Groups with Poly-Context-Free Word Problem

by

Tara Rose Brough

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics

November 2010

Contents

Acknowledgments

Firstly I would like to thank my supervisor, Professor Derek Holt. Derek has been a pleasure to work with, and I am very grateful to him for finding a research project that really interests me, for so many helpful discussions and ideas, and for his support throughout my PhD. My thanks are due to Eamonn O'Brien and Bettina Eick for strongly recommending Derek as a supervisor, as well as for my first experiences of mathematical research.

Thanks to everyone at the Warwick Mathematics Institute for providing an excellent mathematical and social environment. I would especially like to thank all my friends, my mentor Inna Capdebosq, Dave Wood, and all the office and support staff.

Thanks very much to the University of Warwick for the Vice-Chancellor's Studentship which allowed me to study here, and to Carol Wright (the previous postgraduate coordinator) for her help and advice during the application process and when I first arrived.

Special thanks are due to: Rick Thomas, for many helpful discussions, and for interesting me in this area in the first place, when he gave some talks at Auckland University in early 2007; Peter Neumann, for suggesting the approach that led to Theorem 5.7, and for lending me his beautiful spare violin; and Benno Kuckuck, for reading an early version of part of Chapter 4 and helping shorten some of the proofs. I would also like to thank: David Craven, for asking at the Postgraduate Group Theory Conference in 2008 what one could say in general about poly-context-free soluble groups; Sarah Rees and Mike Paterson, for discussions on how to proceed from here; Nik Ruskuc, for pointing out that our original conjecture was clearly wrong; Niamh O'Sullivan, for noticing a mistake in the proof of Theorem 5.7; Michael Brough, Helen Broome and David Howden for proofreading parts of the thesis; Tom Sharland for last-minute help with the thesis template; and Richard Weidmann, for offering me a position for a semester in Kiel, which provided a great incentive to get this thesis finished.

My time at Warwick has been greatly enriched by the friendship and prayers of many people at Westwood Church. I would particularly like to thank James Harrison and Sabrina David, who have both helped out in so many ways. Finally, big thanks to my family, especially my beloved husband Michael and my dear mother Lucy, for all their love and support.

Declarations

To the best of my knowledge and except where otherwise stated, this thesis is my own work. I confirm that this thesis has not been submitted for a degree at any other university. The results in Chapters 4 and 5 (except for Proposition 5.12) have been submitted for publication as a joint paper with Derek Holt.

Abstract

We call a language *poly-context-free* if it is an intersection of finitely many context-free languages. In this thesis, we consider the class of groups with poly-context-free word problem. This is a generalisation of the groups with context-free word problem, which have been shown by Muller and Schupp [17, 3] to be precisely the finitely generated virtually free groups.

We show that any group which is virtually a finitely generated subgroup of a direct product of finitely many free groups has poly-context-free word problem, and conjecture that the converse also holds. We prove our conjecture for several classes of soluble groups, including the metabelian groups and torsion-free soluble groups, and present progress towards resolving the conjecture for soluble groups in general.

Some results in the thesis may be of independent interest in formal language theory or group theory. In Chapter 2 we develop some tools for proving a language not to be poly-context-free, and in Chapter 5 we prove that every finitely generated soluble group which is not virtually abelian has a subgroup of one of a small number of types.

Chapter 1

Introduction and background

1.1 Introduction

The *word problem* of a group G with respect to a finite generating set X is the set of all words in elements of X and their inverses which represent the identity element of G. A *formal language* is just a set of words over some finite alphabet, and so the word problem of a group can be considered as a formal language.

The study of word problems of groups as formal languages has developed slowly since the beginnings of formal language theory in the 1950s. In 1971, Anisimov [1] published a proof that a group has regular word problem if and only if it is finite. The first really significant development in the area was the classification of the groups with context-free word problem by Muller and Schupp in the 1980s [17, 18, 3]: a finitely generated group has context-free word problem if and only if it is virtually free. Since then, research activity in this area has increased, and groups with word problem in various other language classes, generally somewhat related to the context-free languages, have been studied, for example in [8, 9, 10, 14]. The general aim is to determine what implications the formal language Type of a group's word problem has for the structure of the group and vice versa.

In this thesis, we study the class of poly-context-free groups. A group is said to be *poly-context-free* if its word problem is a *poly-context-free language*, that is, an intersection of finitely many context-free languages. The property of being poly-context-free is independent of the choice of finite generating set, and the class of poly-context-free groups is closed under taking finitely generated subgroups, finite index overgroups, and finite direct products.

While we have not been able to even approach a classification of these groups in general, we have come quite close to a classification in the case of soluble groups. In the cases of metabelian and torsion-free soluble groups, we have a complete characterisation of those which are poly-context-free. In both cases, they are precisely the virtually abelian groups, and we conjecture that the virtually abelian groups are the only poly-context-free soluble groups.

The work in this thesis was motivated by the paper of Holt, Rees, Röver and Thomas on groups with context-free co-word problem [9], known as co-context-free groups. (The *co-word problem* of a group is the complement of its word problem.) An unresolved question from that paper was whether the co-context-free groups are closed under taking free products. Derek Holt suggested that this question might be easier to resolve (presumably in the negative) for poly-context-free groups. These are somewhat related to the co-context-free groups, as we shall explain in Section 3.1.

We were unable to determine whether the poly-context-free groups are closed under taking free products, but we obtained several results for poly-context-free groups analogous to some of the results in [9]. Several of these results showed that in various classes of soluble groups, the only poly-context-free groups are the virtually abelian groups. This led to the attempt at a full characterisation of soluble poly-context-free groups.

1.2 Outline of thesis

The remainder of this chapter will be devoted to background material, divided into sections on notation, group theory and algebra, formal language theory, and word problems of groups as formal languages. Throughout the thesis, a knowledge of group theory at the level of an advanced undergraduate course will be assumed, but we include background material on concepts of central importance in the thesis, such as soluble groups.

In Chapter 2, we study the class of poly-context-free languages, with a particular focus on methods for proving languages to be *not* poly-context-free. To this end, we develop several tools based on the correspondence between context-free languages and stratified semilinear sets introduced in Section 1.5.3. In Corollary 2.7, we show that a language satisfying certain properties cannot be poly-context-free, while Theorem 2.19 exhibits sequences of languages $L^{(n,k)}$, where $n, k \in \mathbb{N}$, such that for all n, the language $L^{(n,k)}$ is an intersection of k but not $k-1$ context-free languages.

In Chapter 3, we present the known examples of poly-context-free groups, and conjecture that these are the only ones. We give some evidence for this conjecture (Conjecture 3.2), in the form of results showing that it holds in the classes of nilpotent, Baumslag-Solitar and polycyclic groups. We also introduce a class of 2-generator abelian-by-cyclic groups, each of which is embedded in a semidirect product $\mathbb{Q}^s \rtimes \mathbb{Z}$ for some $s > 0$. We call these *Gc-groups*, and show that they obey our conjecture. The Gc-groups will play an important role in the final two chapters. We say that a Gc-group is *proper* if it is not virtually abelian.

In Chapters 4 and 5, we restrict our attention to soluble groups. Our conjecture restricted to soluble groups says that a soluble group is poly-context-free if and only if it is virtually abelian. Due to the closure of the poly-context-free groups under taking finitely generated subgroups, we can hope to prove our conjecture for soluble groups by showing that every finitely generated soluble subgroup which is not virtually abelian has a finitely generated subgroup which we can prove to be not poly-context-free.

Chapter 4 concentrates on metabelian groups (groups of derived length at most 2). We prove in Theorem 4.14 that every finitely generated metabelian group that is not virtually abelian has a subgroup which is isomorphic to either $\mathbb{Z} \wr \mathbb{Z}$, $C_p \wr \mathbb{Z}$ or a proper Gc-group. Using results from Chapter 3, this allows us to prove our conjecture in the case of metabelian groups.

Finally, in Chapter 5, we prove our conjecture for torsion-free soluble groups, and present progress towards resolving the conjecture for soluble groups in general. The main results are Theorems 5.7 and 5.10. Theorem 5.7 shows that every finitely generated torsion-free soluble group that is not virtually abelian has a subgroup isomorphic to either $\mathbb{Z}^\infty$ or a proper Gc-group. Theorem 5.10 is a little less satisfactory. We show that every finitely generated soluble group that is not virtually abelian has a subgroup H that is either a proper Gc-group or is isomorphic to $\mathbb{Z}^\infty$, or is finitely generated with an infinite normal torsion subgroup U such that H/U is either free abelian or a proper Gc-group. We do not yet know how to prove that the subgroups in the final case are not poly-context-free.

All the material in Chapters 2 to 5 is original unless otherwise indicated.

1.3 Notation

Throughout this thesis, $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ denote the natural numbers, integers and rationals respectively. We denote the natural numbers with zero included by $\mathbb{N}_0$.

For $r \in \mathbb{N}$ and $1 \leq i \leq r$, the vector in $\mathbb{N}_0^r$ with a 1 in the i-th position and zeroes elsewhere will be denoted by e_i. With the exception of these, all vectors will be represented by bold letters. We denote the i-th component of the vector $\mathbf{v}$ by $\mathbf{v}(i)$.

For a set X, we denote the *Kleene star closure* of X, which is the set of all finite length strings (also called words) of elements of X, by X^*. In the special case $X = \{x\}$, we often denote X^* by x^*. For $w = x_1 x_2 \ldots x_n \in X^*$, where $x_i \in X$, the *length* of w, denoted $|w|$, is n. We shall use ϵ for the empty word, which has length 0.

If X is a subset of a group, $X^{-1} := \{x^{-1} \mid x \in X\}$. If $X \subseteq G$, where G is a group, and two words $u, w \in X^*$ are equal as elements of G, then we write $u =_G w$.

For two group elements x and y, we denote the *commutator* $x^{-1}y^{-1}xy$ of x and y by $[x, y]$. The *conjugate* of x by y is $y^{-1}xy$, often denoted x^y.

If G is a group, then the *derived subgroup* of G, denoted G', is the subgroup of G generated by all the commutators of elements of G. So

$$G' = \langle [x, y] \mid x, y \in G \rangle .$$

This is also often called the *commutator subgroup* of G, in which case it is generally denoted by $[G, G]$.

The *centre* of a group G is

$$Z(G) = \{y \in G \mid x^y = x \ \forall x \in G\},$$

which forms a normal subgroup of G. If $H \leq G$, then the *centraliser* of H in G is

$$C_G(H) = \{y \in G \mid x^y = x \ \forall x \in H\}.$$

This is also a subgroup of G, and is normal in G if H is normal in G.

1.4 Group theory and algebra

1.4.1 Finitely generated abelian groups

The isomorphism type of a finitely generated abelian group is easily computed using a matrix formed from a presentation of the group. We shall give more specific details of this after first defining a certain type of normal form for matrices.

Smith normal form

Two $m \times n$ matrices A and B over a ring R are called *Gaussian equivalent* if B can be obtained from A by a sequence of elementary row and column operations.

Proposition 1.1. [26, Theorem 9.58] *Every non-zero $m \times n$ matrix M with entries in a euclidean ring R is Gaussian equivalent to a matrix of the form* $\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$, *where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$ and $\sigma_1 | \sigma_2 | \ldots | \sigma_k$ are non-zero. (The lower blocks of 0's or the right blocks of 0's might not be present.)*

This matrix is called the *Smith normal form* of M and is unique up to multiplication by units. If $R = \mathbb{Z}$, then we can assume the σ_i are all positive.

Proposition 1.2. [26, Theorem 9.64] *Let $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$ be the non-singular diagonal block in the Smith normal form of a matrix M with entries in a euclidean ring R. Let $\gamma_0(M) = 1$, and for $i > 0$ let $\gamma_i(M)$ be the* gcd *of all $i \times i$ minors of M. Then $\sigma_i = \gamma_i(M)/\gamma_{i-1}(M)$ for all $1 \leq i \leq k$.*

Presentation matrices

Any finitely generated abelian group is finitely presentable. Let G be a group with presentation

$$\langle x_1, \ldots, x_n \mid [x_i, x_j] \ (1 \leq i, j \leq n), r_k \ (1 \leq k \leq s) \rangle ,$$

where $s \in \mathbb{N}_0$ and $r_k = x_1^{m_{k1}} \ldots x_n^{m_{kn}}$, with $m_{kl} \in \mathbb{Z}$ for all $1 \leq l \leq n$. We can encode the information given in this presentation into an $s \times n$ matrix M with the i, j-th entry being m_{ij}, called the *presentation matrix*. If the diagonal block in the Smith normal form of M is $\mathrm{diag}(\sigma_1, \ldots, \sigma_t)$, then

$$G \cong \mathbb{Z}_{\sigma_1} \times \ldots \times \mathbb{Z}_{\sigma_t} \times \mathbb{Z}^{n-t}.$$

1.4.2 Virtually χ groups, extensions and χ_1-by-χ_2 groups

If χ is a property of groups (for example finite, abelian or free), then a group G is *virtually* χ if G has a finite index subgroup H having the property χ. This will be an important concept throughout the thesis, as we shall have much to do with virtually free and virtually abelian groups.

A group G is an *extension* of a group N by a group H if G has a normal subgroup $M \cong N$ such that $G/M \cong H$.

If χ_1 and χ_2 are properties of groups, then a group G is called a χ_1*-by-*χ_2 *group* if G is an extension of a group with property χ_1 by a group with property χ_2.

By definition, every χ-by-finite group is virtually χ. If χ is a property inherited by finite index subgroups, then the converse is true, as the core of H in G (the largest normal subgroup of G contained in H) has finite index in G if H has finite index in G (see for example [25, Corollary 4.15]).

The following is probably well-known, but we provide a proof, as we do not have a reference for it.

Lemma 1.3. *Every finitely generated finite-by-abelian group is virtually abelian.*

Proof. Let $G = \langle x_1, \ldots, x_n \rangle$, with finite normal subgroup T such that G/T is abelian. Since T is finite, $C_G(T)$ has finite index in G; so, by replacing G by $C_G(T)$ if necessary, we may assume $T \leq Z(G)$. Let m be the exponent of T and let $A = \langle x_1^m, \ldots, x_n^m \rangle$. Then A has finite index in G and is abelian, since

$$[x_i^m, x_j^m] = [x_i, x_j]^{m^2} \in T^{m^2} = \{1\}$$

for any $i, j \in \{1, \ldots, n\}$. □

1.4.3 Normal, derived and central series

A *(subnormal) series* in a group G is a finite sequence $G_0, G_1, \ldots, G_n$ of subgroups of G such that

$$G = G_0 \rhd G_1 \rhd \ldots \rhd G_n = \{1\}.$$

The *length* of the series is defined to be n, even though the series has $n+1$ terms. The quotient groups G_i/G_{i+1} are called the *factors* of the series. If all the G_i are normal in G, then the series is a *normal series.*

The *derived series* of a group G is the descending normal series

$$G = G^{(0)} \rhd G^{(1)} \rhd G^{(2)} \rhd G^{(i)} \rhd \ldots,$$

where $G^{(i)}$ is defined to be the derived subgroup of $G^{(i-1)}$. Since the derived series does not necessarily terminate in the trivial group, it is not always technically a series. If the derived series does have finite length, then the smallest $n \in \mathbb{N}_0$ such that $G^{(n)} = \{1\}$ is called the *derived length* of G.

A series $G = G_0 \rhd G_1 \rhd \ldots \rhd G_n = \{1\}$ is *central* if each factor G_{i-1}/G_i is contained in the centre of the previous factor G_i/G_{i+1}.

The *upper central series* of a group G is the ascending series

$$\{1\} = \zeta_0 G \lhd \zeta_1 G \lhd \zeta_2 G \lhd \ldots,$$

where $\zeta_{i+1}G/\zeta_i G = Z(G/\zeta_i G)$. Again, this does not necessarily terminate.

There is a related series called the lower central series, but we shall not require it.

1.4.4 Soluble groups

We introduce the class of groups with which we shall be primarily concerned in this thesis. Details of all statements in this section can be found in Sections 5.1 and 5.4 of [22] unless other references are given.

A group G is *soluble* if it has an *abelian series*, that is, a series

$$G = G_0 \rhd G_1 \rhd \ldots \rhd G_n = \{1\}$$

in which all the factors G_i/G_{i+1} are abelian.

If the derived series of G terminates in the trivial group, then it is clearly an abelian series and so G is soluble. The groups of derived length 1 are just the abelian groups. Groups of derived length at most 2 are called *metabelian* groups.

Conversely, if G is soluble, then the derived series of G has finite length, and the derived length of G is the minimal length of an abelian series in G. Thus the soluble groups can equivalently defined to be those groups for which $G^{(n)}$ is trivial for some $n \in \mathbb{N}$.

In the final two chapters, we shall repeatedly use the elementary fact that the class of soluble groups is closed under taking subgroups and homomorphic images. An extension of a group of derived length m by a group of derived length n has derived length at most $m+n$; so the class is also closed under taking extensions.

We now introduce two special classes of soluble groups.

Nilpotent groups

A group G is *nilpotent* if it has a central series, which turns out to be equivalent to the upper central series of G having finite length. The length of the upper central series is called the *nilpotent class* of G.

Any central series is an abelian series, and so nilpotent groups are soluble. The class of nilpotent groups is closed under taking subgroups and homomorphic images, and also under taking finite direct products.

Lemma 1.4. [22, 5.2.18] *A finitely generated nilpotent group has a central series whose factors are cyclic groups with prime or infinite orders.*

Corollary 1.5. *A finitely generated nilpotent group has a torsion-free subgroup of finite index.*

Sketch of proof. Using Lemma 1.3, we can pass the factors of prime order in the central series from Lemma 1.4 to the top of the series. Since the other factors have infinite order, and there are only finitely many factors, this proves the corollary. □

Polycyclic groups

A group is *polycyclic* if it has a series in which every factor is cyclic. Since cyclic groups are abelian, polycyclic groups are soluble. The class of polycyclic groups is closed under taking subgroups, homomorphic images and extensions.

Observation 1.6. *A group is polycyclic if and only if it has a normal series each factor of which is either free abelian of finite rank or finite abelian.*

Finitely generated soluble groups

We finish this section with a few facts about finitely generated soluble groups.

Lemma 1.7. *A finitely generated soluble torsion group is finite.*

In general not much can be said about the structure of finitely generated soluble groups and their subgroups. For example, B.H. Neumann and H. Neumann proved in 1959 [23, Section 4.1.1] that every countable soluble group of derived length d may be embedded in a 2-generator soluble group of derived length at most $d+2$, and P. Hall proved in 1954 [23, Section 4.1.3] that, for any non-trivial countable abelian group A, there are $2^{\aleph_0}$ non-isomorphic 2-generator centre-by-metabelian groups G such that $A \cong Z(G) = G''$.

1.4.5 Some other useful group theory results

We will have several occasions to use the following fundamental theorem from combinatorial group theory. A proof can be found in [19] Chapter 4, Proposition 3.

Proposition 1.8. *Let G be a group with presentation $\langle X \mid R \rangle$ and let H be another group. A mapping $\theta : X \to H$ extends to a homomorphism $\theta' : G \to H$ if and only if, for all $x \in X$ and $r \in R$, the result of substituting $\theta(x)$ for x in r yields the identity of H.*

Next we state a useful lemma about finitely generated groups, due to Philip Hall. See [22, 14.1.3] for a proof.

Lemma 1.9. *Let G be a finitely generated group, let $N \lhd G$ and suppose that G/N is finitely presented. Then N is the normal closure in G of a finite subset of G.*

The following lemma is due to Schur. For a proof, see (for example) [22, 10.1.4].

Lemma 1.10. *Let G be a group with $Z(G)$ of finite index n in G. Then G' is finite and $(G')^n = \{1\}$.*

For given $n \in \mathbb{N}$, there is a bound on the size of finite subgroups of $\mathrm{GL}(n, \mathbb{Z})$. For large n, the precise bounds are known (see [24] and [4]), but as we require a bound for all n and do not need it to be precise, we use the following much older result of Minkowski, for which a proof is given in [24].

Proposition 1.11. *There exists a function $L : \mathbb{N} \to \mathbb{N}$ such that the order of any finite subgroup of $\mathrm{GL}(n, \mathbb{Z})$ divides $L(n)$ and thus is also bounded by $L(n)$.*

1.4.6 Hilbert Basis Theorem

A commutative ring is *Noetherian* if all ascending chains of ideals become constant, which is equivalent to all ideals being finitely generated. Similarly, a module over a ring is *Noetherian* if all ascending chains of submodules become constant or equivalently if all submodules are finitely generated.

A finitely generated module over a Noetherian ring is Noetherian, and hence all of its submodules are finitely generated.

The following result is known as the Hilbert Basis Theorem. For a proof, see [2, Theorem 7.5].

Proposition 1.12. *If the commutative ring R is Noetherian, then so is the polynomial ring $R[x]$. In particular, $\mathbb{Z}[x_1, \ldots, x_n]$ is Noetherian, so all submodules of finitely generated modules over $\mathbb{Z}[x_1, \ldots, x_n]$ are finitely generated.*

1.5 Formal languages and automata

A *formal language*, henceforth referred to as a *language*, is a set of *words* (finite length strings) over a finite set called the *alphabet* of the language. In other words, for a finite set X, a *language over* X is a subset of X^*.

Languages can be described in various ways. For finite languages, we can simply list the words in the language, although this might be very inefficient. Languages are most commonly described either by grammars, which are essentially sets of rules for constructing words in the language, or by automata, which are machines used to recognise whether a word belongs to the language.

In this section we introduce some of the most basic classes of languages, describing them via automata. The material is taken mostly from [12]. Notes on the origin of the results given here can be found at the ends of the relevant chapters in [12].

1.5.1 Finite automata and regular languages

Finite automata

A finite automaton (also called a finite state automaton) is a theoretical model of a system with finitely many states. The automaton takes words as input and reads a word one symbol at a time from left to right, deciding which state to move to next based on the current state and the symbol read. Upon reaching the end of the word, the automaton decides whether or not to accept the word based on the state it has ended in.

Formally, a *(nondeterministic) finite automaton* (NFA) is a tuple $(Q, \Sigma, \delta, q_0, F)$, consisting of

- a finite set of *states* Q,
- a finite *input alphabet* Σ,
- a *transition function* $\delta : Q \times (\Sigma \cup \{\epsilon\}) \to \mathcal{P}(Q)$,
- a distinguished *initial state* $q_0 \in Q$, and
- a distinguished set of *final states* $F \subseteq Q$.

The transition function describes the behaviour of the automaton while reading a word in Σ^*. Upon reading the symbol x in state q, the automaton moves to one of the states in $\delta(q, x)$. This movement is called a *transition*, and if $x = \epsilon$ it is an *ϵ-transition*.

We can think of the automaton as a directed graph, with the vertices identified with the elements of Q and the edges labelled by elements of Σ. Then $\delta(q, x)$ consists of all states $p \in Q$ such that there is a (directed) edge from q to p labelled x.

The automaton *accepts* a word $w = w_1 w_2 \dots w_n \in \Sigma^*$ if there is a directed path from q_0 to some state in F such that the edges in the path are labelled consecutively by $w_1, w_2, \dots, w_n$, with possibly some edges labelled by ϵ interspersed along the path. Note that it is not necessary for *all* directed paths from q_0 labelled by w in this way to end in a final state.

If $A = (Q, \Sigma, \delta, q_0, F)$ is a finite automaton, then *the language accepted by A* is the subset of Σ^* consisting of all words accepted by A. As a trivial example, let $A = (Q, \Sigma, \delta, q_0, F)$ with $Q = F = \{q_0\}$ and $\delta(q_0, x) = q_0$ for all $x \in \Sigma$. Then the language accepted by A is Σ^*.

Regular languages

A language which can be accepted by a finite automaton is called a *regular language*. The regular languages can equivalently be described by regular expressions, but we shall not define those here. See [12, Section 2.5] for details.

We have already seen that for any finite set X, the Kleene star closure of X is a regular language. Any finite subset of X^* is also clearly regular.

For a less trivial example of a regular language, let L be the set of all strings over the alphabet $\{0,1\}$ containing an odd number of 1's. Let A be a finite automaton with input alphabet $\{0,1\}$ and two states q_0 and q_1, with q_0 being the initial state and q_1 the sole final state. The transition function is given by $\delta(q_i, 0) = q_i$ for $i = 1, 2$, $\delta(q_0, 1) = q_1$ and $\delta(q_1, 1) = q_0$. So A changes state when and only when it reads a 1. So when an odd number of 1's has been read, A is in state q_1, while if an even number of 1's has been read, it is in state q_0. Thus the language accepted by A is L and so L is a regular language.

Of course many very natural languages are not regular. Consider for example the language $M = \{0^n 1^n \mid n \in \mathbb{N}_0\}$. Intuitively, a finite automaton accepting this language would have to be able to 'remember' how many 0's it had read in order to check that this matches with the number of 1's. But with only finitely many states, it seems unlikely that the automaton would have the capacity to remember an arbitrary natural number. As we shall see shortly, we can in fact prove that there is no finite automaton accepting M.

The pumping lemma for regular languages

A useful tool for proving languages not to be regular is the pumping lemma. It says roughly that, in any regular language, all words of sufficient length have a short nontrivial subword that can be 'pumped'. This means that the subword can be repeated arbitrarily many times and the resulting word will still be in the language. The idea makes sense, because if A is a finite automaton with n states, then any directed path in A of length n or more must contain a loop.

Lemma 1.13. [12, Lemma 3.1] *Let L be a regular language. Then there is a constant n such that if $z \in L$ with $|z| \geq n$, we may write $z = uvw$ in such a way that $|uv| \leq n$, $|v| \geq 1$, and for all $i \in \mathbb{N}_0$, $uv^iw \in L$. Furthermore, n is no greater than the number of states in the smallest finite automaton accepting L.*

This shows for example that our language M above is not regular, for if n is as in the lemma and $z = 0^n1^n$, then the subword v of z that can be pumped must consist entirely of zeroes. But then $0^{n+i|v|}1^n \in M$ for all $i \in \mathbb{N}_0$, which is not the case.

In the next subsection we will define a type of automaton capable of recognising the language M.

Deterministic finite automata

A *deterministic finite automaton* (DFA) is one in which there are no ϵ-transitions, and each $\delta(q, x)$ consists of a single element. That is, δ is a function from $Q \times \Sigma$ to Q. It is not difficult to show that any NFA A is equivalent to a DFA A', in the sense that A and A' both accept the same language (see [12, Theorem 2.1]), and so DFA's and NFA's both accept the same class of languages. The reason that we have given the nondeterministic definition is that we will be building upon it for the definition of pushdown automata.

1.5.2 Pushdown automata and context-free languages

A pushdown automaton is obtained from a finite automaton by the addition of a *stack*, which is a device for storing a string of symbols, with the restriction that new symbols can only be written onto the top of the stack, and only the top symbol can be read at any time. Symbols may also be erased from the top of the stack. The stack is sometimes called by other names such as a *last-in-first-out store.*

A pushdown automaton reads words from left to right, one symbol at a time. Upon reading a symbol, the automaton can, as well as potentially moving to a new state, write a word onto the top of the stack or delete the symbol currently at the top of the stack. The decisions about which state to move to next and what to do to the stack are both based on the current state and input symbol, as well as the current symbol at the top of the stack.

More formally, a pushdown automaton is a tuple $A = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, where Q, Σ, q_0 and F are as in the definition of a finite automaton, and

- Γ is a finite set called the *stack alphabet*,
- $Z_0 \notin \Gamma$ is the *start symbol*, used to recognise the bottom of the stack, and
- δ is the *transition function* from $Q \times (\Sigma \cup \{\epsilon\}) \times (\Gamma \cup Z_0)$ to the set of finite subsets of $Q \times (\Gamma^* \cup Z_0\Gamma^*)$, with the restriction that the second argument of an element of $\delta(q, x, y)$ is in Γ^* if $y \in \Gamma$, and in $Z_0\Gamma^*$ if $y = Z_0$.

The automaton starts in state q_0, with the single symbol Z_0 on the stack. If the automaton is in state q, then upon reading the symbol $x \in \Sigma \cup \{\epsilon\}$, with the symbol z at the top of the stack, the automaton chooses a pair (p, w) from $\delta(q, x, z)$ and moves to state p, writing w on the top of the stack in place of z. In particular, if $w = \epsilon$, this results in erasing z from the top of the stack, which is called *popping the* *stack*. The restriction on δ ensures that Z_0 only ever appears at the bottom of the stack, and is never removed from the stack.

The pushdown automaton A *accepts* a word $w \in \Sigma^*$ if it is possible for A to be in a final state after having read w and performed the resulting transitions and stack modifications.

Alternatively, we can let A accept *by empty stack*. This means that a word $w \in \Sigma^*$ is accepted if and only if it is possible for the stack to be empty (recognised by the start symbol Z_0 being at the top of the stack) after w has been read. This is equivalent to the usual acceptance by final state, in the sense that for any automaton A accepting by one of these methods, there exists an automaton A' which accepts by the other method and recognises the same language as A [12, Theorem 5.1].

The set of all words in Σ^* accepted by A is called the *language accepted by A*, and a language that can be accepted by a pushdown automaton is called a *context-free language*.

Examples and non-example of context-free languages

Of course every regular language is context-free, since a finite automaton is just a pushdown automaton with no stack (or in which the stack is not used).

As an example of a context-free language which is not regular, consider the language $M = \{0^n 1^n \mid n \in \mathbb{N}_0\}$ introduced in the subsection on regular languages. Let A be a pushdown automaton having three states, with input alphabet $\{0, 1\}$ and stack alphabet $\{X, Z_0\}$. The automaton starts in state q_0, with Z_0 on the stack. For every 0 read, an X is put on the stack. If A ever reads a 1, it records this by moving to state q_1, and pops an X off the stack. For every further 1 read, an X is popped off the stack. If A reads a 1 and Z_0 is at the top of the stack, this means there are more 1's than zeroes, so an X is put on the stack and A moves to the 'fail state' q_2. If A ever reads a 0 in state q_1, then it also puts an X on the stack and moves to q_2. If the automaton reaches q_2, it stays there and makes no further alterations to the stack. Upon having read a word in $w \in M$, the stack is empty, and A is in state q_1 unless w is the empty word, in which case A is in state q_0. After reading any other word in $\{0,1\}^*$, the stack is not empty. Thus A accepts M by empty stack, and so M is context-free.

Now consider the language $M_2 = \{0^m 1^n 0^m 1^n \mid m, n \in \mathbb{N}_0\}$. A pushdown automaton can certainly store the first half of a word in M_2 on its stack, but in order to check that the second 0^m subword matches the first one, it would appear to have to delete all the 1's it has stored, thereby forgetting that information, and not being able to check the first 1^n subword against the second one. Thus M_2 appears to be a good candidate for a non-context-free language.

The pumping lemma for context-free languages

The pumping lemma for context-free languages says roughly that, in any context-free language, every word of sufficient length has two short subwords which can be pumped simultaneously.

Lemma 1.14. [12, Lemma 6.1] *Let L be a context-free language. Then there is a constant n such that if z is in L and $|z| \geq n$, we may write $z = uvwxy$ such that $|vx| \geq 1$, $|vwx| \leq n$, and for all $i \in \mathbb{N}_0$, $uv^iwx^iy \in L$.*

It is easy to show using this lemma that the language M_2 defined above is not context-free, but the details are somewhat tedious, so we will not include them here. See Example 6.2 in [12]. Another good example of a non-context-free language is $L = \{a^n b^n c^n \mid n \in \mathbb{N}_0\}$. Again it is easy to show that this language does not satisfy the condition in the pumping lemma, as we would either end up with a word in L

with too many a's, b's or c's, or else a word in which some a's occur after some b's, or some b's occur after some c's.

Deterministic context-free languages

A pushdown automaton $(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ is *deterministic* if:

- $\delta(q, x, y)$ has at most one element for all $q \in Q$, $x \in \Sigma \cup \{\epsilon\}$, $y \in \Gamma$;
- if $\delta(q, \epsilon, y)$ is non-empty, then $\delta(q, x, y) = \emptyset$ for all $x \in \Sigma$.

This just means that the automaton has no choice at each transition.

A language accepted by a deterministic pushdown automaton is called a *deterministic context-free language.* The classes of context-free and deterministic context-free languages are non-equivalent. For example, the language consisting of all words of the form ww^R (w followed by its mirror image) over an alphabet with more than one symbol is context-free but not deterministic context-free ([12] p.113).

1.5.3 Semilinear sets and bounded context-free languages

The pumping lemma for context-free languages turns out to be not very useful when considering word problems of groups. A more useful tool is a relationship between bounded context-free languages and semilinear sets.

A *linear set* is a subset L of $\mathbb{N}_0^r$ such that there exist a *constant vector* $\mathbf{c} \in \mathbb{N}_0^r$ and a finite set of *periods* $P = \{\mathbf{p}_i \mid 1 \leq i \leq n\} \subseteq \mathbb{N}_0^r$ such that $L = \{\mathbf{c} + \sum_{i=1}^n \alpha_i \mathbf{p}_i \mid \alpha_i \in \mathbb{N}_0\}$. Note that the set of periods P is not uniquely determined: any linear combinations of elements of P with coefficients in $\mathbb{N}_0$ can be added to P without changing the set L.

A *semilinear set* is a union of finitely many linear sets.

A subset P of $\mathbb{N}_0^r$ is *stratified* if it satisfies the following conditions:

(i) each $\mathbf{p} \in P$ has at most two non-zero components, and

(ii) there do not exist $i < j < k < l$ and non-zero $a, b, c, d \in \mathbb{N}$ such that $ae_i + be_k$ and $ce_j + de_l$ are both in P.

A linear set is *stratified* if it can be expressed using a stratified set of periods. A semilinear set is *stratified* if it can be expressed as a union of finitely many stratified linear sets. (We follow Liu and Weiner [15] for this terminology.) Note that stratified linear and semilinear sets are not generally stratified sets in the sense of the previous paragraph.

A language $L \subseteq X^*$ is *bounded* if there exist $w_1, \ldots, w_n \in X^*$ such that $L \subseteq w_1^* \ldots w_n^*$.

Let $L \subseteq w_1^* \ldots w_n^*$ be a bounded language. We define the *commutative image* of L to be the following subset of $\mathbb{N}_0^n$:

$$\Phi(L) = \{(m_1, \ldots, m_n) \mid m_i \in \mathbb{N}_0, w_1^{m_1} \ldots w_n^{m_n} \in L\}.$$

For example, let M_2 be the non-context-free language $\{0^m 1^n 0^m 1^n \mid m, n \in \mathbb{N}_0\}$ defined in the previous section. Then $\Phi(M_2) = \{(m, n, m, n) \mid m, n \in \mathbb{N}_0\}$ is a linear set with constant vector $\mathbf{0}$ and set of periods $P = \{(1, 0, 1, 0), (0, 1, 0, 1)\}$. Since $e_1 + e_3$ and $e_2 + e_4$ are both in P, P is not a stratified subset of $\mathbb{N}_0^4$. It is not difficult to see that no stratified set of periods for M_2 exists. The case $k = 2$ of Theorem 2.16 in the following chapter shows that M_2 is not even a stratified semilinear set.

It was shown by Parikh in [20] that the commutative image of a bounded context-free language is always a semilinear set. This result is known as *Parikh's theorem.* However, as the example of M_2 shows, the commutative image being semilinear is not a sufficient condition for a bounded language to be context-free. Ginsburg and Spanier [7] later strengthened Parikh's result:

Theorem 1.15. [6, Theorem 5.4.2] *Let $W \subseteq w_1^* \dots w_n^*$, each w_i a word. Then W is context-free if and only if $\Phi(W)$ is a stratified semilinear set.*

Ginsburg and Spanier used different notation, which made it more transparent how to get from $\Phi(W)$ back to W. But as we will only require the 'only if' direction, we have kept the tidier notation of Parikh.

To illustrate why the assumption that every period of $\Phi(L)$ has at most two non-zero components is necessary, consider the non-context-free language $\{a^m b^m c^m \mid m \in \mathbb{N}_0\}$ from the previous section. Its commutative image is the linear subset of $\mathbb{N}_0^3$ with constant vector 0 and single period $(1, 1, 1)$.

The class of context-free languages is closed under intersection with regular languages (see Proposition 1.17). Let L be a language over Σ. If we can find words $w_1, \dots, w_n \in \Sigma^*$ such that $\Phi(L \cap w_1^* w_2^* \dots w_n^*)$ is not a stratified semilinear set, then Theorem 1.15 says that L cannot be context-free. The extension of this idea to the poly-context-free languages will be our primary technique for proving languages not to be poly-context-free.

Proving that a given semilinear set is not stratified is by no means straightforward, since there can be many different ways of expressing a semilinear set as a union of finitely many linear sets. Ginsburg [6] mentioned that there was no known decision procedure for determining whether an arbitrary semilinear set is stratified, and it appears that this is still an open problem today.

1.5.4 Operations on languages

One obvious question to ask about a class of languages is which operations it is closed under. Some of the operations we will consider are the standard Boolean operations of union, intersection and complementation (where the *complement* of a language L over Σ is defined to be the complement of L in Σ^*). But we also consider some other operations specific to languages, which we now define.

If L_1 and L_2 are languages, then the *concatenation* of L_1 and L_2, written L_1L_2, is the set of all words uv, where $u \in L_1$, $v \in L_2$.

Let Σ_1 and Σ_2 be alphabets. A *homomorphism* from Σ_1^* to Σ_2^* is a function $h : \Sigma_1^* \to \Sigma_2^*$ given by defining $h(x)$ to be some element of Σ_2^* for each $x \in \Sigma_1$, and if $w = x_1x_2 \dots x_n$ with $x_i \in \Sigma_1$, then $h(w) = h(x_1)h(x_2)\dots h(x_n)$. If L_1 and L_2 are languages over Σ_1 and Σ_2 respectively, then the *homomorphic image* of L_1 is

$$h(L_1) = \bigcup_{w \in L_1} h(w)$$

and the *inverse homomorphic image* of L_2 is

$$h^{-1}(L_2) = \{w \in \Sigma_1^* \mid h(w) \in L_2\}.$$

A *generalised sequential machine (GSM)* is a finite automaton with output. A GSM has the same basic structure as a finite automaton, but it also has an *output alphabet* X, and at each transition it adds a word from X^* onto the end of the current output (based only on the current state and input symbol). See [12, Section 11.2] for a more formal description and some examples.

Let M be a GSM with input alphabet Σ and output alphabet X. If L is a language over Σ, then the *GSM mapping* $M(L)$ of L is given by mapping each $w \in L$ to the word output by M after reading w. If L' is a language over X, then the *inverse GSM mapping* $M^{-1}(L')$ of L' is given by mapping each $w \in L$ to the set of all words in Σ^* that are mapped to w by M. Note that M^{-1} is not necessarily a true inverse, as demonstrated by [12, Example 11.2], which gives a GSM M and a language L such that L is a proper subset of $M^{-1}(M(L))$.

1.5.5 Closure properties

We now summarise some of the closure properties of the classes of regular and context-free languages. Proofs can be found in [12] in sections 3.2 and 6.2 respectively.

Proposition 1.16. *The class of regular languages is closed under union, concatenation, Kleene star closure, complementation, intersection, homomorphisms and inverse homomorphisms.*

Proposition 1.17. *The class of context-free languages is closed under union, concatenation, Kleene star closure, homomorphisms, inverse homomorphisms and intersection with regular languages.*

Our example language $M_2 = \{0^m 1^n 0^m 1^n \mid m, n \in \mathbb{N}_0\}$ shows that the context-free languages are not closed under intersection, since M_2 is the intersection of the context-free languages $\{0^m 1^* 0^m 1^* \mid m \in \mathbb{N}_0\}$ and $\{0^* 1^n 0^* 1^n \mid n \in \mathbb{N}_0\}$, but is not itself context-free. Using this, it is easy to show that the context-free languages are not closed under complementation. The deterministic context-free languages, however, are closed under complementation [12, Theorem 10.1].

A class of languages which is closed under homomorphisms, inverse homomorphisms and intersection with regular sets is called a *trio*. By the above two propositions, the classes of regular and context-free languages are both trios. It turns out that every trio is closed under GSM mappings and inverse generalised sequential machine mappings [12, Theorems 11.1 and 11.2], so we also have

Proposition 1.18. *The classes of regular and context-free languages are both closed under GSM mappings and inverse generalised sequential machine mappings.*

1.5.6 Poly-context-free languages

Since the class of context-free languages is not closed under intersection, it is interesting to consider its closure under the intersection operation. This class does not appear to have been much studied so far or to have a consistent name.

We call a language *k-context-free* (henceforth abbreviated to k-$\mathcal{CF}$) if it can be expressed as an intersection of k context-free languages, and *poly*-$\mathcal{CF}$ if it is k-$\mathcal{CF}$ for some $k \in \mathbb{N}$.

Liu and Weiner showed in [15] that the class of k-$\mathcal{CF}$ languages is properly contained in the class of $(k+1)$-$\mathcal{CF}$ languages for all $k \in \mathbb{N}$. (They call a k-$\mathcal{CF}$ language a 'k-intersection language'.) Note that this implies that the k-$\mathcal{CF}$ languages are not closed under intersection or even under intersection with context-free languages.

As there are a few problems with part of Liu and Weiner's proof, we shall give a complete proof of their result in Chapter 2.

Many closure properties of the classes of k-$\mathcal{CF}$ and poly-$\mathcal{CF}$ languages can be deduced from the similar properties given in Propositions 1.17 and 1.18.

Corollary 1.19. *For any $k \in \mathbb{N}$, the class of k-$\mathcal{CF}$ languages is closed under inverse homomorphisms, inverse GSM mappings, union with context-free languages and intersection with regular languages. The class of poly-$\mathcal{CF}$ languages is closed under all these operations, and also under intersection and union.*

Proof. Let $L = L_1 \cap \ldots \cap L_k$ with each L_i context-free and let Σ be the alphabet of L. Let Γ be an alphabet and let ϕ be a homomorphism from Γ^* to Σ^*, or a GSM mapping with input alphabet Γ and output alphabet Σ. Then

$$\begin{aligned}\phi^{-1}(L) &= \{w \in \Gamma^* \mid \phi(w) \in L_i \ (1 \le i \le k)\} \\ &= \bigcap_{i=1}^{k} \{w \in \Gamma^* \mid \phi(w) \in L_i\} = \bigcap_{i=1}^{k} \phi^{-1}(L_i),\end{aligned}$$

and since the class of context-free languages is closed under inverse homomorphisms and inverse GSM mappings, this implies that $\phi^{-1}(L)$ is k-$\mathcal{CF}$.

For any regular language R, $L_k \cap R$ is context-free, and so

$$L \cap R = L_1 \cap \ldots \cap L_{k-1} \cap (L_k \cap R)$$

is k-$\mathcal{CF}$. For any context-free language M, $L_i \cup M$ is context-free for all $1 \le i \le k$, so $L \cup M = \bigcap_{i=1}^{k}(L_i \cup M)$ is k-$\mathcal{CF}$.

The closure of the class of poly-$\mathcal{CF}$ languages under intersection is obvious, since if L_1 is k_1-$\mathcal{CF}$ and L_2 is k_2-$\mathcal{CF}$, then $L_1 \cap L_2$ is an intersection of $k_1 + k_2$ context-free languages.

If $L = \cap_{i=1}^{m} L_i$ and $M = \cap_{j=1}^{n} M_j$, with each L_i and M_j context-free, then

$$L \cup M = \left(\bigcap_{i=1}^{m} L_i\right) \cup \left(\bigcap_{j=1}^{n} M_j\right) = \bigcap_{i=1}^{m}\bigcap_{j=1}^{n}(L_i \cup L_j)$$

is mn-$\mathcal{CF}$, so the class of poly-$\mathcal{CF}$ languages is also closed under union. $\square$

Any recursively enumerable language can be expressed as a homomorphic image of the intersection of two deterministic context-free languages [5]. The poly-context-free languages are all recursive, while the recursive languages are a proper subclass of the recursively enumerable languages. Thus the poly-$\mathcal{CF}$ languages are not closed under homomorphisms.

1.6 Word problems of groups as languages

The *word problem* of a group G with respect to a finite generating set X is $W(G,X) = \{w \in (X \cup X^{-1})^* \mid w =_G 1\}$, the set of all words in $(X \cup X^{-1})^*$ which represent the identity element of G.

Since the word problem of G with respect to X is a subset of $(X \cup X^{-1})^*$, we can consider $W(G,X)$ as a language.

Consider for example $\mathbb{Z}^2$, the free abelian group of rank 2, with generating set $X = \{a, b\}$. Let $A = a^{-1}$ and $B = b^{-1}$. The word problem of F_2 with respect to X is the set of all words in $(X \cup X^{-1})^*$ which contain exactly as many a's as A's, and exactly as many b's as B's. Intersecting $W(F_2, X)$ with the regular language $a^*b^*A^*B^*$ yields the language consisting of all words of the form $a^m b^n A^m B^n$, which is our non-context-free language M_2 from Section 1.5.3. Thus $\mathbb{Z}^2$ is an example of a group with non-context-free word problem.

A closely related language is the *co-word problem* of G with respect to X, denoted co$W(G,X)$. This is the complement of $W(G,X)$ in $(X \cup X^{-1})^*$, and is therefore the set of all words in $(X \cup X^{-1})^*$ which represent nontrivial elements of G.

A central result in the theory of word problems of groups as languages is the following, for which a proof is given in [9] Lemma 1.

Lemma 1.20. *Let $\mathcal{C}$ be a class of languages closed under inverse homomorphisms and let G be a finitely generated group. Then the following hold.*

(i) $W(G,X) \in \mathcal{C}$ for some finite generating set X if and only if for every finite generating set Y, $W(G,Y) \in \mathcal{C}$.

(ii) co$W(G,X) \in \mathcal{C}$ for some finite generating set X if and only if for every finite generating set Y, co$W(G,Y) \in \mathcal{C}$.

This means that for the classes of languages we are interested in, the formal language type of $W(G, X)$ does not depend on the generating set X, but is a property of the group itself. We can thus say for example that $W(G)$ is context-free, without reference to a generating set. For $\mathcal{C}$ a class of languages, we call a group a $\mathcal{C}$ group if $W(G)$ is in $\mathcal{C}$. So for example, a group is *context-free* if its word problem is context-free. If co$W(G)$ is in $\mathcal{C}$, we call G a co$\mathcal{C}$ group.

Lemma 1.21. [9, Lemma 2] *Let $\mathcal{C}$ be a class of languages closed under inverse homomorphisms and intersection with regular sets. Then the classes of $\mathcal{C}$ groups and co$\mathcal{C}$ groups are closed under taking finitely generated subgroups.*

Lemma 1.22. [9, Lemma 5] *Let $\mathcal{C}$ be a class of languages closed under union with regular sets and inverse GSM mappings. Then the classes of $\mathcal{C}$ groups and co$\mathcal{C}$ groups are closed under passing to finite index overgroups.*

Thus we have:

Proposition 1.23. *The classes of context-free groups, co$\mathcal{CF}$ groups and k-$\mathcal{CF}$ groups are insensitive to choice of generators and closed under passing to finitely generated subgroups and passing to finite index overgroups.*

Proof. Follows immediately from Corollary 1.19 and Lemmas 1.20, 1.21 and 1.22. $\square$

A characterisation of the languages which are word problems of groups was given by Parkes and Thomas:

Proposition 1.24. [21, Proposition 3.3] *Let W be a subset of Σ^*; then W is the word problem of a group if and only if it satisfies the following conditions:*

(i) if $\alpha \in \Sigma^*$ then there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in W$;

(ii) if $\alpha \in W$ and $u\alpha v \in W$ then $uv \in W$.

Of course the conditions (i) and (ii) are necessary for W to be the word problem of a group, but it is interesting that they are also sufficient.

1.6.1 Groups with regular word problem

If G is a finite group and X is a generating set of G, then we can use the Cayley graph of G with respect to $X \cup X^{-1}$ to construct a finite automaton recognising $W(G, X)$. The automaton has the same vertices, directed edges and labels as the Cayley graph, with the vertex corresponding to the identity of G being the initial state and the sole final state. This shows that every finite group has regular word problem.

It is not difficult to show that every group with regular word problem is finite, and so the groups with regular word problem are exactly the finite groups. This result is originally due to Anisimov [1].

1.6.2 Groups with context-free word problem

The obvious examples of context-free groups are the free groups.

Observation 1.25. *A finitely generated virtually free group has context-free word problem.*

Proof. Since the context-free groups are closed under taking finite index overgroups, it suffices to show that a finitely generated free group has context-free word problem.

Let G be the free group on a finite set $X = \{x_1, \ldots, x_n\}$. We construct a pushdown automaton with input alphabet and stack alphabet both $X \cup X^{-1}$. The automaton has a single state, and accepts by empty stack. Upon reading $x_i^{\pm}$ (meaning x_i or x_i^{-1}), the automaton puts $x_i^{\pm}$ on top of the stack, unless the symbol on the top of the stack is $x_i^{\mp}$, in which case the stack is popped. Thus after reading a word $w \in (X \cup X^{-1})^*$, the stack stores the reduced form of w, when read from bottom to top. Hence the only words which result in an empty stack are those which are trivial in $F(X) = G$, and so the automaton recognises $W(G, X)$. $\square$

The context-free groups proved much more difficult to characterise than the groups with regular word problem, and it was not until the 1980s that a characterisation was achieved. Using methods from geometric group theory, Muller and Schupp proved in [17] that a finitely generated group is virtually free if and only if it is context-free and accessible. We shall not go into what accessibility means here. The authors believed that context-free groups were accessible, but were unable to prove it. The accessibility of finitely presented groups was proven by Dunwoody in [3], which appeared two years after [17]. Every context-free group is finitely presented (this is shown for example in [17]), and so we have:

Proposition 1.26. [17, 3] *A finitely generated group is context-free if and only if it is virtually free.*

Every context-free group is deterministic context-free. This was shown by Muller and Schupp in [18], but can also be shown by extending the construction in Observation 1.25 to give a deterministic pushdown automaton accepting the word problem of a virtually free group.

1.6.3 Groups with context-free co-word problem

A language is called *co$\mathcal{CF}$* if it is the complement of a context-free language. Since the class of context-free languages is not closed under complementation, it is worth considering groups whose word problem is co$\mathcal{CF}$, or equivalently, whose co-word problem is context-free.

These groups have been studied by Holt, Rees, Röver and Thomas in [9], and by Lehnert and Schweitzer in [14]. We here summarise some of the main results from both papers.

Results of Holt, Rees, Röver and Thomas

The co$\mathcal{CF}$ groups include all context-free groups, since the context-free groups are determistic context-free and the deterministic context-free languages are closed under complementation.

We have already mentioned some of the properties of the class of co$\mathcal{CF}$ groups in Proposition 1.23, which is simply a generalisation of part of [9, Proposition 6]. Also contained in [9, Proposition 6] is the fact that the co$\mathcal{CF}$ groups are closed under taking finite direct products. (This is not true of the context-free groups, as for example the group $\mathbb{Z}$ is context-free, but $\mathbb{Z}^2$ is not context-free.)

The co$\mathcal{CF}$ groups are also closed under taking restricted standard wreath products with context-free top group:

Proposition 1.27. [9, Theorem 10] *Let G be a co$\mathcal{CF}$ group and let H be a context-free group. Then the restricted standard wreath product, $G \wr H$, of G with H is co$\mathcal{CF}$.*

The next proposition is used throughout [9] in combination with Parikh's theorem (see Section 1.5.3) to prove various classes of groups not to be co$\mathcal{CF}$. Before stating the proposition, we need to introduce some new notation, which we will continue to use throughout the thesis.

If $\mathbf{a}$ and $\mathbf{b}$ are vectors in $\mathbb{N}_0^r$ and $\mathbb{N}_0^s$ respectively, then we denote by $(\mathbf{a}; \mathbf{b})$ the vector in $\mathbb{N}_0^{r+s}$ which consists of all the components of $\mathbf{a}$ in order, followed by those of $\mathbf{b}$ in order. (So $(\mathbf{a};\mathbf{b})(i) = \mathbf{a}(i)$ for $1 \leq i \leq r$ and $(\mathbf{a};\mathbf{b})(r+i) = \mathbf{b}(i)$ for $1 \leq i \leq s$.) When talking about vectors in $\mathbb{N}_0^{r+s}$, if we write $(\mathbf{a};\mathbf{b})$, then it is understood that $\mathbf{a} \in \mathbb{N}_0^r$ and $\mathbf{b} \in \mathbb{N}_0^s$. For $\mathbf{a} \in \mathbb{N}_0^r$, we define $\sigma(\mathbf{a}) = \sum_{i=1}^r \mathbf{a}(i)$.

Proposition 1.28. [9, Proposition 14] *Let $L \subseteq \mathbb{N}_0^{r+s}$ for some $r, s \in \mathbb{N}$. Suppose that for every $k \in \mathbb{N}$, there exists $\mathbf{a} \in \mathbb{N}_0^r \setminus \{\mathbf{0}\}$ such that the following hold:*

(i) There exists a unique $\mathbf{b} \in \mathbb{N}_0^s$ such that $(\mathbf{a};\mathbf{b}) \in L$.

(ii) If $(\mathbf{a};\mathbf{b}) \in L$, then $\mathbf{b}(j) \geq k\sigma(\mathbf{a})$ for $1 \leq j \leq s$.

Then the complement of L in $\mathbb{N}_0^{r+s}$ is not a semilinear set.

The next three results are all proven using Proposition 1.28.

Proposition 1.29. [9, Theorem 12] *A finitely generated nilpotent group has context-free co-word problem if and only if it is virtually abelian.*

For $m, n \in \mathbb{Z} \setminus \{0\}$, the *Baumslag-Solitar group* $\mathsf{BS}(m,n)$ is the group with presentation $\langle x, y \mid y^{-1}x^m y = x^n \rangle$.

Proposition 1.30. [9, Theorem 13] *The Baumslag-Solitar group $\mathsf{BS}(m,n)$ is co$\mathcal{CF}$ if and only if $m = \pm n$.*

The above result is stated incorrectly in [9]. It is claimed that $\mathsf{BS}(m,n)$ is virtually abelian if $m = \pm n$. In Chapter 3, we will show that $\mathsf{BS}(m,n)$ is a co$\mathcal{CF}$ group if $m = \pm n$, completing the proof of this corrected statement of [9, Theorem 13].

Proposition 1.31. [9, Theorem 16] *A polycyclic group has context-free co-word problem if and only if it is virtually abelian.*

One other point of interest from [9] is that the authors were unable to establish whether the class of co$\mathcal{CF}$ groups is closed under taking free products. It is strongly conjectured that it is not, but this problem remains open. To solve this problem will probably require the development of new methods for showing languages not to be context-free, although the correspondence with stratified semilinear sets might still be able to be used.

Results of Lehnert and Schweitzer

Lehnert and Schweitzer's paper consists of further examples of co$\mathcal{CF}$ groups. They show that the Higman-Thompson groups $G_{n,r}$, the Houghton groups H_n ($n \geq 2$) and the groups $\mathrm{Hou}(F_n)$ all have co-context-free word problem. The definitions of these groups are all too involved to be presented in this introduction. The Houghton groups had apparently been proposed as an possible example of groups with indexed (see 1.6.4 below) but not context-free co-word problem.

1.6.4 Word problems in some other language classes

Groups with word problem in several other classes of languages have also been studied.

A *one-counter language* is a context-free language accepted by a pushdown automaton with only one stack symbol. Herbst [8] showed that a group has one-counter word problem if and only if it is virtually cyclic. In [11], Holt, Owens and Thomas extended this result to groups whose word problem is an intersection of finitely many one-counter languages. Considering the relevance of this to our study of poly-$\mathcal{CF}$ groups, we quote the full result.

Proposition 1.32. [11, Theorem 5.2] *The following are equivalent for a finitely generated group G.*

(i) The word problem of G is the intersection of n one-counter languages for some $n \geq 1$.

(ii) The word problem of G is the intersection of n deterministic one-counter languages for some $n \geq 1$.

(iii) The group G is virtually abelian of free abelian rank at most n.

The *free abelian rank* of a virtually abelian group is the rank of a finite index free abelian subgroup.

An *indexed language* is a language accepted by a nested stack automaton, which is a generalisation of a pushdown automaton which we will not define precisely here. The groups with indexed co-word problem were studied by Holt and Röver in [10]. They showed that the Higman-Thompson groups and certain tree-automorphism groups, including the well-known Grigorchuk and Gupta-Sidki groups, have indexed co-word problem.

In the same paper, the authors also considered a more restricted class of groups which they named stack groups, accepted by a nested stack automaton satisfying some extra conditions. All their examples of groups with indexed co-word problem are stack groups, but they were unable to determine whether every group with indexed co-word problem is a stack group. They were able to show that the stack groups are closed under the free product operation.

There are not currently any known examples of stack groups which are not co$\mathcal{CF}$, but Holt and Röver conjecture that the Grigorchuk group is not co$\mathcal{CF}$.

Probably the oldest result on word problems of groups as languages is the famous Novikov-Boone theorem on the existence of finitely presentable groups with undecidable word problem, implying that even the class of recursive languages is not complex enough to contain the word problems of all finitely presented groups.

Chapter 2

Poly-$\mathcal{CF}$ languages

Recall that a k-$\mathcal{CF}$ language is an intersection of k context-free languages, and a poly-$\mathcal{CF}$ language is a language which is k-$\mathcal{CF}$ for some $k \in \mathbb{N}$.

In this chapter, we shall primarily be concerned with proving some results which will assist us in determining that the word problems of certain groups are not poly-$\mathcal{CF}$.

Following Ginsburg [6], we will use the notation $L(\mathbf{c}; \mathbf{p}_1, \ldots, \mathbf{p}_n)$ or $L(\mathbf{c}; P)$ for a linear set with constant $\mathbf{c}$ and set of periods $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. For C a set of constant vectors, we will denote $\bigcup_{\mathbf{c} \in C} L(\mathbf{c}; P)$ by $L(C; P)$. So, for finite subsets C and $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ of $\mathbb{N}_0^r$,

$$L(C; P) = \{\mathbf{c} + \sum_{i=1}^{n} \alpha_i \mathbf{p}_i \mid \mathbf{c} \in C, \alpha_i \in \mathbb{N}_0\}.$$

If $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$, we will also often write $L(\mathbf{c}_1, \ldots, \mathbf{c}_m; \mathbf{p}_1, \ldots, \mathbf{p}_n)$ for $L(C; P)$.

If $L = L(\mathbf{c}; P)$, we define $L^{\mathbb{Q}}$ to be the set $\{\mathbf{c} + \sum_{i=1}^{n} a_i \mathbf{p}_i \mid a_i \in \mathbb{Q}\}$. This is a coset in $\mathbb{Q}^n$ of the $\mathbb{Q}$-subspace spanned by P. We define $L^{\mathbf{0}}$ to be $L(\mathbf{0}; P)$, that is, the linear set having the same periods as L and constant $\mathbf{0}$.

2.1 Semilinear sets and bounded poly-$\mathcal{CF}$ languages

Ginsburg and Spanier's result (Theorem 1.15) on commutative images of bounded context-free languages can easily be generalised to poly-$\mathcal{CF}$ languages.

Corollary 2.1. *If L is an intersection of k context-free languages, then, for any $w_1, \ldots, w_n$, the commutative image $\Phi(L \cap w_1^* \ldots w_n^*)$ is an intersection of k stratified semilinear sets.*

Proof. Let $L = L_1 \cap \ldots \cap L_k$ with each L_i context-free, and let $W = w_1^* \ldots w_n^*$, where each w_i is a word in the alphabet of L. For $1 \leq i \leq k$, let $M_i = L_i \cap W$. Then $L \cap W = L_1 \cap \ldots \cap L_k \cap W = \bigcap_{i=1}^{k} M_i$ and

$$\begin{aligned}
\Phi(L \cap W) &= \{(m_1, \ldots, m_n) \mid m_i \in \mathbb{N}_0, w_1^{m_1} \ldots w_n^{m_n} \in L \cap W\} \\
&= \{(m_1, \ldots, m_n) \mid m_i \in \mathbb{N}_0, w_1^{m_1} \ldots w_n^{m_n} \in M_i \ (1 \leq i \leq k)\} \\
&= \bigcap_{i=1}^{k} \{(m_1, \ldots, m_n) \mid m_i \in \mathbb{N}_0, w_1^{m_1} \ldots w_n^{m_n} \in M_i\} \\
&= \bigcap_{i=1}^{k} \Phi(M_i)
\end{aligned}$$

and each $\Phi(M_i)$ is a stratified semilinear set by Theorem 1.15. □

2.2 Intersections of semilinear sets

In order to apply Corollary 2.1 to prove certain languages not to be poly-$\mathcal{CF}$, it will help to know some more about intersections of semilinear sets.

An intersection of k (stratified) semilinear sets can be expressed as a union of finitely many sets which are each an intersection of k (stratified) linear sets. We shall thus first consider what we can say about intersections of finitely many linear sets.

It turns out that an intersection of finitely many linear sets is always a semilinear set of quite a restricted form. This can be proven as a consequence of the following lemma.

Lemma 2.2. *Let* $L = \{\mathbf{c} + \sum_{i=1}^{k} \alpha_i \mathbf{p}_i \mid \alpha_i \in \mathbb{Z}\}$ *be a coset of a subgroup of* $\mathbb{Z}^r$*. Then* $L \cap \mathbb{N}_0^r$ *is a semilinear subset of* $\mathbb{N}_0^r$*, provided it is non-empty. Moreover,*

$$L \cap \mathbb{N}_0^r = L(\mathbf{a}_1, \ldots, \mathbf{a}_m; \mathbf{q}_1, \ldots, \mathbf{q}_n)$$

for some $\mathbf{a}_i, \mathbf{q}_j \in \mathbb{N}_0^r$*, with the* $\mathbf{q}_j$ *depending only on* $\mathbf{p}_1, \ldots, \mathbf{p}_k$*.*

Proof. Let $\widehat{L} = L \cap \mathbb{N}_0^r$. If the $\mathbf{p}_i$ are all zero, then $\widehat{L}$ is either empty or consists of the single vector $\mathbf{c}$. So we may assume some $\mathbf{p}_i$ is non-zero.

Under the ordering given by $\mathbf{a} \leq \mathbf{b}$ if and only if $\mathbf{a}(i) \leq \mathbf{b}(i)$ for all $1 \leq i \leq r$, any subset of $\mathbb{N}_0^r$ has finitely many minimal elements (see for example [6, Corollary 5.4.1]). Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ be the set of minimal elements of $\widehat{L}$. Clearly $\widehat{L}$ has at least one minimal element. For each $1 \leq i \leq m$, let $S_i = \{\mathbf{v} \in \widehat{L} \mid \mathbf{a}_i \leq \mathbf{v}\}$. Then $\widehat{L} = \cup_{i=1}^m S_i$. Also, define

$$P_i = \{\mathbf{v} - \mathbf{a}_i \mid \mathbf{v} \in S_i\} \setminus \{\mathbf{0}\}.$$

Then $\widehat{L} = \cup_{i=1}^m \{\mathbf{a}_i + \mathbf{q} \mid \mathbf{q} \in P_i\}$.

The difference between any two elements of L is a linear combination of $\mathbf{p}_1, \ldots, \mathbf{p}_k$. Since each $S_i \subseteq \widehat{L}$, this implies $P_i \subseteq \mathsf{span}\{\mathbf{p}_1, \ldots, \mathbf{p}_k\} \cap \mathbb{N}_0^r$ for all $1 \leq i \leq m$. Conversely, for any $\mathbf{w} = \sum_{i=1}^{k} \alpha_i \mathbf{p}_i \in \mathbb{N}_0^r \quad (\alpha_i \in \mathbb{Z})$, adding $\mathbf{w}$ to an element of $\widehat{L}$ gives another element of $\widehat{L}$, so $\mathbf{a}_i + \mathbf{w} \in \widehat{L}$ and $\mathbf{w} \in P_i$ for all i. Thus

$$P_1 = \ldots = P_n = P := \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{p}_i \mid \alpha_i \in \mathbb{Z} \right\} \cap \mathbb{N}_0^r.$$

So $\widehat{L} = \{\mathbf{a}_i + \mathbf{q} \mid 1 \leq i \leq m, \mathbf{q} \in P\}$. Since $P \subseteq \mathbb{N}_0^r$, P has finitely many minimal elements $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. We will show that the minimal elements of P span P. Note that, for any non-minimal element $\mathbf{q} \in P$, if $\mathbf{q}_i < \mathbf{q}$ for some $1 \leq i \leq n$ then $\mathbf{q} - \mathbf{q}_i \in P$.

Let $\mathbf{q} \in P$ be non-minimal. Then there exists some minimal $\mathbf{q}_i$ such that $\mathbf{q}_i < \mathbf{q}$. Let $\mathbf{v}_1 = \mathbf{q} - \mathbf{q}_i \in P$. If $\mathbf{v}_1 = \mathbf{q}_j$ for some $1 \leq j \leq n$, then $\mathbf{q} = \mathbf{q}_i + \mathbf{q}_j$, thus $\mathbf{q} \in \mathsf{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. Otherwise $\mathbf{v}_1$ is not minimal and so there exists some $\mathbf{q}_{j_1}$ with $\mathbf{q}_{j_1} < \mathbf{v}_1$. If we continue in this way, defining $\mathbf{v}_{s+1} = \mathbf{v}_s - \mathbf{q}_{j_s}$ (where $\mathbf{q}_{j_s} < \mathbf{v}_s$) whenever $\mathbf{v}_s$ is not minimal, we must eventually reach some minimal element $\mathbf{q}_t$, since if $d = \max\{\mathbf{q}(i) \mid 1 \leq i \leq r\}$, then $\mathbf{v}_{rd+1} \notin P$. Therefore $\mathbf{v}_s$ is minimal in P for some $s \leq rd$ and so $\mathbf{v}_s = \mathbf{q}_t$ for some $1 \leq i \leq n$. Thus

$$\mathbf{q}_t = \mathbf{v}_s = \mathbf{v}_{s-1} - \mathbf{q}_{j_{s-1}} = \mathbf{q} - (\mathbf{q}_i + \mathbf{q}_{j_1} + \ldots + \mathbf{q}_{j_{s-1}}),$$

thus $\mathbf{q} = \mathbf{q}_t + \mathbf{q}_i + \mathbf{q}_{j_1} + \ldots + \mathbf{q}_{j_{s-1}} \in \mathsf{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. Hence we conclude that $P = \mathsf{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. Since P depends only on $\mathbf{p}_1, \ldots, \mathbf{p}_k$, the $\mathbf{q}_j$ also depend only on $\mathbf{p}_1, \ldots, \mathbf{p}_k$.

We have now shown that $\widehat{L} = L(\mathbf{a}_1, \ldots, \mathbf{a}_m; \mathbf{q}_1, \ldots, \mathbf{q}_n)$, which is a finite union of semilinear sets. Hence $\widehat{L} = L \cap \mathbb{N}_0^r$ is semilinear. □

The following result can be derived from the proof of Theorem 5.6.1 in [6], but since we require a more precise statement, we give a proof, which was obtained independently by the author.

Proposition 2.3. *If L is the nonempty intersection of linear subsets $L_1, \ldots, L_n$ of $\mathbb{N}_0^r$, then L is semilinear. Moreover,*

$$L = L(\mathbf{C}_1, \ldots, \mathbf{C}_k; \mathbf{P}_1, \ldots, \mathbf{P}_m),$$

where $\mathbf{C}_i \in \mathbb{N}_0^r$, and $\mathbf{P}_1, \ldots, \mathbf{P}_m$ are such that $\bigcap_{i=1}^n L_i^{\mathbf{0}} = L(\mathbf{0}; \mathbf{P}_1, \ldots, \mathbf{P}_m)$.
If $L_1, \ldots, L_n$ all have constant vector zero, then L is linear with constant vector zero.

Proof. The statement is certainly true for $n = 1$. We begin by proving the case $n = 2$ and then complete the proof by induction. Let $L = L_1 \cap L_2 \neq \emptyset$, where $L_1 = L(\mathbf{u}; \mathbf{p}_1, \ldots, \mathbf{p}_{k_1})$ and $L_2 = L(\mathbf{v}; \mathbf{q}_1, \ldots, \mathbf{q}_{k_2})$ with $\mathbf{u}, \mathbf{v}, \mathbf{p}_i, \mathbf{q}_i \in \mathbb{N}_0^r$. Let

$$S = \{\mathbf{c}_0 + \sum_{i=1}^{s} \gamma_i \mathbf{c}_i \mid \gamma_i \in \mathbb{Z}\} \subseteq \mathbb{Z}^{k_1+k_2}$$

be the set of integer solutions to the system of linear equations in the k_1+k_2 variables $\alpha_1, \ldots, \alpha_{k_1}, \beta_1, \ldots, \beta_{k_2}$ given by

$$\mathbf{u}(j) + \sum_{i=1}^{k_1} \alpha_i \mathbf{p}_i(j) = \mathbf{v}(j) + \sum_{i=1}^{k_2} \beta_i \mathbf{q}_i(j)$$

for $1 \leq j \leq r$. Note that $S \neq \emptyset$, since $L \neq \emptyset$. Also, $\mathbf{c}_i \in \mathbb{Z}^{k_1+k_2}$ for $1 \leq i \leq s$, and an element $\mathbf{c} \in S$ represents a solution $(\alpha_1, \ldots, \alpha_{k_1}, \beta_1, \ldots, \beta_{k_2})$ to the system of equations. Then

$$L = \{\mathbf{u} + \sum_{i=1}^{k_1} \mathbf{c}(i)\mathbf{p}_i \mid \mathbf{c} \in S \cap \mathbb{N}_0^{k_1+k_2}\}.$$

By Lemma 2.2, $\widehat{S} := S \cap \mathbb{N}_0^{k_1+k_2}$ is semilinear, of the form

$$\{\mathbf{a}_i + \sum_{j=1}^{m} \gamma_j \mathbf{b}_j \mid 1 \leq i \leq k, \gamma_j \in \mathbb{N}_0\},$$

where $\mathbf{b}_1, \ldots, \mathbf{b}_m$ depend only on $\mathbf{c}_1, \ldots, \mathbf{c}_s$. Thus $L = \cup_{i=1}^{k} M_i$, where

$$\begin{aligned} M_i &= \{\mathbf{u} + \sum_{j=1}^{k_1} [\mathbf{a}_i + \gamma_1 \mathbf{b}_1 + \ldots + \gamma_m \mathbf{b}_m](j)\mathbf{p}_j \mid \gamma_1, \ldots, \gamma_m \in \mathbb{N}_0\} \\ &= L(\mathbf{C}_i; \mathbf{P}_1, \ldots, \mathbf{P}_m), \end{aligned}$$

where $\mathbf{C}_i = \mathbf{u} + \sum_{j=1}^{k_1} \mathbf{a}_i(j)\mathbf{p}_j$ and $\mathbf{P}_l = \sum_{j=1}^{k_1} \mathbf{b}_l(j)\mathbf{p}_j$ for $1 \leq l \leq m$. Since $\mathbf{u}, \mathbf{a}_i, \mathbf{b}_l \in \mathbb{N}_0^r$, we see that $\mathbf{C}_i, \mathbf{P}_j \in \mathbb{N}_0^r$ and thus each M_i is a linear subset of $\mathbb{N}_0^r$ and L is semilinear, with $L = L(\mathbf{C}_1, \ldots, \mathbf{C}_k; \mathbf{P}_1, \ldots, \mathbf{P}_m)$.

If $\mathbf{u}$ and $\mathbf{v}$ are both zero, then zero is the unique minimal element of $\widehat{S}$, and so $k = 1$ and $\mathbf{a}_1 = \mathbf{0}$. This gives $\mathbf{C}_i = \mathbf{u} = \mathbf{0}$ for all $1 \leq i \leq k$. Since the M_i already all have the same set of periods, this means that they are all equal and so L is a linear set with constant zero. By induction, the intersection of any finite number of linear sets with constant zero is a linear set with constant zero.

The set $\{\sum_{i=1}^{s} \gamma_i \mathbf{c}_i \mid \gamma_i \in \mathbb{Z}\}$ is the set of integer solutions to the system of linear equations given by

$$\sum_{i=1}^{k_1} \alpha_i \mathbf{p}_i(j) = \sum_{i=1}^{k_2} \beta_i \mathbf{q}_i(j)$$

for $1 \leq j \leq r$. Now the $\mathbf{P}_j$ $(1 \leq j \leq m)$ depend only on $\mathbf{b}_1, \ldots, \mathbf{b}_m$ (which in turn depend only on $\mathbf{c}_1, \ldots, \mathbf{c}_s$) and $\mathbf{p}_1, \ldots, \mathbf{p}_{k_1}$. Thus the $\mathbf{P}_j$ do not depend on $\mathbf{u}$ and $\mathbf{v}$, and so the above argument gives us

$$L_1^{\mathbf{0}} \cap L_2^{\mathbf{0}} = L(\mathbf{0}; \mathbf{p}_1, \ldots, \mathbf{p}_{k_1}) \cap L(\mathbf{0}; \mathbf{q}_1, \ldots, \mathbf{q}_{k_2}) = L(\mathbf{0}; \mathbf{P}_1, \ldots, \mathbf{P}_m).$$

We have now proved the lemma for $n = 2$, and also for all n in the case where all the L_i have constant zero.

For $n \geq 2$, suppose that the lemma holds for intersections of at most n linear sets. Let L be an intersection of n linear subsets $L_1, \ldots, L_n$ of $\mathbb{N}_0^r$ and let L_{n+1} be any other linear subset of $\mathbb{N}_0^r$. We can write $L = L(\mathbf{C}_1, \ldots, \mathbf{C}_k; \mathbf{P}_1, \ldots, \mathbf{P}_m)$ as in the statement of the lemma. As before, let $M_i = L(\mathbf{C}_i; \mathbf{P}_1, \ldots, \mathbf{P}_m)$ for $1 \leq i \leq k$. Each M_i is a linear set, so by the induction hypothesis, for each $1 \leq i \leq k$ there exist finite subsets D_i and Q_i of $\mathbb{N}_0^r$ such that $M_i \cap L_{n+1} = L(D_i; Q_i)$, with $M_i^{\mathbf{0}} \cap L_{n+1}^{\mathbf{0}} = L(\mathbf{0}; Q_i)$. But $M_i^{\mathbf{0}} = L(\mathbf{0}; \mathbf{P}_1, \ldots, \mathbf{P}_m) = L^{\mathbf{0}}$ for all i, so the Q_i are all equal, say $Q_i = Q$ for all i. Thus

$$\begin{aligned} L \cap L_{n+1} &= (\cup_{i=1}^{k} M_i) \cap L_{n+1} = \cup_{i=1}^{k} (M_i \cap L_{n+1}) \\ &= \cup_{i=1}^{k} L(D_i; Q) = L(\cup_{i=1}^{k} D_i; Q) \end{aligned}$$

as required. Moreover, by the induction hypothesis $L_1^{\mathbf{0}} \cap \ldots \cap L_n^{\mathbf{0}} = L^{\mathbf{0}}$, so

$$L_1^{\mathbf{0}} \cap \ldots \cap L_{n+1}^{\mathbf{0}} = L^{\mathbf{0}} \cap L_{n+1}^{\mathbf{0}} = L(\mathbf{0}; Q). \qquad \square$$

Corollary 2.4. *Let L be an intersection of finitely many semilinear sets. Then L is a semilinear set.*

Proof. L can be expressed as a union of finitely many sets, each of which is an intersection of finitely many linear sets. By Proposition 2.3, an intersection of finitely many linear sets is semilinear, hence L is itself semilinear. $\square$

2.3 A criterion for a language to be not poly-$\mathcal{CF}$

In Chapter 1 we mentioned that many of the results in [9] used Proposition 1.28 (Proposition 14 in [9]), which gives a technical condition for a subset L of $\mathbb{N}_0^r$ not to be the complement of a semilinear set. We now show that a strictly weaker set of conditions imply that the set L itself is not an intersection of finitely many semilinear sets. This is very useful, as it will allow us to apply the proofs of some of the results in [9] to show that the same groups are not poly-$\mathcal{CF}$.

We first require a lemma extracted from the proof of Proposition 11 in [9]. We call a vector $\mathbf{v} \in \mathbb{N}_0^{r+s}$ *simple* if its first r components are all zero, and *complex* otherwise. The proof is quoted from [9] with only minor modifications.

Lemma 2.5. *Let $L = L_1 \cup \ldots \cup L_n$, with each L_i a linear subset of $\mathbb{N}_0^{r+s}$. Then there exists a constant $C \in \mathbb{N}$ such that if $(\mathbf{a}; \mathbf{b}) \in L$ can be expressed using only complex periods, then $\mathbf{b}(j) < C\sigma(\mathbf{a})$ for all $1 \leq j \leq s$.*

Proof. Fix some $i \in \{1, \ldots, n\}$ and let $L_i = L(\mathbf{c}_i; P_i)$. If $(\mathbf{p}; \mathbf{q}) \in P_i$ is a complex period, then $\sigma(\mathbf{p}) \neq 0$, so there exists t such that $\mathbf{q}(j) < t\sigma(\mathbf{p})$ for all $1 \leq j \leq s$. Since P_i is finite, we can choose the same t for all $(\mathbf{p}; \mathbf{q}) \in P_i$. If $(\mathbf{a}_1; \mathbf{b}_1), (\mathbf{a}_2; \mathbf{b}_2) \in \mathbb{N}_0^{r+s}$ satisfy $\mathbf{b}_k(j) < t\sigma(\mathbf{a}_k)$ $(k = 1, 2)$, then $(\mathbf{b}_1 + \mathbf{b}_2)(j) < t\sigma(\mathbf{a}_1 + \mathbf{a}_2)$. Thus there is a constant $q \in \mathbb{N}_0$, which can be taken to be $\max\{\mathbf{c}_i(j) \mid 1 \leq j \leq r\}$, such that if

$(\mathbf{a};\mathbf{b}) \in L_i$ can be expressed using only complex periods, then $\mathbf{b}(j) < t\sigma(\mathbf{a}) + q$ for all $1 \leq j \leq s$.

Now let $C \in \mathbb{N}$ be twice the maximum of all of the constants t, q that arise for all L_i. Then, for any $(\mathbf{a};\mathbf{b}) \in L$ which can be expressed using only complex periods, $\mathbf{b}(j) < C\sigma(\mathbf{a})$ for all $1 \leq j \leq s$. □

Proposition 2.6. *Let $L \subseteq \mathbb{N}_0^{r+s}$ for some $r, s \in \mathbb{N}$. Let $f : \mathbb{N} \to \mathbb{N}$ be an unbounded function and suppose that, for every $k \in \mathbb{N}$, there exists $\mathbf{a} \in \mathbb{N}_0^r \setminus \{\mathbf{0}\}$ such that the following hold:*

(i) There exists $\mathbf{b} \in \mathbb{N}_0^s$ such that $(\mathbf{a};\mathbf{b}) \in L$.

(ii) If $(\mathbf{a};\mathbf{b}) \in L$, then $\mathbf{b}(j) \geq k\sigma(\mathbf{a})$ for $1 \leq j \leq s$.

(iii) If $(\mathbf{a};\mathbf{b}), (\mathbf{a};\mathbf{b}') \in L$ with $\mathbf{b} \neq \mathbf{b}'$, then $|\mathbf{b}(j) - \mathbf{b}'(j)| \geq f(k)$ for some $1 \leq j \leq s$.

Then L is not an intersection of finitely many semilinear sets.

Proof. By Corollary 2.4, it suffices to show that L is not a semilinear set.

Let L be as in the statement of the proposition and suppose $L = L_1 \cup \ldots \cup L_n$, where each $L_i = L(\mathbf{c}_i; P_i)$ is a linear subset of $\mathbb{N}_0^{r+s}$.

By Lemma 2.5, there exists a constant $C \in \mathbb{N}$ such that if $(\mathbf{a};\mathbf{b}) \in L$ can be expressed using only complex periods in some L_i, then $\mathbf{b}(j) < C\sigma(\mathbf{a})$ for all $1 \leq j \leq s$.

Choose $k > C$, and suppose $\mathbf{a}$ satisfies the hypotheses of the proposition with respect to k. If $(\mathbf{a};\mathbf{b}) \in L$, then $(\mathbf{a};\mathbf{b})$ cannot be expressed using only complex periods, so some P_i must contain a simple period $(\mathbf{0};\mathbf{v})$ with $\mathbf{v}$ non-zero. But then $(\mathbf{a};\mathbf{b}+\mathbf{v}) \in L_i \subseteq L$ and so, for some $1 \leq j \leq s$,

$$|\mathbf{v}(j)| = |(\mathbf{b}+\mathbf{v})(j) - \mathbf{b}(j)| \geq f(k).$$

So for all $k > C$, there is a non-zero simple period $\mathbf{v}_k$ in $\cup_{i=1}^{n} P_i$, with some component of $\mathbf{v}_k$ being at least $f(k)$. But since $\cup_{i=1}^{n} P_i$ is finite and $f(k)$ is unbounded, this is impossible. Thus L is not a semilinear set. $\square$

For $\mathbf{v} = (n_1, \ldots, n_r) \in \mathbb{N}_0^r$ and τ a permutation of $\{1, \ldots, r\}$, we define

$$\tau(\mathbf{v}) = (n_{\tau(1)}, n_{\tau(2)}, \ldots, n_{\tau(r)}).$$

We extend this to a subset L of $\mathbb{N}_0^r$ by defining

$$\tau(L) = \{\tau(\mathbf{v}) \mid \mathbf{v} \in L\}.$$

If L is a linear set $L(\mathbf{c}; \mathbf{p}_1, \ldots, \mathbf{p}_k)$, then $\tau(L) = (\tau(\mathbf{c}); \tau(\mathbf{p}_1), \ldots, \tau(\mathbf{p}_k))$, so the property of being a linear set, or indeed an intersection of k semilinear sets, is preserved by τ.

We shall make significant use of the following corollary to Propositions 1.28 and 2.6 in Chapter 3.

Corollary 2.7. *Let $L \subseteq w_1^* \ldots w_k^*$ be a bounded language over an alphabet X with $w_i \in X^*$, and let τ be a permutation of $\{1, \ldots, k\}$. Let*

$$S = \tau(\Phi(L)) = \{\tau(n_1, \ldots, n_k) \mid n_i \in \mathbb{N}_0 \ (1 \leq i \leq k), w_1^{n_1} \ldots w_k^{n_k} \in L\}.$$

If S satisfies the hypothesis of Proposition 1.28, then L is neither co$\mathcal{CF}$ nor poly-$\mathcal{CF}$. If S satisfies the hypothesis of Proposition 2.6, then L is not poly-$\mathcal{CF}$.

Proof. Since any set satisfying the hypothesis of Proposition 1.28 also satisfies the hypothesis of Proposition 2.6, and since τ preserves semilinearity, this follows immediately from Propositions 1.28 and 2.6, Theorem 1.15 and Corollary 2.1. $\square$

2.4 Dimension of linear sets

If V is a subspace of a vector space W with $\dim(V) < \dim(W)$, then the *dimension* of a coset of V in W is defined to be the dimension of V. The *dimension* of a linear set L is defined to be the dimension of $L^{\mathbb{Q}}$ or, equivalently, the dimension of the vector space over $\mathbb{Q}$ spanned by the periods of L.

We record here a result about the dimension of linear sets which will be useful later. This is a known result, but we provide a proof, as the one given in [15] is incorrect. We make use of a result about intersections of cosets of vector spaces, which is an easy corollary of the following lemma. (Again, these are known results.)

Lemma 2.8. *Let H_1 and H_2 be subgroups of a group G. Then, for any $g_1, g_2 \in G$, if $H_1g_1 \cap H_2g_2$ is non-empty, then it is a coset of $H_1 \cap H_2$.*

Proof. Suppose $H_1g_1 \cap H_2g_2 \neq \emptyset$ and fix an arbitrary $x \in H_1g_1 \cap H_2g_2$. We will show that $H_1g_1 \cap H_2g_2 = (H_1 \cap H_2)x$.

Write $x = h_1g_1 = h_2g_2$, where $h_i \in H_i$ for $i = 1, 2$. Now for any $y \in H_1g_1 \cap H_2g_2$, write $y = h_1'g_1 = h_2'g_2$, where $h_i' \in H_i$. Then, for $i = 1, 2$,

$$y = h_i'h_i^{-1}h_ig_i = h_i'h_i^{-1}x \in H_ix.$$

But then $h_1'h_1^{-1}x = h_2'h_2^{-1}x$, so $h_1'h_1^{-1} = h_2'h_2^{-1}$. Thus $h_1'h_1^{-1} \in H_1 \cap H_2$ and so $y = h_1'h_1^{-1}x \in (H_1 \cap H_2)x$. Hence $H_1g_1 \cap H_2g_2 \subseteq (H_1 \cap H_2)x$.

Conversely, for any $h \in H_1 \cap H_2$, we have $hx = hh_ig_i \in H_ig_i$ for $i = 1, 2$, so $hx \in H_1g_1 \cap H_2g_2$. Hence $(H_1 \cap H_2)x \subseteq H_1g_1 \cap H_2g_2$, concluding the proof. $\square$

Corollary 2.9. *Let M be a vector space and let $V, W \subseteq M$ be subspaces of dimension n and $\mathbf{v}, \mathbf{w} \in M$. If the cosets $V + \mathbf{v}$ and $W + \mathbf{w}$ are not equal, then their intersection has dimension at most $n - 1$.*

Proof. Suppose $V + \mathbf{v} \neq W + \mathbf{w}$. By Lemma 2.8, if $(V + \mathbf{v}) \cap (W + \mathbf{w})$ is non-empty, then it is a coset of $V \cap W$. Since V and W are n-dimensional vector spaces, the dimension of $V \cap W$ is at most $n - 1$ unless $V = W$. But two cosets of the same subspace are either disjoint or equal. Thus either $(V + \mathbf{v}) \cap (W + \mathbf{w}) = \emptyset$, or $V \neq W$ and $(V + \mathbf{v}) \cap (W + \mathbf{w})$ is a coset of $V \cap W$ and hence has dimension at most $n - 1$. $\square$

Proposition 2.10. *A linear set of dimension $n + 1$ cannot be expressed as a union of finitely many linear sets of dimension n or less.*

Proof. A 1-dimensional linear set contains infinitely many points, and thus cannot be expressed as a union of finitely many 0-dimensional linear sets, which are points.

Assume that for all $k \leq n$, a linear set of dimension k cannot be expressed as a union of finitely many linear sets of dimension $k - 1$ or less.

Let L be a linear set of dimension $n + 1$, having constant vector $\mathbf{c}$ and periods $P = \{\mathbf{p}_0, \ldots, \mathbf{p}_r\}$, where $\{\mathbf{p}_0, \ldots, \mathbf{p}_n\}$ is a maximal linearly independent subset of P over $\mathbb{Q}$. Suppose $L = L_1 \cup \ldots \cup L_m$, where each L_i is a linear set of dimension n or less. For $j \in \mathbb{N}_0$, let

$$M_j = L(\mathbf{c} + j\mathbf{p}_0; \mathbf{p}_1, \ldots, \mathbf{p}_n).$$

Then the M_j are pairwise disjoint n-dimensional linear subsets of L.

Let $\mathcal{L}_i = L_i^{\mathbb{Q}}$ and $\mathcal{M}_j = M_j^{\mathbb{Q}}$. The $\mathcal{M}_j$ are pairwise disjoint; so, if $\mathcal{L}_i \subset \mathcal{M}_j$, then $\mathcal{L}_i \cap \mathcal{M}_{j'} = \emptyset$ for all $j' \neq j$.

If $\mathcal{L}_i \not\subseteq \mathcal{M}_j$, then $\dim(\mathcal{L}_i \cap \mathcal{M}_j) \leq n - 1$, since $\mathcal{L}_i$ and $\mathcal{M}_j$ both have dimension at most n and are not equal. By Lemma 2.2, $L_i \cap M_j$ is a semilinear set, and since $L_i \cap M_j \subseteq \mathcal{L}_i \cap \mathcal{M}_j$, its linear subsets can have dimension at most $n - 1$.

Since there are infinitely many $\mathcal{M}_j$ and only finitely many $\mathcal{L}_i$, some $\mathcal{M}_J$ does not contain any $\mathcal{L}_i$. But then $M_J = L \cap M_J = (L_1 \cap M_J) \cup \ldots \cup (L_n \cap M_J)$, which is a union of finitely many linear sets of dimension $n-1$ or less, contradicting the induction hypothesis. $\square$

2.5 The languages $L^{(k)}$

A $(k-1)$-$\mathcal{CF}$ language is clearly also n-$\mathcal{CF}$ for all $n \geq k$. In [15], Liu and Weiner showed that the class of k-$\mathcal{CF}$ languages properly contains the class of $(k-1)$-$\mathcal{CF}$ languages, thus exhibiting an infinite heirarchy of languages in between the context-free and context-sensitive languages. The *context-sensitive* languages are those which can be recognised by a non-deterministic Turing machine using linear space (see [12, Section 9.3]). They include the context-free languages and are closed under intersection, thus they also include the poly-context-free languages.

There are some problems with Liu and Weiner's proof, particularly in the proof of their Theorem 10. In this section, we provide an improved proof.

Following Liu and Weiner, we define a sequence of languages $L^{(k)}$ and corresponding subsets $S^{(k)}$ of $\mathbb{N}_0^{2k}$.

For $k \in \mathbb{N}$, let $a_1, \ldots, a_{2k}$ be $2k$ distinct symbols, and define the language

$$L^{(k)} = \{a_1^{n_1} \ldots a_k^{n_k} a_{n+1}^{n_1} \ldots a_{2k}^{n_k} \mid n_i \in \mathbb{N}_0\}.$$

Define $S^{(k)}$ to be the commutative image of $L^{(k)}$. That is,

$$S^{(k)} = \{v \in \mathbb{N}_0^{(2k)} \mid v(i) = v(k+i)\ (1 \leq i \leq k)\}.$$

The following lemma gives a condition which implies a linear set is not an intersection of $k-1$ stratified semilinear sets. The proof is assembled primarily from the proof of Lemma 4 in [15], but the result is stated differently here because it will also be useful in proving a generalisation of Liu and Weiner's result.

Lemma 2.11. *Let $S = L(\mathbf{0}; P)$ be a k-dimensional linear subset of $\mathbb{N}_0^r$ such that P is linearly independent over $\mathbb{Q}$. Suppose that any subset of S which can be expressed as an intersection of $k-1$ stratified linear sets with constant vector zero has dimension at most $k-1$. Then S is not an intersection of $k-1$ stratified semilinear sets.*

Proof. If S is an intersection of $k-1$ stratified semilinear sets, then S is a finite union of intersections of $k-1$ stratified linear sets.

Let $L = \bigcap_{i=1}^{k-1} L_i$ be a subset of S with each $L_i = L(\mathbf{c}_i; \mathbf{p}_{i1}, \ldots, \mathbf{p}_{im_i})$ a stratified linear set. Let $M = \bigcap_{i=1}^{k-1} L_i^{\mathbf{0}}$ and write $M = L(\mathbf{0}; \mathbf{p}_1, \ldots, \mathbf{p}_m)$. By Proposition 2.3, there exists a finite subset C of $\mathbb{N}_0^r$ such that

$$L = \bigcup_{\mathbf{c}_i \in C} L(\mathbf{c}_i; \mathbf{p}_1, \ldots, \mathbf{p}_m).$$

For any $\mathbf{c}, \mathbf{p} \in \mathbb{N}_0^r$ such that $\mathbf{c} + n\mathbf{p} \in L$ for all $n \in \mathbb{N}_0$, we have $\mathbf{p} \in L$, since P is linearly independent over $\mathbb{Q}$. Thus $M \subseteq S$, since $L(\mathbf{c}_1; \mathbf{p}_1, \ldots, \mathbf{p}_m) \subseteq S$.

Since $M \subseteq S$ is an intersection of $k-1$ stratified linear sets with constant zero, M has dimension at most $k-1$ by the hypothesis of the lemma. Each $L(\mathbf{c}_i; \mathbf{p}_1, \ldots, \mathbf{p}_m)$ is a coset of M and thus has the same dimension as M. Thus L is a union of finitely many linear sets of dimension at most $k-1$. This implies that S itself is a union of finitely many linear sets of dimension at most $k-1$, but by Proposition 2.10, this cannot happen since $\dim(S) = k$. $\square$

2.5.1 The new part of the proof

This subsection contains a new proof of the result which is Theorem 10 in [15], namely that $S^{(k)}$ satisfies the hypothesis of Lemma 2.11. Since the proof turned out be rather long, we break most of it up into three lemmas, which then come together to give a relatively simple proof of the proposition itself (which here is Proposition 2.15).

Lemma 2.12. *Let $S = L_1 \cap \ldots \cap L_k$, where each L_i is a linear subset of $\mathbb{N}_0^r$ with constant vector zero and periods $P_i = \{\mathbf{p}_{i1}, \ldots, \mathbf{p}_{im_i}\}$. For each $1 \leq i \leq k$, let $\mathcal{L}_i = L_i^{\mathbb{Q}}$. If $\dim(S) < \dim(\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k)$, then there exist $1 \leq i \leq k$, $1 \leq j \leq m_i$, such that removing $\mathbf{p}_{ij}$ from P_i does not change the set S.*

Proof. Suppose that $\dim(S) < \dim(\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k)$ and that, for all $1 \leq i \leq k$, removing any $\mathbf{p}_{ij}$ from P_i changes the set S. Then, for all $1 \leq i \leq k$, $1 \leq j \leq m_i$, there must exist $\mathbf{v}_{ij} = \alpha^j_{i1}\mathbf{p}_{i1} + \ldots + \alpha^j_{im_i}\mathbf{p}_{im_i} \in S$ with $\alpha^j_{ij} \geq 1$.

Let $\{\mathbf{q}_1, \ldots, \mathbf{q}_s\}$ be a basis for $\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k$. Since $\mathbf{q}_1, \ldots, \mathbf{q}_s \in \mathcal{L}_i$ for all $1 \leq i \leq k$, we can write $\mathbf{q}_l = \sum_{j=1}^{m_i} \beta^l_{ij}\mathbf{p}_{ij}$, where $\beta^l_{ij} \in \mathbb{Q}$. Now for $1 \leq i \leq k$, $1 \leq j \leq m_i$, let $c_{ij} = \min\{\beta^l_{ij} \mid 1 \leq l \leq s\}$, and let

$$\Lambda_i = \{j \mid 1 \leq j \leq m_i,\ c_{ij} < 0\}.$$

Then, if $\mathbf{w}_i := \sum_{j \in \Lambda_i} -c_{ij}\mathbf{v}_{ij}$, we have $\mathbf{w}_i \in S$, since $\mathbf{v}_{ij} \in S$ and $-c_{ij} \in \mathbb{N}$ for all $j \in \Lambda_i$. Each $\mathbf{w}_i$ can thus be expressed in L_i as $\sum_{j=1}^{m_i} \gamma_{ij}\mathbf{p}_{ij}$, where

$$\gamma_{ij} = \sum_{j' \in \Lambda_i} -c_{ij'}\alpha^{j'}_{ij}.$$

Also, since $\mathbf{w}_i$ is in S, it also has an expression $\mathbf{w}_i = \sum_{j=1}^{m_{i'}} \gamma^i_{i'j}\mathbf{p}_{i'j}$, for all $i' \neq i$ in $\{1, \ldots, k\}$, where $\gamma^i_{i'j} \in \mathbb{N}_0$. For convenience, let $\gamma^i_{ij} = \gamma_{ij}$.

Let $\mathbf{w} = \sum_{i=1}^{k} \mathbf{w}_i$. Then $\mathbf{w} \in S$ and, for each $1 \leq i \leq k$, we can write

$$\mathbf{w} = \sum_{i'=1}^{k} \sum_{j=1}^{m_i} \gamma^{i'}_{ij}\mathbf{p}_{ij}.$$

For all $j \in \Lambda_i$, the coefficient of $\mathbf{p}_{ij}$ in this expression for $\mathbf{w}$ is

$$\sum_{i'=1}^{k} \gamma^{i'}_{ij} \geq \gamma^i_{ij} = \sum_{j' \in \Lambda_i} -c_{ij'}\alpha^{j'}_{ij} \geq -c_{ij}\alpha^j_{ij} \geq -c_{ij},$$

since $\alpha^j_{ij} \geq 1$. Thus we have shown that for each $1 \leq i \leq k$, we can express $\mathbf{w}$ in the form $\sum_{j=1}^{m_i} a_{ij}\mathbf{p}_{ij}$, where $a_{ij} \geq -c_{ij}$ for all $j \in \Lambda_i$.

For any $\mathbf{q}_l$ in the basis for $\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k$, and any $1 \leq i \leq k$, we have

$$\mathbf{w} + \mathbf{q}_l = \sum_{j=1}^{m_i} (a_{ij} + \beta_{ij}^l)\mathbf{p}_{ij} \in L_i,$$

since $a_{ij} + \beta_{ij}^l \geq a_{ij} + c_{ij} \geq 0$ for all $j \in \Lambda_i$, and $c_{ij} \geq 0$ for $j \notin \Lambda_i$. Thus $\mathbf{w} + \mathbf{q}_l \in S$ for all $1 \leq l \leq s$. Let $M = \{\mathbf{w}, \mathbf{w} + \mathbf{q}_1, \ldots, \mathbf{w} + \mathbf{q}_s\} \subset S$. Then $\mathbf{q}_1, \ldots, \mathbf{q}_s$ are in the subspace of $\mathbb{Q}^r$ generated by M, which is contained in $S^{\mathbb{Q}}$. Since $\{\mathbf{q}_1, \ldots, \mathbf{q}_s\}$ is a basis for $\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k$, it is a linearly independent set over $\mathbb{Q}$. Thus $S^{\mathbb{Q}}$ has at least s linearly dependent elements, contradicting

$$\dim(S) = \dim(S^{\mathbb{Q}}) < \dim(\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_k) = s. \qquad \square$$

For a stratified linear set $L \subseteq \mathbb{N}_0^r$, let ρ_L be the relation on $\{1, \ldots, r\}$ given by $m\rho_L n$ if there exist non-zero α, β such that $\alpha e_m + \beta e_n \in P$. Note that ρ_L is symmetric. Now define a relation $\sim_L$ on $\{1, \ldots, r\}$ as follows: $m \sim_L n$ if $m = n$ or there exist $m_1, \ldots, m_t \in \{1, \ldots, 2k\}$ with $m = m_1 \rho_L m_2 \rho_L \ldots \rho_L m_t = n$. This is the reflexive and transitive closure of ρ_L; so $\sim_L$ is an equivalence relation. This gives a partition Π_L of $\{1, \ldots, r\}$ into equivalence classes under $\sim_L$. Note that since L is stratified, if $m_1 < n_1 < m_2 < n_2$, then at most one of $m_1 \rho_L m_2$ and $n_1 \rho_L n_2$ is true. A similar property applies to $\sim_L$:

Lemma 2.13. *Let $L \subseteq \mathbb{N}_0^r$ be a stratified linear set with constant vector zero. If $m_1, n_1, m_2, n_2 \in \{1, \ldots, r\}$ with $m_1 < n_1 < m_2 < n_2$ and $m_1 \not\sim_L n_1$, $m_2 \not\sim_L n_2$, then $m_1 \sim_L m_2$ and $n_1 \sim_L n_2$ cannot both occur.*

Proof. Suppose $m_1 \sim_L m_2$ and $n_1 \sim_L n_2$. Then there exist $i_1, \ldots, i_s, j_1, \ldots, j_t$ in $\{1, \ldots, r\}$ such that

$$m_1 = i_1 \rho_L i_2 \rho_L \ldots \rho_L i_s = m_2$$

and

$$n_1 = j_1 \rho_L j_2 \rho_L \ldots \rho_L j_t = n_2.$$

Let $\Lambda \in \Pi_L$ such that $m_1, m_2 \in \Lambda$. Then, since $m_1 < n_1 < m_2 < n_2$ and $n_1, n_2 \notin \Lambda$, there must exist k such that either $m_1 < j_k < m_2 < j_{k+1}$, or $j_{k+1} < m_1 < j_k < m_2$. Since $j_k \rho_L j_{k+1}$, this forces i_l to lie between j_k and j_{k+1} for all $1 \leq l \leq s$. But either $m_1 (= i_1)$ or $m_2 (= i_s)$ does not lie between j_k and j_{k+1}, thus we have a contradiction and in fact at most one of $m_1 \sim_L m_2$ or $n_1 \sim_L n_2$ can happen. $\square$

For $\mathcal{L} \subseteq \mathbb{Q}^n$, the *orthogonal complement* of $\mathcal{L}$ is defined to be the set of all vectors in $\mathbb{Q}^n$ which are orthogonal to every element of $\mathcal{L}$. That is,

$$\mathcal{L}^\perp = \{\mathbf{v} \in \mathbb{Q}^n \mid \mathbf{v} \cdot \mathbf{w} = 0 \; \forall \mathbf{w} \in \mathcal{L}\}.$$

The following result gives a relationship between Π_L and the orthogonal complement of $L^{\mathbb{Q}}$.

Lemma 2.14. *Let $L \subseteq \mathbb{N}_0^r$ be a stratified linear set, with $\Pi_L = \{\Lambda_1, \ldots, \Lambda_t\}$, and let $\mathcal{L} = L^{\mathbb{Q}}$. Then $\mathcal{L}^\perp$ has a basis of the form $\{\mathbf{x}_i = \sum_{j \in \Lambda_i} \gamma_j e_j \mid i \in M\}$, where $M \subseteq \{1, \ldots, t\}$. In particular, $\dim(\mathcal{L}^\perp) = |M| \leq t$.*

Proof. Let M be the set of all $i \in \{1, \ldots, t\}$ such that $\mathbf{x}(j) \neq 0$ for some $\mathbf{x} \in \mathcal{L}^\perp$ and $j \in \Lambda_i$. For each $i \in M$, fix some non-zero $\mathbf{x}^{(i)} \in \mathcal{L}^\perp$ with $\mathbf{x}^{(i)}(j) \neq 0$ for some $j \in \Lambda_i$. We can write $\mathbf{x}^{(i)} = \sum_{j=1}^r \gamma_{ij} e_j = \sum_{s=1}^t \mathbf{x}_s^{(i)}$, where $\mathbf{x}_s^{(i)} = \sum_{j \in \Lambda_s} \gamma_{ij} e_j$, since $\{1, \ldots, r\}$ is the disjoint union of $\Lambda_1, \ldots, \Lambda_t$. For $i \in M$, let $\mathbf{x}_i = \mathbf{x}_i^{(i)}$. Then $\{\mathbf{x}_i \mid i \in M\}$ is a linearly independent set, since $\mathbf{x}_i \neq 0$ by the choice of $\mathbf{x}^{(i)}$, and $\mathbf{x}_i(j) = 0$ for all $j \notin \Lambda_i$.

Let P be the set of periods of L, and for $1 \leq i \leq t$, let

$$P_i = \{\alpha_m e_m + \alpha_n e_n \in P \mid m, n \in \Lambda_i\},$$

where one of α_m or α_n may be zero. Then $\{P_1, \ldots, P_t\}$ is a partition of P. Now, if $\mathbf{p} \in P_i$, then $\mathbf{p} \cdot \mathbf{x}_{i'}^{(i)} = 0$ for all $i' \neq i$, since $\mathbf{x}_{i'}^{(i)}(j) = 0$ for all $j \in \Lambda_i$. Thus

$$\mathbf{p} \cdot \mathbf{x}^{(i)} = \mathbf{p} \cdot (\mathbf{x}_1^{(i)} + \ldots + \mathbf{x}_t^{(i)}) = \mathbf{p} \cdot \mathbf{x}_i^{(i)} = \mathbf{p} \cdot \mathbf{x}_i.$$

But $\mathbf{x}^{(i)} \in \mathcal{L}^\perp$, thus $\mathbf{p} \cdot \mathbf{x}_i = 0$. Since also $\mathbf{p} \cdot \mathbf{x}_i = 0$ for all $\mathbf{p} \in P_{i'}$ with $i' \neq i$, we have $\mathbf{p} \cdot \mathbf{x}_i = 0$ for all $\mathbf{p} \in P$ and hence $\mathbf{x}_i \in \mathcal{L}^\perp$, for all $i \in M$.

It remains to show that $\{\mathbf{x}_i \mid i \in M\}$ spans $\mathcal{L}^\perp$. Recall that $\mathbf{x}_i = \sum_{j \in \Lambda_i} \gamma_{ij} e_j$. First we show that $\gamma_{ij} \neq 0$ for all $i \in M$, $j \in \Lambda_i$. For $i \in M$, certainly $\gamma_{im} \neq 0$ for some $m \in \Lambda_i$, since $\mathbf{x}_i \neq \mathbf{0}$. For any $n \in \Lambda_i$ there exist $m_1, \ldots, m_l \in \Lambda_i$ such that $m = m_1 \rho_L m_2 \rho_L \ldots \rho_L m_l = n$, which implies the existence of periods $\alpha_{m_1} e_{m_1} + \alpha_{m_2} e_{m_2}, \ldots, \alpha_{m_{l-1}} e_{m_{l-1}} + \alpha_{m_l} e_{m_l} \in P_i$ with non-zero α_{m_j} for all $1 \leq j \leq l$. Now

$$\mathbf{x}_i \cdot (\alpha_{m_j} e_{m_j} + \alpha_{m_{j+1}} e_{m_{j+1}}) = \gamma_{im_j} \alpha_{m_j} + \gamma_{im_{j+1}} \alpha_{m_{j+1}} = 0$$

for all $1 \leq j \leq l-1$, since $\mathbf{x}_i \in \mathcal{L}^\perp$. Thus $\gamma_{im_{j+1}} = -\gamma_{im_j} \frac{\alpha_{im_j}}{\alpha_{im_{j+1}}}$ and so by induction $\gamma_{in} = \gamma_{im_l} \neq 0$, since $\gamma_{im} = \gamma_{i1} \neq 0$. Moreover, for all $n \in \Lambda_i$, γ_{in} is uniquely determined by γ_{im}. (If two different paths between m and n gave different values for γ_{in}, then our non-zero $\mathbf{x}_i \in \mathcal{L}^\perp$ could not exist.)

Finally, let $\mathbf{y} \in \mathcal{L}^\perp$ and write $\mathbf{y} = \sum_{j=1}^{r} c_j e_j = \sum_{i=1}^{t} \mathbf{y}_i$, where $\mathbf{y}_i = \sum_{j \in \Lambda_i} c_j e_j$. If $\mathbf{y}_i \neq \mathbf{0}$, then choose $j \in \Lambda_i$ with $c_j \neq 0$. Since $\gamma_{ij} \neq 0$, we can write $c_j = q\gamma_{ij}$, where $q \in \mathbb{Q}$. By exactly the same argument as we used for $\mathbf{x}_i$, we can conclude that $\mathbf{p} \cdot \mathbf{y}_i = 0$ for all $\mathbf{p} \in P$. Now for any $\alpha_j e_j + \alpha_{j'} e_{j'} \in P_i$, we have

$$\mathbf{y}_i \cdot (\alpha_j e_j + \alpha_{j'} e_{j'}) = c_j \alpha_j + c_{j'} \alpha_{j'},$$

thus $c_{j'} = -c_j \frac{\alpha_j}{\alpha_{j'}}$. But also $\gamma_{ij'} = -\gamma_{ij} \frac{\alpha_j}{\alpha_{j'}}$. Thus $c_{j'} = -q\gamma_{ij} \frac{\alpha_j}{\alpha_{j'}} = q\gamma_{ij'}$, and we can extend this to show that $c_n = q\gamma_{in}$ for all $n \in \Lambda_i$, thus $\mathbf{y}_i = q\mathbf{x}_i$. Since this applies to all $i \in M$ with $\mathbf{y}_i \neq \mathbf{0}$, we can conclude that $\mathbf{y}$ is a linear combination of the elements of $\{\mathbf{x}_i \mid i \in M\}$, and thus this set spans $\mathcal{L}^\perp$. □

We are now ready to prove Theorem 10 of [15].

Proposition 2.15. *For $1 \leq i \leq k-1$, let L_i be a stratified linear set with constant vector zero, and let $L_1 \cap \ldots \cap L_{k-1} = S \subseteq S^{(k)}$. Then S is a linear set of dimension at most $k-1$.*

Proof. S is a linear set with constant vector zero by Proposition 2.3. Let $\mathcal{L}_i = L^{\mathbb{Q}}$ for all $1 \leq i \leq k-1$, and let $\mathcal{S} = \mathcal{L}_1 \cap \ldots \cap \mathcal{L}_{k-1}$. By Lemma 2.12 we can assume that $\dim(\mathcal{S}) = \dim(S)$. Since $S \subseteq \mathcal{S}$, this implies that any maximal linearly independent subset of the periods of S is a basis for $\mathcal{S}$. Thus, since $\mathbf{v}(i) = \mathbf{v}(k+i)$ for all $\mathbf{v} \in S$, we also have $\mathbf{v}(i) = \mathbf{v}(k+i)$ for all $\mathbf{v} \in \mathcal{S}$. For all $1 \leq i \leq k$, we have $e_i - e_{k+i} \in \mathcal{S}^{\perp}$, since $\mathbf{v} \cdot (e_i - e_{k+i}) = \mathbf{v}(i) - \mathbf{v}(k+i) = 0$ for all $\mathbf{v} \in \mathcal{S}$.

Assume $\{e_i - e_{k+i} \mid 1 \leq i \leq k\}$ spans $\mathcal{S}^{\perp}$, since otherwise $\dim(\mathcal{S}^{\perp}) \geq k+1$ and thus $\dim(\mathcal{S}) \leq 2k - (k+1) = k-1$.

If $\mathcal{L}_i^{\perp} \neq \{\mathbf{0}\}$, let $\Pi_{L_i} = \{\Lambda_1, \ldots, \Lambda_t\}$. Then, by Lemma 2.14, $\mathcal{L}_i^{\perp}$ has a basis of the form $\{\mathbf{x}_s \mid s \in M\}$, where $M \subseteq \{1, \ldots, t\}$ and $\mathbf{x}_s = \sum_{j \in \Lambda_s} \gamma_j e_j$. If $s \in M$ then, since $\mathbf{x}_s \in \mathcal{L}_i^{\perp} \subseteq \mathcal{L}^{\perp}$, we can write

$$\mathbf{x}_s = \sum_{j=1}^{k} \gamma_j (e_j - e_{k+j}) = \sum_{j \in \Gamma_s} \gamma_j (e_j - e_{k+j}),$$

where $\Gamma_s = \Lambda_s \cap \{1, \ldots, k\}$. Certainly some γ_j is non-zero, implying $j, (k+j) \in \Lambda_s$. Thus if $s, s' \in M$ then we would have some $j, (k+j) \in \Lambda_s$, $l, (k+l) \in \Lambda_{s'}$. But either $j < l < (k+j) < (k+l)$ or $l < j < (k+l) < (k+j)$, thus this would contradict Lemma 2.13. Therefore at most one $s \in M$, and so $\dim(\mathcal{L}_i^{\perp}) \leq 1$. This holds for all $1 \leq i \leq k-1$.

But if each $\mathcal{L}_i^{\perp}$ is at most one dimensional, then since $\mathcal{S}^{\perp} = \mathcal{L}_1^{\perp} + \ldots + \mathcal{L}_{k-1}^{\perp}$, $\dim(\mathcal{S}^{\perp})$ cannot exceed $k-1$, contradicting the fact that $e_j - e_{k+j} \in \mathcal{S}^{\perp}$ for all $1 \leq j \leq k$. Thus our assumption that $\{e_j - e_{k+j} \mid 1 \leq j \leq k\}$ spans $\mathcal{S}^{\perp}$ was false, and so in fact $\dim(S) \leq k-1$. □

2.5.2 The rest of the proof

Theorem 2.16. *$L^{(k)}$ is k-$\mathcal{CF}$, but not $(k-1)$-$\mathcal{CF}$. Thus the class of k-$\mathcal{CF}$ languages properly contains the class of $(k-1)$-$\mathcal{CF}$ languages.*

Proof. By Corollary 2.1, it suffices to show that $S^{(k)}$ is an intersection of k but not $k-1$ stratified semilinear sets. For $1 \leq i \leq k$, define

$$S_i = \mathsf{span}\left\{e_i + e_{k+i}, e_j \mid 1 \leq j \leq 2k, j \notin \{i, k+i\}\right\}.$$

Then each S_i is a stratified linear set and $S^{(k)} = \bigcap_{i=1}^{k} S_i$, so $S^{(k)}$ is an intersection of k stratified semilinear sets.

Also, $S^{(k)}$ has constant vector zero and dimension k, since $\{e_i + e_{k+i} \mid 1 \leq i \leq k\}$ is a linearly independent subset which spans $S^{(k)}$. Hence, by Proposition 2.15, $S^{(k)}$ satisfies the hypothesis of Lemma 2.11, so cannot be expressed as an intersection of $k-1$ stratified semilinear sets. □

2.6 The languages $L^{(n,k)}$

We can extend Theorem 2.16 to a larger, but very similar, class of languages. The extended result will be used to prove that certain groups, for example the restricted standard wreath products $C_p \wr \mathbb{Z}$ (for any $p > 1$), are not poly-$\mathcal{CF}$.

For each $n, k \in \mathbb{N}$, let $a_1, a_2, \ldots, a_{2nk}$ be $2nk$ distinct symbols and define

$$\begin{aligned} L^{(n,k)} = \{a_1^{m_1} a_2^{m_2} \ldots a_{2nk}^{m_{2nk}} \mid \quad & m_i \in \mathbb{N}_0, m_i = m_{nk+i} \ (1 \leq i \leq nk), \\ & m_{nj+1} = m_{nj+l} \ (0 \leq j \leq k-1,\ 2 \leq l \leq n)\}. \end{aligned}$$

For example, $L^{(2,2)} = \{a_1^m a_2^m a_3^n a_4^n a_5^m a_6^m a_7^n a_8^n \mid m, n \in \mathbb{N}_0\}$.

Define $S^{(n,k)}$ to be the commutative image of $L^{(n,k)}$. Then

$$\begin{aligned} S^{(n,k)} = \{\mathbf{v} \in \mathbb{N}_0^{2nk} \mid \quad & \mathbf{v}(i) = \mathbf{v}(i+nk) \ (1 \leq i \leq nk), \\ & \mathbf{v}(nj+1) = \mathbf{v}\,(nj+l) \ \ (0 \leq j \leq k-1,\ 2 \leq l \leq n)\}. \end{aligned}$$

These sets are like $S^{(k)}$, except with each entry being repeated n times. Thus $S^{(1,k)}$ is just $S^{(k)}$. For any $n \in \mathbb{N}$, the set $S^{(n,k)}$ has dimension k, so it is not surprising that the following result does not depend on n.

Proposition 2.17. *For $1 \leq i \leq k-1$, let L_i be a stratified linear set with constant vector zero, and let $L_1 \cap \ldots \cap L_{k-1} = S \subseteq S^{(n,k)}$. Then S is a linear set of dimension at most $k-1$.*

Proof. The proof follows the idea of the proof of Proposition 2.15, but is a good deal more complicated.

S is a linear set with constant vector zero by Proposition 2.3. Let $\mathcal{L}_i = L_i^{\mathbb{Q}}$ for $1 \leq i \leq k-1$, and let $\mathcal{S} = \mathcal{L}_1 \cap \ldots \cap \mathcal{L}_{k-1}$. By Lemma 2.12, we can assume that $\dim(\mathcal{S}) = \dim(S)$. Since $S \subseteq \mathcal{S}$, this implies that any maximal linearly independent subset of the periods of S is a basis for $\mathcal{S}$. Thus, since $\mathbf{v}(i) = \mathbf{v}(nk+i)$ for all $\mathbf{v} \in S$, we also have $\mathbf{v}(i) = \mathbf{v}(nk+i)$ for all $\mathbf{v} \in \mathcal{S}$, $1 \leq i \leq nk$. Moreover, for all $\mathbf{v} \in \mathcal{S}$ we have $\mathbf{v}(nj+l) = \mathbf{v}(nj+l+1)$ for all $0 \leq j \leq k-1$, $1 \leq l \leq n-1$.

For all $1 \leq i \leq nk$, we have $e_i - e_{nk+i} \in \mathcal{S}^{\perp}$, since

$$\mathbf{v} \cdot (e_i - e_{nk+i}) = \mathbf{v}(i) - \mathbf{v}(nk+i) = 0$$

for all $\mathbf{v} \in \mathcal{S}$. Similarly, $e_{nj+l} - e_{nj+l+1} \in \mathcal{S}^{\perp}$ for all $0 \leq j \leq k-1$ and $1 \leq l \leq n-1$. Thus we know of $nk + (n-1)k = (2n-1)k$ linearly independent elements of $\mathcal{S}^{\perp}$.

Assume that these $(2n-1)k$ elements form a basis of $\mathcal{S}^{\perp}$, since otherwise we have $\dim(S) = \dim(\mathcal{S}) < 2nk - (2n-1)k = k$, as we require. We will now derive a contradiction, using the fact that $\mathcal{S}^{\perp} = (\mathcal{L}_1 \cap \ldots \cap \mathcal{L}_{k-1})^{\perp} = \mathcal{L}_1^{\perp} + \ldots + \mathcal{L}_{k-1}^{\perp}$.

For $0 \leq j \leq k-1$ and $\epsilon \in \{0, 1\}$, define

$$\Delta_j^{\epsilon} = \{n(\epsilon k + j) + l \mid 1 \leq l \leq n\}$$

and $\Delta_j = \Delta_j^0 \cup \Delta_j^1$. Let $\mathcal{S}_j$ be the image of the projection of $\mathcal{S}^{\perp}$ onto the coordinates in Δ_j. Since every vector in the basis of $\mathcal{S}^{\perp}$ above is contained in some $\mathcal{S}_j$, and the Δ_j are disjoint, $\mathcal{S}^{\perp}$ is the direct sum of $\mathcal{S}_0, \ldots, \mathcal{S}_{k-1}$.

Call $\mathbf{x} \in \mathcal{S}^{\perp}$ a *j-bridge* if there exist $l \in \Delta_j^0$, $l' \in \Delta_j^1$ such that $\mathbf{x}(l)$ and $\mathbf{x}(l')$ are both non-zero. By extension, for $\Gamma \subseteq \{0, \ldots, k-1\}$, call $\mathbf{x}$ a *Γ-bridge* if $\mathbf{x}$ is a j-bridge for all $j \in \Gamma$.

For $0 \leq j \leq k-1$, let Ω_j be the $2(n-1)$-dimensional subspace of $\mathcal{S}_j^{\perp}$ generated by

$$\{e_{n(\epsilon k+j)+l} - e_{n(\epsilon k+j)+l+1} \mid \epsilon \in \{0,1\}, 1 \leq l \leq n-1\}$$

and let $\Omega = \Omega_0 + \ldots + \Omega_{k-1}$.

Suppose $\mathbf{x}$ is not a j-bridge for any $0 \leq j \leq k-1$. We will show that $\mathbf{x}$ must be in Ω. Write $\mathbf{x} = \sum_{j=0}^{k-1} \mathbf{y}_j$, where $\mathbf{y}_j \in \mathcal{S}_j$. For $j' \neq j$, all entries of $\mathbf{y}_{j'}$ on Δ_j are zero, so $\mathbf{x}$ is a j-bridge if and only if $\mathbf{y}_j$ is a j-bridge. Thus no $\mathbf{y}_j$ is a j-bridge, since $\mathbf{x}$ is not a j-bridge for any j.

Let $j \in \{0, \ldots, k-1\}$. Since $\mathbf{y}_j$ is in $\mathcal{S}_j$ and is not a j-bridge, its non-zero coordinates are either all in Δ_j^0 or all in Δ_j^1. For any $\mathbf{v} \in \mathcal{S}^{\perp}$, the sum of the entries of $\mathbf{v}$ is zero, as can be seen by considering the basis vectors of $\mathcal{S}^{\perp}$. Thus the subspace of $\mathcal{S}^{\perp}$ consisting of vectors whose non-zero coordinates all lie in Δ_j^{ϵ} is spanned by

$$\{e_{n(\epsilon k+j)+l} - e_{n(\epsilon k+j)+l+1} \mid 1 \leq l \leq n-1\} \subseteq \Omega_j,$$

for $\epsilon \in \{0,1\}$. Hence $\mathbf{y}_j \in \Omega_j$, and since this applies for all $0 \leq j \leq k-1$, we conclude that $\mathbf{x} = \sum_{j=0}^{k-1} \mathbf{y}_j \in \Omega$.

If $\mathcal{L}_i^{\perp} \neq \{\mathbf{0}\}$, let $\Pi_{L_i} = \{\Lambda_1, \ldots, \Lambda_t\}$. Then, by Lemma 2.14, $\mathcal{L}_i^{\perp}$ has a basis of the form $B_i = \{\mathbf{x}_s \mid s \in M\}$, where $M \subseteq \{1, \ldots, t\}$ and $\mathbf{x}_s = \sum_{j \in \Lambda_s} \gamma_{sj} e_j$. Note that if $\mathbf{x}_s$ is a j-bridge and $s' \neq s$, $j' \neq j$, then $\mathbf{x}_{s'}$ cannot be a j'-bridge, since this would imply the existence of $l_1, l_2, l_1', l_2' \in \{1, \ldots, n\}$ such that

$$nj + l_1, n(k+j) + l_2 \in \Lambda_s, \qquad nj' + l_1', n(k+j') + l_2' \in \Lambda_{s'},$$

but without loss of generality $j < j'$ and hence

$$nj + l_1 < nj' + l_1' < n(k+j) + l_2 < n(k+j') + l_2',$$

contradicting Lemma 2.13.

If B_i contains no Γ-bridges for any non-empty Γ, then in particular it contains no j-bridges for any $0 \leq j \leq k-1$, and so every $\mathbf{x}_s \in B_i$ is in Ω, hence $\mathcal{L}_i^{\perp} \subseteq \Omega$.

If the largest Γ such that $\mathbf{x}_s$ is a Γ-bridge is a singleton $\{j\}$, then B_i may contain other j-bridges. For example, we could have $\mathbf{x}_s = e_1 - e_{nk+2}$ and $\mathbf{x}_{s'} = e_2 - e_{nk+1}$. But, as already observed, for $j' \neq j$, B_i contains no j'-bridges.

If Γ has at least two elements and $\mathbf{x}_s$ is a Γ-bridge, then B_i contains no other Γ'-bridges, even for $\Gamma' = \Gamma$, since this would again imply a situation contradicting Lemma 2.13.

Thus there is at most one $\Gamma \subseteq \{0, \ldots, k-1\}$ such that B_i contains one or more Γ-bridges. If such Γ exists, call it Γ_i.

For each i, we have $\mathcal{L}_i^{\perp} = \mathcal{M}_i + \mathcal{N}_i$, where $\mathcal{M}_i$ is the subspace generated by the Γ_i-bridge(s) and $\mathcal{N}_i$ is the subspace generated by the remaining elements of B_i.

Now consider

$$\mathcal{S}^{\perp} = \mathcal{L}_1^{\perp} + \ldots + \mathcal{L}_{k-1}^{\perp} = \mathcal{M}_1 + \ldots + \mathcal{M}_{k-1} + \mathcal{N}_1 + \ldots + \mathcal{N}_{k-1}.$$

Since the $\mathcal{N}_i$ are generated by elements which are not Γ-bridges for any non-empty Γ, they are all subspaces of Ω. Thus $\mathcal{S}^{\perp} \subseteq \mathcal{M}_1 + \ldots + \mathcal{M}_{k-1} + \Omega$.

If Γ_i contains at least two elements, then B_i has a single Γ_i-bridge, so $\mathcal{M}_i$ has dimension one. If $\Gamma_i = \{j\}$, then even though $\mathcal{M}_i$ can have dimension up to n, $\Omega_j + \mathcal{M}_i$ has to be contained in $\mathcal{S}_j$, so can have dimension at most $2n-1$, which is one more than the dimension of Ω_j. Thus each $\mathcal{M}_i$ contributes at most one extra dimension to the set $\Omega + \mathcal{M}_1 + \ldots + \mathcal{M}_{k-1}$, and so

$$\begin{aligned} \dim(\mathcal{S}^{\perp}) &\leq \dim(\Omega + \mathcal{M}_1 + \ldots + \mathcal{M}_{k-1}) \\ &\leq 2k(n-1) + k - 1 = (2n-1)k - 1, \end{aligned}$$

giving a contradiction. Thus our assumption that $\mathcal{S}^\perp$ was spanned by $(2n-1)k$ elements is incorrect, and so $\dim(\mathcal{S}^\perp) \geq (2n-1)k+1$. This implies that

$$\dim(S) = \dim(\mathcal{S}) \leq 2nk - ((2n-1)k+1) = k-1. \qquad \square$$

Corollary 2.18. *A k-dimensional linear subset of $S^{(n,k)}$ cannot be expressed as an intersection of $k-1$ stratified semilinear sets.*

Proof. Suppose $L \subseteq S^{(n,k)}$ is k-dimensional and can be expressed as an intersection of $k-1$ stratified semilinear sets. Then we can write $L = S_1 \cup \ldots \cup S_l$, where each S_i is an intersection of $k-1$ stratified linear sets. By Proposition 2.3, there exist finite subsets C_i and P_i of $\mathbb{N}_0^{2nk}$ such that $S_i = L(C_i; P_i)$ for $1 \leq i \leq l$. By Proposition 2.10, there must exist $1 \leq i \leq l$ and $\mathbf{c} \in C_i$ such that $L(\mathbf{c}; P_i)$ has dimension k, and hence $L(\mathbf{0}; P_i)$ has dimension k. Writing $S_i = \cap_{i=1}^{k-1} N_i$, where each N_i is a stratified linear set, from Proposition 2.3 we have $L(\mathbf{0}; P_i) = \cap_{i=1}^{k-1} N_i^{\mathbf{0}}$. But $L(\mathbf{0}; P_i)$ is a k-dimensional linear subset of $S^{(n,k)}$ with constant zero, while each $N_i^{\mathbf{0}}$ is a stratified linear set, contradicting Proposition 2.17. $\square$

Theorem 2.19. *For any $k, n \in \mathbb{N}$, the set $S^{(n,k)}$ is not an intersection of $k-1$ stratified semilinear sets, and so the language $L^{(n,k)}$ is not $(k-1)$-$\mathcal{CF}$.*

Proof. Recall from the proof of Proposition 2.17 the notation

$$\Delta_j = \{nj + l \mid 1 \leq l \leq n\} \cup \{n(k+j) + l \mid 1 \leq l \leq n\}.$$

For $0 \leq j \leq k-1$, let $\mathbf{u}_j = \sum_{i \in \Delta_j} e_i$. Then $\{\mathbf{u}_j \mid 0 \leq j \leq k-1\}$ is a linearly independent set which spans $S^{(n,k)}$, so $S^{(n,k)}$ is k-dimensional. Since $S^{(n,k)}$ has constant vector zero, it follows from Lemma 2.11 and Proposition 2.17 that $S^{(n,k)}$ cannot be an intersection of $k-1$ stratified semilinear sets and thus $L^{(n,k)}$ cannot be a $(k-1)$-$\mathcal{CF}$ language. $\square$

Chapter 3

Some results on poly-$\mathcal{CF}$ groups

3.1 An observation and a conjecture

We begin with a simple observation.

Observation 3.1. *The class of poly-$\mathcal{CF}$ groups is closed under taking finite direct products. The direct product of a k_1-$\mathcal{CF}$ group and a k_2-$\mathcal{CF}$ group is $(k_1 + k_2)$-$\mathcal{CF}$.*

Proof. It suffices to prove the second statement.

Let G_i be a k_i-$\mathcal{CF}$ group for $i = 1, 2$. Let $A_{i1}, \ldots, A_{ik_i}$ be pushdown automata with input alphabet X_i such that a word is in $W(G_i, X_i)$ if and only if it is accepted by all A_{ij}. We may assume that X_1 and X_2 are disjoint. Now modify the automata A_{ij} so that their input alphabet is $X = X_1 \cup X_2$, but each A_{1j} ignores the symbols in X_2 and A_{2j} ignores the symbols in X_1. Let $h_1 : X^* \to X_1^*$ be the homomorphism sending every symbol in X_2 to the empty word, and define h_2 similarly. Then a word w in $(X \cup X^{-1})^*$ is accepted by all of the modified automata A_{ij} if and only if $h_i(w) \in W(G_i, X_i)$ for $i = 1, 2$. Thus the intersection of the languages accepted by all the A_{ij} is precisely $W(G_1 \times G_2, X)$, and hence $G_1 \times G_2$ is $(k_1 + k_2)$-$\mathcal{CF}$. $\square$

Since free groups are context-free (Observation 1.25), this implies that a direct product of k free groups is k-$\mathcal{CF}$. Since the k-$\mathcal{CF}$ groups are closed under taking finite index overgroups and finitely generated subgroups, any finitely generated group which is a subgroup of a direct product of k free groups, and any finite index overgroup of such a group, is k-$\mathcal{CF}$. These are the only known k-$\mathcal{CF}$ groups, and we conjecture that they are the only ones.

Conjecture 3.2. *Let G be a finitely generated group. Then G is poly-$\mathcal{CF}$ if and only if G is virtually a finitely generated subgroup of a finite direct product of free groups.*

The rest of this chapter is therefore entirely devoted to proving certain classes of groups to be *not* poly-$\mathcal{CF}$.

Note that the truth of Conjecture 3.2 would imply that the word problem of a poly-$\mathcal{CF}$ group is always an intersection of finitely many deterministic context-free languages. Since the deterministic context-free languages are closed under complementation, this would imply that the co-word problem of a poly-$\mathcal{CF}$ group is a union of finitely many context-free languages, hence itself context-free, and so the poly-$\mathcal{CF}$ groups would be a subclass of the co$\mathcal{CF}$ groups.

3.2 Some groups which are not poly-$\mathcal{CF}$

In this section we present some results which require very little further effort to prove, since we may use Corollary 2.7 and the proofs of Propositions 1.29, 1.30 and 1.31 (Theorems 12, 13 and 16 in [9]).

Proposition 3.3. *A finitely generated nilpotent group is poly-$\mathcal{CF}$ if and only if it is virtually abelian.*

Proof. A finitely generated virtually abelian group is poly-$\mathcal{CF}$ by Observation 3.1. The rest of the proof is copied almost word-for-word from the proof of Theorem 12 in [9].

Assume that G is a finitely generated nilpotent but not virtually abelian group. By Corollary 1.5, G has a torsion-free nilpotent subgroup of finite index, which cannot be virtually abelian since G is not virtually abelian. Furthermore, every non-abelian torsion-free nilpotent group has a subgroup isomorphic to the Heisenberg group

$$H = \langle A, B, C \mid [A, B] = C, [A, C] = [B, C] = 1 \rangle .$$

To see this let A be a non-central element of the second term of the upper central series of G and let B be some element not commuting with A.

Since the poly-$\mathcal{CF}$ groups are closed under taking finitely generated subgroups, it suffices to show that H is not poly-$\mathcal{CF}$. Since $[A^m, B^m] = C^{m^2}$ holds in H for all $m \geq 0$, the commutative image of the intersection of $(A^{-1})^*(B^{-1})^*A^*B^*(C^{-1})^*$ with $W(H, \{A, B, C\})$ satisfies the condition of Proposition 1.28. (Given $k \in \mathbb{N}$, consider $\mathbf{a} = (m, m, m, m)$ with $m \geq 4k$.) Thus H, and hence G, is not poly-$\mathcal{CF}$ by Corollary 2.7. □

Note that this shows that the class of poly-$\mathcal{CF}$ groups is not closed under taking semidirect products, even with context-free top group, since the Heisenberg group is a semidirect product $\mathbb{Z}^2 \rtimes \mathbb{Z}$.

Next we will give an analogue of Proposition 1.30 for poly-$\mathcal{CF}$ groups, but first we need the promised completion of the proof of Proposition 1.30. This uses the following result, known as the Kurosh Subgroup Theorem - a special case of the Grushko-Neumann Theorem for free products. A proof can be found in [16, III.3.6].

Proposition 3.4. *Let G be the free product of groups G_i, $i \in I$, where I is an index set. Let H be a subgroup of G. Then H is the free product of a free group together* *with groups that are conjugates of subgroups of the free factors G_i of G.*

Proposition 3.5. *For $m \in \mathbb{Z} \setminus \{0\}$, the Baumslag-Solitar group $\mathsf{BS}(m, \pm m)$ is virtually a direct product of two free groups and is thus both co$\mathcal{CF}$ and 2-$\mathcal{CF}$.*

Proof. First let $G = \mathsf{BS}(m, m) = \left\langle x, y \mid y^{-1}x^m y = x^m \right\rangle$. Then $x^m \in Z(G)$ and

$$G/\left\langle x^m \right\rangle = \left\langle x, y \mid x^m \right\rangle = C_m * \mathbb{Z}.$$

Let $H/\left\langle x^m \right\rangle$ be the normal closure in $G/\left\langle x^m \right\rangle$ of $\langle y \rangle$. Then

$$\left|G/\left\langle x^m \right\rangle : H/\left\langle x^m \right\rangle\right| = m$$

and hence $|G : H| = m$. Since $H/\left\langle x^m \right\rangle$ does not intersect any conjugate of C_m, by the Kurosh Subgroup Theorem (Proposition 3.4), $H/\left\langle x^m \right\rangle$ is the free product of a free group with conjugates of $\mathbb{Z}$, and is thus free. Also, $H \cong H/\left\langle x^m \right\rangle \times \left\langle x^m \right\rangle$, since $x^m \in Z(G)$. Thus G is virtually a direct product of two free groups.

Now let $G = \mathsf{BS}(m, -m) = \left\langle x, y \mid y^{-1}x^m y = x^{-m} \right\rangle$. Let K be the normal closure in G of $\left\langle x, y^2 \right\rangle$, which has index 2 in G. Setting $a = x$, $b = y^{-1}x^{-1}y$ and $c = y^2$ gives

$$K = \left\langle a, b, c \mid a^m = b^m, [a^m, c] \right\rangle,$$

with $a^m \in Z(K)$, since $(a^m)^b = (b^m)^b = b^m = a^m$. Now take

$$H := K/\left\langle a^m \right\rangle = \left\langle a, b, c \mid a^m = b^m = 1 \right\rangle = C_m * C_m * \mathbb{Z}.$$

Let ϕ be the homomorphism from H to $C_m \times C_m$ given by mapping a onto a generator of the first C_m and b onto a generator of the second C_m, and c onto the identity. Then the intersection of $\ker \phi$ with every conjugate of $\langle a \rangle$ and $\langle b \rangle$ is trivial. Thus by the Kurosh Subgroup Theorem, $\ker \phi$ is a free product of a free group and conjugates of $\mathbb{Z}$, and is hence itself free. Also, $|H : \ker \phi| = |C_m \times C_m| = m^2$. Let K_1 be the preimage of $\ker \phi$ in K. Since $\ker \phi$ is free and $\left\langle a^m \right\rangle \in Z(H)$, K_1 is isomorphic to $\ker \phi \times \left\langle a^m \right\rangle$. Also, K_1 has finite index in K, and hence also in G, since $\ker\phi$ has finite index in $H = K/\langle a^m \rangle$. Thus G is virtually a direct product of two free groups.

Hence G is 2-$\mathcal{CF}$ by Observation 3.1, and co$\mathcal{CF}$ by the fact that the co$\mathcal{CF}$ groups are closed under taking finite direct products [9, Proposition 6]. $\square$

We can now determine which Baumslag-Solitar groups are poly-$\mathcal{CF}$.

Proposition 3.6. *The Baumslag-Solitar group* $\mathsf{BS}(m,n)$ *is poly-*$\mathcal{CF}$ *if and only if* $m = \pm n$.

Proof. Let $G = \mathsf{BS}(m,n) = \langle x, y \mid y^{-1}x^m y = x^n \rangle$, where $m,n \in \mathbb{Z}\setminus\{0\}$. By Proposition 3.5, if $m = \pm m$, then G is poly-$\mathcal{CF}$. The rest of the proof is copied from the proof of Theorem 13 in [9].

We deal here with the case $0 < m < n$, the other cases being similar. Let L be the commutative image of $W(G) \cap (y^{-1})^*(x^{-1})^*y^*x^*$. Then $(k, m^k, k, n^k) \in L$ for all $l \in \mathbb{N}$, and n^k is the only value of x for which $(k, m^k, k, x) \in L$. Moreover, since $n > m$, for every given l there exists k with $l(2k + m^k) \leq n^k$. This simply means that L satisfies the hypothesis of Proposition 1.28, so $W(G)$ is not poly-$\mathcal{CF}$ by Corollary 2.7. $\square$

Proposition 1.31 (Theorem 16 in [9]) also has its analogue for poly-$\mathcal{CF}$ groups. We do not give the proof in full, since the proof of Theorem 16 in [9] is quite long.

Proposition 3.7. *If G is a finitely generated polycyclic group, then G is poly-$\mathcal{CF}$ if and only if G is virtually abelian.*

Proof. The proof of Theorem 16 in [9] shows that if G is a finitely generated polycyclic group which is not virtually abelian, then $W(G)$ can be intersected with a regular language to give a sublanguage satisfying the hypothesis of Proposition 1.28, and so $W(G)$ is neither co$\mathcal{CF}$ nor poly-$\mathcal{CF}$ by Corollary 2.7. $\square$

3.3 Free abelian groups and wreath products

The obvious application of Proposition 2.16 to word problems of groups is to the free abelian groups. This will be useful in proving some other groups not to be poly-$\mathcal{CF}$.

Lemma 3.8. *A free abelian group of rank k is k-$\mathcal{CF}$ but not $(k-1)$-$\mathcal{CF}$.*

Proof. The group $\mathbb{Z}^k$ is a direct product of k free groups, and so is k-$\mathcal{CF}$. Let $\{x_1, \ldots, x_k\}$ be a generating set for $\mathbb{Z}^k$ and let X_i denote the inverse of x_i. Consider $L = W(\mathbb{Z}^k, \{x_1, \ldots, x_k\}) \cap (x_1^* \ldots x_k^* X_1^* \ldots X_k^*)$. This is precisely the language $L^{(k)} = \{x_1^{n_1} \ldots x_k^{n_k} X_1^{n_1} \ldots X_k^{n_k} \mid n_i \in \mathbb{N}_0\}$ defined in Section 2.5. Thus, by Proposition 2.16, L is not $(k-1)$-$\mathcal{CF}$. Since L is the intersection of $W(\mathbb{Z}^k)$ with a regular language, this implies that $\mathbb{Z}^k$ is not $(k-1)$-$\mathcal{CF}$. $\square$

We saw in Section 1.6.3 that the class of co$\mathcal{CF}$ groups is closed under taking restricted standard wreath products with context-free top group (Proposition 1.27). In contrast, we have the following result for poly-$\mathcal{CF}$ groups.

Proposition 3.9. *The restricted standard wreath product $\mathbb{Z} \wr \mathbb{Z}$ is not poly-$\mathcal{CF}$.*

Proof. Since $\mathbb{Z} \wr \mathbb{Z}$ contains free abelian subgroups of rank k for all $k \in \mathbb{N}$, this follows immediately from Lemma 3.8 and the fact that the poly-$\mathcal{CF}$ groups are closed under taking finitely generated subgroups. $\square$

A further result on wreath products will be useful when we come to consider metabelian groups. It is our first application of Theorem 2.19.

Proposition 3.10. *For any $p \in \mathbb{N} \setminus \{1\}$, the restricted standard wreath product $C_p \wr \mathbb{Z}$ is not poly-$\mathcal{CF}$.*

Proof. Let $G = \langle b \rangle \wr \langle a \rangle = C_p \wr \mathbb{Z}$, with $p > 1$. Let A and B be the inverses of a and b respectively.

For $k \in \mathbb{N}$, let $W_k = (A^* b a^*)^k (A^* B a^*)^k$ and let M_k be the sublanguage of W_k consisting of all those words

$$w = (A^{m_1} b a^{n_1}) \dots (A^{m_k} b a^{n_k})(A^{m_{k+1}} B a^{n_{k+1}}) \dots (A^{m_{2k}} B a^{n_{2k}})$$

satisfying the following:

(i) $m_i = n_i$ for all i;

(ii) $n_i < m_{i+1}$ for $i \notin \{k, 2k\}$.

Each of (i) and (ii) can be checked by a pushdown automaton, so M_k is the intersection of two context-free languages and the regular language W_k and is thus 2-$\mathcal{CF}$.

Now let $L_k = W(G, \{a, b\}) \cap M_k$. Then L_k consists of all words of the form

$$b^{a^{m_1}} \cdots b^{a^{m_k}} B^{a^{m_{2k+1}}} \cdots B^{a^{m_{2k}}} =_G 1,$$

with $m_i \in \mathbb{N}_0$ for all i, and $m_i < m_{i+1}$ for $i \notin \{k, 2k\}$. Since the conjugates of b in such a word are all distinct, for each $1 \leq i \leq k$ we must have some $1 \leq j \leq k$ such that $m_{k+j} = m_i$. But since $m_i < m_{i+1}$ and $m_{k+i} < m_{k+i+1}$ for all $1 \leq i \leq k-1$, this means $m_i = m_{k+i}$ for all $1 \leq i \leq k-1$.

When we take the commutative image of L_k, we can ignore the b's and B's, since these would contribute nothing to the aspects of the structure of the resulting subset of $\mathbb{N}_0^{6k}$ that interest us. It is equivalent and more straightforward to consider the commutative image as a subset of $\mathbb{N}_0^{4k}$, thus:

$$\begin{aligned}
\Phi(L_k) &= \{(m_1, m_1, \dots, m_k, m_k, m_1, m_1, \dots, m_k, m_k) \mid m_i \in \mathbb{N}_0, m_i < m_{i+1}\} \\
&= \{(n_1, n_2, \dots, n_{2k}, n_1, n_2 \dots, n_{2k}) \mid n_i = n_{i+1},\ n_{i-1} < n_i \ (i \text{ odd})\}.
\end{aligned}$$

We see that $\Phi(L_k)$ is a k-dimensional subset of the set $S^{(2,k)}$ studied in Section 2.6. Thus $\Phi(L_k)$ cannot be expressed as an intersection of $k-1$ stratified semilinear sets, by Corollary 2.18.

Hence L_k is not $(k-1)$-$\mathcal{CF}$ by Corollary 2.1. Since L_k is the intersection of $W(G)$ with a 2-$\mathcal{CF}$ language, this implies that $W(G)$ is not $(k-3)$-$\mathcal{CF}$ for any $k \in \mathbb{N}$ and so G is not poly-$\mathcal{CF}$. $\square$

3.4 The groups $G(\mathbf{c})$

We introduce a class of two-generator groups and present some of their properties. These groups will be important in our discussion of metabelian groups in the next chapter.

For $\mathbf{c} = (c_0, \ldots, c_s) \in \mathbb{Z}^{s+1}$ with $s \geq 1$, $c_0, c_s \neq 0$ and $\gcd(c_0, \ldots, c_s) = 1$, let $G(\mathbf{c})$ be the group defined by the presentation $\langle a, b \mid \mathcal{R}_{\mathbf{c}}(a, b)\rangle$, where

$$\mathcal{R}_{\mathbf{c}}(a, b) := \left\{[b, b^{a^i}]\ (i \in \mathbb{Z}),\ b^{c_0}(b^a)^{c_1} \cdots (b^{a^s})^{c_s}\right\}.$$

We shall call such groups *Gc-groups*, and when we refer to 'the Gc-group $G(\mathbf{c})$', we will assume that $\mathbf{c} \in \mathbb{Z}^{s+1}$ satisfies the above conditions. We will use the shorthand $\langle x, y\rangle_{\mathbf{c}}$ for $\langle x, y \mid \mathcal{R}_{\mathbf{c}}(x, y)\rangle$. Note that here we depart from our usual convention of denoting the i-th component of $\mathbf{c}$ by $\mathbf{c}(i)$, as it makes the notation more pleasant.

As an example, the soluble Baumslag-Solitar groups $\mathsf{BS}(1, m)$ for $m \in \mathbb{Z} \setminus \{0\}$ are all Gc-groups, since if $\mathbf{c} = (-m, 1)$ then

$$\begin{aligned} G(\mathbf{c}) &= \left\langle a, b \mid [b, b^{a^i}]\ (i \in \mathbb{Z}),\ b^{-m}b^a\right\rangle \\ &= \left\langle a, b \mid a^{-1}ba = b^m\right\rangle = \mathsf{BS}(1, m). \end{aligned}$$

The main result in this section will be that a Gc-group is poly-$\mathcal{CF}$ if and only if it is virtually abelian. We begin with some simple observations about Gc-groups.

We shall simplify the notation by setting $b_i = b^{a^i}$ for all $i \in \mathbb{Z}$ and $B = \langle b_i \mid i \in \mathbb{Z} \rangle$. Since B is an abelian normal subgroup of $G(\mathbf{c})$ and $G(\mathbf{c})/B \cong \langle a \rangle$, we see that Gc-groups have derived length at most 2.

Lemma 3.11. *Let $G = G(\mathbf{c})$ be a Gc-group with $|c_0| = |c_s| = 1$. Then G is polycyclic.*

Proof. Since $b_0^{c_0} b_1^{c_1} \cdots b_s^{c_s} = 1$ in G, either b_0 or its inverse is equal to $b_1^{c_1} \cdots b_s^{c_s}$, while b_{s+1} or its inverse is equal to $b_1^{c_0} b_2^{c_1} \ldots b_s^{c_{s-1}}$. Thus b_0 and b_{s+1}, and hence also all b_i for $i \leq 0$ and for $i \geq s+1$, are contained in $\langle b_1, b_2, \ldots, b_s \rangle$. Hence $B = \langle b_1, \ldots, b_s \rangle$, and so $G \rhd B \rhd \{1\}$ is a normal series for G with finitely generated abelian factors and G is polycyclic by Observation 1.6. □

Next we show that different elements of $\mathbb{Z}^{s+1}$ can produce isomorphic Gc-groups.

Lemma 3.12. *Let $G = G(\mathbf{c})$, where $\mathbf{c} = (c_0, \ldots, c_s)$ and let $\mathbf{c}' = (c_s, c_{s-1}, \ldots, c_0)$. Then $G(\mathbf{c}) \cong G(\mathbf{c}')$.*

Proof. Let $G = G(\mathbf{c}) = \langle a, b \rangle_{\mathbf{c}}$ and let $x = a^{-1}$ and $y = b^{a^s}$. Then $y^{x^i} = b^{a^{s-i}}$ for $i \in \mathbb{Z}$, so

$$b^{c_0}(b^a)^{c_1} \cdots (b^{a^s})^{c_s} = y^{c_s}(y^x)^{c_{s-1}} \cdots (y^{x^s})^{c_0}.$$

Hence $G(\mathbf{c}) \cong \langle x, y \rangle_{\mathbf{c}'} = G(\mathbf{c}')$. □

The following proposition gives a useful embedding of a Gc-group in a semidirect product $\mathbb{Q}^s \rtimes \mathbb{Z}$. We delay the proof until the next chapter (between Proposition 4.7 and Lemma 4.8).

Proposition 3.13. *Let $G = G(\mathbf{c})$ be a Gc-group. Let $\{x_1, \ldots, x_s\}$ be a basis for $\mathbb{Q}^s$ over $\mathbb{Q}$ (the rationals under addition), and let $\mathbb{Z} = \langle y \rangle$. Let $Q = \mathbb{Q}^s \rtimes \mathbb{Z}$, with the* *action of y on $\mathbb{Q}^s$ being given by the (columns of the) matrix*

$$A(\mathbf{c}) = \left(\begin{array}{cccc} 0 & \ldots & 0 & -c_0/c_s \\ & & & -c_1/c_s \\ & & & . \\ & I_{s-1} & & . \\ & & & . \\ & & & -c_{s-1}/c_s \end{array} \right).$$

Then G is isomorphic to the subgroup $\langle x_1, y \rangle$ of Q.

This implies that Gc-groups are torsion-free. Actually, the torsion-freeness of Gc-groups will be proved in the next chapter and then used in the proof of Proposition 3.13, but for now we simply note that they are torsion-free, in order to allow us to determine which Gc-groups are virtually abelian.

Lemma 3.14. *A Gc-group $G = G(\mathbf{c}) = \langle a, b \rangle_{\mathbf{c}}$ is virtually abelian if and only if some power of a centralises $B = \langle b_i \mid i \in \mathbb{Z} \rangle$. Hence $G(\mathbf{c})$ is virtually abelian if and only if the matrix $A(\mathbf{c})$ given in Proposition 3.13 has finite order.*

Proof. Suppose that A is a finite index abelian subgroup of G. Then there exist $k, m \in \mathbb{N}$ such that $a^k, b^m \in A$. Let $B_k = \langle b_{ik} \mid i \in \mathbb{Z} \rangle = \left\langle b^{a^{ik}} \mid i \in \mathbb{Z} \right\rangle$. Then $B_k^m = \left\langle a^k, b^m \right\rangle \leq A$. Since G is torsion-free and B_k^m is abelian,

$$(b_{ik}^m)^{a^k} = b_{ik}^m \;\Rightarrow\; (b_{ik}^{a^k})^m = b_{ik}^m \;\Rightarrow\; b_{ik}^{a^k} = b_{ik}.$$

Thus $a^k \in C_G(B_k)$. But then

$$b_{ik}^{a^k} = b_{ik} \;\Rightarrow\; (b^{a^{ik}})^{a^k} = b^{a^{ik}} \;\Rightarrow\; (b^{a^k})^{a^{ik}} = b^{a^{ik}} \;\Rightarrow\; b^{a^k} = b,$$

and hence $a^k \in C_G(B)$.

Conversely, if $a^k \in C_G(B)$, then $\left\langle a^k, b \right\rangle$ is abelian and has index k in G.

In the proof of Lemma 3.13 in the next chapter, we will see that y is the image of a in the embedding of $G(\mathbf{c}) = \langle a, b \rangle_{\mathbf{c}}$ in Q. Thus some power of a centralises B if and only if some power of y centralises $\left\langle x_1^{y^i} \mid i \in \mathbb{Z} \right\rangle$. Since $\left\langle x_1^{y^i} \mid i \in \mathbb{Z} \right\rangle$ contains $x_1, \ldots, x_s$, which form a basis for $\mathbb{Q}^s$, this means that a^k centralises B if and only if y^k centralises $\mathbb{Q}^s$, which happens if and only if $(A(\mathbf{c}))^k$ is the identity matrix. $\square$

Before proving our main theorem in this section, we first need a lemma about powers of the matrix $A(\mathbf{c})$ in Lemma 3.13. For the purposes of the proof of this lemma, we introduce some new notation.

Let p be a prime. The *p-adic valuation* $v_p : \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ is given by

- $v_p(0) = \infty$;
- $v_p(m/n) = d_m - d_n$ for $m, n \in \mathbb{Z},\ n \neq 0$, where $d_k := \max\{i \in \mathbb{N}_0 \mid p^i | k\}$ for all $k \in \mathbb{Z}$.

It can easily be seen that the p-adic valuation satisfies the following, which are the defining properties of a valuation.

(i) $v_p(a) = \infty$ if and only if $a = 0$;

(ii) $v_p(ab) = v_p(a) + v_p(b)$;

(iii) $v_p(a + b) \geq \min\{v_p(a), v_p(b)\}$.

In the lemma, we shall be concerned with powers of a prime occuring in the denominator of various rational numbers. Therefore, rather than v_p, we shall always be using $-v_p$, which, because of the frequency of its occurence, we shall denote by $\bar{v}_p$. Note that $\bar{v}_p$ satisfies the following:

(i) $\bar{v}_p(a) = -\infty$ if and only if $a = 0$;

(ii) $\bar{v}_p(ab) = \bar{v}_p(a) + \bar{v}_p(b)$;

(iii) $\bar{v}_p(a+b) \leq \max\{\bar{v}_p(a), \bar{v}_p(b)\}$.

(iv) If $\bar{v}_p(a) < \bar{v}_p(b)$, then $\bar{v}_p(a+b) = \bar{v}_p(b)$.

We now proceed to the lemma. It is stated in slightly more generality than we require, as it is just as easy to prove the more general result.

Lemma 3.15. *Let M be a matrix of the form*

$$\begin{pmatrix} 0\ldots0 & a_1 \\ & a_2 \\ I_{s-1} & . \\ & . \\ & a_s \end{pmatrix},$$

where all $a_i \in \mathbb{Q}$ and at least one $a_i \notin \mathbb{Z}$. Write $M^k = (m_{ij}^{(k)})$ for $k \in \mathbb{N}$. Then there exist $N \in \{1, \ldots, s\}$ and a prime p such that for every $k \in \mathbb{N}$ there exists some $i_k \leq ks$ with $\bar{v}_p(m_{Ns}^{(i_k)}) \geq k$.

Proof. Choose some $a_j \notin \mathbb{Z}$, and let p be a prime such that $\bar{v}_p(a_j) > 0$. For $1 \leq i \leq s$, let $n_i = \bar{v}_p(a_i)$. Let $n = \max\{n_i \mid 1 \leq i \leq s\}$ and let $N = \max\{i \mid n_i = n\}$. For $k \in \mathbb{N}$, denote the entry in the N-th row and s-th column of M^k by m_k.

Note that for $k \geq 2$ and $1 \leq i \leq s-1$, the i-th column of M^k is the same as the $(i+1)$-th column of M^{k-1}. Thus the N-th row of M^k is

$$(\epsilon_1, \ldots, \epsilon_{s-k}, m_1, \ldots, m_k)$$

with $\epsilon_i \in \{0, 1\}$ if $k < s$, and

$$(m_{k-s+1}, \ldots, m_{k-1}, m_k)$$

if $k \geq s$. For convenience, rename $\epsilon_1, \ldots, \epsilon_{s-k}$ as $m_{k-s+1}, \ldots, m_0$, so that we can write the N-th row of M^k in the second form in both cases. Notice that m_k is in the $N, s-i$ position in M^{k+i}. In particular, letting $N' = s - N$, we have m_k in the N, N position of $M^{k+N'}$ for all $k \in \mathbb{N}$.

For $k \in \mathbb{N}$, define i_k to be the minimal natural number such that $\bar{v}_p(m_{i_k}) \geq k$ if such a number exists, or ∞ otherwise. To begin with,

$$\bar{v}_p(m_1) = \bar{v}_p(a_N) = n \geq 1,$$

so $i_1 = 1$. We shall show by induction on k that $i_k \leq ks$ for all $k \in \mathbb{N}$, hence proving the lemma.

Fix $k \in \mathbb{N}$ and suppose that $i_k \leq ks$. Let $j_k = i_k + N'$ and consider M^{j_k}. The N-th row of this matrix is $(m_{j_k-s+1}, \ldots, m_{i_k}, \ldots, m_{j_k-1}, m_{j_k})$. Note that $\bar{v}_p(m_{i_k}) \geq k$ and $\bar{v}_p(m_i) < k$ for $1 \leq i < i_k$, by the minimality of i_k. For $i \leq 0$, we have $m_i \in \{0, 1\}$ and so $\bar{v}_p(m_i) \in \{0, -\infty\}$. Thus $\bar{v}_p(m_{i_k}) < k$ for all $i < i_k$.

If $\bar{v}_p(m_i) \geq k+1$ for some $i \leq j_k$, then

$$i_{k+1} \leq i \leq j_k = i_k + N' \leq ks + s - N \leq (k+1)s.$$

So we may assume that $\bar{v}_p(m_i) \leq k$ for all $i \leq j_k$. Then

$$\begin{aligned} m_{j_k+1} &= (m_{j_k-s+1}, \ldots, m_{i_k}, \ldots, m_{j_k}) \cdot (a_1, \ldots, a_N, \ldots, a_s) \\ &= \sum_{i=1}^{s} m_{j_k-s+i} a_i \\ &= \sum_{i=1}^{s} m_{i_k-N+i} a_i \\ &= m_{i_k} a_N + S_1 + S_2, \qquad\qquad (*) \end{aligned}$$

where $S_1 := \sum_{i=1}^{N-1} m_{i_k-N+i} a_i$ and $S_2 := \sum_{i=N+1}^{s} m_{i_k-N+i} a_i$.

We have

$$\bar{v}_p(m_{i_k-N+i} a_i) = \bar{v}_p(m_{i_k-N+i}) + n_i$$

for $1 \le i \le s$. In particular, $\bar{v}_p(m_{i_k} a_N) = \bar{v}_p(m_{i_k}) + n \ge k + n$.

Now consider $\bar{v}_p(S_1)$. Firstly, $\bar{v}_p(m_{i_k - N + i}) < k$ for $1 \le i \le N - 1$, since

$$i_k - N + i \le i_k - N + N - 1 < i_k.$$

Thus, since $n_i \le n$, we have $\bar{v}_p(m_{i_k - N + i} a_i) < k + n$ for $1 \le i \le N - 1$. Hence $\bar{v}_p(S_1) < k + n$.

Next, consider $\bar{v}_p(S_2)$. For $i \in \{N + 1, \dots, s\}$, we have $\bar{v}_p(m_{i_k - N + i}) \le k$ by our assumption, since $i \le j_k$. Also, $n_i < n$ by the maximality of N. Thus for $N + 1 \le i \le s$ we have $\bar{v}_p(m_{i_k - N + i} a_i) < k + n$, implying $\bar{v}_p(S_2) < k + n$.

Since $\bar{v}_p(S_1), \bar{v}_p(S_2) < k + n \le \bar{v}_p(m_{i_k} a_N)$, by $(*)$ we conclude that

$$\bar{v}_p(m_{j_k + 1}) = \bar{v}_p(m_{i_k} a_N) \ge k + n \ge k + 1,$$

and hence

$$i_{k+1} \le j_k + 1 = i_k + N' + 1 \le ks + s - N + 1 \le ks + s = (k+1)s. \qquad \square$$

We are now ready to prove the main result of this section.

Proposition 3.16. *A Gc-group is poly-$\mathcal{CF}$ if and only if it is virtually abelian.*

Proof. Let $G = G(\mathbf{c})$ be a Gc-group with $\mathbf{c} \in \mathbb{Z}^{s+1}$ and suppose that G is not virtually abelian. If $|c_0| = |c_s| = 1$, then G is polycyclic and hence not poly-$\mathcal{CF}$ by Proposition 3.7. Hence, if $|c_s| = 1$, we may assume $|c_0| \ne 1$. By Lemma 3.12, G is isomorphic to $G(\mathbf{c}')$, where $\mathbf{c}' = (c_s, c_{s-1}, \dots, c_0)$. Thus we may assume that $|c_s| \ne 1$.

By Lemma 3.13, we can identify G with the subgroup $\langle x_1, y \rangle$ of $Q = \mathbb{Q}^s \rtimes \mathbb{Z}$, where $\{x_1, \dots, x_s\}$ is a basis for $\mathbb{Q}^s$ over $\mathbb{Q}$, $\mathbb{Z} = \langle y \rangle$, and y acts on $\mathbb{Q}^s$ by the matrix $A(\mathbf{c})$ given in the lemma.

Let $M = A(\mathbf{c})$ and use the notation of Lemma 3.15 for entries of M^k. Since $|c_s| \neq 1$ and $\gcd(c_0, \ldots, c_s) = 1$, some c_i/c_s for $0 \leq i \leq s-1$ is not an integer. Thus M satisfies the hypothesis of Lemma 3.15. Hence there exist $I \in \{1, \ldots, s\}$ and a prime p such that for every $k \in \mathbb{N}$ there exists some $\iota_k \leq ks$ such that $\bar{v}_p(m_{Is}^{(\iota_k)})$ is at least k.

For $k \in \mathbb{N}$, let

$$\ell_k = \min \left\{ \ell \in \mathbb{N} \mid \ell m_{is}^{(k)} \in \mathbb{Z} \ (1 \leq i \leq s) \right\}.$$

This is the smallest nonnegative integer ℓ such that the final column of ℓM^k has all integer entries. We are especially interested in the matrices M^{ι_k}, and so it will be convenient to set $\lambda_k = \ell_{\iota_k}$. Since $\bar{v}_p(m_{Is}^{(\iota_k)}) \geq k$, we have $\lambda_k \geq p^k$ for all $k \in \mathbb{N}$.

For $1 \leq i \leq s$, define functions $h_i : \mathbb{N} \to \mathbb{Z}$ by setting $h_i(n)$ to be the i-th entry of the final column of $\lambda_n M^{\iota_n}$. That is,

$$h_i(n) = \lambda_n m_{is}^{(\iota_n)}.$$

We will construct a sequence $n_1, n_2, \ldots$ of natural numbers such we can specify certain properties of the sequences $h_i(n_1), h_i(n_2), \ldots$.

Define sequences of natural numbers $n_1^i, n_2^i, \ldots$ for $0 \leq i \leq s$ inductively as follows. Let $n_k^0 = k$ for all $k \in \mathbb{N}$. For $1 \leq i \leq s$, the sequence $n_1^{i-1}, n_2^{i-1}, \ldots$ must have an infinite subsequence $n_1^i, n_2^i, \ldots$ which satisfies the following properties.

- Either $h_i(n_k^i) \geq 0$ for all $k \in \mathbb{N}$ or $h_i(n_k^i) < 0$ for all $k \in \mathbb{N}$. In the first case we say that i is of Type 1, while in the second case i is of Type 2.

- Either for every quadratic $f : \mathbb{N} \to \mathbb{N}$, there exists some $N \in \mathbb{N}$ such that $|h_i(n_k^i)| < f(n_k^i)$ for all $k \geq N$, or else there exists some quadratic $f : \mathbb{N} \to \mathbb{N}$ such that $f(n_k^i) \leq |h_i(n_k^i)|$ for all $k \in \mathbb{N}$. In the first case we say that i is of Type A, while in the second case i is of Type B.

Let $n_1, n_2, \ldots$ be the final sequence obtained, so $n_k = n_k^s$ for all $k \in \mathbb{N}$. Since for all $1 \leq i \leq s-1$ the sequence $n_1, n_2, \ldots$ is a subsequence of $n_1^i, n_2^i, \ldots$, we have

- For $i \in \{1, \ldots, s\}$, if i is of Type 1, then $h_i(n_k) \geq 0$ for all $k \in \mathbb{N}$, and if i is of Type 2, then $h_i(n_k) < 0$ for all $k \in \mathbb{N}$.
- For every quadratic $f : \mathbb{N} \to \mathbb{N}$, there exists some $N \in \mathbb{N}$ with $|h_i(n_k)| < f(n_k)$ for all $k \geq N$ and all i of Type A.
- There exists a quadratic $g : \mathbb{N} \to \mathbb{N}$ such that $g(n_k) \leq |h_i(n_k)|$ for all $k \in \mathbb{N}$ and all i of Type B.

We are now ready to define a bounded sublanguage of $W(G)$ which we can show to be not poly-$\mathcal{CF}$ using Corollary 2.7.

Let $X = \{x_1, \ldots, x_s, y\}$ and consider the intersection of $W(G, X)$ with the bounded context-free language

$$L' = \cup_{k \in \mathbb{N}_0} (y^{-1})^k x_s^* y^k (x_1^{\epsilon_1})^* (x_2^{\epsilon_2})^* \ldots (x_s^{\epsilon_s})^*,$$

where $\epsilon_i = -1$ if i is of Type 1 and $\epsilon_i = 1$ if i is of Type 2. Let L be the commutative image of $W(G, X) \cap L'$.

The final column of M^k represents the action of y^k on x_s. Specifically,

$$x_s^{y^k} = x_1^{m_{1s}^{(k)}} \cdots x_I^{m_{Is}^{(k)}} \cdots x_s^{m_{ss}^{(k)}}.$$

For $\lambda \in \mathbb{Z}$ and $k \in \mathbb{N}$, the element $\left((x_s^\lambda)^{y^k}\right)^{-1}$ of G can be expressed as a word in $(x_1^{\epsilon_1})^*(x_2^{\epsilon_2})^* \ldots (x_s^{\epsilon_s})^*$ if and only if $\ell_k | \lambda$.

For all $k \in \mathbb{N}$, we thus have $(\iota_k, \lambda, \iota_k; \mathbf{v}) \in L$, where $\mathbf{v} \in \mathbb{N}_0^s$, if and only if $\ell_{\iota_k} = \lambda_k | \lambda$ and $\mathbf{v}(i) = \lambda |m_{is}^{(\iota_k)}|$ for $1 \leq i \leq s$.

In order to apply Corollary 2.7 to $W(G) \cap L'$, we need to first find a suitable permutation τ to apply to L. Let Γ_A be the set of all $i \in \{1, \ldots, s\}$ of Type A and let $r = |\Gamma_A|$. Let ρ be a permutation of $\{4, \ldots, s+3\}$ which sends all $i+3$ for $i \in \Gamma_A$ to $\{4, \ldots, r+3\}$. Now define

$$\tau = (s+3, s+2, \ldots, 3, 2) \circ \rho.$$

Applying τ to $\mathbf{w} = (\iota_k, \lambda, \iota_k; \mathbf{v}) \in L$ places all the 'smaller' entries of $\mathbf{w}$ in the first $r+2$ components of $\tau(\mathbf{w})$, and places the second entry, λ, in the final position of $\tau(\mathbf{w})$.

For $1 \leq i \leq s$, define $i' = \tau^{-1}(i)$. Then, for all $k \in \mathbb{N}$, we have $(\iota_k, \iota_k; \mathbf{v}) \in \tau(L)$, where $\mathbf{v} \in \mathbb{N}_0^{s+1}$, if and only if $\lambda_k | \mathbf{v}(s+1)$ and

$$\mathbf{v}(i) = \mathbf{v}(s+1)|m_{i's}^{(\iota_k)}| = \frac{1}{\lambda_k}\mathbf{v}(s+1)|h_{i'}(k)|$$

for $1 \leq i \leq s$.

For $k \in \mathbb{N}$, let $\mathbf{a}_k = ((\iota_{n_k}, \iota_{n_k}); \mathbf{u}_k)$, where $\mathbf{u}_k \in \mathbb{N}_0^r$ with $\mathbf{u}_k(i) = |h_{i'}(n_k)|$. Let $\mathbf{b}_k \in \mathbb{N}_0^{s-r+1}$ with

$$\mathbf{b}_k(j) = |h_{(r+j)'}(n_k)|$$

for $1 \leq j \leq s-r$ and $\mathbf{b}_k(s-r+1) = \lambda_{n_k}$.

As we have already observed, we can find a quadratic $g : \mathbb{N} \to \mathbb{N}$ such that

$$g(n_k) \leq |h_{(r+j)'}(n_k)| = \mathbf{b}_k(j)$$

for all $1 \leq j \leq s-r$, since all $(r+j)'$ are of Type B. Since also

$$\mathbf{b}_k(s-r+1) = \lambda_{n_k} \geq p^{n_k},$$

we can take g to be such that $g(n_k) \leq \mathbf{b}_k(i)$ for all $1 \leq i \leq s-r+1$ for all $k \in \mathbb{N}$.

If two sequences are eventually dominated by every quadratic $f : \mathbb{N} \to \mathbb{N}$, then the same is true of their sum, and also of a constant multiple of one of the sequences. Thus, for any $t \in \mathbb{N}$, there exists $N \in \mathbb{N}$ such that

$$t\sigma(\mathbf{a}_k) = t\left(2\iota_{n_k} + \sum_{i=1}^{r} \mathbf{u}(i)\right) \leq t\left(2sn_k + \sum_{i=1}^{r} |h_{i'}(n_k)|\right) < g(n_k) \leq \mathbf{b}_k(j)$$

for all $k \geq N$ and $1 \leq j \leq s-r+1$, because all i' are of Type A for $1 \leq i \leq r$.

If $\Gamma_A \neq \emptyset$, then $(\mathbf{a}_k; \mathbf{b}) \in \tau(L)$ if and only if $\mathbf{b} = \mathbf{b}_k$. We have thus shown that $\tau(L)$ satisfies the hypothesis of Proposition 1.28.

If $\Gamma_A = \emptyset$, then $\mathbf{a}_k = (\iota_{n_k}, \iota_{n_k})$, and $(\mathbf{a}_k; \mathbf{b}) \in \tau(L)$ if and only if $\mathbf{b} = \lambda \mathbf{b}_k$ for some $\lambda \in \mathbb{N}_0$. We have thus shown that for all $t \in \mathbb{N}$, there is an $\mathbf{a}_k$ satisfying the first two conditions of Proposition 2.6 with respect to t. We can take k such that $n_k \geq t$. Since $(\mathbf{a}_k; \mathbf{b}) \in \tau(L)$ if and only if $\mathbf{b}$ is a nonnegative integer multiple of $\mathbf{b}_k$, for any two distinct $\mathbf{b}$ and $\mathbf{b}'$ such that $(\mathbf{a}; \mathbf{b}), (\mathbf{a}; \mathbf{b}') \in \tau(L)$, there are distinct $\lambda_1, \lambda_2 \in \mathbb{N}_0$ such that

$$\begin{aligned} |\mathbf{b}(s+1) - \mathbf{b}'(s+1)| &= |\lambda_1 \mathbf{b}_k(s+1) - \lambda_2 \mathbf{b}_k(s+1)| \\ &= |\lambda_1 - \lambda_2| \lambda_{n_k} \geq p^{n_k} \geq p^t. \end{aligned}$$

Since $f(t) = p^t$ is an unbounded function, this shows that $\mathbf{a}_k$ also satisfies the third condition of Proposition 2.6 with respect to t.

Thus in either case $W(G) \cap L'$ is not poly-$\mathcal{CF}$, by Corollary 2.7. Since L' is context-free, this implies that $W(G)$ is not poly-$\mathcal{CF}$. $\square$

Chapter 4

Poly-$\mathcal{CF}$ metabelian groups

A group G is *metabelian* if it has an abelian normal subgroup N such that G/N is also abelian. This is equivalent to saying that G has derived length at most 2. The metabelian groups are thus the most basic examples of soluble groups after the abelian groups.

We conjectured in the previous chapter that the only poly-$\mathcal{CF}$ groups are those which are virtually a finitely generated subgroup of a direct product of free groups.

A finitely generated virtually abelian group is virtually a finite direct product of free groups of rank 1. Thus every finitely generated virtually abelian group is poly-$\mathcal{CF}$ by Observation 3.1. Any group containing a free subgroup of rank 2 is not soluble, so the only soluble groups which are virtually a direct product of free groups are the virtually abelian groups.

If G is a direct product of free groups and H is a finitely generated soluble subgroup of G, then the image of the projection of H onto a direct factor of G must be trivial or isomorphic to $\mathbb{Z}$. Since H is a subgroup of the direct product of the images of these projections, which is abelian, H is itself abelian. Thus, in the class of soluble groups, Conjecture 3.2 becomes very straightforward:

Conjecture 4.1. *A finitely generated soluble group is poly-$\mathcal{CF}$ if and only if it is virtually abelian.*

In this chapter we prove that the conjecture holds in the case of metabelian groups. In order to do this, we prove a purely group theoretic result. We show that every finitely generated metabelian group which is not virtually abelian or polycyclic has a subgroup isomorphic to one of a few types of groups. Each of these types of groups has already been shown in the previous chapter to be not poly-$\mathcal{CF}$.

4.1 Two-generator metabelian groups

The proof our main result in this chapter will rely largely on some facts about a certain type of two-generator metabelian group. These are groups H satisfying the following hypothesis.

Hypothesis $(*)$**:** $H = \langle a, b \rangle$, where a has infinite order, the subgroup

$$B := \left\langle b^{a^i} \mid i \in \mathbb{Z} \right\rangle$$

is abelian, and H is not virtually abelian.

Throughout our discussion of metabelian groups, H and B will have these definitions, and for convenience we will use the notation b_i for b^{a^i}. We will also sometimes use additive notation within B.

We will first consider two types of groups satisfying Hypothesis $(*)$: (i) H is torsion-free; (ii) b is a torsion element; before going on to the general case. We will have several occasions to use the following observation, which we record here for reference:

Observation 4.2. *Let $H = \langle a, b \rangle$ and assume Hypothesis $(*)$. Then $H = B \rtimes \langle a \rangle$.*

Proof. We have $H = B\langle a\rangle$, since $B \lhd H$. Also $a^m \notin B$ for all $m \in \mathbb{Z} \setminus \{0\}$, since otherwise $H/B \cong C_n$ for some n dividing m, but H is not virtually abelian. $\square$

4.1.1 Case 1: H is torsion-free

The following lemma gives a useful starting point in considering torsion-free groups satisfying Hypothesis $(*)$.

Lemma 4.3. *Suppose $H = \langle a, b\rangle$ is torsion-free and assume Hypothesis $(*)$. Then either $H \cong \mathbb{Z} \wr \mathbb{Z}$ or there exist $c_0, \ldots, c_s \in \mathbb{Z}$ with $\gcd(c_0, \ldots, c_s) = 1$ and $c_0, c_s \neq 0$ such that $\sum_{i=0}^{s} c_i b_i = 0$.*

Proof. If B is free abelian on $\{b_i \mid i \in \mathbb{Z}\}$, then H is isomorphic to $\mathbb{Z} \wr \mathbb{Z}$. If not, then there exist integers r, s, c_i with $r \leq s$ and the c_i not all zero such that $\sum_{i=r}^{s} c_i b_i = 0$. Since we are assuming that H is torsion-free and not virtually abelian, we must have $r < s$.

By conjugating by a^{-r}, we may assume $r = 0$. Thus we can choose $s \in \mathbb{N}$ minimal such that there exist $c_i \in \mathbb{Z}$ with $\sum_{i=0}^{s} c_i b_i = 0$. We may assume $\gcd(c_0, \ldots, c_s) = 1$, by dividing through by a common factor if necessary, which we can do because H is torsion-free. The minimality of s implies that $c_0, c_s \neq 0$. $\square$

We will eventually show that this relation $\sum_{i=0}^{s} c_i b_i = 0$, together with the fact that B is abelian, fully determines H, and thus H is isomorphic to a Gc-group.

Recall from Section 3.4 that the group $G(\mathbf{c})$ has a presentation $\langle \alpha, \beta \mid \mathcal{R}_{\mathbf{c}}(\alpha, \beta)\rangle$ (or in shorthand $\langle \alpha, \beta\rangle_{\mathbf{c}}$), where

$$\mathcal{R}_{\mathbf{c}}(\alpha, \beta) = \left\{[\beta, \beta^{\alpha^i}]\ (i \in \mathbb{Z}),\ \beta^{c_0}(\beta^{\alpha})^{c_1} \cdots (\beta^{\alpha^s})^{c_s}\right\},$$

with $\mathbf{c} = (c_0, \ldots, c_s) \in \mathbb{Z}^{s+1}$, $c_0, c_s \neq 0$ and $\gcd(c_0, \ldots, c_s) = 1$.

If $\langle X \mid R\rangle$ is a group presentation, denote the abelianisation of the group with this presentation by $\mathrm{Ab}\langle X \mid R\rangle$. This enables us to write shorter presentations for abelian groups, by omitting the commutators of generators from the relator set. We call such a presentation an *abelian presentation*, and we often use additive notation for its relators.

Lemma 4.4. *Let $G = G(\mathbf{c}) = \langle \alpha, \beta\rangle_{\mathbf{c}}$ be a Gc-group, and let $J = \left\langle \beta^{\alpha^i} \mid i \in \mathbb{Z}\right\rangle$. Then $G = J \rtimes \langle\alpha\rangle$,*

$$J \cong \mathrm{Ab}\left\langle \beta_i \ (i \in \mathbb{Z}) \mid \sum_{i=0}^{s} c_i \beta_{k+i} \ (k \in \mathbb{Z})\right\rangle,$$

and G is torsion-free.

Proof. Clearly $J \lhd G$, and so $G = J\langle\alpha\rangle$. Also $\alpha^m \notin J$ for all $m \in \mathbb{Z}\setminus\{0\}$, since $G/J \cong \langle \alpha \mid \rangle \cong \mathbb{Z}$. Hence $G = J \rtimes \langle\alpha\rangle$.

Let $\mathcal{J} = \mathrm{Ab}\langle \beta_i \ (i\in\mathbb{Z}) \mid \sum_{i=0}^{s} c_i\beta_{k+i} \ (k \in \mathbb{Z})\rangle$. There is an automorphism ψ of $\mathcal{J}$ given by $\psi(\beta_i) = \beta_{i+1}$. This corresponds to the automorphism of J given by conjugation by α. We thus have

$$\mathcal{J} \rtimes \langle\alpha\rangle = \langle \beta_i \ (i \in \mathbb{Z}), \alpha \mid [\beta_0, \beta_i],\ \beta_i^{\alpha} = \beta_{i+1} \ (i \in \mathbb{Z}),\ \beta_0^{c_0}\cdots\beta_s^{c_s}\rangle,$$

since all the other relators of $\mathcal{J}$ are consequences of those listed. We can delete the generators β_i for $i \neq 0$ using $\beta_i = \beta_0^{\alpha^i}$, yielding

$$\mathcal{J} \rtimes \langle\alpha\rangle = \left\langle \beta_0, \alpha \mid [\beta_0, \beta_0^{\alpha^i}],\ \beta_0^{c_0}\cdots\beta_s^{c_s}\right\rangle.$$

Finally, replacing β_0 by β, we arrive at a presentation for G, and so $G \cong \mathcal{J} \rtimes \langle\alpha\rangle$.

The argument above shows that there is an isomorphism $\phi : G \to \mathcal{J} \rtimes \langle\alpha\rangle$ with $\phi(\alpha) = \alpha$, $\phi(\beta^{\alpha^i}) = \beta_i$. Moreover, ϕ maps J onto $\mathcal{J}$, so $J \cong \mathcal{J}$, and hence J has an abelian presentation as claimed. Let $r_k = \sum_{i=0}^{s} c_i\beta_{k+i}$ for $k \in \mathbb{Z}$; so $J \cong \mathrm{Ab}\langle \beta_i \ (i \in \mathbb{Z}) \mid r_k \ (k \in \mathbb{Z})\rangle$.

It remains to show that G is torsion-free. Since $G/J \cong \mathbb{Z}$, it suffices to show that J is torsion-free. Any element of J is contained in some subgroup of the form $J_I := \langle \beta_k, \ldots, \beta_{k'} \rangle$, where $I = \{k, \ldots, k'\}$. Thus it is enough to show that every subgroup of this form is torsion-free.

To show this, we first derive a presentation for J_I. Suppose that $r := \sum_{i=k}^{k'} \delta_i \beta_i$ with $\delta_i \in \mathbb{Z}$ is a relator of J_I. Since r is also a relator of J, it must be equal in the free abelian group on the β_i to an element of the form

$$\sum_{i=l}^{l'} \gamma_i r_i = \sum_{i=l}^{l'} \gamma_i \sum_{j=0}^{s} c_j \beta_{j+i},$$

where $\gamma_l, \gamma_{l'} \neq 0$. So we must have $k \leq l$ and $l' + s \leq k'$. Conversely, any r_j with $k \leq j \leq k' - s$ is a relator of J_I. We thus have an abelian presentation for J_I:

$$J_I \cong \mathrm{Ab} \left\langle \beta_i \ (i \in I) \mid r_j \ (k \leq j \leq k' - s) \right\rangle .$$

Let $m = k' - s - k$. As discussed in Section 1.4.1, since J_I is a finitely generated abelian group, we can find its isomorphism type by computing the Smith normal form of its presentation matrix, which is a matrix with m rows and $m + s$ columns of the following form, where all the entries to the left of the c_0's and to the right of the c_s's are 0:

$$M(\mathbf{c}, m) := \begin{pmatrix} c_0 & c_1 & \ldots & c_s & & & & \\ & c_0 & c_1 & \ldots & c_s & & & \\ & & . & \ldots & \ldots & . & & \\ & & & . & \ldots & \ldots & . & \\ & & & & c_0 & c_1 & \ldots & c_s \end{pmatrix}.$$

We will show in Corollary 4.6 that the Smith normal form of this matrix is $(I_m \mid 0_{m,s})$, and so J_I is a free abelian group, and hence J_I, J and G are torsion-free as claimed. $\square$

The matrices $M(\mathbf{c}, m)$

We complete the proof of Lemma 4.4 by calculating the Smith normal form of the matrices $M(\mathbf{c}, m)$.

Lemma 4.5. *Let $\mathbf{c} = (c_0, c_1, \ldots, c_s) \in \mathbb{Z}^{s+1}$ with $\gcd(c_0, \ldots, c_s) = 1$ and $c_0, c_s \neq 0$. Let $m \geq 1$, and let M be the $m \times (m+s)$ matrix $M(\mathbf{c}, m)$ as defined above in the proof of Lemma 4.4. Then the $m \times m$ minors of M are relatively prime.*

Proof. It suffices to find a subset of $m \times m$ minors of M which are relatively prime. Label the columns of M by $C_0, C_1, \ldots, C_{s+m-1}$, and for $i \notin \{0, \ldots, s\}$ set $c_i = 0$. Then $C_i = (c_i, c_{i-1}, \ldots, c_{i-m+1})^{\mathrm{T}}$. Now for $i \in \{0, \ldots, s\}$, let $M^{(i)} = (C_i, C_{i+1}, \ldots, C_{i+m-1})$, and let $d_i = \det(M^{(i)})$.

Then $M^{(0)}$ is an upper triangular matrix with c_0 in every position on the diagonal, so $d_0 = c_0^m$. Also, $M^{(s)}$ is a lower triangular matrix with c_s in every position on the diagonal, so $d_s = c_s^m$. In general, $M^{(i)}$ has c_i in every position on the diagonal, and every entry below the diagonal is either 0 or c_j for some $j < i$. Denote the j, k entry of $M^{(i)}$ by $m_{jk}^{(i)}$.

For a permutation $\sigma \in S_m$, let $\pi_\sigma = \operatorname{sgn}(\sigma) \prod_{j=1}^m m_{j\sigma(j)}^{(i)}$. Then

$$d_i = c_i^m + \sum_{\sigma \in S_m \setminus \{1\}} \pi_\sigma.$$

For $\sigma \in S_m \setminus \{1\}$, there must be some $j \in \{1, \ldots, m\}$ with $\sigma(j) < j$. Thus each π_σ is a multiple of some entry below the diagonal of $M^{(i)}$; that is, of some c_j with $j < i$. Thus we can write

$$d_i = c_i^m + \sum_{j=0}^{i-1} a_{ij} c_j.$$

Let $\delta_i = \gcd(d_0, \ldots, d_i)$ and $\gamma_i = \gcd(c_0, \ldots, c_i)$. We claim that the prime divisors of δ_i all divide γ_i. Firstly, $\delta_0 = c_0^m$ and any prime dividing c_0^m divides $c_0 = \gamma_0$. Now suppose that, for some k, every prime divisor of δ_k divides γ_k. Since

$$\delta_{k+1} = \gcd(\delta_k, c_{k+1}^m + a_{k+1,k}c_k + \ldots + a_{k+1,0}c_0),$$

if a prime p divides δ_{k+1}, then it must divide γ_k (and hence $c_0, \ldots, c_k$) as well as dividing $c_{k+1}^m + a_{k+1,k}c_k + \ldots + a_{k+1,0}c_0$. But this implies that p divides c_{k+1} and hence p divides $\gamma_{k+1} = \gcd(\gamma_k, c_{k+1})$. Thus, in particular, the prime divisors of δ_s all divide $\gamma_s = 1$, so $\delta_s = \gcd(d_0, \ldots, d_s) = 1$ and the result follows. $\square$

Corollary 4.6. *Let M be a matrix as in Lemma 4.5. Then the Smith normal form of M is $(I_m \mid 0_{m,s})$.*

Proof. Let $\gamma_i(M)$ and σ_i be defined as in Proposition 1.1. Then $\gamma_m(M) = 1$ by Lemma 4.5. By Proposition 1.2, $\sigma_i = \gamma_i(M)/\gamma_{i-1}(M)$ and all $\sigma_i \in \mathbb{Z}$, and so $\gamma_i(M) = 1$ and $\sigma_i = 1$ for all $1 \leq i \leq m$. Hence, by Proposition 1.1, the Smith normal form of M is $(\mathrm{diag}(\sigma_1, \ldots, \sigma_m) \mid 0_{m,s}) = (I_m \mid 0_{m,s})$. $\square$

Conclusion for Case 1

We are now ready to classify the groups we have been considering in this subsection.

Proposition 4.7. *Let $H = \langle a, b \rangle$ be torsion-free and satisfy Hypothesis $(*)$. Then H is isomorphic to either $\mathbb{Z} \wr \mathbb{Z}$ or a Gc-group.*

Proof. Suppose H is not isomorphic to $\mathbb{Z} \wr \mathbb{Z}$. By Lemma 4.3, there exist $s \geq 1$ and $\mathbf{c} = (c_0, \ldots, c_s) \in \mathbb{Z}^{s+1}$ with $c_0, c_s \neq 0$ and $\gcd(c_0, \ldots, c_s) = 1$ such that $\sum_{i=0}^{s} c_i b_i = 0$ (or in multiplicative notation $b_0^{c_0} \cdots b_s^{c_s} = 1$). We choose $\mathbf{c}$ with s minimal.

Let $G = G(\mathbf{c}) = \langle \alpha, \beta \rangle_{\mathbf{c}}$. Let $\phi' : \{\alpha, \beta\} \to H$ be given by $\phi'(\alpha) = a$ and $\phi'(\beta) = b$. Substituting a for α and b for β in each relation of G yields the identity of H; so, by

Proposition 1.8, ϕ' extends to a homomorphism $\phi : G \to H$, which is an epimorphism since $H = \langle a, b \rangle$. We shall show that ϕ is an isomorphism.

For convenience we shall write β_i for β^{α^i} and b_i for b^{a^i}. Thus $\mathcal{R}_{\mathbf{c}}(\alpha, \beta)$ can be written more simply as

$$\{[\beta_0, \beta_i]\ (i \in \mathbb{Z}),\ \beta_0^{c_0} \cdots \beta_s^{c_s}\}.$$

Let $J = \langle \beta_i \mid i \in \mathbb{Z} \rangle$ and $K = \langle \beta_0, \beta_1, \ldots, \beta_{s-1} \rangle$. We have already shown in Lemma 4.4 that J (and hence also K) is abelian, $G = J \rtimes \langle \alpha \rangle$ and G is torsion-free.

If $\beta_0, \beta_1, \ldots, \beta_{s-1}$ were not linearly independent over $\mathbb{Z}$, then their images under ϕ, namely $b_0, b_1, \ldots, b_{s-1}$, would not be linearly independent either, contradicting the minimality of s. Thus K is free abelian of rank s.

We now show that J/K is a torsion group. We claim that for any $j \in \mathbb{Z}$, there exists some non-zero $\lambda_j \in \mathbb{Z}$ such that $\lambda_j \beta_j \in K$. This is clearly true for $0 \leq j \leq s-1$, and also for $j = s$ with $\lambda_s = c_s$, since $c_s \beta_s = -\sum_{i=0}^{s-1} c_i \beta_i$. Now, if $\lambda_j \beta_j \in K$ for some $j \in \mathbb{Z}$ with $\lambda_j \neq 0$, then we can write

$$\lambda_j \beta_j = \sum_{i=0}^{s-1} \delta_i \beta_i. \qquad (*)$$

Multiplying $(*)$ by c_s and conjugating by α:

$$c_s \lambda_j \beta_{j+1} = c_s \sum_{i=1}^{s} \delta_{i-1} \beta_i = c_s \sum_{i=1}^{s-1} \delta_{i-1} \beta_i - \delta_{s-1} \sum_{i=0}^{s-1} c_i \beta_i \in K,$$

hence by induction our claim is true for all $j \geq 0$. Now conjugating $(*)$ by α^{-1}:

$$\lambda_j \beta_{j-1} = \sum_{i=0}^{s-1} \delta_i \beta_{i-1} = \delta_0 \beta_{-1} + \sum_{i=1}^{s-1} \delta_i \beta_{i-1},$$

and since $c_0\beta = -\sum_{i=1}^{s} c_i\beta_i$,

$$c_0\beta_{-1} = -\sum_{i=1}^{s} c_i\beta_{i-1} \in K,$$

thus $c_0\lambda_j\beta_{j-1} \in K$. Since $c_0 \neq 0$, our claim is thus also true for all $j < 0$. Thus all the generators of J are torsion elements modulo K, so J/K is a torsion group as claimed. So for each $g \in J$ there exists $t_g \in \mathbb{N}$ such that $t_g g \in K$.

Let $g \in \ker\phi$. Since $G = J \rtimes \langle\alpha\rangle$, we can write $g = g'\alpha^n$, where $g' \in J$, $n \in \mathbb{Z}$. So $1 = \phi(g) = \phi(g')a^n$, which (by Observation 4.2) implies $\phi(g') = a^n = 1$, so $n = 0$ and $g \in J$. Note that ϕ is injective on K, since $\phi(\beta_0), \ldots, \phi(\beta_{s-1})$ are linearly independent. Now $\phi(t_g g) = t_g\phi(g) = 0$ and since $t_g g \in K$ and ϕ is injective on K, this implies $t_g g = 0$ and hence $g = 0$, since G is torsion-free. Thus $\ker\phi$ is trivial and H is isomorphic to G. $\square$

More about Gc-groups

We finish this section by proving the embedding result for Gc-groups stated in the previous chapter (Proposition 3.13), together with two useful lemmas that follow from it.

Proof of Proposition 3.13. Define a map $\theta' : \{a, b\} \to Q$ by $\theta'(a) = y$ and $\theta'(b) = x_1$. We shall show that substituting y for a and x_1 for b in each relator in $\mathcal{R}_{\mathbf{c}}(a, b)$ yields the identity of Q, and so θ' extends to a homomorphism from G to Q by Proposition 1.8. Note that making this substitution in the word b^{a^i} yields $x_1^{y^i} = x_{i+1}$.

Firstly, for the relators $[b, b^{a^i}]$, the substitution gives $[x_1, x_1^{y_i}] = 1$, since $x_1^{y^i} \in \mathbb{Q}^s$ for all $i \in \mathbb{Z}$ and $\mathbb{Q}^s$ is abelian. The only other relator in $\mathcal{R}_{\mathbf{c}}(a, b)$ is $\sum_{i=0}^{s} c_i b^{a^i}$, which becomes

$$\begin{aligned}
\sum_{i=0}^{s} c_i x_1^{y^i} &= c_s x_s^y + \sum_{i=1}^{s} c_i x_i \\
&= -c_s \left(\sum_{i=1}^{s} \frac{c_i}{c_s} x_i \right) + \sum_{i=1}^{s} c_i x_i \\
&= \sum_{i=1}^{s} (-c_i x_i + c_i x_i) = 0.
\end{aligned}$$

Thus θ' extends to a homomorphism $\theta : G \to Q$.

It remains to show that θ is injective. As in the proof of Proposition 4.7, let $K = \left\langle \beta, \beta^{\alpha}, \dots, \beta^{\alpha^{s-1}} \right\rangle$. Then θ is injective on K, since $\{\theta(\beta_0), \dots, \theta(\beta_{s-1})\} = \{x_1, \dots, x_s\}$ is a linearly independent subset of $\mathbb{Q}^s$. Given this, the proof that θ is injective is the same as the proof that ϕ is injective in Proposition 4.7. □

The following two results will be valuable in the next chapter.

Lemma 4.8. *Let $G = G(\mathbf{c}) = \langle \alpha, \beta \rangle_{\mathbf{c}}$ be a Gc-group. Then, for any $t \in \mathbb{N}$, the subgroup $H = \langle \alpha, \beta^t \rangle$ has finite index in G and is isomorphic to G.*

Proof. By Proposition 3.13, we can identify G with a subgroup of $\mathbb{Q}^s \rtimes \mathbb{Z}$ for some s, with $\beta \in \mathbb{Q}^s$. Multiplication by t in $\mathbb{Q}^s$ is a $\mathbb{Z}$-module isomorphism of $\mathbb{Q}^s$, and induces an isomorphism $G \to H$. Let $N = G \cap \mathbb{Q}^s$ and $M = H \cap \mathbb{Q}^s$. Then $|G : H| = |N : M|$, because $G \cap \mathbb{Z} = H \cap \mathbb{Z} = \mathbb{Z}$.

Let $\beta_i = \beta^{\alpha^i}$ for $i \in \mathbb{Z}$. Since N is isomorphic to the subgroup of G generated by β_i for all $i \in \mathbb{Z}$ and M is isomorphic to the subgroup of G generated by β_i^t for all $i \in \mathbb{Z}$, and N is abelian, we see that every element of N/M has finite order dividing t.

Now a finitely generated (additive) subgroup K of $\mathbb{Q}^s$, and hence of N, is isomorphic to a subgroup of $\mathbb{Z}^s$. (Let m be the least common multiple of the denominators of all entries in the vectors generating K. Then multiplication by m is an isomorphism of $\mathbb{Q}^s$ which maps K into $\mathbb{Z}^s$.)

So a finitely generated subgroup of N requires at most s elements to generate it. Thus every finitely generated subgroup of N/M is finite of order at most t^s. This implies that N/M is finite of order at most t^s, so $|G : H|$ is finite. $\square$

Lemma 4.9. *Let G be a finitely generated group with finite normal subgroup T such that G/T is isomorphic to the Gc-group $G(\mathbf{c})$. Then G has a finite index subgroup isomorphic to $G(\mathbf{c})$.*

Proof. Suppose that G/T is isomorphic to the Gc-group $G(\mathbf{c}) = \langle \alpha T, \beta T \rangle_{\mathbf{c}}$. By the previous lemma, for any $t \in \mathbb{N}$, $\left\langle \alpha T, \beta^t T \right\rangle$ is isomorphic to $G(\mathbf{c})$ and has finite index in G/T, so by replacing β by a suitable power we may assume $\beta \in C_G(T)$. Define $\beta_i = \beta^{\alpha^i}$ for $i \in \mathbb{Z}$, and denote the exponent of T by m. Then $[\beta_i^m, \beta_j^m] = [\beta_i, \beta_j]^{m^2} = 1$ for all i, j, so the elements β_i^m all commute. Hence, by passing to a finite index subgroup again, we may assume that the β_i all commute. For $\mathbf{v} = (v_0, \ldots, v_r) \in \mathbb{Z}^{r+1}$, let $r_{\mathbf{v}} = \beta_0^{v_0} \beta_1^{v_1} \cdots \beta_r^{v_r}$. Then

$$(\beta_0^m)^{c_0} (\beta_1^m)^{c_1} \cdots (\beta_s^m)^{c_s} = (\beta_0^{c_0} \beta_1^{c_1} \cdots \beta_s^{c_s})^m = r_{\mathbf{c}}^m \in T^m = \{1\}.$$

So by passing to a finite index subgroup a third time we may assume $r_{\mathbf{c}} = 1$. Since any relation in G also holds in G/T, s must be minimal such that $r_{\mathbf{v}} = 1$ for some $\mathbf{v} \in \mathbb{Z}^{s+1}$. Since $\langle \alpha, \beta \rangle$ satisfies Hypothesis $(*)$, Proposition 4.7 tells us that $\langle \alpha, \beta \rangle \cong G(\mathbf{c})$. Thus $\langle \alpha, \beta \rangle$ is torsion-free and so has trivial intersection with T. This means that $\langle \alpha, \beta \rangle \cong G(\mathbf{c})$ is a complement of T in G and hence has finite index in G. $\square$

Henceforth we shall only be interested in Gc-groups which are not virtually abelian. We shall call these *proper Gc-groups*.

4.1.2 Case 2: b is a torsion element

Proposition 4.10. *Let $H = \langle a, b \rangle$ with b a torsion element and assume Hypothesis $(*)$. If H is not virtually abelian, then H has a subgroup isomorphic to $C_p \wr \mathbb{Z}$ for some prime p.*

Proof. The proof is by induction on the order of b. If b has prime order p, then $H \cong C_p \wr \mathbb{Z}$. For if not, then some relation $\sum_{i=r}^{s} \gamma_i b_i = 0$ with $p \not| \gamma_r, \gamma_s$ is true in H. Multiplying by the inverse of γ_s modulo p gives an expression for b_s as a linear combination of $b_r, \ldots, b_{s-1}$, while conjugating by a^{-1} and multiplying by the inverse of γ_r modulo p gives b_{r-1} as a linear combination of $b_r, \ldots, b_{s-1}$, implying that B is generated by $\{b_r, \ldots, b_{s-1}\}$ and H is virtually abelian.

Suppose b to have order np, where p is a prime and $n > 1$. Let $H_1 = \langle a, b^n \rangle$. Since b^n has order p, either $H_1 \cong C_p \wr \mathbb{Z}$ or H_1 is virtually abelian. In the first case we are done, so suppose that H_1 is virtually abelian. Then $\langle b_i^n \mid i \in \mathbb{Z} \rangle$ is finitely generated, hence finite, and so some power of a, say a^k, centralises b^n. If $\left\langle a^k, b \right\rangle$ is virtually abelian, then $\langle b_{ik} \mid i \in \mathbb{Z} \rangle$ is finitely generated, hence finite. But then some power of a^k centralises b and hence B, implying that H is virtually abelian. So we can replace a by a^k and thereby assume that a itself centralises b^n. We then have $b_i^n = b_j^n$ for all $i, j \in \mathbb{Z}$.

Let $b' = b_0 b_1^{-1}$ and $H_2 = \langle a, b' \rangle$. Letting $b'_i = (b')^{a^i}$ and $\widehat{B} = \langle b'_i \mid i \in \mathbb{Z} \rangle$, we have $H_2 = \widehat{B} \rtimes \langle a \rangle$. Observe that

$$a^b = b^{-1} a b = a (b^{-1})^a b = a b_1^{-1} b_0 = a b_0 b_1^{-1} = a b'.$$

Thus $H_2^a = \langle a, (b')^a \rangle = \langle a, b' \rangle = H_2$ and $H_2^b = \left\langle a^b, b' \right\rangle = \langle ab', b' \rangle = \langle a, b' \rangle = H_2$, therefore $H_2 \lhd H$.

For any $g = \sum_{i=r}^{r'} \gamma_i b_i \in B$,

$$g = \sum_{i=r}^{r'} \gamma_i b_0 - \sum_{i=r}^{r'} \gamma_i (b_0 - b_i).$$

Since

$$b_0 - b_i = (b_0 - b_1) + (b_1 - b_2) + \ldots + (b_{i-1} - b_i) = \sum_{j=0}^{i} b'_j \in \widehat{B},$$

we conclude that $B/\widehat{B} \cong \left\langle b\widehat{B} \right\rangle = C_m$ for some m dividing np. So

$$\begin{aligned} H &= \{a^k h \mid k \in \mathbb{Z}, h \in B\} \\ &= \{a^k h' b^j \mid k \in \mathbb{Z}, h' \in \widehat{B}, 0 \leq j \leq m-1\} \\ &= \{g b^j \mid g \in H_2, 0 \leq j \leq m-1\}. \end{aligned}$$

Thus $|H : H_2| \leq m$ and so H_2 cannot be virtually abelian. Also,

$$(b')^n = b_0^n b_1^{-n} = 1,$$

so the order of b' divides n, which strictly divides the order of b. Hence, by induction, there exists $b^\star \in B$ of prime order q such that $\langle a, b^\star \rangle$ is not virtually abelian and is thus isomorphic to $C_q \wr \mathbb{Z}$. □

4.1.3 The general case

In general, we can always reduce to one of the first two cases, using the Hilbert Basis Theorem and Lemma 1.3.

Proposition 4.11. *Let $H = \langle a, b \rangle$ and assume H satisfies Hypothesis $(*)$. Then one of the following holds:*

(i) H has a subgroup isomorphic to $C_p \wr \mathbb{Z}$ for some prime p;

(ii) H has a subgroup that is isomorphic to $\mathbb{Z} \wr \mathbb{Z}$ or to a proper Gc-group.

Proof. If b has finite order, then we are done by Proposition 4.10, so assume b has infinite order. Any torsion elements of H are contained in B, since $H/B \cong \langle a \rangle$ is torsion-free. Since B is abelian, its torsion elements form a subgroup T, which is normal in G.

We can regard B as a $\mathbb{Z}[a]$-module. Since any element of B can be expressed as $b^{p(a)}$ for some polynomial $p(a) \in \mathbb{Z}[a]$, B is finitely generated (by b) as a $\mathbb{Z}[a]$-module. Hence, by the Hilbert Basis Theorem, all submodules of B are finitely generated. In particular, this implies that T is finitely generated as a $\mathbb{Z}[a]$ module. So there is a finite subset $\{t_1, \ldots, t_n\}$ of T such that

$$T = \left\langle t_i^{p(a)} \mid p(a) \in \mathbb{Z}[a], 1 \leq i \leq n \right\rangle = \left\langle t_i^{ka^j} \mid j, k \in \mathbb{Z}, 1 \leq i \leq n \right\rangle.$$

Thus if m_i is the order of t_i and $m = \operatorname{lcm}(m_1, \ldots, m_n)$, any element of T has order dividing m, so T has finite exponent.

If $\langle a, t \rangle$ is not virtually abelian for some $t \in T$, then from Proposition 4.10 we know that H has a subgroup isomorphic to $C_p \wr \mathbb{Z}$ for some prime p. Thus we can assume that $\langle a, t \rangle$ is virtually abelian for all $t \in T$. Then $\left\langle t^{a^i} \mid i \in \mathbb{Z} \right\rangle$ is finitely generated for all $t \in T$, and since T is finitely generated as an $\langle a \rangle$-module, this implies T is finitely generated, hence finite.

Let $H_1 = H/T$. By Lemma 1.3, H_1 cannot be virtually abelian, since H is not virtually abelian. So H_1 is a torsion-free group satisfying Hypothesis $(*)$. Hence, by Proposition 4.7, H_1 is isomorphic to $\mathbb{Z} \wr \mathbb{Z}$ or to a proper Gc-group. In the latter case the result follows immediately from Lemma 4.9; so suppose that $H_1 \cong \mathbb{Z} \wr \mathbb{Z} = \langle \alpha T, \beta T \rangle$ with βT in the base group of the wreath product. By replacing β by a positive power we may assume that $\beta \in C_H(T)$ and then (as in the proof of Lemma 4.9), setting $\beta_i = \beta^{\alpha^i}$ for $i \in \mathbb{Z}$ and $m = |T|$, we have $[\beta_i^m, \beta_j^m] = 1$ for all i, j and so $\langle \alpha, \beta^m \rangle \cong \mathbb{Z} \wr \mathbb{Z}$. $\square$

4.2 The main result

We require two further results in order to deal with metabelian groups that have no subgroups satisfying Hypothesis $(*)$.

Proposition 4.12. *Let G be a finitely generated group with a normal abelian subgroup N such that G/N is free abelian. If for all $a \in G$, $b \in N$, the subgroup $\langle a, b\rangle$ is virtually abelian, then G is polycyclic.*

Proof. Suppose that $\langle a, b\rangle$ virtually abelian for all $a \in G$, $b \in N$. Then every subgroup of $\langle a, b\rangle$ is finitely generated. In particular $\left\langle b^{a^i} \mid i \in \mathbb{Z}\right\rangle$ is finitely generated for any $a \in G$, $b \in N$. We will show that N is finitely generated and thus G is polycyclic.

Since G/N is a quotient of a finitely generated group, it is finitely generated, say by $\{a_1N, \ldots, a_kN\}$. Thus, since G/N is abelian, any $g \in G$ can be written in the form $g = ba_1^{r_1} \ldots a_k^{r_k}$, where $b \in N, r_i \in \mathbb{Z}$. Now, for $b \in N$, consider

$$\begin{aligned} H_b &:= \langle b^g \mid g \in G\rangle \\ &= \left\langle b^{b' a_1^{r_1} \ldots a_k^{r_k}} \mid b' \in N, r_i \in \mathbb{Z}\right\rangle \\ &= \left\langle b^{a_1^{r_1} \ldots a_k^{r_k}} \mid r_i \in \mathbb{Z}\right\rangle. \end{aligned}$$

We know that $\left\langle b^{a_1^{r_1}} \mid r_1 \in \mathbb{Z}\right\rangle$ is finitely generated. Suppose $\left\langle b^{a_1^{r_1} \ldots a_{s-1}^{r_{s-1}}} \mid r_i \in \mathbb{Z}\right\rangle$ is finitely generated, say by $\{y_1, \ldots, y_t\}$. Since all the y_i are in N, $\left\langle y_i^{a_s^{r_s}} \mid r_s \in \mathbb{Z}\right\rangle$ is also finitely generated for all $1 \leq i \leq t$, say

$$\left\langle y_i^{a_s^{r_s}} \mid r_s \in \mathbb{Z}\right\rangle = \langle y_{i1}, \ldots, y_{it_i}\rangle .$$

Then

$$\left\langle b^{a_1^{r_1} \ldots a_s^{r_s}} \mid r_i \in \mathbb{Z}\right\rangle = \langle y_{i1}, \ldots, y_{it_i} \mid 1 \leq i \leq t\rangle$$

and so $\left\langle b^{a_1^{r_1} \ldots a_s^{r_s}} \mid r_i \in \mathbb{Z}\right\rangle$ is also finitely generated. Hence by induction H_b is finitely generated for all $b \in N$.

Since G/N is free abelian of finite rank, it is finitely presented, and hence, by Lemma 1.9, N is the normal closure in G of a finite set of elements. Say $N = \{n_i^g \mid g \in G, 1 \leq i \leq l\} = \{H_{n_i} \mid 1 \leq i \leq l\}$. Each of the finitely many H_{n_i} is finitely generated, and so N itself is finitely generated. Thus G is polycyclic by Observation 1.6, since $G \rhd N \rhd \{1\}$ is a subnormal series for G with finitely generated abelian factors. □

Lemma 4.13. *If a group G is polycyclic and not virtually abelian, then G has a subgroup that is isomorphic to a proper Gc-group.*

Proof. By the second and third paragraphs of the proof of Theorem 16 in [9], there exist $a \in G$ and a non-trivial free abelian subgroup $N \lhd G$ such that $\langle a, N \rangle$ is not virtually abelian. Let $N = \langle n_1, \ldots, n_k \rangle$. If $\langle a, n_i \rangle$ is virtually abelian for some i, then there exist $s_i, t_i \in \mathbb{N}$ such that $a^{s_i} \in C_G(n_i^{t_i})$. If this were true for all i with $1 \leq i \leq k$, then, putting $s = \max\{s_i\}$ and $t = \max\{t_i\}$, $\left\langle a^s, N^t \right\rangle$ would be an abelian subgroup of finite index in $\langle a, N \rangle$, contrary to assumption. So at least one of the subgroups $\langle a, n_i \rangle$ is not virtually abelian. Then since $n_i \in N$, which is a free abelian normal subgroup of G, $\langle a, n_i \rangle$ is torsion-free and satisfies Hypothesis $(*)$ and is thus isomorphic to $\mathbb{Z} \wr \mathbb{Z}$ or a proper Gc-group by Proposition 4.7. But since all subgroups of a polycyclic group are polycyclic, a polycyclic group has no subgroups isomorphic to $\mathbb{Z} \wr \mathbb{Z}$, hence $\langle a, n_i \rangle$ is isomorphic to a proper Gc-group. □

We are now ready to prove that every finitely generated metabelian group which is not virtually abelian has a subgroup which we already know is not poly-$\mathcal{CF}$.

Theorem 4.14. *Let G be a finitely generated metabelian group that is not virtually abelian. Then G has a finitely generated subgroup isomorphic to one of the following:*

(i) $\mathbb{Z} \wr \mathbb{Z}$;

(ii) $C_p \wr \mathbb{Z}$ *for some prime p;*

(iii) A proper Gc-group.

Proof. Since G/G' is a finitely generated abelian group, it is virtually free abelian, so we can replace G by a finite index (and hence finitely generated) subgroup which is again a finitely generated metabelian group, with a normal abelian subgroup N (the old G') such that G/N is free abelian.

If for all $a \in G$, $b \in N$, the subgroup $\langle a, b \rangle$ is virtually abelian, then, by Proposition 4.12 and Lemma 4.13, G has a subgroup isomorphic to a proper Gc-group. So assume that, for some $a \in G$ and $b \in N$, $H := \langle a, b \rangle$ is not virtually abelian. Then H satisfies Hypothesis $(*)$, and the result follows from Proposition 4.11. $\square$

This leads to the proof of Conjecture 4.1 in the case of metabelian groups.

Theorem 4.15. *A finitely generated metabelian group has poly-$\mathcal{CF}$ word problem if and only if it is virtually abelian.*

Proof. We have observed already that all finitely generated virtually abelian groups are poly-$\mathcal{CF}$.

We have shown in Chapter 3 that $\mathbb{Z} \wr \mathbb{Z}$, $C_p \wr \mathbb{Z}$ and proper Gc-groups are not poly-$\mathcal{CF}$ (Propositions 3.9, 3.10 and 3.16). These groups are all finitely generated; so, by Theorem 4.14, every finitely generated metabelian group which is not virtually abelian contains a finitely generated subgroup which is not poly-$\mathcal{CF}$. Since the class of poly-$\mathcal{CF}$ groups is closed under taking finitely generated subgroups (Proposition 1.23), this means that a finitely generated metabelian group which is not virtually abelian cannot have poly-$\mathcal{CF}$ word problem. $\square$

Chapter 5

Poly-$\mathcal{CF}$ soluble groups

In this chapter we prove the torsion-free case of Conjecture 4.1, and present progress towards completing the proof in the general case.

5.1 The torsion-free case

In order to prove that Conjecture 4.1 is true in the torsion-free case, we prove a result for finitely generated torsion-free soluble groups similar to Theorem 4.14.

For G a finitely generated group of derived length n, we shall apply induction to finitely generated subgroups of G'. The result will follow easily, unless all finitely generated subgroups of G' are virtually abelian. Since G' is not necessarily finitely generated, we need some preparatory results about countable locally virtually abelian groups. A group is *locally virtually abelian* if all its finitely generated subgroups are virtually abelian.

5.1.1 Countable locally virtually abelian groups

Proposition 5.1. *Let G be a torsion-free group and A an abelian subgroup of finite index in G. Then $C_G(A)$ is abelian.*

Proof. Since A is abelian, $A \leq Z(C_G(A))$, and hence $Z(C_G(A))$ has finite index in $C_G(A)$. It follows from Lemma 1.10 that $C_G(A)'$ is finite and thus in fact trivial, since G is torsion-free. $\square$

A subgroup H of a group G is *characteristic* in G if H is preserved by all automorphisms of G.

Proposition 5.2. *Let G be a countable torsion-free locally virtually abelian group and suppose that G does not contain free abelian subgroups of arbitrarily high finite rank. Then G has a finite index normal abelian subgroup that is characteristic.*

Proof. Let $k \in \mathbb{N}$ be maximal such that G has a free abelian subgroup of rank k. We can choose elements $g_i \in G$ for $i \in \mathbb{N}$ such that $G = \langle g_i \mid i \in \mathbb{N} \rangle$ and $\langle g_1, \ldots, g_k \rangle \cong \mathbb{Z}^k$.

For $i \geq k$, let $H_i = \langle g_1, \ldots, g_i \rangle$ and let A_i be a subgroup of H_i that is maximal subject to having finite index in H_i and being isomorphic to $\mathbb{Z}^k$. By Proposition 5.1, $C_{H_i}(A_i)$ is abelian, and so the maximality of A_i implies $C_{H_i}(A_i) = A_i$.

Suppose B_i is another rank k abelian subgroup of H_i. Then $A_i \cap B_i$ has finite index in B_i, so has rank k, so has finite index in A_i and hence in H_i. Proposition 5.1 implies that $C_{H_i}(A_i \cap B_i)$ is abelian. But $A_i, B_i \leq C_{H_i}(A_i \cap B_i)$, and so, by the maximality of A_i, we have $C_{H_i}(A_i \cap B_i) = A_i$, so $B_i \leq A_i$. This proves that A_i is the *unique* maximal abelian subgroup of rank k in H_i. It follows that $A_i \trianglelefteq H_i$ and $A_i \leq A_{i+1}$ for all $i \geq k$. Since A_i is self-centralising, H_i/A_i is isomorphic to a subgroup of $\mathrm{Aut}(A_i) \cong \mathrm{GL}(k, \mathbb{Z})$.

We will now show that for each $i \in \mathbb{N}$, H_i/A_i is isomorphic to a subgroup of H_{i+1}/A_{i+1}. For $i \leq k$ this is obvious, since $A_i = H_i \leq H_{i+1} = A_{i+1}$.

For $i \geq k$, let ι_i be the natural embedding of H_i in H_{i+1}, ϕ_i the natural homomorphism from H_i to H_i/A_i and let $\psi_i : H_i/A_i \to H_{i+1}/A_{i+1}$ be given by $xA_i \mapsto xA_{i+1}$. Then $\psi_i \circ \phi_i = \phi_{i+1} \circ \iota_i$ and so $\ker \psi_i \circ \phi_i = \ker \phi_{i+1} \circ \iota_i$. Now

$$\ker \phi_{i+1} \circ \iota_i = H_i \cap \ker \phi_{i+1} = H_i \cap A_{i+1} = A_i.$$

The last equality is due to the maximality of A_i and the fact that $H_i \cap A_{i+1}$ is abelian and contains A_i. Thus $\ker \psi_i \circ \phi_i = A_i$. Since already $\ker \phi_i = A_i$, this implies that ψ_i is injective and so H_i/A_i is isomorphic to $\operatorname{im} \psi_i \leq H_{i+1}/A_{i+1}$.

Hence, since each H_i/A_i is isomorphic to a finite subgroup of $\mathrm{GL}(m, \mathbb{Z})$, Proposition 1.11 implies that there exists $m \in \mathbb{N}$ such that H_i/A_i is isomorphic to H_m/A_m for all $i \geq m$.

We have a commutative diagram:

$$\begin{array}{ccccccccc}
H_m & \xrightarrow{\iota_m} & H_{m+1} & \xrightarrow{\iota_{m+1}} & \dots & \xrightarrow{\iota_{i-1}} & H_i & \xrightarrow{\iota_i} & H_{i+1} \\
\downarrow{\scriptstyle \phi_m} & & \downarrow{\scriptstyle \phi_{m+1}} & & \downarrow & & \downarrow{\scriptstyle \phi_i} & & \downarrow{\scriptstyle \phi_{i+1}} \\
H_m/A_m & \xrightarrow{\psi_m} & H_{m+1}/A_{m+1} & \xrightarrow{\psi_{m+1}} & \dots & \xrightarrow{\psi_{i-1}} & H_i/A_i & \xrightarrow{\psi_i} & H_{i+1}/A_{i+1}
\end{array}$$

For $i \geq m$, define $\theta_i : H_i \to H_m/A_m$ by $\theta_m = \phi_m$ and for $i > m$

$$\theta_i = \psi_m^{-1} \circ \psi_{m+1}^{-1} \dots \circ \psi_{i-1}^{-1} \circ \phi_i.$$

Since each ϕ_i is an epimorphism and all the ψ_j^{-1} $(j \geq m)$ are isomorphisms, each θ_i is an epimorphism. Moreover, since $\phi_i = \psi_i^{-1} \circ \phi_{i+1} \circ \iota_i$, we have

$$\begin{aligned}
\theta_{i+1} \circ \iota_i &= \psi_m^{-1} \circ \psi_{m+1}^{-1} \dots \circ \psi_i^{-1} \circ \phi_{i+1} \circ \iota_i \\
&= \psi_m^{-1} \circ \psi_{m+1}^{-1} \dots \circ \psi_{i-1}^{-1} \circ \phi_i = \theta_i,
\end{aligned}$$

which means that $\theta_{i+1}(x) = \theta_i(x)$ for $x \in H_i$.

Define $\theta : G \to H_m/A_m$ by $\theta(x) = \theta_i(x)$ where $i = \min\{j \geq m \mid x \in H_j\}$. Then θ is a well-defined epimorphism with kernel $A := \cup_{i\in\mathbb{N}} A_i$. So A is abelian of finite index in G.

Finally, we show that A is characteristic in G. Let B be an abelian subgroup of G containing a subgroup isomorphic to $\mathbb{Z}^k$, with generating set $\{h_i \mid i \in \mathbb{N}\}$, where $\langle h_1, \ldots, h_k\rangle \cong \mathbb{Z}^k$. Let $B_i = \langle h_1, \ldots, h_i\rangle$ for $i \in \mathbb{N}$. Each B_i, being finitely generated, is contained in some H_{j_i}, and hence in A_{j_i}. Now since $B_i \leq B_{i+1}$ for all $i \in \mathbb{N}$,

$$B = \cup_{i\in\mathbb{N}} B_i \leq \cup_{i\in\mathbb{N}} A_{j_i} \leq \cup_{i\in\mathbb{N}} A_i = A.$$

Thus A is the unique maximal abelian subgroup of G containing a $\mathbb{Z}^k$ subgroup and so is characteristic in G. □

Next we deal with groups which do have subgroups isomorphic to $\mathbb{Z}^k$ for all $k \in \mathbb{N}$.

Observation 5.3. *An automorphism of a free abelian group of finite rank that centralises a subgroup of finite index is trivial.*

Proof. Let G be a free abelian group of rank k and let H be a finite index subgroup of G. Let α be an automorphism of G which centralises H. We can regard G and H as subgroups of $\mathbb{Q}^k$, so α can thus be considered as an element of $\mathrm{GL}(k, \mathbb{Q})$. A finite index subgroup of a free abelian group of rank k is also free abelian of rank k, so any generating set for H is a basis for $\mathbb{Q}^k$ over $\mathbb{Q}$. Since α centralises H, it fixes a basis of $\mathbb{Q}^k$ and thus also fixes the whole of $\mathbb{Q}^k$ and in particular is trivial on G. □

Lemma 5.4. *Let G be a countable torsion-free locally virtually abelian group and let N be an abelian normal subgroup of G such that $G/N \cong \mathbb{Z}^\infty$. Then G has a subgroup isomorphic to $\mathbb{Z}^\infty$.*

(We recall that $\mathbb{Z}^\infty$ denotes a free abelian group of countably infinite rank.)

Proof. If N contains $\mathbb{Z}^k$ for all $k \in \mathbb{N}$ then we are done, since N is abelian; so assume t is maximal with $\mathbb{Z}^t \leq N$. Write $G = \bigcup_{i \in \mathbb{N}} G_i$, with each G_i finitely generated (and hence virtually abelian), $G_i < G_{i+1}$ for all $i \in \mathbb{N}$, and each $N_i := G_i \cap N$ isomorphic to $\mathbb{Z}^t$.

Let $A_i = C_{G_i}(N_i)$ and let B_i be an abelian subgroup of finite index in G_i. Then B_i centralises the finite index subgroup $B_i \cap N_i$ of N_i. Hence, by Observation 5.3, B_i must centralise N_i. Thus $B_i \leq A_i$, so $|G_i : A_i|$ is finite.

Each N_i has finite index in N_{i+1}; so, by Observation 5.3, for all $i \in \mathbb{N}$ we have $A_i \leq C_{G_{i+1}}(N_{i+1}) = A_{i+1}$.

If A_i is not abelian, then there exist $x, y \in A_i$ with $[x, y] = z \in N_i \setminus \{1\}$. But then, since $z \in Z(A_i)$, the subgroup of A_i generated by x, y and z is

$$\langle x, y, z \mid [x, y] = z, [x, z] = [y, z] = 1 \rangle ,$$

which is the Heisenberg group, which is not virtually abelian. But this contradicts the hypothesis that all finitely generated subgroups of G are virtually abelian. Thus A_i must be abelian. Hence $\bigcup_{i \in \mathbb{N}} A_i$ is abelian and contains finitely generated subgroups of arbitrarily high rank; so it must contain $\mathbb{Z}^\infty$. $\square$

Lemma 5.5. *Let G be a locally virtually abelian group with a normal torsion subgroup T such that $G/T \cong \mathbb{Z}^\infty$. Then G has a subgroup isomorphic to $\mathbb{Z}^\infty$.*

Proof. Let $g_1, g_2, \ldots$ be an irredundant generating set for G modulo T. Then, for all $i \in \mathbb{N}$, $G_i := \langle g_1, \ldots, g_i \rangle$ is virtually abelian; so $T_i := G_i \cap T$ is finite. Also $G_i/T_i \cong \mathbb{Z}^i$.

We shall construct a chain of subgroups $H_1 < H_2 < \cdots < H_i < \cdots$ such that $H_i \leq G_i$ with $|G_i : H_i|$ finite and $H_i \cong \mathbb{Z}^i$. Suppose that we have defined $H_1, \ldots, H_i$ for some $i \geq 0$. Then $g_{i+1} \in G_{i+1}$ centralises $H_i T_{i+1}/T_{i+1}$, while some power g_{i+1}^k

of g_{i+1} centralises T_{i+1}. If $l = k|T_{i+1}|$, then, for any $h \in H_i$, there is some $t \in T_{i+1}$ such that $h^{g_{i+1}} = ht$, and so

$$\begin{aligned} h^{g_{i+1}^l} &= htt^{g_{i+1}}t^{g_{i+1}^2}\dots t^{g_{i+1}^l} \\ &= h\left(tt^{g_{i+1}}\dots t^{g_{i+1}^{k-1}}\right)^{|T_{i+1}|} = h, \end{aligned}$$

hence g_{i+1}^l centralises H_i. Thus $H_{i+1} := \left\langle H_i, g_{i+1}^l \right\rangle$ has finite index in G_{i+1} and is isomorphic to $\mathbb{Z}^{i+1}$.

So we can construct the chain as claimed, and $\bigcup_{i\in\mathbb{N}} H_i \cong \mathbb{Z}^\infty$. □

Proposition 5.6. *Let G be a soluble group with all finitely generated subgroups virtually abelian, and suppose that G has subgroups isomorphic to $\mathbb{Z}^k$ for all $k \in \mathbb{N}$. Then G has a subgroup isomorphic to $\mathbb{Z}^\infty$.*

Proof. The proof is by induction on the derived length of G. The statement is true for abelian groups. Now suppose it is true for groups of derived length at most n, and let G be locally virtually abelian of derived length $n+1$, with $\mathbb{Z}^k \leq G$ for all $k \in \mathbb{N}$.

Let $N = G^{(n)}$. If $\mathbb{Z}^k \leq N$ for all $k \in \mathbb{N}$ then we are done, since N is abelian. So suppose there exists $t \in \mathbb{N}$ maximal such that $\mathbb{Z}^t \in N$. This means that G/N must contain $\mathbb{Z}^k$ for all $k \in \mathbb{N}$. By the induction hypothesis, this means that G/N has a subgroup isomorphic to $\mathbb{Z}^\infty$. We can assume without loss of generality that $G/N \cong \mathbb{Z}^\infty$.

Let T be the torsion subgroup of N. Then N/T is torsion-free, and so G/T and N/T satisfy the hypothesis of Lemma 5.4. Thus we can replace G by a subgroup and assume $G/T \cong \mathbb{Z}^\infty$. Then, by Lemma 5.5, G has a subgroup isomorphic to $\mathbb{Z}^\infty$. □

5.1.2 Finitely generated torsion-free soluble groups

We can now prove our main result for torsion-free finitely generated soluble groups.

Theorem 5.7. *Let G be a finitely generated torsion-free soluble group. Then at least one of the following holds:*

(i) G is virtually abelian;

(ii) G' has a subgroup isomorphic to $\mathbb{Z}^\infty$;

(iii) G has a subgroup isomorphic to a proper Gc-group.

Proof. Let n be the derived length of G. The proof is by induction on n. The statement is vacuously true for $n = 1$, and true for $n = 2$ (G metabelian) by Theorem 4.14. So assume that $n \geq 2$ and that the result is true for groups with derived length less than n. By applying the result inductively to finitely generated subgroups of G', we may assume that G' is locally virtually abelian.

If G' contains $\mathbb{Z}^k$ for all $k \in \mathbb{N}$, then G' has a subgroup isomorphic to $\mathbb{Z}^\infty$ by Proposition 5.6.

Now suppose that G' does not contain $\mathbb{Z}^k$ for all $k \in \mathbb{N}$. Since G' is countable, Proposition 5.2 shows that G' has a finite index normal abelian subgroup A which is characteristic in G' and hence $A \lhd G$. Now G/A is finitely generated with $(G/A)/(G'/A) \cong G/G'$ abelian, while G'/A is finite. So G/A satisfies the hypotheses of Lemma 1.3, with $T = G'/A$, and is hence virtually abelian. But since A is an abelian normal subgroup of G, this implies that G itself is virtually metabelian, and the result follows from Theorem 4.14. $\square$

This gives us the torsion-free case of Conjecture 4.1.

Theorem 5.8. *If G is a finitely generated torsion-free soluble group, then G is poly-$\mathcal{CF}$ if and only if G is virtually abelian.*

Proof. Recall that the poly-$\mathcal{CF}$ groups are closed under taking finitely generated subgroups (Proposition 1.23). If G' has a $\mathbb{Z}^\infty$ subgroup, then G is not poly-$\mathcal{CF}$ by Lemma 3.8. If G has a subgroup isomorphic to a proper Gc-group, then G is not poly-$\mathcal{CF}$ by Proposition 3.16. $\square$

5.2 The main result

We require just one further lemma before proceeding to prove our result for finitely generated soluble groups in general.

Lemma 5.9. *Let G be a group with normal subgroup N and let H be a subgroup of $(G/N)'$. Then G' has a subgroup K such that $K/(N \cap K) \cong H$.*

Proof. Firstly,

$$\begin{aligned}(G/N)' &= \langle [xN, yN] \mid x, y \in G \rangle \\ &= \langle [x, y]N \mid x, y \in G \rangle \\ &= \big\langle hnN \mid h \in G', n \in N \big\rangle = G'N/N.\end{aligned}$$

And $G'N/N \cong G'/(G' \cap N)$, so H is isomorphic to a subgroup of $G'/(G' \cap N)$. Let K be a preimage of this subgroup in G'. Then $K/(N \cap K)$, which is the image of K under the natural homomorphism from G' to $G'/(G' \cap N)$, is isomorphic to H. $\square$

Theorem 5.10. *Let G be a finitely generated soluble group. Then at least one of the following holds:*

(i) G is virtually abelian;

(ii) G' *has a subgroup isomorphic to* $\mathbb{Z}^\infty$*;*

(iii) G *has a subgroup isomorphic to a proper Gc-group;*

(iv) G *has a finitely generated subgroup* H *with an infinite normal torsion subgroup* U*, such that* H/U *is either free abelian or a proper Gc-group.*

Proof. The proof is by induction on the derived length of G. We have already proved it for groups of derived length at most two and when G is torsion-free (Theorems 4.14 and 5.7). Suppose the result holds for groups of derived length at most $n \geq 2$, and let G have derived length $n+1$ and be not virtually abelian. So the result holds for G/N, where $N = G^{(n)}$, and we may assume that G' is locally virtually abelian. Let T be the torsion subgroup of N. For the remainder of the proof, we shall only use the fact that N is abelian and G/N has derived length n, and not that $N = G^{(n)}$. This enables us to replace G by a subgroup when convenient.

If (i) holds for G/N, then G is virtually of derived length 2, and the result follows from Theorem 4.14.

Suppose that (ii) holds for G/N; that is, $(G/N)'$ contains a subgroup isomorphic to $\mathbb{Z}^\infty$. Applying Lemma 5.9, let K be a subgroup of G' such that $K/(N \cap K) \cong \mathbb{Z}^\infty$. Then $K/(T \cap K)$ is torsion-free and so, by Lemma 5.4, there is also a $\mathbb{Z}^\infty$ subgroup in $K/(T \cap K)$. By Lemma 5.5, this implies that K (and hence G') has a subgroup isomorphic to $\mathbb{Z}^\infty$ and so (ii) holds for G.

Suppose that (iii) holds for G/N. Then, by replacing G by a subgroup, we may assume that G/N is isomorphic to a proper Gc-group. So G/T is torsion-free and not virtually abelian, and hence, by Theorem 5.7, one of (ii), (iii) holds for G/T.

If (ii) holds for G/T, then (ii) holds for G by Lemma 5.5. So suppose that (iii) holds for G/T. By passing to a subgroup we may assume that G/T is a proper Gc-group. If T is finite, then the result follows from Lemma 4.9, whereas if T is infinite, then

G satisfies (iv).

Finally, suppose that G/N satisfies (iv). We may assume that G/N has an infinite normal torsion subgroup U/N, where G/U is either free abelian or a proper Gc-group. If $N = T$, then U is a torsion group and G satisfies (iv); so suppose not. Since N is abelian, if N/T contains subgroups isomorphic to $\mathbb{Z}^k$ for all $k \in \mathbb{N}$, then N contains a $\mathbb{Z}^\infty$ subgroup; so we may assume that there is a largest $k > 0$ such that $\mathbb{Z}^k \leq N/T$.

Suppose that U is generated by elements g_i ($i \in \mathbb{N}$) and, for $i \in \mathbb{N}$, define $U_i = \langle g_1, \ldots, g_i \rangle$ and $N_i = U_i \cap N$. A finitely generated soluble torsion group is finite (Lemma 1.7), so the groups $U_i/N_i \cong U_iN/N$ are all finite with $\cup_{i\in\mathbb{N}} U_iN/N = U/N$. The groups $N_iT/T \cong N_i/(N_i \cap T)$ are all finitely generated free abelian groups, since $N_i \cap T$ is the torsion subgroup of the finitely generated abelian group N_i. Since $\bigcup_{i\in\mathbb{N}} N_iT/T = N/T$, for all sufficiently large i we have $N_iT/T \cong \mathbb{Z}^k$.

For $i \in \mathbb{N}$, define the subgroup C_i by $U_i \cap T \leq C_i \leq U_i$ and

$$C_iT/T = C_{U_iT/T}(N_iT/T).$$

By Proposition 1.11, there is a constant L such that $|U_i : C_i| \leq L$ for all i. Since $N_i \leq N_{i+1}$ for each i, we have $C_{i+1} \cap U_i \leq C_i$, and hence $|U_i : C_i| \leq |U_{i+1} : C_{i+1}|$. So, for sufficiently large i (say $i \geq \kappa$), we have $N_iT/T \cong \mathbb{Z}^k$ and $|U_i : C_i| = L$ for some constant L, and so we must also have $C_{i+1} \cap U_i = C_i$. Let $C := \cup_{i\geq\kappa} C_i$. Then $T < C$, since $U_i \cap T \leq C_i$ and $U = \cup_{i\geq\kappa} U_i$. Also,

$$C/T = \bigcup_{i\geq\kappa} C_iT/T = \bigcup_{i\geq\kappa} C_{U_iT/T}(N_iT/T) = C_{U/T}(N/T),$$

and $|U : C| = L$.

Now since C/T is the centraliser in U/T of the normal subgroup N/T of G/T, we have $C/T \lhd G/T$ and hence $C \lhd G$. By applying Lemma 1.3 to G/C if G/U is free abelian, and applying Lemma 4.9 to G/C if G/U is a proper Gc-group, we can replace G by a finite index subgroup containing C such that $G \cap U = C$, and thereby assume that $U = C$ and hence $U_i = C_i$ for all i. So $N_iT/T \leq Z(U_iT/T)$ for each i, thus $U_i'T/T$ is finite by Lemma 1.10. Since $U' = \cup_{i \in \mathbb{N}} U_i'$, $U'T/T$ and hence also $U'T$ is a torsion group. Let $V/U'T$ be the torsion subgroup of the abelian group $U/U'T$. So U/V and hence also G/V is torsion-free, and hence by Theorem 5.7 one of (i), (ii), (iii) holds for G/V.

If G/V is virtually abelian then, since G is not virtually abelian, V must be infinite by Lemma 1.3, and so G satisfies (iv). If G/V satisfies (ii) then so does G by Lemmas 5.9 and 5.5. If (iii) holds for G/V then, by replacing G by a subgroup, we may assume that G/V is isomorphic to a proper Gc-group. Then the result follows by Lemma 4.9 if V is finite, and G satisfies (iv) if V is infinite. $\square$

We have a slightly unsatisfactory corollary to this theorem, which encompasses Theorems 4.15 and 5.8.

Corollary 5.11. *If G is a finitely generated poly-$\mathcal{CF}$ soluble group, then either G is virtually abelian, or G has finitely generated subgroup H with an infinite normal torsion subgroup U, such that H/U is either free abelian or isomorphic to a proper Gc-group. The second case does not occur for metabelian or torsion-free soluble groups.*

We conjecture that the second case does not occur at all, but have been unable to prove this so far.

5.3 An example of case (iv) in Theorem 5.10

We give a proof of non-poly-$\mathcal{CF}$-ness in a specific example of the fourth case of Theorem 5.10.

Proposition 5.12. *Let p be a prime and let G be the group given by the following presentation.*

$$\begin{aligned}\langle a, b_i\ (i \in \mathbb{Z}), c_j\ (j>0) \mid\ & b_i^a = b_{i+1}\ (i \in \mathbb{Z}),\ [b_i, b_{i+j}] = c_j\ (i \in \mathbb{Z}, j > 0),\\ & b_i^p = c_j^p = 1\ (i \in \mathbb{Z}, j>0),\ c_j \text{ central } (j>0)\rangle.\end{aligned}$$

Then G has derived length 3 and satisfies (iv) of Theorem 5.10, and is not poly-$\mathcal{CF}$.

Proof. In this proof, we shall always assume that the indices on the right hand side of a presentation run over all available values (specified on the left hand side). This prevents the presentations from becoming too cluttered. With this convention, the presentation for G is simplified to

$$\left\langle b_i\ (i \in \mathbb{Z}), c_j\ (j>0) \mid b_i^a = b_{i+1},\ [b_i, b_{i+j}] = c_j, b_i^p = c_j^p = 1,\ c_j \text{ central}\right\rangle.$$

Let H be the group defined by the subpresentation

$$\left\langle b_i\ (i \in \mathbb{Z}), c_j\ (j>0) \mid [b_i, b_{i+j}] = c_j, b_i^p = c_j^p = 1,\ c_j \text{ central}\right\rangle.$$

Then a acts on H by conjugation as an automorphism of infinite order, and so $G \cong H \rtimes \langle a \rangle$ and $G/H \cong \mathbb{Z}$. Thus G satisfies the fourth case of Theorem 5.10, with $U = H$.

Since $G \rhd H \rhd \langle c_j\ (j>0)\rangle \rhd \{1\}$ is a normal series for G with abelian factors, G has derived length at most 3.

By results on 'Darstellungsgruppen' (covering groups) in [13, Chapter V.23], in the group E_n given by the presentation

$$\left\langle b_i\ (-n \leq i \leq n), c_{ij}\ (-n \leq i < j \leq n) \mid [b_i, b_j] = c_{ij}, b_i^p = c_{ij}^p = 1, c_{ij} \text{ central}\right\rangle,$$

the subgroup generated by all the c_{ij} (which is E_n') has the abelian presentation $\mathrm{Ab}\left\langle c_{ij}\ (-n \leq i < j \leq n) \mid c_{ij}^p\right\rangle$.

Let E be the union of the ascending sequence of groups $E_1, E_2, E_3, \ldots$. Then $E' = \cup_{i \in \mathbb{N}} E'_n$, with presentation $\mathrm{Ab}\left\langle c_{ij}\ (i, j \in \mathbb{Z},\ i < j) \mid c_{ij}^p \right\rangle$.

Our subgroup H of G is obtained from E by quotienting out the subgroup $N := \left\langle c_{0,j-i} c_{ij}^{-1} \mid i < j \right\rangle$ and setting $c_j = c_{0j}$ for all $j > 0$. The subgroup of H generated by all the c_j is isomorphic to E'/N, and thus has abelian presentation $\mathrm{Ab}\left\langle c_j\ (j > 0) \mid c_j^p \right\rangle$. In particular, all c_j are non-trivial and so H is not abelian, and therefore G has derived length 3.

Let $b = b_0$, $B = B_0$ and let M_k be the sublanguage of

$$W_k = (BA^*Ba^*bA^*ba^*)^k (BA^*ba^*bA^*Ba^*)^k$$

consisting of all those words

$$(BA^{m_1}Ba^{n_1}bA^{\mu_1}ba^{\nu_1}) \ldots (BA^{m_k}Ba^{n_k}bA^{\mu_k}ba^{\nu_k})(BA^{m_{k+1}}ba^{n_{k+1}}bA^{\mu_{k+1}}Ba^{\nu_{k+1}})$$
$$\ldots (BA^{m_{2k}}ba^{n_{2k}}bA^{\mu_{2k}}Ba^{\nu_{2k}})$$

such that

(i) $m_i = n_i = \mu_i = \nu_i$ for all i;

(ii) $m_i < m_{i+1}$ for $i \notin \{k, 2k\}$.

The first condition can be checked by two pushdown automata, one checking that $m_i = n_i$ and $\mu_i = \nu_i$ for all i, and the other checking that $m_i = \mu_i$ for all i. The second condition can be checked by a single pushdown automaton. Thus M_k is the intersection of 2-$\mathcal{CF}$ language, a 1-$\mathcal{CF}$ language and a regular language, and is hence 3-$\mathcal{CF}$.

Now a word in M_k is of the form

$$(BA^{m_1}Ba^{m_1}bA^{m_1}ba^{m_1}) \ldots (BA^{m_k}Ba^{m_k}bA^{m_k}ba^{m_k})(BA^{m_{k+1}}ba^{m_{k+1}}bA^{m_{k+1}}Ba^{m^{k+1}})$$
$$\ldots (BA^{m_{2k}}ba^{m_{2k}}bA^{m_{2k}}Ba^{m^{2k}}),$$

with $m_i < m_{i+1}$ and $m_{k+i} < m_{k+i+1}$ for $1 \leq i \leq k-1$. This is equal in G to

$$[b, b_{m_1}] \cdots [b, b_{m_k}][b, B_{m_{k+1}}] \cdots [b, B_{m_{2k}}] = c_{m_1} \cdots c_{m_k} c_{m_{k+1}}^{-1} \cdots c_{m_{2k}}^{-1}.$$

Let L_k be the commutative image of the intersection of $W(G)$ with M_k. As in the proof of Proposition 3.10, we can ignore the b's and B's and take L_k to be a subset of $\mathbb{N}_0^{8k}$. Since the c_{m_i} are distinct for $1 \leq i \leq k$ and

$$\langle c_j \mid j > 0 \rangle = \mathrm{Ab} \left\langle c_j \ (j > 0) \mid c_j^p \ (j > 0) \right\rangle,$$

the only way that a word in M_k can be in $W(G)$ is if some $m_{k+j} = m_i$ for each $1 \leq i \leq k$. But since $m_i < m_{i+1}$ and $m_{k+i} < m_{k+i+1}$ for $1 \leq i \leq k-1$, this implies that $m_i = m_{k+i}$ for $1 \leq i \leq k$ and so L_k is the set of all $8k$-tuples of the form

$$(m_1, m_1, m_1, m_1, \ldots, m_k, m_k, m_k, m_k, m_1, m_1, m_1, m_1, \ldots, m_k, m_k, m_k, m_k),$$

with $m_i \in \mathbb{N}_0$, and $m_i < m_{i+1}$ for $1 \leq i \leq k-1$. Thus L_k is a k-dimensional linear subset of the set $S^{(4,k)}$ introduced in Section 2.6, and is therefore not an intersection of $k-1$ stratified semilinear sets by Corollary 2.18.

By Corollary 2.1, this means that $W(G) \cap M_k$ is not $(k-1)$-$\mathcal{CF}$. Since M_k is 3-$\mathcal{CF}$, this implies that $W(G)$ is not $(k-4)$-$\mathcal{CF}$ for any $k \in \mathbb{N}$. Hence G is not poly-$\mathcal{CF}$. $\square$

Quotienting out a proper subgroup of $\langle c_j \ (j > 0) \rangle$ in the group G in Proposition 5.12 results in another group of derived length 3 satisfying (iv) of Theorem 5.10. We do not know how to show that such quotients are not poly-context-free except in some very specific cases.

5.4 Future directions

In order to complete the proof of Conjecture 4.1, we need only show that a finitely generated soluble group having an infinite torsion subgroup U such that G/U is either free abelian or isomorphic to a proper Gc-group is not poly-$\mathcal{CF}$.

One way of approaching this which looks promising would be to show that a poly-$\mathcal{CF}$ group cannot have an infinite torsion subgroup. We know that context-free groups cannot have infinite torsion subgroups, because they are virtually free.

Actually, we conjecture something stronger, which again is true in the case of context-free groups.

Conjecture 5.13. *If a group G is poly-$\mathcal{CF}$, then G does not have arbitrarily large finite subgroups.*

So far, the author's approaches towards this conjecture, from the perspective of automata theory, have not succeeded. It may be that an approach using grammars would be more fruitful.

Bibliography

[1] V.A. Anisimov, 'The group languages', *Kibernetika* 4 (1971) 18–24.

[2] M.F. Atiyah and I.G. Macdonald, *Introduction to commutative algebra* (Addison-Wesley, 1969).

[3] M. Dunwoody, 'The accessibility of finitely presented groups', *Invent. Math.* 81 (1985) 449–457.

[4] S. Friedland, 'The maximal orders of finite subgroups in $GL_n(\mathbb{Q})$', *Proc. Amer. Math. Soc.* 125 (1997) 3519–3526.

[5] S. Ginsburg, S.A. Greibach and M.A. Harrison, 'One-way stack automata', *J. ACM* 14 (1967) 389–418.

[6] S. Ginsburg, *The mathematical theory of context-free languages*, (McGraw-Hill 1966).

[7] S. Ginsburg and E.H. Spanier, 'Semigroups, Presburger formulas, and languages', *Pacif. J. Math.* 16 (1966) 285–296.

[8] T. Herbst, 'On a subclass of context-free groups', *Theoret. Inform. Appl.* 25 (1991) 255–272.

[9] D.F. Holt, S. Rees, C.E. Röver and R.M. Thomas, 'Groups with context-free co-word problem', *J. London Math. Soc.* (2) 71 (2005) 643–657.

[10] D.F. Holt and C.E. Röver, 'Groups with indexed co-word problem', *Internat. J. Algebra Comput.* 16 (2006), no.5, 985–1014.

[11] D.F. Holt, M.D. Owens and R.M. Thomas, 'Groups and semigroups with a one-counter word problem', *J. Aust. Math. Soc.* 85 (2008) 197–209.

[12] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, 1979).

[13] B. Huppert, *Endliche Gruppen I*, (Springer-Verlag, 1967).

[14] J. Lehnert and P. Schweitzer, 'The co-word problem for the Higman-Thompson group is context-free', *Bull. Lond. Math. Soc.* 39 (2007), no.2, 235–241.

[15] L.Y. Liu and P. Weiner, 'An infinite heirarchy of intersections of context-free languages', *Math. Systems Theory* 7 (1973) 185–192.

[16] R.C. Lyndon and P.E. Schupp, *Combinatorial Group Theory*, (Springer-Verlag, 1977).

[17] D.E. Muller and P.E. Schupp, 'Groups, the theory of ends, and context-free languages', *J. Comput. System Sci.* 26 (1983) 295–310.

[18] D.E. Muller and P.E. Schupp, 'The theory of ends, pushdown automata, and second-order logic', *Theoret. Comput. Sci.* 37 (1985) 51–75.

[19] D.L. Johnson, *Presentations of Groups* (London Mathematical Society, 1997).

[20] R.J. Parikh, 'Language-generating devices', *MIT Res. Lab. Electron. Quart. Prog. Rep.* 60 (1961) 199–212.

[21] D.W. Parkes and R.M. Thomas, 'Groups with context-free reduced word problem', *Comm. Algebra* 39 (2002), no.7, 3143–3156.

[22] D.J.S. Robinson, *A course in the theory of groups* (Springer-Verlag, New York 1982).

[23] J.C. Lennox and D.J.S. Robinson. *The Theory of Infinite Soluble Groups.* (Clarendon Press, Oxford 2004).

[24] D.N. Rockmore and K.S. Tan, 'A note on the order of finite subgroups of $GL(n, \mathbb{Z})$', *Arch. Math.* 64 (1995) 283–288.

[25] J.S. Rose, *A course in group theory* (Cambridge University Press, 1978).

[26] J.J. Rotman, *Advanced Modern Algebra* (Prentice Hall, 2002).