Illinois Journal of Mathematics
Volume 54, Number 1, Spring 2010, Pages 11–18
S 0019-2082

A Strengthened Freiheitssatz

Paul E. Schupp*
Department of Mathematics, University of Illinois, Urbana, Ill. 61801, USA

The basic result in the theory of one relator groups is, of course, the Freiheitssatz of Magnus [5]. In the case where the defining relator is a proper power, the "Spelling Theorem" of Newman [8, 9] gives sharper results. At the 1974 Calgary Conference on infinite group theory, Steve Pride told me that Gurevich [1] had strengthened Newman's theorem. Steve asked whether this result could be further improved. Reflection on the matter led to the discovery that there is a single theorem which strengthens both the Freiheitssatz for one relator groups in general and Newman's results in the torsion case.

We state the general theorem below. Perhaps the most interesting consequence is the following. Let $G=\langle a,b,c,\dots;r\rangle$ where r is cyclically reduced. Let R^* be the symmetrized set generated by r, that is, R^* consists of all cyclic permutations of $r^{\pm 1}$. If u is a non-trivial freely reduced word such that $u=1$ in G, then u has a subword s which contains all the generators occurring in r, and such that s is also a subword of an element of R^*.

We state the general theorem in an "equational form".

Theorem. *Let $G=\langle a,b,c,\dots;r\rangle$ where r is cyclically reduced. Write $r=z^n$, $n\geqq 1$, where z is not a proper power in the free group on $a,b,c,\dots$. (Elements of the symmetrized set R^* generated by r thus have the form $(z^*)^n$ where z^* is a cyclic permutation of $z^{\pm 1}$.) If an equation $u=v$ holds in G where u and v are freely reduced words and v omits a generator which occurs in both r and u, then u contains a subword t of an element of R^* such that $t\equiv(z^*)^{n-1}s$ and s contains every generator which occurs in r but not v. (If $n=1$, then t is simply s.)*

The case $n>1$ is announced in Gurevich [1].

The amount of additional information which the theorem yields depends, of course, on the form of the defining relator r. It is interesting to note that a "small cancellation" type of conclusion follows for some groups purely by one-relator methods. For example, if $G=\langle a_1,\dots,a_g;\, a_1^2\dots a_g^2=1\rangle$, then a subword containing all the generators must be of length at least $2(g-1)$.

* This research was supported by the National Science Foundation and a sabbatical leave from the University of Illinois.

The following consequence about free subgroups is different from but in the spirit of Theorem 1.3.11 of Newman [9]. Let $G=\langle a, b, c, d, \ldots; r\rangle$ where r is cyclically reduced and involves at least the generators a, b, and c. Then there is an integer m such that $a^m, b^m, c^m, d, \ldots$ freely generate a free subgroup of G. Indeed, all that is necessary is to take m greater than the maximum absolute value of any power of a, b, or c occuring in a permutation of r. For then any part of a word of the form $w(a^m, b^m, c^m, d, \ldots)$ which contains occurrences of a, b, and c must contain an m-th power and thus could not be a subword of a permutation of $r^{\pm 1}$.

Our proof is geometric and makes essential use of Lyndon's "maximum modulus" approach to the Freiheitssatz [3]. It is interesting to note that efforts to find a non-geometric proof have so far failed. The difficulty is with the case where no generator has exponent sum zero. This case is not exceptional in the geometric approach.

If all of the generators of G do not occur in r, we can write $G=K*L$ where K is the free group on all the generators not occuring in r. The normal form theorem for free products quickly shows that if the theorem is true for L then it is also true for G. We will therefore assume from now on that all the generators of G occur in the defining relator r.

We turn to the necessary notation and terminology. The reader familiar with a non-geometric proof of the Freiheitssatz knows that the basic idea is to rewrite the defining relator. For example, if $G=\langle a, b, c;\, b^2a^{-2}b^2c^2a^2\rangle$, we can view G as an HNN extension of the one relator group

$$H=\langle b_{-2}, b_{-1}, b_0, c_i, i\in\mathbb{Z};\, b_0^2b_{-2}^2c_{-2}^2\rangle$$

since

$$G=\langle H, a;\, ab_{-2}a^{-1}=b_{-1}, ab_{-1}a^{-1}=b_0, ac_ia^{-1}=c_{i+1}, i\in\mathbb{Z}\rangle\,.$$

(Compare [6].)

In short, one now has many generators and relators with subscripts. We formalize this situation, since it also arises in the geometric approach, although in quite a different manner.

Consider a presentation $G=\langle X; R\rangle$. We assume that X is a disjoint union, say $X=\cup X_\tau$, and say that the generators in each X_τ are of the same *type*. We also assume that there is an integer valued function assigning a *subscript* to each generator in X. (In our example above, a type would consist of a letter with subscripts, say all the c_i, $i\in\mathbb{Z}$.) We also require that there is a unique *subscript* assigned to each relator in R, say $R=\{r_j : j\in J\subseteq\mathbb{Z}\}$. Furthermore, we required that there is a fixed integer $n\geqq 1$ such that each $r_j=z_j^n$ where z_j is not a proper power in the free group on X.

If w is a word on $X^{\pm 1}$ and τ is a type, then $\max_\tau(w)$ will denote the generator of type τ with maximum subscript which occurs in w. (We will use this notation only when w actually contains a generator of type τ.) Define $\min_\tau(w)$ similarly. Let X and R be as above. We say that the presentation $\langle X; R\rangle$ is *staggered* if every relator in R contains at least one generator of each type and the following condition holds for every type τ:

If μ can be chosen so that $\hat{R}$ contains at least two different relators, then we have succeeded in reducing the proof to the previous case. Now Lyndon shows that such a choice of μ is possible provided that we make one preliminary modification of the labelling on the diagram M. Call two positions p_i and p_j *immediately related*, written $p_i \sim p_j$, if $p_i = p_j$ or if there are regions D_1 and D_2 of M with an edge $e \in \partial D_1 \cap \partial D_2$ such that the label on e occurs at position p_i in the label on ∂D_1 and occurs at position p_j in the label on ∂D_2.

Call two positions p_i and p_j *related*, written $p_i \approx p_j$, if there is a sequence of positions such that

$$p_i \sim p_k \sim \cdots \sim p_l \sim p_j.$$

What we desire in the original relator r is that any two positions p_i and p_j such that y occurs in p_i and $y^{\pm 1}$ occurs in p_j are related. If this is not the case, take each equivalence class under $\approx$ and relabel it with a new generator. This procedure is easily seen to be consistent. Let M' be the diagram obtained after this modification of the labelling. It is clear that the modulus principle for M' is, if anything, stronger than the result for M. Under the assumption that any two positions labelled by the same $y^{\pm 1}$ are related, Lyndon now shows how to choose μ so that at least two distinct relators will be present after subscripting.

We now state a geometric "omission principle" which implies the theorem we wish to prove.

Theorem. *Let $\langle X; R\rangle$ be a staggered presentation. Let M be a connected simply connected reduced R^*-diagram with at least one region and no vertices of degree one. Suppose that $\alpha\beta$ is a boundary cycle of M where $\varphi(\beta)$ omits* $\max_\sigma(\partial M)$ *for some type σ. Then M has a region D such that $\alpha' = \alpha \cap \partial D$ is a consecutive part of both α and ∂D and $\varphi(\alpha') \equiv (z^*)^{n-1}s$ where $(z^*)^n$ is a boundary label of D and s contains an occurrence of a generator of type τ for every τ such that $\varphi(\beta)$ omits* $\max_\tau(\partial M)$. *The same statement holds with* "max" *replaced by* "min".

Proof. Note that the geometric version implies the non-geometric version, for suppose that an equation $u = v$ holds in G where u and v are freely reduced and v omits $\max_\sigma(u, v)$ for some type σ. As noted previously, we may assume that uv^{-1} is cyclically reduced without cancellation. Apply the present theorem to a reduced R^*-diagram with boundary label uv^{-1}.

The proof is by induction on the number of regions of M with the induction being applied simultaneously over all staggered presentations. If M has only one region the result clearly holds. We carry out the argument assuming that $\varphi(\beta)$ omits $\max_\sigma(\partial M)$. The other case will follow by interchanging "max" and "min".

We first show how the result follows from the induction hypothesis if ∂M is not a simple closed path. In this situation pick an extremal disk J of M. (See Lemma 3 of [7].) Let J be attached to the rest of M at the vertex v_2, and let λ be the boundary cycle of J beginning at v_2. (See Fig. 1.) Let v_1 be the vertex preceeding v_2 and closest to v_2 so that, if η is the stem running from v_1 to v_2, the removal of $\eta\lambda\eta^{-1}$ (and its interior) from M leaves a map K with no vertices of degree one. (It is possible that $v_1 = v_2$ and η is empty.)

label on a region of M. Pick a region D labelled by r_k and let K be the maximal connected submap of M satisfying:

K contains D and all edges of K are on the boundaries of regions labelled by relators with subscript k.

By the maximality of K, if a region E which is labelled by an r_k has a boundary edge in K, then E is in K.

Now K is connected by definition. We claim that K is also simply connected. If not, let J be a bounded component of $M-K$. Now J contains a region labelled by some r_j with $j<k$, and J has fewer regions than M. By the induction hypothesis, ∂J contains edges labelled by $\min_\tau(J)$ for every type τ. But all the edges of ∂J are on the boundaries of regions labelled by relators with subscript k. Since $\langle X;\, R\rangle$ is staggered, $\min_\tau(J) < \min_\tau(r_k)$, which is a contradiction establishing the claim.

Since K is not all of M, the induction hypothesis says that ∂K contains edges labelled by $\max_\tau(r_k)$. By the maximality of K, these edges must also be in ∂M. Minimum modulus follows by interchanging "max" and "min".

Thus we are left with the case where there is one relator r such that all the regions of M are labelled by cyclic permutations of $r^{\pm 1}$. We recall Lyndon's ingenious procedure for reducing this case to the previous case. We change notation and consider each distinct generator in r to be of a different type and to be without subscript. Let Y be the set of generators occuring in r, and let F be the free group on Y. Recall that each edge of M is labelled by a generator or its inverse. Let $r=z^n$ where z is not a proper power in F.

Consider a function $\mu\colon F\to\mathbb{Z}$ with $\mu(z)=0$. Then μ induces a function on paths α in M by defining $\mu(\alpha)=\mu(\varphi(\alpha))$. Since μ vanishes on boundary cycles of regions of M, we have $\mu(\beta)=0$ for every closed path β in M. Pick a vertex v_0 on the boundary of a region D_0 of M. For any vertex v of M, define the "potential function" $\varrho(v)$ by $\varrho(v)=\mu(\alpha)$ where α is any path from v_0 to v. Since μ vanishes on closed paths, ϱ is independent of path and is thus well defined. Let e be an edge of M, say with label $\varphi(e)=y\in Y^{\pm 1}$, and let v_1 and v_2 be the endpoints of e. We subscript this occurrence of y by assigning to it the sum of the potentials at the endpoints, that is, we give y the subscript $\varrho(v_1)+\varrho(v_2)$. (Our definition of this subscript differs slightly but inessentially from Lyndon's definition.)

Write $z=y_1 \dots y_t$ with each $y_i\in Y^{\pm 1}$. Since z is not a proper power, starting at each y_i determines a distinct cyclic permutation of z. We thus think of the letters in $r=z^n$ as being in one of the distinct *positions* $p_1, \dots, p_t$. (Each position occurs n times in r.) Before subscripting, we may assume that r was the boundary label on D_0 beginning at v_0. If E is a region of M, pick a vertex $v_1\in\partial E$ and an orientation so that r is the boundary label on ∂E beginning at v_1 and reading in the chosen direction. Let $\hat{r}$ and $\hat{r}^*$ be the subscripted relators obtained from ∂D_0 and ∂E read as above. Let $d=\mu(\alpha)$ where α is a path from v_0 to v_1. Then the subscript on a generator occurring in position p_i in $\hat{r}^*$ differs from the subscript on a generator occurring in position p_i in $\hat{r}$ by $2d$.

Thus we have the following situation. Let $\hat{Y}$ be the set of generators with subscripts which now occur on edges of M. Let $\hat{R}$ be the set of relators obtained by reading (as above) the boundaries of regions of M. Pick a type σ of generator in $\hat{Y}$. Assign to each $\hat{r}\in\hat{R}$ the minimum of the subscripts on generators of type σ which occur in $\hat{r}$. Then the presentation $\langle\hat{Y};\hat{R}\rangle$ is staggered.

If $i<j$, then $\max_\tau(r_i)<\max_\tau(r_j)$ and $\min_\tau(r_i)<\min_\tau(r_j)$. (We have used an obvious extension of notation, writing $x_i<x_j$ if x_i and x_j are generators of the same type with $i<j$.)

We must, of course, work with the symmetrized set R^* generated by R. If some $r\in R^*$ is a cyclic permutation of $r_j^{\pm 1}$, we also consider r as having the subscript j assigned to it. We now formulate our theorem as an "omission principle" in terms of staggered presentations.

Theorem. *Let G have the staggered presentation $\langle X\,;R\rangle$. If an equation $u=v$ holds in G where u and v are freely reduced words and, for some type σ, v omits* $\max_\sigma(u,v)$, *then u contains a subword t of some $r_j^*\in R^*$ where $t\equiv(z^*)^{n-1}s$ and s containes a generator of type τ for every τ such that v omits* $\max_\tau(u,v)$. *The same statement also holds with* "max" *replaced by* "min".

This statement about staggered presentation clearly implies the desired result for one relator presentation. For, if $G=\langle X;r\rangle$, we can regard each distinct generator as being of a different type and having the subscript zero assigned to it.

One further comment is required at this point. Suppose that an equation $u=v$ holds in G with u and v as above. If u and v have a non-trivial common initial segment, say $u=wu_1$ and $v=wv_1$, then $u_1=v_1$ in G, u_1 and v_1 satisfy the hypothesis of the theorem, and the conclusion applied to u_1 and v_1 yields the conclusion for u and v. The same remark applies if u and v have a common terminal segment. Thus from now on, we need consider only equations $u=v$ where uv^{-1} is cyclically reduced without cancellation.

We assume that the reader is familiar with cancellation diagrams; for example, with Lyndon [2] or [3], or Miller and Schupp [7]. Let $\langle X;\ R\rangle$ be a staggered presentation with R^* the symmetrized set generated by R. If M is an R^*-diagram, the boundary of M will be denoted by ∂M, and the labelling function for M will always be denoted by φ. If α and β are paths, we use the notation $\alpha\subseteqq\beta$ to mean that α is a subpath of β, that is, the edges in α are a consecutive subsequence of the edges in β. Recall that a diagram M is *reduced* if it is not the case that there are regions D_1 and D_2 of M with an edge $e\subseteqq\partial D_1\bigcap\partial D_2$ such that if we remove the edge e, combining D_1 and D_2 into a single region D, the resulting region D has boundary label equal to 1 in the free group on X. For our purposes, we will consider the edges of M subdivided so that each edge is labelled by a generator or its inverse. For any type τ, $\max_\tau(M)$ will denote the generator of type τ with maximum subscript which occurs as the label on an edge of M. Define $\max_\tau(\partial M)$ similarly, but with the maximum taken over edges in ∂M. Define $\min_\tau(M)$ and $\min_\tau(\partial M)$ analogously.

In [3] Lyndon proves the *Maximum-minimum Modulus Principle:* Let M be a connected simply-connected reduced R^*-diagram. Then for each type τ, there are edges in ∂M labelled by $\max_\tau(M)$ and $\min_\tau(M)$.

Since we shall need not only this result but a firm grasp on the details of the proof, we sketch the proof, which is by induction on the number of regions of M. The induction is applied simultaneously over all staggered presentations. If M has only one region the result holds.

We first suppose that M contains regions labelled by relators with different subscripts. Let k be the maximum subscript on any relator which occurs as the

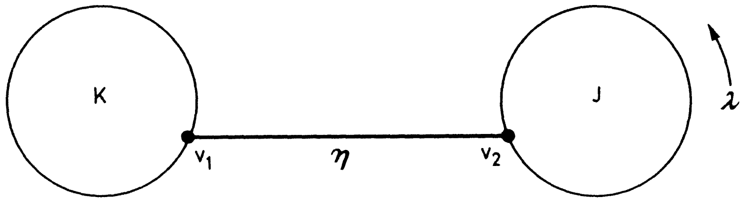

Fig. 1

We consider three cases. First suppose that $\lambda \subseteqq \alpha$. By the maximum modulus principle applied to J, $\varphi(\lambda)$ contains $\max_\tau(J)$ for every type τ. Let γ be the empty path at v_2. Clearly then, $1 = \varphi(\gamma)$ omits $\max_\tau(\varphi(\lambda))$ for every τ. By the induction hypothesis, the theorem holds for J and the boundary cycle $\lambda\gamma$. Since $\lambda \subseteqq \alpha$, the theorem holds for M and $\alpha\beta$. Suppose next that $\eta\lambda\eta^{-1} \supseteqq \beta$. Let α_1 be the boundary cycle of K beginning at v_1, and note that $\alpha_1 \subseteqq \alpha$ in view of the assumption on β. Let γ_1 be the empty path at v_1. The result follows as in the previous case by applying the induction hypothesis to K and the boundary cycle $\alpha_1\gamma_1$.

Finally, suppose that neither of the above situations apply, that is, both λ and ∂K contain edges of β. Thus K has a boundary cycle $\alpha_2\beta_2$ where $\alpha_2 \subseteqq \alpha$ and $\beta_2 \subseteqq \beta$. Similarly, J has a boundary cycle $\alpha_1\beta_1$ where $\alpha_1 \subseteqq \alpha$ and $\beta_1 \subseteqq \beta$. (It is possible that one of α_1 or α_2 is empty.) Now $\max_\sigma(\partial M)$ cannot occur as the label on an edge in η. For, such an edge could not be in β and we would have $\beta \subseteqq \eta\lambda\eta^{-1}$, contrary to assumption. Thus at least one of $\varphi(\alpha_1)$ or $\varphi(\alpha_2)$ contains an occurrence of $\max_\sigma(\partial M)$ and the result follows by induction.

We are thus left with considering only the situation where ∂M is a simple closed path. Following Lyndon's dichotomy, we first consider the case where M contains regions labelled by relators with different subscripts. Let k be the maximum subscript on relators labelling regions of M. By the maximum modulus principle,

$$\max_\sigma(M) = \max_\sigma(\partial M) = \max_\sigma(r_k),$$

so it is $\max_\sigma(r_k)$ which $\varphi(\beta)$ omits.

We now show how the result follows from the induction hypothesis if any region D_1 labelled by some r_j with $j<k$ has even so much as a single vertex v_1 of its boundary in common with β. (We include here the case that β is the empty arc based at v_1.) For, we can excise D_1 to form a new diagram M_1 as follows.

Pull apart the boundary of D_1 at the vertex v_1. Form M_1 by deleting from M the region D_1 and any edges in the component of $\partial D_1 \cap \partial M$ which contains v_1. (See Fig. 2)

Now M_1 has a boundary cycle $\beta_1\alpha_1$ where $\alpha_1 \subseteqq \alpha$ and all edges of β_1 are edges which were originally in β or in ∂D_1. Since the presentation $\langle X; R\rangle$ is staggered, $\max_\sigma(r_j) < \max_\sigma(r_k)$ and thus $\varphi(\beta_1)$ omits $\max_\sigma(\partial M_1)$. The result now follows from the induction hypothesis applied to M_1 and the boundary cycle $\alpha_1\beta_1$.

We may now suppose that all regions having any part of their boundary on β are labelled by a cyclic permutation of $r_k^{\pm 1}$. Let K be a maximal connected submap of M such that:

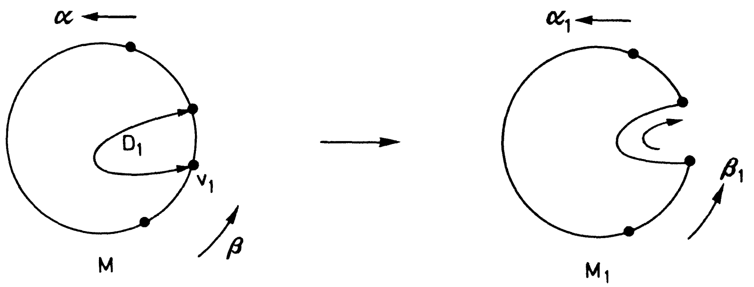

Fig. 2

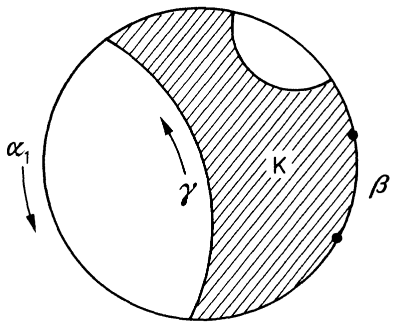

Fig. 3

K contains β and all edges in K are on the boundaries of regions labelled by relators with subscript k.

Now K is connected by definition, and, as in the proof of the modulus principle, K is also simply connected. Some component of $\partial K \cap \partial M$ contains β. Thus there is a simple path γ lying entirely in K and a subpath $\alpha_1 \subseteqq \alpha$ so that $\alpha_1\gamma$ bounds a disk J containing all regions of M not labelled by relators of subscript k. (This property of J is possible since K is simply connected. See Fig. 3.)

Thus J contains at least one region labelled by some r_j with $j<k$. By the minimum modulus principle, $\min_\tau(J)$ equals $\min_\tau(\partial J)$. Since γ lies entirely in K and the presentation $\langle X; R\rangle$ is staggered, $\varphi(\gamma)$ omits $\min_\tau(\partial J)$ for every type τ. The theorem now follows from the induction hypothesis applied to J and the boundary cycle $\alpha_1\gamma$.

Finally, we are left with the case where all regions of M are labelled by cyclic permutations of a single $r^{\pm 1}$. To conclude our proof, it remains only to note that if, as in the establishment of the modulus principle, we follow Lyndon's preliminary modification and subscripting procedures, the conclusion for the relabelled diagram M implies the conclusion for the original M.

Original article: Schupp, P. E.: A Strengthened Freiheitssatz. Math. Ann. 221, 73–80 (1976), reprinted with kind permission of Springer Science + Business Media.

References

1. Gurevich, G. A.: On the conjugacy problem for one relator groups. Dokladi Akademii Nauk SSSR, Ser. Mat. **207**, 18—20 (1972). (English translation in Soviet Math. Dokl. **13**, 1436—1439 (1972))
2. Lyndon, R. C.: On Dehn's Algorithm. Math. Ann. **166**, 208—228 (1966)
3. Lyndon, R. C.: On the Freiheitssatz. J. London Math. Soc. **5**, 95—101 (1972)
4. Lyndon, R. C., Schupp, P. E.: Combinatorial Group Theory. Springer, Berlin, Heidelberg, New York: Springer 1976
5. Magnus, W.: Über diskontinuierliche Gruppen mit einer definierenden Relation (Der Freiheitssatz). J. reine u. angew. Math. **163**, 141—165 (1930)
6. McCool, J., Schupp, P. E.: On one relator groups and HNN extensions. J. Australian Math. Soc. **16**, 249—256 (1973)
7. Miller, C. F., III, Schupp, P. E.: The geometry of HNN extensions. Comm. Pure Appl. Math. **26**, 787—802 (1973)
8. Newman, B. B.: Some results on one relator groups. Bull. Amer. Math. Soc. **74**, 568—571 (1968)
9. Newman, B. B.: Some aspects of one relator groups. Ph. D. Thesis, University of Queensland 1968

Received August 12, 1975

Note added in Proof. Gurevich has recently informed me that his methods of proof will also yield the theorem of this paper when the relator is not a proper power.