GROUPS WHOSE GEODESICS ARE LOCALLY TESTABLE

SUSAN HERMILLER, DEREK F. HOLT, AND SARAH REES

Abstract. A regular set of words is (k-)locally testable if membership of a word in the set is determined by the nature of its subwords of some bounded length k. In this article we study groups for which the set of all geodesic words with respect to some generating set is (k-)locally testable, and we call such groups (k-)locally testable. We show that a group is 1-locally testable if and only if it is free abelian. We show that the class of (k-)locally testable groups is closed under taking finite direct products. We show also that a locally testable group has finitely many conjugacy classes of torsion elements.

Our work involved computer investigations of specific groups, for which purpose we implemented an algorithm in GAP to compute a finite state automaton with language equal to the set of all geodesics of a group (assuming that this language is regular), starting from a shortlex automatic structure. We provide a brief description of that algorithm.

1. Introduction

Let G be a finitely generated group and let X be a finite symmetric (that is, inverse-closed) generating set of G. Let $\mathrm{Geo}(G, X)$ be the set of all words over X that label geodesic paths in the corresponding Cayley graph for G. For many classes of groups, including word hyperbolic groups [3, Theorem 3.4.5], abelian groups, geometrically finite hyperbolic groups [13], Coxeter groups [10], and Garside groups [2], there are generating sets X for which the language $\mathrm{Geo}(G, X)$ is regular.

Starting only from the assumption that $\mathrm{Geo}(G, X)$ is regular, we know very little about G. We do however have the following result, which applies in the more general situation when $\mathrm{Geo}(G, X)$ is recursive.

Proposition 1.1. *With the notation above, if* $\mathrm{Geo}(G, X)$ *is recursive and* G *has a recursively enumerable presentation, then* G *has solvable word problem.*

2000 *Mathematics Subject Classification.* 20F65 (primary), 20F65, 68Q45 (secondary).

Proof. Let w be a word over X. Since $\mathrm{Geo}(G,X)$ is recursive, we can decide whether or not $w \in \mathrm{Geo}(G,X)$. If so, then $w =_G 1$ if and only if w is the empty word. If not, then there exists a word v with $v =_G w$ and $l(v) < l(w)$, and so $v^{-1}w$ is a relator of G. But since G has a recursively enumerable presentation, we can enumerate the relators of G and hence find such a v. We now repeat the process with v in place of w. □

We are motivated by the question of whether there might be interesting subclasses of the class of regular languages for which we could say more about groups with $\mathrm{Geo}(G,X)$ in that subclass. In [7], we considered groups for which $\mathrm{Geo}(G,X)$ is a *star-free* regular language (see also [14, Chapter 4, Definition 2.1] for the definition), and although we obtained some conditions under which groups have this property, we were not able to obtain any classifications. In [6], we considered groups satisfying the much more restrictive hypothesis that $\mathrm{Geo}(G,X)$ is *locally excluding* for some symmetric generating set X of G, which means that there exists a finite set $\mathcal{W}$ of words over X with the property that a word $w \in \mathrm{Geo}(G,X)$ if and only if w does not have a subword equal to a word in $\mathcal{W}$. We proved that this is the case if and only if G is virtually free.

In this paper, we consider groups G for which $\mathrm{Geo}(G,X)$ is a *locally testable* language for some X. The locally testable languages form a class lying between locally excluding and star-free regular languages.

Informally, where k is a positive integer, a set of words is k-locally testable if membership of a word in the set depends on the nature of its subwords of length k. By a *subword* of a word $a_1a_2\cdots a_n$, we mean either the empty word or a contiguous substring $a_ia_{i+1}\cdots a_j$ for some $1 \le i \le j \le n$.

Examples of locally testable sets are given by the sets of all geodesics in both the free group on 2 generators and the free abelian group on 2 generators, using standard generating sets. The set of geodesics in the free group on generators a, b is 2-locally testable, since a word in those generators is geodesic if and only if it contains no subword equal to any of aa^{-1}, $a^{-1}a$, bb^{-1} or $b^{-1}b$. The set of geodesics in the free abelian group on generators a, b is 1-locally testable since a word in those generators is geodesic if and only if it contains neither both a and a^{-1} nor both b and b^{-1}. Note that the first example is locally excluding but the second is not. We can show similarly that the geodesics of free groups and free abelian groups on any number of generators are respectively 2- and 1-locally testable.

A formal definition of a locally testable language is given in Section 2. We shall see that membership of a word w in such a language may also depend on the prefix and suffix of the word of length $k-1$ as well as its subwords of length k. In other words, it depends on the subwords of length k in the word $\diamond w \diamond$, where $\diamond$ is a symbol not lying in X.

From a language theoretic point of view, the locally testable languages have been well-studied. Various characterizations, provided by Brzozowksi and Simon, and by McNaughton, are described in Section 2.

The local testability of $\mathrm{Geo}(G, X)$ will certainly depend on X in general. For example, the free group on four generators has $\mathrm{Geo}(G, X)$ locally excluding and hence locally testable when X is a set of free generators for G, but in [7] we exhibited a generating set of size 6 for which $\mathrm{Geo}(G, X)$ is not star-free and hence not locally testable.

We shall say that the group G is k-locally testable if $\mathrm{Geo}(G, X)$ is k-locally testable for some generating set X of G, and G is locally testable if it is k-locally testable for some k.

The 2-generator Artin groups A_k are defined by the presentations

$$\langle\, a, b \mid (ab)^{k/2} = (ba)^{k/2} \,\rangle \ (k \text{ even}),$$

$$\langle\, a, b \mid (ab)^{(k-1)/2}a = (ba)^{(k-1)/2}b \,\rangle \ (k \text{ odd}).$$

It is proved in [11] that $\mathrm{Geo}(G, X)$ is regular with $X = \{a, b, a^{-1}, b^{-1}\}$. In fact it follows directly from Proposition 4.3 of that paper that $\mathrm{Geo}(G, X)$ is k-locally testable, so this provides a further series of examples of k-locally testable groups.

It can be shown that the kernel of the natural homomorphism of A_n onto the dihedral group of order $2n$ in which the images of a and b have order 2 is a direct product of an infinite cyclic group and a free group of rank $k-1$. We conjecture that any locally testable group is virtually a direct product of free groups.

We are able to prove a complete characterization of 1-locally testable groups. Specifically, in Section 3, we show the following.

Theorem 3.2. *Let G be a finitely generated group. Then the following are equivalent.*

1. *G is free abelian.*
2. *G is 1-locally testable.*
3. *There is a finite symmetric generating set X for G such that the syntactic semigroup associated to* $\mathrm{Geo}(G, X)$ *is idempotent.*

In Section 4, we prove the following.

Theorem 4.1. *For any $k > 0$ the class of k-locally testable groups is closed under taking finite direct products.*

In fact, this is deduced as a corollary of the more general result:

Theorem 4.2. *Let $\mathcal{L}$ be a class of languages that is closed under the operations of union and taking inverse images under length preserving morphisms. Then the class of $\mathcal{L}$-groups is closed under taking finite direct products.*

In Section 5, we show that the local testability restriction on the geodesics results in an algebraic restriction on the group.

Theorem 5.1. *A locally testable group has finitely many conjugacy classes of torsion elements.*

Finally, in Section 6, we describe a method which, upon input of a finite presentation of a shortlex automatic group G with symmetric generating set X, will attempt to construct a finite state automaton with language equal to $\mathrm{Geo}(G, X)$. This process is guaranteed to succeed eventually if $\mathrm{Geo}(G, X)$ is a regular language. This has been implemented using the GAP [5] interface to KBMAG [9].

Using our program to construct a minimal automaton for $\mathrm{Geo}(G, X)$ when this language is regular, we can then use conditions equivalent to those described in Section 2 to decide whether $\mathrm{Geo}(G, X)$ is also (k-)locally testable. These conditions, as well as our results, are stated throughout this paper in terms of the syntactic semigroup of the regular language, but for computational purposes it is generally more efficient to work with the the transition semigroup of the minimal automaton M; these two semigroups are isomorphic (see for example [14, p. 18]). It is shown in [17] that local testability can be tested in polynomial time from the input M. We conclude the paper by listing some examples of the use of our program to test $\mathrm{Geo}(G, X)$ first for being regular and then for being (k-)locally testable.

2. Definition and characterizations of local testability

Let X be any finite set. Let X^* be the free monoid over X, that is the set of all strings over X, and let X^+ be the free semigroup, the set of all non-empty strings.

Let $k > 0$ be a natural number. For $u \in X^*$ of length at least k, let $\mathrm{pre}_k(u)$ be the prefix of u of length k, let $\mathrm{suf}_k(u)$ be the suffix of u of length k, and let $\mathrm{sub}_k(u)$ be the set of all subwords of u of length k. If $l(u) < k$, then we define $\mathrm{pre}_k(u) = u$, $\mathrm{suf}_k(u) = u$, and $\mathrm{sub}_k(u) = \emptyset$. Two words $u, v \in X^*$ satisfy $u \sim_k v$ if and only if $\mathrm{pre}_{k-1}(u) = \mathrm{pre}_{k-1}(v)$, $\mathrm{suf}_{k-1}(u) = \mathrm{suf}_{k-1}(v)$, and $\mathrm{sub}_k(u) = \mathrm{sub}_k(v)$.

A subset $L \subseteq X^*$ is defined to be *k-locally testable* [1, p. 247] if L is a union of equivalence classes of $\sim_k$. The set L is called *locally testable* if L is k-locally testable for some k.

Let LT denote the class of locally testable languages and LT(k) the class of all k-locally testable languages. From the definition as a union of equivalence classes, it follows immediately that LT(k) is closed under the Boolean operations of union, intersection and complementation. It also follows from the definition that any k-locally testable language is also m-locally testable for all $m \geq k$, and hence LT is also closed under the three Boolean operations.

In Proposition 2.1 we list two characterizations of local testability due to Brzozowski and Simon, and McNaughton.

From the first, it follows immediately that LT is strictly contained in the class of regular languages over X. (In fact it shows that locally testable languages are star-free.)

The second characterization involves the syntactic semigroup of L, SS(L), which is defined to be the quotient $X^+/\sim_L$, where $\sim_L$ is the congruence defined as follows. We have $u \sim_L v$ if and only if, for all words s and $t \in X^*$, sut and svt are either both in L or both not in L. Note that the syntactic semigroup should be distinguished from the syntactic monoid, defined to be the quotient of X^* by the same congruence. A semigroup S is said to be *locally idempotent and commutative* if, for every idempotent $e \in S$, eSe is an idempotent commutative subsemigroup of S.

Proposition 2.1. (1) [1, Theorem 2.1(iv)] LT *is the Boolean closure of the languages of the form X^*wX^*, wX^*, and X^*w, where $w \in X^*$.*
(2) [1, Theorem 6.2], [12, Main Theorem, p. 63] *L is locally testable if and only if its syntactic semigroup* SS(L) *is locally idempotent and commutative.*

Using Proposition 2.1, Brzozowski and Simon derived the following characterization of the locally testable languages. They stated it in terms of the associated minimal automata and transition semigroups, but we prefer to phrase it in terms of syntactic semigroups.

Proposition 2.2. [1, Theorem 6.2] *A language L is k-locally testable if and only if its syntactic semigroup S satisfies the following:*
For all $x, y, z \in X^*$ such that $l(x) = k - 1$,
(1) *$xyxzx =_S xzxyx$, and*
(2) *whenever $xy = zx$, then $xy =_S xy^2$.*

We apply this characterization in the case $k = 1$ in Section 3 below in order to characterize 1-locally testable groups. More generally this characterization provides a test for k-local testability that can be used algorithmically.

3. 1-locally testable groups

Given a word w defined over X, we define the *support* of w, $\mathrm{supp}(w)$, to be the subset of X consisting of the symbols in X that appear in the word w. A set L of words over an alphabet X is 1-locally testable if membership of a word in L can be determined simply from examination of which symbols appear in the word, irrespective of where they appear; that is, for some finite set $\mathcal{S}$ of subsets of X, w is in L precisely if $\mathrm{supp}(w) \in \mathcal{S}$.

From Proposition 2.2 we see that a language L is 1-locally testable if and only if the following two conditions hold in the syntactic semigroup $S = \mathrm{SS}(L)$, for all representatives in $y, z \in X^*$: (1) $yz =_S zy$ and (2) $y =_S y^2$. Note that there is no x in these conditions because the x of the conditions of Proposition 2.2 has length 0 in this case. This gives us the following.

Corollary 3.1. *A language L is 1-locally testable if and only if its syntactic semigroup* $\mathrm{SS}(L)$ *is idempotent and commutative.*

Using this result we can characterize 1-locally testable groups.

Theorem 3.2. *Let G be a finitely generated group. Then the following are equivalent.*

(1) *G is free abelian.*
(2) *G is 1-locally testable.*
(3) *There is a finite symmetric generating set X of G such that the syntactic semigroup of* $\mathrm{Geo}(G, X)$ *is idempotent.*

Proof. Suppose that G is finitely generated by $Y = \{y_1, \dots y_n\}$ and let $X = Y \cup Y^{-1}$.

If G is free abelian on Y, then a word w over X is geodesic in G if and only if its support is a subset of one of the 2^n distinct sets

$$\{y_1^{\epsilon_1}, y_2^{\epsilon_2}, \dots, y_n^{\epsilon_n}\}$$

defined by sequences $(\epsilon_1, \epsilon_2, \dots, \epsilon_n) \in \{1, -1\}^n$. Hence G is 1-locally testable.

Corollary 3.1 shows that if G is 1-locally testable, then (3) holds.

Finally suppose that the syntactic semigroup $S := \mathrm{SS}(L)$ is idempotent, where $L := \mathrm{Geo}(G, X)$. For any word w, we have $w =_S w^2$, and so (by the definition of the congruence $\sim_L$) for all words $u, v \in X^*$ we have $uwv \in L \iff uw^2v \in L$. In particular, $w^m \in L \iff w^{m+1} \in L$ for all positive integers m. Hence if w is geodesic, then w^m must also be geodesic for all $m > 0$. This implies in particular that G must be torsion-free.

Now let $x \in X$ be any generator, let $g \in G$ be any element, and let w be a geodesic word representing $g^{-1}xg$. Then, since w^m must be a geodesic representative of $g^{-1}x^m g$ for all $m > 1$, we see that w must have length 1. So x has finitely many conjugates, and hence its centralizer has finite index. Therefore the center $Z(G)$ has finite index in G.

Next we apply [15, Thm 10.1.4] to deduce that the derived subgroup G' must be finite. Since G is also torsion-free this implies that G' is trivial, and so G is abelian. Thus G is free abelian. □

We note that we have definitely made use of the group structure here. In general, an idempotent syntactic semigroup is not necessarily commutative. As an example consider the 2-locally testable language over $\{a, b\}$ consisting of strings that do not start with b. Its syntactic semigroup S is idempotent but $ab \neq_S ba$. However, it is a consequence of the above theorem that if the syntactic semigroup of the language of geodesics of a group is idempotent, then the group is commutative, and hence so is the syntactic semigroup.

4. Closure under finite direct products

In this section we prove the following theorem.

Theorem 4.1. *For any $k > 0$, the class of k-locally testable groups is closed under taking finite direct products.*

In fact the theorem is a special case of the following more general result. For a class of languages $\mathcal{L}$, we understand an $\mathcal{L}$-group to be a group G for which $\mathrm{Geo}(G, X)$ is in $\mathcal{L}$, for some finite symmetric generating set X.

Theorem 4.2. *Let $\mathcal{L}$ be a class of languages that is closed under the operations of union and taking inverse images under length preserving morphisms. Then the class of $\mathcal{L}$-groups is closed under taking finite direct products.*

Proof. Suppose that the geodesic languages L_1 and L_2 of the groups G and H over the symmetric generating sets X and Y are in $\mathcal{L}$. We adjoin generators to each of X and Y that are equal to the identity elements of the respective groups; since any words containing those generators are non-geodesic, this action does not change L_1 or L_2, and hence it does not affect their membership in $\mathcal{L}$.

Define $Z := \{(x, y) : x \in X, y \in Y\} \subseteq G \times H$. Then Z generates $G \times H$ (since X and Y each contain an identity). We define L to be the geodesic language for $G \times H$ over Z.

For any word w over Z, define $\pi_1(w)$ to be the projection of w onto X^* and $\pi_2(w)$ to be its projection onto Y^*. Then, for $w \in Z^*$, we see that if w is to be non-geodesic, both of its projections must be non-geodesic, and conversely. In other words, the following holds.

$$w \in L \iff \pi_1(w) \in L_1 \vee \pi_2(w) \in L_2.$$

Since π_1 and π_2 are length preserving morphisms, the result now follows immediately. □

In order to deduce Theorem 4.1 from Theorem 4.2 we simply need the following lemma.

Lemma 4.3. *The class of k-locally testable languages is closed under the operations of union and taking inverse images under length preserving morphisms.*

Proof. Closure under union is immediate from the definition. The second claim follows from the fact that, if $\phi : A^* \to B^*$ is a length preserving morphism and $v, w \in A^*$ with $v \sim_k w$, then $\phi(v) \sim_k \phi(w)$ (where $\sim_k$ is as in the definition of local testability in Section 2). □

Note that Theorem 4.2 can also be applied to the class of star-free languages. That result is also a special case of [7, Theorem 5.3], which is proved using a different construction, and different generating set.

At this point it is natural to ask whether the class of groups with a locally testable language of all geodesics is closed under the free product operation. Although we do not have a definitive answer to that question, we suspect that the answer is no, and the following example shows that the natural generating set does not work.

Consider the group $G = \mathbb{Z}^2 * \mathbb{Z} = (\langle a \rangle \times \langle b \rangle) * \langle c \rangle$. Giving the $\mathbb{Z}^2$ subgroup the generating set from the proof of Theorem 4.2, consider the generating set

$$\begin{aligned} X \;=\; & \{(1,1), (1,b), (1,b^{-1}), (a,1), (a,b), (a,b^{-1}), \\ & (a^{-1},1), (a^{-1},b), (a^{-1},b^{-1}), c, c^{-1}\} \end{aligned}$$

for G. For every natural number n, the language $\mathrm{Geo}(G, X)$ of geodesics contains the word $u = (1,b)^n(a,b)^n c^n (a,b)^n (a,b^{-1})^n (a,b)^n$ but does not contain the word $v = (1,b)^n(a,b)^n(a,b^{-1})^n(a,b)^n c^n (a,b)^n$. However, $u \sim_n v$. So $\mathrm{Geo}(G, X)$ is not a union of $\sim_n$ equivalence classes for any n, and hence is not locally testable.

5. Locally testable geodesics and torsion conjugacy classes

In this section we prove the following.

Theorem 5.1. *A locally testable group has finitely many conjugacy classes of torsion elements.*

First we have a useful lemma.

Lemma 5.2. *Let X be a set and let $k > 0$ be a natural number. There exists a natural number N such that whenever w is a word over X of length at least N, then $w^2 \sim_k w^j$ for all $j \geq 2$. Moreover, there exists a cyclic permutation $\widetilde{w}$ of w satisfying $\widetilde{w} \sim_k \widetilde{w}^j$ for all $j \geq 1$.*

Proof. Let $N := 2k(|X|^{2k} + 1)$, and suppose that w is a word with length $l(w) \geq N$. Since $l(w) > k$, for $j \geq 2$ we have $\mathrm{pre}_{k-1}(w) = \mathrm{pre}_{k-1}(w^2) = \mathrm{pre}_{k-1}(w^j)$, $\mathrm{suf}_{k-1}(w) = \mathrm{suf}_{k-1}(w^2) = \mathrm{suf}_{k-1}(w^j)$ and $\mathrm{sub}_k(w^2) = \mathrm{sub}_k(w^j)$, as required.

By our choice of N, we can write $w = w_1 w_2 \cdots w_{|X|^{2k}+1} w'$, where each subword w_i has length $2k$. There are only $|X|^{2k}$ distinct words over X of length $2k$, so for some $i \neq j$, we have $w_i = w_j$. Hence we can write $w = xuyuz$ for some word u of length $2k$ and $x, y, z \in X^*$. Also write $u = st$ for words $s, t \in X^*$ of length k each. Then $w = xstystz$.

Next let $\widetilde{w} := tystzxs$, a cyclic permutation of the word w. By analyzing prefixes, suffixes, and subwords of $\widetilde{w}$ and $\widetilde{w}^j$ as above, we find that $\widetilde{w} \sim_k \widetilde{w}^j$ for all $j \geq 1$. ☐

Now we prove Theorem 5.1.

Proof. Suppose that X is a finite symmetric generating set for G and that $L := \mathrm{Geo}(G, X)$ is k-locally testable. Let $N := 2k(|X|^{2k}+1)$ as in the proof of Lemma 5.2. Suppose that w is a word of length $l(w) \geq N$, and that w represents a torsion element of G.

Lemma 5.2 says that there is a cyclic permutation (and hence a conjugate) $\widetilde{w}$ of the word w such that $\widetilde{w} \sim_k \widetilde{w}^i$ for all $i > 0$. Then $\widetilde{w} \in L \iff \widetilde{w}^i \in L$, so $\widetilde{w}$ is geodesic if and only if $\widetilde{w}^i$ is. Since w represents a torsion element, $\widetilde{w}$ must also be torsion in G, and so for some i, and hence for all i, the word $\widetilde{w}^i$ is not geodesic. Therefore $\widetilde{w} =_G v$ for some word v satisfying $l(v) < l(w)$ and representing an element of G in the conjugacy class of w.

If $l(v) \geq N$, then repeat this argument to obtain words representing conjugates of w of successively strictly shorter length. Eventually this process must end with a word u of length $l(u) < N$ such that u and w are in the same conjugacy class. Therefore there are only finitely many conjugacy classes of torsion elements. ☐

6. Constructing a word acceptor for regular geodesics

In this section, we describe a method for constructing a finite state automaton that accepts the set of all geodesic words in a finitely presented shortlex automatic group, in the case that this set is a regular language and hence such an automaton exists. For Coxeter groups, a method for doing this is described in [10], which is considerably faster than the general approach presented here.

We suppose that our group G is generated by the symmetric set X, and as usual let $\mathrm{Geo}(G,X) \subset X^*$ be the set of all geodesic words. We also suppose that we have successfully computed the shortlex automatic structure of G with respect to some ordering of X. We shall assume throughout this section that X^* is ordered by the associated shortlex ordering.

The procedure to be described here, which attempts to construct a finite state automaton GW accepting a language $L(\mathrm{GW})$ equal to $\mathrm{Geo}(G,X)$, will succeed eventually if $\mathrm{Geo}(G,X)$ is a regular language. If G is word-hyperbolic, then the method described in Section 3 of [4] will construct GW. Otherwise, we (repeatedly, if necessary) construct candidates for GW, and try to prove their correctness. We shall first discuss our method for proving correctness of GW, which will succeed if and only and GW really is correct (i.e. $L(\mathrm{GW}) = \mathrm{Geo}(G,X)$), and then discuss how to come up with suitable candidates.

The automatic structure computed includes the shortlex word acceptor W, the multiplier automata M_x for $x \in X \cup \{\varepsilon\}$ (which we shall not need), and also a *word-difference* automaton D with the following properties:

- (D1) $(u,v)^+ \in L(D) \Rightarrow u =_G v$;
- (D2) If $u, v \in X^*$ with $u, v \in L(W)$, $x \in X \cup \{\varepsilon\}$ and $ux =_G v$, then $(ux, v)^+ \in L(D)$.

Here $(u,v)^+$ denotes the padded pair corresponding to $u, v \in X^*$. In fact, the construction of D is such that its start state σ is its only accepting state, and D contains transitions labeled (a,a) from σ to σ for all $a \in X$. It therefore has the additional property:

- (D3) If $(u,v)^+ \in L(D)$ and $w, w' \in X^*$ then $(wu, wv)^+ \in L(D)$ and, if $l(u) = l(v)$, then $(wuw', wvw') \in L(D)$.

For a candidate GW for a finite state automaton with language $\mathrm{Geo}(G,X)$, we use standard operations on finite state automata, as described in Chapter 1 of [3] or Section 13.1 of [8], to check whether GW satisfies the hypotheses of the following theorem. If so, then the theorem tells us that GW is correct (that is, $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$). It is easy to see that these conditions are also necessary for the correctness of GW, so our procedure will succeed if and only if GW is correct.

Theorem 6.1. *Let* GW *be a finite state automaton over* X *which satisfies the following conditions:*

(i) $L(\mathrm{GW})$ *is prefix-closed;*
(ii) $L(W) \subseteq L(\mathrm{GW})$*;*
(iii) *If* $(u, v)^+ \in L(D)$ *with* $l(u) = l(v)$*, then* $u \in L(\mathrm{GW}) \iff v \in L(\mathrm{GW})$*;*
(iv) *If* $(u, v)^+ \in L(D)$ *with* $l(u) > l(v)$*, then* $u \notin L(\mathrm{GW})$*.*

Then $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$*.*

Proof. Suppose that $\mathrm{Geo}(G, X) \not\subseteq L(\mathrm{GW})$ and let w be the shortlex least element of $\mathrm{Geo}(G, X) \setminus L(\mathrm{GW})$. Then, by (ii), $w \notin L(W)$. Let $w = w'uw''$, where u is of minimal length with $u \notin L(W)$. Then, by (D2), we have $(u, v)^+ \in L(D)$ for some $v \in L(W)$. Since by (D1) $u =_G v$, and $w \in \mathrm{Geo}(G, X)$, we must have $l(u) = l(v)$ and hence, by (D3), $(w'uw'', w'vw'') \in L(D)$. Moreover $w'vw'' \in \mathrm{Geo}(G, X)$. Now, since $v \in L(W)$, we have $u > v$ (recall that we are using the shortlex ordering on X^*) and hence $w = w'uw'' > w'vw''$, and then $w'vw'' \in L(\mathrm{GW})$ by choice of w. But this contradicts (iii).

Suppose, on the other hand, that $L(\mathrm{GW}) \not\subseteq \mathrm{Geo}(G, X)$ and let w be the shortlex least element of $L(\mathrm{GW}) \setminus \mathrm{Geo}(G, X)$. Then $w \notin \mathrm{Geo}(G, X)$ implies $w \notin L(W)$, and again we can write $w = w'uw''$, where u is of minimal length with $u \notin L(W)$. Again we have $(u, v)^+ \in L(D)$ for some $v \in L(W)$. If $l(u) > l(v)$ then, by (i) and the minimality of w, we must have $w = w'u$, and hence, by (D3), $(w'u, w'v)^+ \in L(D)$. But this contradicts (iv). On the other hand, if $l(u) = l(v)$ then, by (D3), $(w'uw'', w'vw'') \in L(D)$, which contradicts (iii) again. $\square$

This gives the method of testing, for a candidate automaton GW, whether $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$. Now we turn to the problem of constructing potential candidates for GW.

In the methods to be described below, we can prove that if $\mathrm{Geo}(G, X)$ is regular and if we run the relevant programs for sufficiently long, then GW will be correctly calculated. Although we have no means of estimating for how long we actually need to run the programs, we can test a sequence of candidates for correctness, and thereby produce a terminating algorithm to compute GW with $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$, provided of course that $\mathrm{Geo}(G, X)$ is regular.

As mentioned previously, if G is word-hyperbolic, then the procedure described in Section 3 of [4] will construct the correct automaton GW. This method can be summarized as follows. Starting from the automaton $\mathrm{GW}_0 := W$, the shortlex word acceptor in the automatic structure for G, we construct the automaton GW_1 which accepts all words v for which there exists a $w \in L(\mathrm{GW}_0)$ with $l(w) = l(v)$ and with (v, w) in the language of the word-difference machine D. So $L(\mathrm{GW}_0) \subseteq L(\mathrm{GW}_1) \subseteq \mathrm{Geo}(G, X)$. We can construct GW_i from GW_{i-1} in the same way, for all $i > 0$. It can be shown that, if $L(\mathrm{GW}_i) = L(\mathrm{GW}_{i-1})$ for some i, then $L(\mathrm{GW}_i) = \mathrm{Geo}(G, X)$, and that this is the case if and only G is word-hyperbolic.

If applied to a group that is not word-hyperbolic, then this procedure will not terminate, and will construct a sequence GW_i ($i \in \mathbb{N}$) of automata with $L(\mathrm{GW}_i) \subset \mathrm{Geo}(G, X)$, $L(\mathrm{GW}_i) \subset L(\mathrm{GW}_{i+1})$, and $\mathrm{Geo}(G, X) = \cup_{i=1}^{\infty} L(\mathrm{GW}_i)$. So, for any $k > 0$, there exists an n such that, for all $i \geq n$, $L(\mathrm{GW}_i)$ has the following property:

(P_k) the set of words of length at most k in $L(\mathrm{GW}_i)$ is equal to the set of words of length at most k in $\mathrm{Geo}(G, X)$.

We apply a method described by Trakhtenbrot and Barzdin in [16, Section IV.2] to construct candidates GW for an automaton that satisfies $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$. This procedure, as presented in [16], assumes the existence of an automaton M, and takes as input a set of pairs (w, b), where w is a word and b is true or false, with $\mathsf{b} = \mathsf{true}$ if and only if $w \in L(M)$. It attempts to construct M from this information. It is proved in [16, Theorem 2.16] that, if M is known to have at most n states and the procedure is given the membership of all words of length at most $2n - 1$ in $L(M)$ as input, then it will successfully construct M.

We shall apply a variant of this procedure using an automaton M' and an integer $k > 0$ as input, and use the membership of words of length at most k in $L(M')$ as our criterion to determine whether they lie in $L(M)$. It will either abort or output a new automaton $c(M', k)$; specifically, we apply it to GW_i, for some i and k.

It follows from [16, Theorem 2.16] that, for any finite state automaton M, there exists an integer $B(M) > 0$, such that if $k \geq B(M)$ and the sets of words of length at most k in $L(M)$ and $L(M')$ are equal, then the procedure applied to M' will output M as $c(M', k)$. If M has n states, then this is true with $B(M) = 2n - 1$, but it is proved in [16, Theorem 5.10] that there is a constant C such that, for almost all automata M, $B(M) \leq C \log_{|A|}(n)$, where A is the alphabet of M.

Unfortunately, we know no upper bound on the number of states of M, so we have no way of estimating $B(M)$ in advance. However, since we know that, for any k, P_k is satisfied by GW_i for all sufficiently large i, the following is true. If $\mathrm{Geo}(G, X)$ is a regular language and if we calculate GW_i for sufficiently large i and then apply the procedure for sufficiently large k, then an automaton GW with $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$ will be output as $c(\mathrm{GW}_i, k)$.

Here is a summary of our version of the Trakhtenbrot and Barzdin procedure. For a state σ of a deterministic finite state automaton M with start state σ_0, we define the *depth* of σ to be the length of a shortest word w with $\sigma_0^w = \sigma$ (where σ_0^w is the state reached from σ_0 on reading the string w). For an integer $k > 0$, define $M(\sigma, k)$ to be the set of words w with $l(w) \leq k$ such that σ^w is an accept state of M. For two states σ, τ of M, let d be maximum of the depths of σ and τ and, for $k > d$, define σ and τ to be *k-equivalent* if $M(\sigma, k-d) = M(\tau, k-d)$. Note that, since the value of d depends on the pair of states σ, τ, k-equivalence is not necessarily an equivalence relation on the set of all states of M, although it will be for sufficiently large k.

With input M' and k, the procedure attempts to define an automaton $c(M', k)$ in which the set of states is the set of k-equivalence classes of states of M', and the transitions are induced from those of M'. If k-equivalence turns out not to be a equivalence relation, or if transitions with a given label from k-equivalent states do not lead to k-equivalent states, then the procedure aborts. Otherwise it outputs $c(M', k)$.

Unfortunately, we do not know $B(\mathrm{GW})$ in advance, and neither do we know how large i must be to guarantee that GW_i satisfies P_k. We have not yet implemented any heuristics for estimating these values; we have simply guessed at them.

We finish with some examples in which we have successfully computed GW with $L(\mathrm{GW}) = \mathrm{Geo}(G, X)$, together with the values of k and i used, and the number of states of GW. The first example is a 5-generator presentation of the wreath product of an infinite cyclic group with the group of order 2. There are presentations on fewer generators with regular sets of geodesics, but with this presentation $\mathrm{Geo}(G, X)$ also turns out to be 2-locally testable. The second, third and fourth examples are 2-generator Artin groups and they each have locally testable sets of geodesics. Indeed, we checked, using the conditions in Proposition 2.2, that they are k-locally testable for $k = 3, 4$ and 5, respectively, a fact which we know to be true from [11, Proposition 4.3].

The final example is a 4-generator Coxeter group for which $\mathrm{Geo}(G, X)$ is regular but not locally testable; in fact it is not even star-free. As mentioned at the beginning of this section, it is more efficient to use the method described in [10] to compute GW for this example.

1. $G = \langle\, a, b, t, u, v \mid ab = ba,\ t^2 = 1,\ tat = b,\ at = u,\ bt = v\,\rangle$; $i = 3$, $k = 4$, 10 states.

2. $G = \langle\, a, b \mid aba = bab\,\rangle$; $i = 3$, $k = 7$, 28 states.

3. $G = \langle\, a, b \mid abab = baba\,\rangle$; $i = 4$, $k = 10$, 61 states.

4. $G = \langle\, a, b \mid ababa = babab\,\rangle$; $i = 5$, $k = 12$, 115 states.

5. $G = \langle\, a, b, c, d \mid a^2 = b^2 = c^2 = d^2 = (ab)^3 = (bc)^3 = (cd)^3 = (da)^3 = (ac)^2 = (bd)^2 = 1\,\rangle$; $i = 6$, $k = 14$, 125 states.

References

[1] J. A. Brzozowski and I. Simon, Characterizations of locally testable events, Disc. Math 4 (1973) 243–271.

[2] R. Charney and J. Meier, The language of geodesics for Garside groups, Math. Zeitschrift 248 (2004) 495–509.

[3] D.B.A. Epstein, J. Cannon, D. Holt, S. Levy, M. Paterson, and W. Thurston, *Word Processing in Groups*, Jones and Bartlett, Boston, 1992.

[4] D.B.A. Epstein and D.F. Holt, Computation in word-hyperbolic groups, Internat. J. Algebra Comput. 11 (2001), 467–487.

[5] The GAP Group. *GAP — Groups, Algorithms, and Programming, Version 4.4*, 2004. (`http://www.gap-system.org`).

[6] R. Gilman, S. Hermiller, D. F. Holt and S. Rees, A characterisation of virtually free groups, Arch. Math. 89 (2007), 289–295.

[7] S. Hermiller, D. F. Holt and S. Rees, Star-free geodesic languages for groups, Internat. J. Algebra Comput. 17 (2007) 329–345.

[8] D.F. Holt, B. Eick and E.A. O'Brien, *Handbook of Computational Group Theory*, Chapman & Hall/CRC, 2005.

[9] D.F. Holt, KBMAG—*Knuth-Bendix in Monoids and Automatic Groups*, software package (1995), available from (`http://www.maths.warwick.ac.uk/~dfh/download/kbmag2/`).

[10] R.B. Howlett, Miscellaneous facts about Coxeter groups, lecture notes (1993), available from (`http://www.maths.usyd.edu.au/res/Algebra/How/anucox.html`).

[11] J. Mairesse and F. Mathéus, Growth series for Artin groups of dihedral type, Internat. J. Algebra Comput. 16 (2006), 1087–1107.

[12] R. McNaughton, Algebraic decision procedures for local testability, Math. Systems Theory 8 (1974), 60-76.

[13] W. Neumann and M. Shapiro, Automatic structures, rational growth, and geometrically finite hyperbolic groups, Invent. Math. 120 (1995), 259–287.

[14] J.-E. Pin, *Varieties of Formal Languages*, Plenum Publishing Corp., New York, 1986.

[15] Derek J. S. Robinson, *A course in the theory of groups*, Second edition, Graduate Texts in Mathematics **80**, Springer-Verlag, New York, 1996.

[16] B.A. Trakhtenbrot and Ya. M. Barzdin, *Finite automata. Behavior and synthesis.* Translated from the Russian by D. Louvish. English translation edited by

E. Shamir and L. H. Landweber. Fundamental Studies in Computer Science, Vol. 1. North-Holland Publishing Co., Amsterdam, 1973.

[17] A.N. Trahtman, A polynomial time algorithm for local testability and its level, Internat. J. Algebra Comput., 9 (1999), 31–39.

Department of Mathematics, University of Nebraska, Lincoln NE 68588-0130, USA
E-mail address: smh@math.unl.edu

Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK
E-mail address: dfh@maths.warwick.ac.uk

School of Mathematics and Statistics, University of Newcastle, Newcastle NE1 7RU, UK
E-mail address: Sarah.Rees@ncl.ac.uk