thesis, University of Warwick, 1981).

Department of Mathematics
University of Glasgow
Glasgow G12 8QW