On bounded languages and the geometry of nilpotent groups

MARTIN R. BRIDSON AND ROBERT H. GILMAN [1]

Abstract

Bounded languages are a class of formal languages which includes all context free languages of polynomial growth. We prove that if a finitely generated group G admits a combing by a bounded language and this combing satisfies the asynchronous fellow traveller property, then either G is virtually abelian, or else G contains an element g of infinite order such that g^n and g^m are conjugate for some $0 < n < m$.

The introduction of automatic groups [5] has precipitated a host of questions about the roles which formal language theory and geometry play in the study of normal forms for finitely generated groups, particularly groups which arise in geometric settings. For example, when a group G is given as the fundamental group of a compact Riemannian manifold, words in a fixed set of generators for G have a natural interpretation as paths in the universal cover of the manifold; it is natural to ask how the geometry of the manifold is reflected in the linguistic complexity of normal forms for elements of G. The results presented here and in [3] can be interpreted as providing a partial answer to this question in the case where the manifold under consideration is a quotient of a nilpotent Lie group.

It has become customary in geometric group theory to refer to a set of normal forms for elements in a finitely generated group as a *combing* of the group. There is much work to be done on the problem of determining how various geometric and linguistic constraints on the type of combings which a group admits are reflected in the structure of the group. The results presented here contribute to this task. These results arose in the course of our work on the structure of normal forms for elements in 3-manifold groups [3].

Theorem A *If a finitely generated group G admits a combing by a bounded language, and if this combing satisfies the asynchronous fellow traveller property, then either*

1. *G is virtually abelian, or*
2. *there is an element $g \in G$ of infinite order such that for some m, n with $0 < m < n$, g^m and g^n are conjugate in G.*

[1] The first author was supported in part by NSF grant DMS-9203500 and FNRS (Suisse). The second author thanks the Institute for Advanced Study for its hospitality while this paper was being written

Bounded languages are defined in Section 1, as is the asynchronous fellow traveller property.

We do not know of an example for which the above possibility (2) occurs. Certainly, one can exclude possibility (2) by placing restrictions on the class of groups considered. For example, it is shown in [1] that semihyperbolic groups, which are defined in terms of the type of combings which they admit, do not contain elements of the type described in possibility (2). (The class of semihyperbolic groups includes all biautomatic groups [8], and all groups which act properly and cocompactly by isometries on any 1-connected space of non-positive curvature, as all finitely generated virtually abelian groups do.)

Corollary B *A semihyperbolic group G admits a combing by a bounded language with the asynchronous fellow traveller property if and only if G is virtually abelian.*

Theorem A is proved by reducing it to:

Theorem C *If a finitely generated virtually nilpotent group G admits a combing by a bounded language, and if this combing satisfies the asynchronous fellow traveller property, then G is virtually abelian.*

Theorem C plays an important role in [3], where it is used to show that a virtually nilpotent group with a context free combing satisfying the asynchronous fellow traveller property is virtually abelian. In [3] we presented a purely algebraic proof of Theorem C, but this result is essentially a fact about the *geometry* of nilpotent groups. Here we try to present as accessible an account of this geometry as possible.

The results of this article were presented by one of the authors in April 1993 at the meeting on Geometric Methods in Group Theory hosted by the ICMS in Edinburgh. We would like to thank Andrew Duncan, Nick Gilbert and Jim Howie, not only for arranging such an enjoyable conference, but also for the courteous and efficient way in which they have behaved as editors of these proceedings.

1. Definitions and Preliminary Results

Throughout this paper A stands for a finite set and A^* for the free monoid on A. The length of a word $w \in A^*$ shall be denoted $|w|$. The empty word is denoted ϵ. A *formal language* is just a subset $L \subset A^*$. A language L is said to be *bounded* if there are words $w_1, \ldots, w_n$ in A^* such that every $w \in L$ can be written $w = w_1^{m_1} \ldots w_n^{m_n}$ for some choice of non-negative integers $m_i \in \mathbb{N}$. Bounded languages were introduced by Ginsburg and Spanier [7].

Let G be a finitely generated group. A *choice of generators* for G is a map $\mu : A \to G$ from a finite set which extends to a surjective monoid homomorphism $\mu : A^* \twoheadrightarrow G$. We assume that A is closed under formal inverse and that formal inverses are extended to A^* in the usual way. We shall usually write $\overline{w}$ rather than $\mu(w)$ for the image in G of a word w, and more generally $\overline{X}$ for the image of a set of words X.

A *combing* of G is a language $L \subset A^*$ which projects bijectively to G. We shall often use the letter $\mathcal{C}$ to denote a language which is a combing. If, in addition, $\mathcal{C}$ is a bounded language, then we shall refer to it as a combing of G by a bounded language. It is an easy exercise to check that if a finitely generated group admits a combing by a bounded language with respect to one choice of generators, then it admits such a combing with respect to every choice of generators.

Combings by bounded languages arise naturally when considering normal forms for polycyclic groups. For example, if $\{1\} = G_0 \subset \cdots \subset G_n = G$ is a normal tower for G, with each G_i/G_{i-1} cyclic, then arguing by induction on n one may assume that G_{n-1} admits a combing by a bounded language, and if we fix an element $a \in G - G_{n-1}$ whose image generates G_n/G_{n-1}, then by appending suitable powers of a to the words in the combing for G_{n-1} we obtain a combing of G by a bounded language.

Any choice of generators determines a word metric on G by $d(g,h) = \min\{|w| \mid w \in A^*, \mu(w) = g^{-1}h\}$. It is straightforward to check that d is indeed a metric and that it is left-invariant in the sense that $d(gh_1, gh_2) = d(h_1,h_2)$ for all $g, h_1, h_2 \in G$. The metrics d_1, d_2 determined by different choices of generators are Lipschitz equivalent; that is, $(1/c)\, d_1(g,h) \leq d_2(g,h) \leq c\, d_1(g,h)$ for some constant c.

For a given choice of generators, each word $w = a_1 \ldots a_n$ determines a discrete path $p_w : \mathbb{N} \to G$ given by

$$p_w(t) = \begin{cases} 1 & \text{if } n = 0 \\ \overline{a_1 \ldots a_t} & \text{if } 1 \leq t < n \\ \overline{w} & \text{if } n \geq t. \end{cases}$$

Henceforth we shall identify w with p_w and often talk of words as discrete paths in G. The synchronous distance D_s between two words $w, v \in A$ is the maximum separation of points (fellow-travellers) traversing the two corresponding paths at unit speed.

$$D_s(w,v) = \max_t \{d(p_w(t), p_v(t))\}.$$

There is also a notion of asynchronous distance in which each point is allowed to stop for a while. The standard technical device to encode this idea is the set $\mathcal{R}$ of all unbounded maps $\rho : \mathbb{N} \to \mathbb{N}$ such that $\rho(0) = 0$, and $\rho(n+1) = \rho(n)$ or $\rho(n) + 1$. The asynchronous distance is defined to be:

$$D_a(w,v) = \min_{\rho_1,\rho_2 \in \mathcal{R}} \{ \max_{n \geq 0} \{ d(p_w(\rho_1(n)), p_v(\rho_2(n))) \} \}. \tag{1.1}$$

We shall be extensively concerned with constraints of the form $D_a(w,v) \leq K$. This inequality can be rephrased as follows: There exist sequences of prefixes

$$x_0 = \epsilon,\, x_1,\, \ldots,\, x_N = w \quad y_0 = \epsilon,\, y_1,\, \ldots,\, y_N = v$$

of w and v respectively, such that for all i,

$$|x_i| \leq |x_{i+1}| \leq |x_i| + 1 \quad |y_i| \leq |y_{i+1}| \leq |y_i| + 1 \quad \text{and} \quad d(\overline{x_i}, \overline{y_i}) \leq K. \quad (1.2)$$

A language $L \subset A$ is said to satisfy the *synchronous fellow traveller property* if there exists a constant $K > 0$ such that for all $w, v \in L$, if $d(\overline{w}, \overline{v}) \leq 1$ then $D_s(w,v) \leq K$. The *asynchronous fellow traveller property* is defined analogously. Throughout this paper *we shall retain the symbol K to denote the constant in the definition of the (synchronous or asynchronous) fellow traveller property.*

Remark. Notice that if $L \subset A^*$ satisfies the asynchronous fellow traveller property, than so does every sublanguage of it. Since a sublanguage of a bounded language is itself bounded, we see that the hypothesized existence of the combing in Theorem A could be replaced by the existence of any bounded sublanguage of A^* which maps onto G and satisfies the asynchronous fellow traveller property.

We claim that D_s and D_a are pseudometrics on A^*; that is, they satisfy all the requirements of a metric except the requirement that distinct points be a positive distance apart. This assertion is an immediate consequence of the following lemma, whose proof we leave to the reader.

Lemma 1.1

1. *$\mathcal{R}$ is closed under composition.*
2. *If $D_a(w,v)$ is realized by ρ_1 and ρ_2 as in (1.1), then it is also realized by $\rho_1 \circ \rho$ and $\rho_2 \circ \rho$, where $\rho \in \mathcal{R}$ is arbitrary.*
3. *Given $\rho_1, \rho_2 \in \mathcal{R}$, there exist $\rho', \rho'' \in \mathcal{R}$ such that $\rho_1 \circ \rho' = \rho_2 \circ \rho''$.*

We note some elementary properties of D_a.

Lemma 1.2

1. *If $\overline{w} = \overline{v}$, then $D_a(wu, vu') \leq \max\{D_a(w,v), D_a(u,u')\}$ for all $u, u' \in A^*$.*
2. *If v is obtained from w by deleting any number of disjoint subwords xx^{-1}, then $D_a(w,v) \leq |x|$.*

3. *If* $\overline{w} = \overline{x}\,\overline{v}\,\overline{x}^{-1}$, *then* $D_a(w^n, xv^nx^{-1}) \leq D_a(w, xvx^{-1}) + |x|$ *for all* $n \in \mathbb{N}$.

Proof. From a geometric viewpoint, assertion (1) is clear. An algebraic proof can be obtained using (1.2): Let $K = \max\{D_a(w,v), D_a(u,u')\}$; multiply the sequences of prefixes used to compute $D_a(u,u')$ on the left by w and v, then append them to the sequences for $D_a(w,v)$ to get sequences of prefixes for wu and vu'; if u_i, u_i' is a pair of prefixes for u, u' then $d(\overline{wu_i}, \overline{vu_i'}) = d(\overline{u_i}, \overline{u}_i') \leq K$, because $\overline{w} = \overline{v}$ and d is left invariant.

(2) is proved by estimating $D_a(w,v)$ using the unit speed parametrization $\rho(n) = n$ for p_w and the parametrization for p_v which causes a point traversing the image of p_v to move with unit speed except for remaining stationary during the period in which p_w traces out the subword xx^{-1}. Part (3) follows from (1) , (2) and the triangle inequality for D_a.

It is shown in [3], Proposition 1.3 that every context free language of polynomial growth is bounded. We will need the following additional result on bounded languages [3], Lemma 1.4.

Lemma 1.3 *Every bounded language* $L \subset A^*$ *can be expressed as the union of finitely many bounded sublanguages* L_l *such that for each* l *there exists an integer* r *and a choice of words* $u_{l,1}, \ldots, u_{l,r}, v_{l,0}, \ldots, v_{l,r} \in A^*$ *with*

$$L_l = \{v_{l,0}u_{l,1}^{n_1}v_{l,1}\ldots u_{l,r}^{n_r}v_{l,r} \mid (n_1, \ldots, n_r) \in S_l\},$$

where $S_l \subset \mathbb{N}^r$ *is empty if* $r = 0$ *(in which case* $L_l = \{v_{l,0}\}$*) and otherwise there exist* $r-$*tuples in* S_l *whose smallest entry is arbitrarily large.*

2. Reduction of Theorem A to Theorem C

Throughout this section $\mathcal{C}$ will denote the combing hypothesized in Theorem A. We begin by showing that we are free to assume that $\mathcal{C}$ enjoys certain extra properties. Since $\mathcal{C}$ is bounded, there are words $w_i \in A^*$ such that

$$\mathcal{C} = \{w = w_1^{n_1}\ldots w_r^{n_r} \mid (n_1, \ldots, n_r) \in S\} \qquad (2.1)$$

for some subset S of $\mathbb{N}^r$. We work in terms of this fixed decomposition of $\mathcal{C}$. We say that w_i has *bounded exponent* if n_i is bounded as w ranges over $\mathcal{C}$. Otherwise w_i is said to have *unbounded exponent.*

Lemma 2.1 *Without loss of generality, the combing* $\mathcal{C}$ *may be assumed to have the following additional properties:*

1. *the exponent to which each* w_i *appears in any word of* $\mathcal{C}$ *is less than the order of* $\overline{w_i}$ *in* G*;*

2. *if w_i and w_j are both of unbounded exponent, and if some positive powers of $\overline{w_i}$ and $\overline{w_j}$ are conjugate in G, then $w_i = w_j$.*

Proof. We suppose that $\mathcal{C}$ has been chosen so that the number of distinct words w_i of unbounded exponent is minimal amongst all combings of G which are bounded languages and satisfy the asynchronous fellow traveller property. We emphasize that we have chosen our meaning carefully here; we have minimized the cardinality of the subset of $\{w_i\}$ consisting of words with unbounded exponent without counting multiplicities to account for the case where $w_i = w_j$ for some $i \neq j$.

Suppose that $\overline{w_i}$ has order m. If $m = 1$, we delete occurrences of w_i so as to decrease the integer r in the definition of $\mathcal{C}$. Otherwise, for each $w \in \mathcal{C}$, we replace the subword $w_i^{n_i}$ by w_i^q where $n_i = mp + q$ with $0 \leq q < m$. The image of w in G is not changed by this procedure, so $\mathcal{C}$, thus modified, is still a combing; and the number of distinct words of unbounded exponent has not increased. Moreover, by repeated application of Lemma 1.2 (1) , we have that $D_a(xw_i^{n_i}y, xw_i^q y) \leq D_a(w_i^m, \epsilon)$, and hence, by the triangle inequality for D_a, we see that $\mathcal{C}$ is still asynchronously bounded after being modified as above.

For (2) suppose for $0 < m \leq n$ and for some $x \in A^*$, $xw_i^m x^{-1}$ and w_j^n have the same image in G. If $w_i \neq w_j$, then in every word $w \in \mathcal{C}$ we replace $w_j^{n_j}$ by $xw_i^{mp}x^{-1}w_j^q$, where $n_j = np + q$, with $0 \leq q < n$. With these changes $\mathcal{C}$ is transformed into a combing $\mathcal{C}'$ which is still a bounded language. It again follows from Lemma 1.2 that $\mathcal{C}'$ is asynchronously bounded. We deduce that $w_i = w_j$, for otherwise $\mathcal{C}'$ would have fewer subwords w_i of unbounded exponent than $\mathcal{C}$, contradicting the minimality of $\mathcal{C}$.

Suppose that $\mathcal{C}$ has been chosen so as to satisfy the conclusions of Lemma 2.1. For each integer i with $1 \leq i \leq r$ we add a new generator a_i and its formal inverse to A, and define $\overline{a_i} = \overline{w_i}$. We then redefine $\mathcal{C}$ by replacing each of the words w_i in its definition by the corresponding new generator a_i. (Notice that since the index i was encoded in the definition of a_i, we have that $a_i \neq a_j$ even if $w_i = w_j$.) It follows from Lemma 1.2 and the fact that any two word metrics on G are Lipschitz equivalent that $\mathcal{C}$, redefined in this way, still satisfies the asynchronous fellow traveller property. The introduction of the new generators a_i allows us to refine Lemma 2.1:

Lemma 2.2 *Without loss of generality, the combing $\mathcal{C}$ may be assumed to have the following properties:*

1. *$\mathcal{C} = \{w = a_1^{n_1} \ldots a_r^{n_r} \mid (n_1, \ldots n_r) \in S\}$ for some subset S of $\mathbb{N}^r$ and distinct generators $a_i \in A$.*

2. *If $w = a_1^{n_1} \ldots a_r^{n_r} \in \mathcal{C}$, then n_i is less than the order of $\overline{a_i}$ in G; in particular if a_i is of unbounded exponent, then $\overline{a_i}$ has infinite order.*

3. *If a_i and a_j have unbounded exponent and some positive powers of their images are conjugate in G, then $\overline{a_i} = \overline{a_j}$.*

Next we consider the restrictions which the asynchronous fellow traveller property places on bounded languages. We consider the situation $D_a(w,v) \leq K$ described by conditions (1.2), and maintain the notation established there. In (1.2) we allowed the possibility that for some i we have both $x_i = x_{i+1}$ and $y_i = y_{i+1}$. But clearly this possibility can be avoided simply by deleting all such pairs and reindexing. Likewise the simultaneous inequalities $x_i \neq x_{i+1}$ and $y_i \neq y_{i+1}$ may be avoided by interpolating an extra copy of x_i before x_{i+1} and an extra copy of y_{i+1} after y_i. One must also increase the constant K by 1. With these changes we have (with the notations of (1.2)): for all i,

$$\text{Either } x_i = x_{i+1} \text{ and } y_i \neq y_{i+1}, \text{ or vice versa.} \tag{2.2}$$

From now on we assume that the choices of prefixes expressing the condition $D_a(w,v) \leq K$ for words $w, v \in \mathcal{C}$ with $D(w,v) = 1$, satisfy condition (2.2). It is also convenient to introduce the following notation: Given $0 \leq i \leq j \leq N$ define $x_{i,j} \in A^*$ by $x_i x_{i,j} = x_j$, and define $y_{i,j}$ likewise. Thus, given any partition $0 \leq i_1 \leq i_2 \leq \cdots \leq i_s = N$ we have $w = x_{0,i_1} x_{i_1,i_2} \ldots x_{i_{s-1},N}$. Also, from (2.2) we have

$$|x_{i,j}| + |y_{i,j}| = j - i. \tag{2.3}$$

In particular $N = |w| + |v|$.

Lemma 2.3 *Suppose that G satisfies the hypotheses of Theorem A, suppose that $\mathcal{C}$ has been chosen as in Lemma 2.2, and suppose that Theorem A (2) does not hold. Then, there is a constant M such that for all $w, v \in \mathcal{C}$ with $d(\overline{w}, \overline{v}) \leq 1$, the prefixes of w and v described above satisfy $\big| |x_i| - |y_i| \big| \leq M$.*

Proof. We assume that the prefixes x_i for w, and y_i for v, are as in the preceding discussion. Because w can be written as in Lemma 2.2(1), we may write $w = x_{0,i_1} x_{i_1,i_2} \ldots x_{i_{r-1},i_r}$, where each x_{i_{p-1},i_p} is either a positive power of a_p or the empty word. We refine this decomposition a little: Each y_{i_{p-1},i_p} is a product of powers of a_1 to a_r in order, and we decompose it as such, then refine the above decomposition of w so as to make the comparison of prefixes notationally simple. Thus we factor w and v into products of powers of the generators a_λ.

$$w = x_{0,k_1} x_{k_1,k_2} \ldots x_{k_{r2-1},k_{r2}} \qquad v = y_{0,k_1} y_{k_1,k_2} \ldots y_{k_{r2-1},k_{r2}} \tag{2.4}$$

Define $\Delta(i) := \big| |x_i| - |y_i| \big|$. Clearly $\Delta(0) = 0$, and by (2.2) $\Delta(i+1) \leq \Delta(i) + 1$. To complete the proof of the lemma it suffices to show that there is a constant M' such that if $x_{i,j}$ and $y_{i,j}$ are both powers of generators then

$\Delta(j) \leq \Delta(i) + M'$. Indeed, if we exhibit such a constant, then we can set $M = r^2 M'$.

Assume $x_{i,j}$ and $y_{i,j}$ are powers of the generators a and b respectively. Let M_1 be an upper bound on the exponents of all generators a_λ of finite exponent. If a and b are both of bounded exponent, then it follows from (2.3) and Lemma 2.2(2) that $j - i < 2M_1$, whence $\Delta(j) < \Delta(i) + 2M_1$. Thus we may assume that a has unbounded exponent. Notice also that there is no loss of generality in assuming that $j - i$ is greater than a convenient constant. In the next stage of the proof we show that when $j - i$ is large neither $\overline{a}$ nor $\overline{b}$ can be of finite order.

For $i \leq k \leq j$ define $g_k := \overline{y_k}^{-1}\overline{x_k}$. As $d(1, g_j) = d(\overline{y_k}, \overline{x_k}) \leq K$, there are only a finite number of possibilities, say M_2, for g_k. Suppose b is of finite exponent; then $|y_{i,j}| < M_1$. If $j - i$ is large enough ($j - i > M_1 M_2$ suffices), then for some k, k' with $i \leq k \leq k' \leq j$ and $k' - k > M_2$, $y_k = y_{k+1} = \ldots = y_{k'}$. By choice of M_2, $g_{k_1} = g_{k_2}$ for some k_1, k_2 with $k \leq k_1 < k_2 \leq k'$ whence some power of $\overline{a}$ is conjugate to the identity in G. But this is impossible, by Lemma 2.2 (2) , because we are assuming that a is of unbounded exponent.

It remains to consider the case where both $\overline{a}$ and $\overline{b}$ have infinite order. We may assume $j - i > M_2$. Consequently, for some k, k' with $i \leq k \leq k' \leq j$, we have $g_k = g_{k'}$. Consider any such k, k'. The images in G of $x_{k,k'}$ and $y_{k,k'}$ are conjugate. In other words $\overline{a}$ raised to the power $|x_{k,k'}|$ is conjugate to $\overline{b}$ raised to the power $|y_{k,k'}|$. Since $k < k'$, condition (2.2) implies that at least one of these powers is nontrivial, and since neither $\overline{a}$ nor $\overline{b}$ is trivial, this implies that the other power is also nontrivial, hence we may apply Lemma 2.2(3) to deduce that $\overline{a} = \overline{b}$. We are assuming that Theorem A (2) does not hold, so we conclude that $|x_{k,k'}| = |y_{k,k'}|$ whenever $g_k = g_{k'}$. Hence $\Delta(k) = \Delta(k')$. Thus the number of different values of $\Delta(k)$ for $i \leq k \leq j$ is at most M_2. Condition (2.2) ensures that $\Delta(k)$ assumes every value between $\Delta(i)$ and $\Delta(j)$ as k ranges from i to j, consequently $\Delta(j) \leq \Delta(i) + M_2$.

A well known and very beautiful result of Gromov [6] states that a group with polynomial growth is virtually nilpotent. In light of Gromov's theorem, the following lemma completes the reduction of Theorem A to Theorem C.

Lemma 2.4 *Suppose that Theorem A (2) does not hold. Then:*

1. *$\mathcal{C}$ satisfies the synchronous fellow traveller property.*
2. *There exist constants M and Q such that for all $w \in \mathcal{C}$,*

$$|w| \leq Md(1, \overline{w}) + Q.$$

3. *G has polynomial growth.*

Proof. Consider $w, v \in \mathcal{C}$ with $d(\overline{w}, \overline{v}) = 1$ and let the sequences of prefixes $x_0, x_1 \ldots$ and $y_0, y_1, \ldots$ be as in the discussion prior to Lemma 2.3. If w_1 is any prefix of w, then $w_1 = x_i$ for at least one value of i. Thus $d(\overline{w_1}, \overline{y_i}) \leq K$. By Lemma 2.3, the difference between $|w_1| = |x_i|$ and $|y_i|$ is at most M. It follows that if v_1 is the prefix of v with length $\min\{|w_1|, |v|\}$, then $d(\overline{w_1}, \overline{v_1}) \leq d(\overline{w_1}, \overline{y_i}) + d(\overline{y_i}, \overline{v_1}) \leq K + M$. Thus (1) holds.

By Lemma 2.3 we have that $|\,|w| - |v|\,| = |\,|x_N| - |y_N|\,| \leq M$. If we let Q denote the length of the word in $\mathcal{C}$ representing the identity, then a simple induction on $d(1, \overline{w})$ establishes (2). Finally, (2) implies that every $g \in G$ with $d(1, g) \leq m$ is represented in $\mathcal{C}$ by a word of length at most $Mm + Q$. By Lemma 2.2 (1) each such word is determined by an r-tuple $(n_1 \dots n_r)$ of integers each between 0 and $Mm+Q$. Thus there are at most $(Mm+Q+1)^r$ such words.

3. The proof of Theorem C

Let G be a finitely generated virtually nilpotent group and let $\mathcal{C}$ be a combing of G by a bounded language. Suppose that $\mathcal{C}$ satisfies the asynchronous fellow traveller property and the conditions of Lemma 2.3. We claim that Theorem A (2) does not hold in G, so in particular we may apply Lemma 2.4. In order to see that this is the case, we consider a nilpotent subgroup H of finite index in G and suppose $g \in G$ is of infinite order with g^m conjugate to g^n for some m, n with $0 < m < n$. Since the finitely many conjugates of H intersect in a normal subgroup of finite index, there is an integer p such that all p-th powers lie in H. If g_1 conjugates g^m to g^n, then g_1^p conjugates g^{pm^p} to g^{pn^p}. Consequently, it suffices to show that Theorem A (2) cannot hold if G is nilpotent. Clearly G cannot be abelian. More generally, the subgroup generated by g^m must intersect $Z(G)$, the center of G, trivially whence the image of g in $G/Z(G)$ is of infinite order and we are done by induction on the nilpotency class of G.

At its core, our proof of Theorem C depends upon the geometry of conjugation in nilpotent groups. However, this geometry is somewhat obscured by the surrounding technicalities, so to clarify our exposition we concentrate on a case of particular interest, the 3-dimensional integral Heisenberg group:

$$\mathcal{H}_3 = \langle x, y, z \mid [x, y] = z, [x, z] = [y, z] = 1 \rangle.$$

After completing the proof in this case we shall see that the extension to arbitrary virtually nilpotent groups requires only a few observations about the structure of the proof in the 3-dimensional case, together with some elementary facts about nilpotent groups, in particular the structure of centralizers of elements in the penultimate term of the upper central series.

Remark. The import of Theorem C to 3-manifold topology is essentially contained in the case $G = \mathcal{H}_3$, because any group G which acts properly and cocompactly by isometries on the 3-dimensional geometry *Nil* contains $\mathcal{H}_3$ as a subgroup of finite index (see [9]).

The Case $G = \mathcal{H}_3$**.** Let $A \to G$ be a choice of generators for G and let $\mathcal{C} \subset A^*$ be a combing which satisfies Lemma 2.2 (1) - (3) and Lemma 2.4 (1) - (2) . In particular there is a constant K such that for all $w, v \in \mathcal{C}$ with $d(\overline{w}, \overline{v}) = 1$, the synchronous distance between the paths w and v is at most K. In other words, if w_1 is any prefix of w and v_1 is the prefix of v of length $\min\{|w_1|, |v|\}$, then $d(\overline{w_1}, \overline{v_1}) \leq K$.

Because $\mathcal{C}$ is a bounded language, Lemma 1.3 allows us to write it as the union of finitely many sublanguages $\mathcal{C}_i$ of the form:

$$\mathcal{C}_i = \{w = u_{i,0} b_{i,1}^{n_1} u_{i,1} \cdots b_{i,r}^{n_r} u_{i,r} \mid (n_1, \cdots, n_r) \in S_i\}$$

where $r = r(i) \in \mathbb{N}$, $b_{i,j} \in A$, and $u_{i,j} \in A^*$. (Each $b_{i,j}$ is one of the a_i's of Lemma 2.2 (1) .) The set S_i is empty if $r(i) = 0$ and otherwise $S_i \subset \mathbb{N}^r$ has the property that there exist r-tuples in S_i whose smallest entry is arbitrarily large. Given $w \in \mathcal{C}_i$ we define

$$\ell(w) = \min\{n_j\}.$$

In what follows, when referring to $w \in \mathcal{C}_i$ we shall assume that it is decomposed as in the above definition of $\mathcal{C}_i$.

We wish to use Euclidean geometry as a tool to analyze the geometry of the language $\mathcal{C}$. In order to do so, we consider $\widehat{G} = G/Z(G)$, the quotient of $G = \mathcal{H}_3$ by its center. $\widehat{G}$ is isomorphic to $\mathbb{Z}^2$ and so may be identified with the integer lattice of the Euclidean plane $\mathbb{E}^2$. Let $\pi : G \to \widehat{G}$ be the projection. Define $\widehat{g} = \pi(g)$ and $\widehat{w} = \pi(\overline{w})$; $\widehat{g}$ and $\widehat{w}$ are vectors in $\mathbb{E}^2$. In particular $\widehat{A} = \{\widehat{a} \mid a \in A\}$ spans $\mathbb{E}^2$. The integral translates of vectors in $\widehat{A}$ form the edges of a realization of the Cayley graph $\widehat{\Gamma}$ of $\widehat{G}$ corresponding to the choice of generators $\widehat{A} \to \widehat{G}$ induced by the choice of generators $A \to G$.

Let $\widehat{d}$ be the word metric in $\widehat{G}$ corresponding to the choice of generators $\widehat{A}$. Clearly $d(g_1, g_2) \geq \widehat{d}(\widehat{g_1}, \widehat{g_2})$, and it is straightforward to see that the ball of radius n around 1 in G projects onto the ball of radius n around $\widehat{1} = 0$ in $\widehat{G}$. Notice that if $\| \; \|$ denotes the usual Euclidean norm on $\mathbb{E}^2$, then there is a constant $\lambda > 1$ such that

$$\frac{1}{\lambda}\|\widehat{g_1} - \widehat{g_2}\| \leq \widehat{d}(\widehat{g_1}, \widehat{g_2}) \leq \lambda \|\widehat{g_1} - \widehat{g_2}\|.$$

The path in the Cayley graph Γ of G determined by $w \in A^*$ projects to a polygonal path $P(w)$ in $\widehat{\Gamma}$. If $\ell(w)$ is large, then qualitatively (to the distant observer) $P(w)$ looks like a concatentation of at most $r(i)$ long line segments corresponding to the $b_{i,j}$'s with $\widehat{b_{i,j}}$ nontrivial. Of course, upon closer examination one would see that these long segments of $P(w)$ were in fact interspersed with short line segments, translates of the $\widehat{u_{i,j}}$.