PRESENTATIONS FOR SUBSEMIGROUPS OF GROUPS

Alan James Cain

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2005

ALAN JAMES CAIN
M.Sci. (Glasguensis)

PRESENTATIONS FOR SUBSEMIGROUPS OF GROUPS

Ph.D. Thesis
University of St Andrews
11th July 2005

Th F28

To many absent friends.

DECLARATIONS

I, Alan James Cain, hereby certify that this thesis, which is approximately 57 000 words in length, has been written by me, that it is the record of work carried out by me, and that it has not been submitted in any previous application for a higher degree.

DATE: 11/07/2005 SIGNATURE OF CANDIDATE:
NAME: Alan J. Cain

I was admitted as a research student in September 2002 and as a candidate for the degree of Doctor of Philosophy in September 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2002 and 2005.

DATE: 11/07/2005 SIGNATURE OF CANDIDATE:
NAME: Alan J. Cain

We hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

DATE: 11|07|05 SIGNATURE OF SUPERVISOR:
NAME: Prof. Edmund F. Robertson

DATE: 29/06/05 SIGNATURE OF SUPERVISOR:
NAME: Prof. Nikola Ruškuc

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

DATE: 11/07/2005 SIGNATURE OF CANDIDATE:
NAME: Alan J. Cain

ABSTRACT

THIS THESIS studies subsemigroups of groups from three perspectives: automatic structures, ordinary semigroup presentations, and Malcev presentaions. [A Malcev presentation is a presentation of a special type for a semigroup that can be embedded into a group. A group-embeddable semigroup is Malcev coherent if all of its finitely generated subsemigroups admit finite Malcev presentations.]

The theory of synchronous and asynchronous automatic structures for semigroups is expounded, particularly for group-embeddable semigroups. In particular, automatic semigroups embeddable into groups are shown to inherit many of the pleasant geometric properties of automatic groups. It is proved that group-embeddable automatic semigroups admit finite Malcev presentations, and such presentations can be found effectively. An algorithm is exhibited to test whether an automatic semigroup is a free semigroup. Cancellativity of automatic semigroups is proved to be undecidable.

Study is made of several classes of groups: virtually free groups; groups that satisfy semigroup laws (in particular [virtually] nilpotent and [virtually] abelian groups); polycyclic groups; free and direct products of certain groups; and one-relator groups. For each of these classes, the question of Malcev coherence is considered, together with the problems of whether finitely generated subsemigroups are finitely presented or automatic. This study yields closure and containment results regarding the class of Malcev coherent groups.

The property of having a finite Malcev presentation is shown to be preserved under finite Rees index extensions and subsemigroups. Other concepts of index are also studied.

CONTENTS

PREFACE

It is not my intention to detain the reader by expatiating on the variety, or the importance of the subject, which I have undertaken to treat; since the merit of the choice would serve to render the weakness of the execution still more apparent, and still less excusable.

— Edward Gibbon,
Decline and Fall of the Roman Empire (1776–88),
vol. i, preface

THIS THESIS is largely concerned with subsemigroups of groups, and from its pages a reader may discover much of their character and a little of their history. The title is perhaps a little restrictive: the body of the thesis approaches subsemigroups of groups from three directions: 'ordinary' semigroup presentations, Malcev presentations, and automatic structures.

Malcev presentations are semigroup presentations of a special type for group-embeddable semigroups, introduced by Spehner (1977). Informally, whilst an 'ordinary' semigroup presentation defines a semigroup by means of generators and defining relations, a Malcev presentation defines a semigroup using generators, defining relations, and a rule of group-embeddability. This rule of group-embeddability is worth an infinite number of defining relations, in the sense that a semigroup may admit a finite Malcev presentation but no finite ordinary presentation. During the three decades since Spehner's definition, little research was carried out in the area. This thesis should convince the reader that this neglect has been unfair. In preparation for the main body of the thesis, Chapter 1 formally defines Malcev presentations and establishes their basic properties.

Campbell, Robertson, Ruškuc & Thomas (2001) extended the theory of synchronous automatic structures from groups to semigroups, and Hoffmann, Kuske, Otto & Thomas (2002*a*) defined asynchronous automatic structures for monoids, generalizing such structures for groups. Their investigations showed that many of the pleasant properties of synchronous and asynchronous automatic groups are lost when one passes to semigroups or monoids. In particular, automatic structures for groups have an elegant geometric characterization known as the 'fellow traveller property'. This characterization does not extend to automatic structures for arbitrary semigroups. Chapter 2 treats of synchronous and asynchronous automatic semigroups, and shows that if such semigroups embed into groups, then they retain the geometric properties just mentioned. An important link between Malcev presentations and automatic structures is then established: every automatic semigroup embeddable into a group admits a finite Malcev presentation (Theorem 2.5.1). Chapter 3 then considers various algorithmic questions for automatic semigroups. In particular, Section 3.5 shows that left-cancellativity is undecidable for automatic semigroups.

The core of the thesis, Chapters 4–8, takes a structural approach. In successive chapters are considered: virtually free groups (Chapter 4), groups that satisfy semigroup laws (Chapter 5), free products of groups (Chapter 6), direct products (Chapter 7), and certain one-relator semigroups and groups (Chapter 8). Each chapter considers whether finitely generated subsemigroups of the subject are finitely presented, have finite Malcev presentations, or admit automatic structures. The parallel results of these investigations, together with various facts regarding the closure of certain classes of groups established during these chapters, are collected and summarized in the Précis at the end of the thesis.

The other chapter in the main body of the thesis, Chapter 9, investigates, for group-embeddable semigroups, the interaction of subsemigroups and extensions that are finite with regard to certain concepts of index. In particular, the property of having a finite Malcev presentation is shown to be preserved under extensions and subsemigroups of finite Rees index. Section 9.5 considers the problem of whether the class of groups all of whose subsemigroups have finite Malcev presentations is closed under forming finite extensions.

The remaining chapters are of a secondary nature: Chapter 0 is mainly concerned with making definitions and establishing notation, although Subsection 0.9.1 surveys conditions for the group-embeddability of semigroups, and Appendix A gathers various necessary results from the theory of formal languages and automata.

Many — if not most — of the results and examples herein have appeared in the work of Cain, Robertson *&* Ruškuc (2005*a*, 2005*b*) and Cain (2005*a*, 2005*b*, 2005*c*). Nevertheless, the mode of exposition in this thesis is probably superior: those articles reflect the chronological order of the various discoveries, which does not coincide with the order that is best for understanding the corpus of work as a whole.

The opening sentence of this preface paraphrases that of the prologue to J. R. R. Tolkien's *The Lord of the Rings* [Second edition. London: Allen *&* Unwin, 1965–6].

ACKNOWLEDGEMENTS

... li buoni consigli... conviene naschino dalla prudenzia del principe, e non la prudenza del principe da' buoni consigli.

— Niccolò Machiavelli,
Il Principe (1513), ch. xxiii

THREE YEARS AGO, there was no thesis; now it is close to completion. Although it is not yet 7AM, the high latitude and the nearness of midsummer mean that the sun is already high over the ruins of St Andrews Cathedral, pouring light into my room at Deans Court. The air and sea are still; only the crying of the gulls breaks the silence. As the rest of the town awakens, I turn to the duty and the pleasure of acknowledging many debts of gratitude.

The friendships formed in this hall transcend cultural and national boundaries. Amongst people of a dozen lands, I must mention a few names in particular. Firstly, I owe a great deal to Robert Kevis, Alexa Royden, Monica Stensland, Pedro Vasconcelos, and Jacqueline Vaughan, all of whom were present during 2003–04; the significance of that statement is known by all who need to. Rachel Brearley, one of the first people I met on arriving in St Andrews in 2002, deserves my thanks, both as a friend and as an honourable chess opponent. I also thank Paul von Bünau — who made valuable observations on the first chapter of this thesis — and Michal Horvath.

In the School of Mathematics and Statistics at St Andrews, the warmest thanks go to my supervisors, Nik Ruškuc and Edmund Robertson[†] for their aid and toleration over the past three years and, in particular, for reading and commenting on a draft of this thesis. My thanks also go to Kenneth Falconer and to Steve Linton.

Amongst my fellow research students, I must express particular gratitude to Erzsébet Dombi. I hope she found the several conversations we had about automatic semigroups as profitable and as enjoyable as I did. I also thank Peter Campbell,

[†]Publishing with Edmund means that my Erdős number is 3. A shortest path is Cain, Robertson & Ruškuc (2005*a*) ⟶ Brooksbank, Qin, Robertson & Seress (2004) ⟶ Erdős, Füredi, Hajnal, Komjáth, Rödl & Seress (1986).

Catarina Carvalho, Peter Gallagher, Robert Gray, Elizabeth Kimber, Dale Sutherland and Maja Waldhausen.

My time as an undergraduate at the University of Glasgow obliges me to give thanks to Ian Anderson, Andrew Baker, Kenneth Brown, and to Iain Gordon, whom I must also thank for being the first person to invite me to deliver a seminar. Thanks are also due to Christopher Hogarty and Clare McGrory, respectively Vice-President and Treasurer of the Maclaurin Society during my tenure as President; and to Laura Isbister.

Finally, my thanks go to the Carnegie Trust for the Universities of Scotland, who awarded me a scholarship to fund my doctoral studies at St Andrews.

I know I risk invidiousness in mentioning some names and not others. The present section would doubtless dwarf the remainder of the thesis if I gave everyone the thanks he or she deserves. Too much has happened; too many debts have I incurred. Too much has changed. Time inevitably colours one's outlook, but I think that there are few experiences that can alter one more than pursuing a doctorate. I cannot now fully remember the time before I began my Ph.D., but I know I am not the same person who arrived in St Andrews three winters — and a thousand years — ago.

Deans Court, St Andrews,
26th June 2005.

A. J. C. קין

CHAPTER ZERO

ELEMENTARY THEORY OF SEMIGROUPS & GROUPS

Amid the vastness of the things among which we live,
the existence of nothingness holds the first place.
— Leonardo da Vinci, Notebook xix
(trans. J. P. Richter)

0.1. INTRODUCTION

AREAS of group and semigroup theory necessary for the remainder of this thesis are briefly covered in this chapter, which also makes definitions and establishes notation. Subsection 0.9.1 contains a brief survey of conditions for a semigroup to be embeddable into a group.

For further background information on the general theory of semigroups, see Clifford & Preston, *The Algebraic Theory of Semigroups* (Vols. I & II) [Providence: American Mathematical Society, 1961 & 1967] or Howie, *Fundamentals of Semigroup Theory* [Oxford University Press, 1995].

0.2. WORDS, PREFIXES, AND SUFFIXES

A SET OF FORMAL SYMBOLS forms an *alphabet*. A *language* over an alphabet A is a set of strings made up of the symbols in A. These strings are known as *words*, and the symbols within them as *letters*. The set of all strings over the alphabet A includes an *empty word*, denoted ε_A. Formally, one should distinguish the empty words ε_A and ε_B, where B is another alphabet. Usually, however, this distinction is unimportant and ε denotes the empty word over any alphabet.

The set of all words over A is denoted A^* and is the free monoid with basis A. The multiplication in this monoid is concatenation of words; the identity element is the empty word. The set of all non-empty words, $A^+ = A^* - \{\varepsilon\}$, is the free semigroup with basis A.

Let $w = a_1 \cdots a_n$ be a word over the alphabet A, with $a_i \in A$ for each $i = 1, \ldots, n$, where $n \in \mathbb{N}$. The length of the word w, denoted $|w|$, is the number of symbols it contains, namely n. The empty word has length $|\varepsilon| = 0$, and it is the unique word with this property. The set of words of length 1 is identified with the alphabet A.

A *subword* of w is a word of the form $a_i \cdots a_j$, where $1 \leq i \leq j \leq n$. A *prefix* of w is such a subword of the form $a_1 \cdots a_j$, where $1 \leq j \leq n$. A *suffix* of w is subword $a_i \cdots a_n$, where $1 \leq i \leq n$. Additionally, every word includes the empty word as a subword, a prefix, and a suffix.

For $t \in \mathbb{N} \cup \{0\}$, let

$$w(t) = \begin{cases} \varepsilon & \text{if } t = 0, \\ a_1 \cdots a_t & \text{if } 0 < t \leq n, \\ a_1 \cdots a_n & \text{if } n < t, \end{cases}$$

and let

$$w[t] = \begin{cases} a_{t+1} \cdots a_n & \text{if } 0 \leq t < n, \\ \varepsilon & \text{if } n \leq t. \end{cases}$$

So $w(t)$ is the prefix of u up to and including the t-th letter whilst $w[t]$ is the suffix of w after and not including the t-th letter. Observe that for all $t \in \mathbb{N} \cup \{0\}$, $w = w(t)w[t]$, and that if one formally assumes that $a_t = \varepsilon$ for $t > n$, then $w(t+1) = w(t)a_{t+1}$ and $w[t] = a_{t+1}w[t+1]$.

0.3. GENERATORS AND REPRESENTATIVES

FOLLOWING Epstein et al. (1992), the notation used in this thesis distinguishes a word over a set of generators from the element of the semigroup or group it represents. This is slightly unusual: in combinatorial semigroup theory, the same notation is normally used for the word and the element of the semigroup, and words are described as being 'identically equal' or 'equal in the semigroup'. However, this would lead to confusion in some of the more complex arguments ahead, notably the proof of Theorem 4.4.1.

Let S be a semigroup and let X be a generating set for S. Let A be an alphabet and $\phi : A \to X$ be a surjective mapping. The letter $a \in A$ is said to *represent*, or *be a representative of*, the element $a\phi \in X$. Observe that two different letters of A may represent the same element of X.

Extend the mapping ϕ to a surjective homomorphism from the free semigroup A^+ over A onto S. The word w *represents*, or *is a representative of*, the element $w\phi \in S$. Unless the particular homomorphism ϕ is important, denote $w\phi$ by $\overline{w}$. If W is a set of words, then $\overline{W}$ is the image of W under ϕ — the set of elements of S represented by at least one word in W.

Let G be a group, and A an alphabet representing a set of group generators for G. Let A^{-1} be a set in bijection with A under the mapping $a \mapsto a^{-1}$ for each $a \in A$. Extend this mapping to an involution on $(A \cup A^{-1})^*$ by defining $a_1 \cdots a_n \mapsto a_n^{-1} \cdots a_1^{-1}$ for $a_i \in A$. To extend ϕ to a homomorphism from $(A \cup A^{-1})^*$ onto G, let a^{-1} map to $(a\phi)^{-1}$ and the empty word ε map to 1_G. This extended mapping ϕ factors through the free group over the alphabet A, which is denoted $\mathrm{FG}(A)$:

$$\begin{tikzcd} (A \cup A^{-1})^* \arrow[r] \arrow[dr, "\phi"'] & \mathrm{FG}(A) \arrow[d] \\ & G \end{tikzcd}$$

If X is a subset of a semigroup S, then $\mathrm{Sg}\langle X\rangle$ is the subsemigroup of S that X generates. Similarly, if S is a monoid, then $\mathrm{Mon}\langle X\rangle$ is the submonoid of S that X generates, and if S is a group, the subgroup of S generated by X is $\mathrm{Gp}\langle X\rangle$. As noted above, the free group over the alphabet [or with basis] A is denoted $\mathrm{FG}(A)$.

0.4. RELATIONS

LET S be a semigroup, and let A be an alphabet representing a set of generators for S. The term 'relation' has two different usages in this thesis:

i.) A *relation* is a pair $(u, v) \in A^+ \times A^+$. Such a relation *holds* or *is valid* in S if u and v represent the same element of S.

ii.) A [*binary*] *relation* is a *set* of pairs $(u, v) \in A^+ \times A^+$.

Let $\mathcal{R}$ be the kernel of the mapping ϕ that takes a word to the element of S it represents. That is,

$$\mathcal{R} = \ker \phi = \left\{(u, v) \in A^+ \times A^+ : u\phi = v\phi\right\}.$$

Then $\mathcal{R}$ is the set of all relations that hold in S, and $S \simeq A^+/\mathcal{R}$.

For any binary relation $\mathcal{Q} \subseteq A^+ \times A^+$, define the inverse relation $\mathcal{Q}^{-1}$ by

$$\mathcal{Q}^{-1} = \{(v, u) : (u, v) \in \mathcal{Q}\}.$$

0.5. PRESENTATIONS

DEFINITIONS and basic results regarding semigroup and group presentations needed elsewhere in this thesis are stated below. For further reading on group presentations, see Lyndon & Schupp, *Combinatorial Group Theory* [Berlin: Springer, 1977]; for semigroup presentations, see Ruškuc, *Semigroup Presentations* [Ph.D. Thesis, University of St Andrews, 1995].

Let A be an alphabet, and let ρ be a binary relation on A^+ — that is, a subset of $A^+ \times A^+$. The congruence generated by ρ, denoted $\rho^\#$, is the smallest congruence on A^+ containing ρ. More formally,

$$\rho^\# = \bigcap \left\{\sigma : \sigma \supseteq \rho,\ \sigma \text{ is a congruence on } A^+\right\}.$$

Then $\mathrm{Sg}\langle A|\rho\rangle$ is a *presentation* for, or *presents*, or *defines*, [any semigroup isomorphic to] the factor semigroup $A^+/\rho^\#$. Each word over the alphabet A represents its $\rho^\#$-congruence class: $\overline{w} = [w]_{\rho^\#}$ for each $w \in A^+$. The binary relation ρ is a set of *defining relations* for S. The presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ is finite if A and ρ are both finite. A semigroup is *finitely presented* if it admits a finite presentation.

Two words $u, v \in A^+$ represent the same element of $\mathrm{Sg}\langle A \mid \rho\rangle$ if and only if there is a sequence

$$u = u_0 \to u_1 \to \ldots \to u_n = v,$$

with $n \geq 0$, where, for each $i \in \{0, \ldots, n-1\}$, there exist $p_i, q_i, q_i', r_i \in A^*$ such that $u_i = p_i q_i r_i$, $u_{i+1} = p_i q_i' r_i$, and $(q_i, q_i') \in \rho$ or $(q_i', q_i) \in \rho$. That is, $\overline{u} = \overline{v}$ if and only if it is possible to get from u to v by a finite number of replacements of a subword that forms one side of a defining relation in ρ with the word on the other side of that defining relation. Such a sequence is called a *ρ-chain* from u to v, or, where there is no risk of confusion, simply a *chain*. If $(u, v) \in \rho^\#$ — and u and v are thus linked by a ρ-chain — then (u, v) is said to be a *consequence* of ρ.

The definitions above apply to monoids: the monoid presentation $\mathrm{Mon}\langle A \mid \rho\rangle$ defines the monoid $A^*/\rho^\#$, where $\rho^\#$ is the smallest congruence on A^* containing ρ. In this case, ρ is a binary relation on A^* — a defining relation may have the empty word on one side.

The group presentation $\mathrm{Gp}\langle A \mid \rho\rangle$ defines the factor group $\mathrm{FG}(A)/N(\rho)$, where $\mathrm{FG}(A)$ is the free group on the alphabet A and $N(\rho)$ is the normal closure of the set $\{uv^{-1} : (u, v) \in \rho\}$. In particular, the free group on the alphabet A is presented by $\mathrm{Gp}\langle A \mid \emptyset\rangle$. For group presentations, defining relations are drawn from the set $(A \cup A^{-1})^* \times (A \cup A^{-1})^*$.

If a semigroup, monoid, or group is finitely presented, and $\langle A \mid \rho\rangle$ is a presentation for it with A being finite, then it admits a finite presentation $\langle A \mid \sigma\rangle$ with $\sigma \subseteq \rho$.

Of course, any group has a monoid presentation:

$$\mathrm{Gp}\langle A \mid \rho\rangle = \mathrm{Mon}\langle A \cup A^{-1} \mid \rho \cup \{(aa^{-1}, \varepsilon), (a^{-1}a, \varepsilon) : a \in A\}\rangle. \tag{1}$$

Similarly, every monoid has a semigroup presentation:

$$\mathrm{Mon}\langle A \mid \rho\rangle = \mathrm{Sg}\langle A \cup \{e\} \mid \rho' \cup \{(ae, a), (ea, a) : a \in A\}\rangle, \tag{2}$$

where ρ' is obtained from ρ by replacing ε by e whenever the former occurs as one side of a defining relation.

0.6. STRING-REWRITING SYSTEMS

SOME FACTS about string rewriting will be needed in Section 3.5. For further background information, see Book *&* Otto (1993).

Definition 0.6.1. A *string rewriting system*, or simply a *rewriting system*, is a pair $(A, \mathcal{R})$, where A is a finite alphabet and $\mathcal{R}$ is a set of pairs (l, r), known as *rewriting rules*, drawn from $A^* \times A^*$. The single reduction relation $\Rightarrow_\mathcal{R}$ is defined as follows: $u \Rightarrow_\mathcal{R} v$ (where $u, v \in A^*$) if there exists a rewriting rule $(l, r) \in \mathcal{R}$ and words $x, y \in A^*$ such that $u = xly$ and $v = xry$. That is, $u \Rightarrow_\mathcal{R} v$ if one can obtain v from u by substituting the word r for a subword l of u, where (l, r) is a rewriting rule. The reduction relation $\stackrel{*}{\Rightarrow}_{\mathcal{R}}$ is the reflexive and transitive closure of $\Rightarrow_{\mathcal{R}}$. [Where there is no possibility of confusion, '$\Rightarrow$' and '$\stackrel{*}{\Rightarrow}$' are used, rather than '$\Rightarrow_{\mathcal{R}}$' and '$\stackrel{*}{\Rightarrow}_{\mathcal{R}}$'.] The process of replacing a subword l by a word r, where $(l, r) \in \mathcal{R}$, is called *reduction*, as is the iteration of this process.

A word $w \in A^*$ is *reducible* if it contains a subword l that forms the left-hand side of a rewriting rule in $\mathcal{R}$; it is otherwise called *irreducible*.

The string rewriting system $(A, \mathcal{R})$ is *noetherian* if there is no infinite sequence $u_1, u_2, \ldots \in A^*$ such that $u_i \Rightarrow u_{i+1}$ for all $i \in \mathbb{N}$. That is, $(A, \mathcal{R})$ is noetherian if any process of reduction must eventually terminate with an irreducible word. The rewriting system $(A, \mathcal{R})$ is *confluent* if, for any words $u, u', u'' \in A^*$ with $u \stackrel{*}{\Rightarrow} u'$ and $u \stackrel{*}{\Rightarrow} u''$, there exists a word $v \in A^*$ such that $u' \stackrel{*}{\Rightarrow} v$ and $u'' \stackrel{*}{\Rightarrow} v$.

The following results will be called upon in the proof of Theorem 3.5.6.

Proposition 0.6.2 (Book & Otto 1993, p. 50). *Any string rewriting system $(A, \mathcal{R})$ is confluent if there are no overlaps between left-hand sides of rewriting rules in $\mathcal{R}$ — that is, if there are no rules (l_1, r_1) and (l_2, r_2) in $\mathcal{R}$ with either*

i.) *$l_1 = pq$ and $l_2 = qr$, where p, q, and r are words over A with q non-empty; or*

ii.) *$l_1 = pl_2q$, where p and q are words over A.* $\boxed{0.6.2}$

Theorem 0.6.3 (Book & Otto 1993, Theorem 1.1.12). *Let $(A, \mathcal{R})$ be a string rewriting system and suppose that it is both confluent and noetherian. Then for any word $u \in A^*$, there is a unique irreducible word $v \in A^*$ with $u \stackrel{*}{\Rightarrow} v$.* $\boxed{0.6.3}$

Let $(A, \mathcal{R})$ be a confluent noetherian string rewriting system. The set of irreducible words are said to be in *normal form*, and the unique normal form to which a word w can be reduced is denoted $\mathrm{NF}(w)$. The semigroup presented by $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$ may be identified with the set of normal form words under the operation of 'concatenation plus reduction to normal form'.

0.7. CAYLEY GRAPHS AND GEOMETRY

THROUGHOUT this section, let S be a semigroup and let A be an alphabet representing a generating set for S.

Definition 0.7.1. The [*right*] *Cayley graph* of S with respect to A, denoted $\Gamma(S, A)$, is a labelled directed graph with vertex set S and, for each $s, t \in S$, an edge from s to t labelled by a if and only if $s\overline{a} = t$.

The Cayley graph of the free monoid on two letters is shown in Figure 0.1.

Definition 0.7.2. Define a metric d_A on $\Gamma(S, A)$ as follows. For any two elements s, t of S, let $d_A(s, t)$ — the distance between s and t — be the infimum of the lengths of the undirected paths from s to t, or ∞ if there are no such paths.

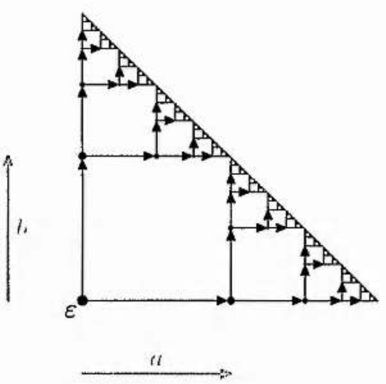

Figure 0.1. The Cayley graph of the free monoid $\{a,b\}^*$. Every horizontal edge is labelled a and every vertical edge by b. The identity ε serves as basepoint.

If the semigroup S happens to be a group, then $\Gamma(S,A)$ is always connected. Furthermore, in this circumstance, $\Gamma(S,A)$ is homogeneous: for any $g \in S$, the transformation defined on vertices by $s \mapsto gs$ and edges by $(s,t) \to (gs,gt)$ is an automorphism of the graph. This means that any neighbourhood of a vertex 'looks like' the corresponding neighbourhood of any other vertex.

If S is a monoid, the vertex 1_S is a natural *basepoint* for the Cayley graph $\Gamma(S,A)$: the path labelled by any word $w \in A^+$ starting at 1_S ends at $\overline{w}$. If S is not a monoid, one can add a vertex ω that behaves like an identity to serve as a basepoint: for all $a \in A$, there is an edge from ω to $\overline{a}$ labelled by a. This thesis assumes that every Cayley graph is equipped with a basepoint. Denote by $\widehat{w}$ the unique path labelled by w starting at the basepoint (and ending at $\overline{w}$). [In graph theory, the term 'path' is usually reserved for a walk that visits no vertex twice. For the purposes of Cayley graphs, such a distinction is unimportant.]

Definition 0.7.3. Let $r \in \mathbb{N}$. The *open ball* of radius r around the element $s \in S$ is the set

$$B_r(s) = \{t \in S : d_A(s,t) < r\}.$$

The following result relates distance in a semigroup and distance in a subsemigroup:

Proposition 0.7.4. *Let B be an alphabet representing a generating set for a subsemigroup T of S. For each $b \in B$, choose a word over A representing $\overline{b}$. Let m be the maximum length of these words. Then, for all elements s and t of T:*

$$d_A(s,t) \le m \times d_B(s,t).$$

That is, the distance from s to t with respect to A is bounded by the distance from s to t with respect to B multiplied by m. $\boxed{0.7.4}$

0.8. SEMIGROUPS AND UNIVERSAL GROUPS

MANY OF THE basic properties of semigroups embeddable into groups rely on the results regarding universal groups of semigroups contained in this section. For further background on this subject, see Chapter 12 of Clifford & Preston (1967).

Definition 0.8.1. Let S be a semigroup. An S-group is a pair (G,γ), where G is a group and γ is a homomorphism from S to G whose image generates G as a group. A *universal group of S* (or *free S-group*) is an S-group (G,γ) with the property that if (H,η) is an S-group, then there exists a homomorphism $\alpha : G \to H$ satisfying $\gamma\alpha = \eta$, or equivalently such that the following diagram commutes:

$$\begin{tikzcd} S \arrow[r, "\gamma"] \arrow[dr, "\eta"'] & G \arrow[d, dotted, "\alpha"] \\ & H \end{tikzcd}$$

[Note that Clifford & Preston prefer the term 'free S-group' to 'universal group of S'.]

Proposition 0.8.2. *Let S be a semigroup. Suppose S is presented by $\mathrm{Sg}\langle A \mid \rho\rangle$ for some alphabet A and set of defining relations ρ. Then $(\mathrm{Gp}\langle A \mid \rho\rangle, \gamma)$ is a universal group of S, where γ is the homomorphism extending the identity mapping on $\overline{A}$.*

Proof of 0.8.2. See Clifford & Preston (1967, Construction 12.6). $\boxed{0.8.2}$

Proposition 0.8.3. *Let S be a semigroup, and let (G,γ) and (H,η) both be universal groups of S. Then there is an isomorphism from G to H that maps $S\gamma$ to $S\eta$.*

Proof of 0.8.3. Since (G,γ) is a universal group of S and (H,η) is an S-group, there is a homomorphism $\alpha : G \to H$ with $\gamma\alpha = \eta$. Similarly, there is a homomorphism $\beta : H \to G$ with $\eta\beta = \gamma$.

$$\begin{tikzcd} & S \arrow[dl, "\gamma"'] \arrow[dr, "\eta"] & \\ G \arrow[rr, dotted, bend left, "\alpha"] & & H \arrow[ll, dotted, bend left, "\beta"] \end{tikzcd}$$

Observe that $s\gamma\alpha\beta = s\eta\beta = s\gamma$ for $s \in S$. Therefore $\alpha\beta$ is the identity map on $S\gamma$. Since $S\gamma$ generates G as a group, $\alpha\beta$ is the identity map on G. Similarly, $\beta\alpha$ is the identity map on H. Therefore $\alpha : G \to H$ and $\beta : H \to G$ are mutually inverse isomorphisms that map $S\gamma$ to $S\eta$ and vice versa. $\boxed{0.8.3}$

In light of Proposition 0.8.3, it is sensible to write of *the* universal group of a semigroup. Furthermore, although the universal group (G,γ) of the semigroup S is formally a pair, one normally writes of G being the universal group of S and suppresses mention of the homomorphism γ.

0.9. EMBEDDING SEMIGROUPS INTO GROUPS

A SEMIGROUP S embeds in a group G if there is a monomorphism from S into G. A semigroup S is *embeddable into a group*, or is *group-embeddable*, if there exists some group G into which S embeds.

Proposition 0.9.1. *Let S be a semigroup and let (G, γ) be its universal group. Suppose S embeds in a group. Then $\gamma : S \to G$ is a monomorphism: S is embedded into its universal group G by γ.*

Proof of 0.9.1. Let (G, γ) be the universal group of S. Let H be a group in which S embeds by means of a monomorphism η. By restricting, if necessary, to the subgroup generated by $S\eta$, assume that the image of S generates H. Therefore (H, η) is an S-group. By the definition of a universal group, there is a homomorphism $\alpha : G \to H$ satisfying $\gamma\alpha = \eta$. Since η is injective, γ is injective. Therefore γ embeds S into G. $\boxed{0.9.1}$

Suppose that S is a subsemigroup of a group G, and that H is the subgroup of G generated by S. In general, the universal group of S is *not* isomorphic to H, as the following example illustrates:

Example 0.9.2. Let G be the free group on the letters x and y. Identify elements of G with reduced words on x and y. Let the alphabet $A = \{a, b, c, d\}$ represent elements of G in the following way:

$$\overline{a} = x^2,\ \overline{b} = xy,\ \overline{c} = yx,\ \overline{d} = y^2.$$

Let S be the subsemigroup and H the subgroup of G generated by $\overline{A}$.

The Nielsen–Schreier Theorem asserts that every subgroup of a free group is free (see Lyndon & Schupp 1977, Proposition I.2.6). Therefore H is a free group. The rank of H is at most 4. If it were free of rank 4, it would be free with the given generating set as a basis (Lyndon & Schupp 1977, Proposition I.2.7). However, the relation $(bd^{-1}c, a)$ holds: the group H cannot be free of rank 4. It must therefore be free of rank at most 3.

However, it is easy to prove that S is a free semigroup of rank 4. (Either reason directly or use one of the algorithms described in Subsection 4.2.1.) By Proposition 0.8.2, its universal group is the free group of rank 4, which is not isomorphic to H.

Furthermore, the universal group of a semigroup may be finitely generated even if the semigroup itself is not finitely generated:

Example 0.9.3. Let $G = \mathbb{Z} \times \mathbb{Z}$, and let S be the subsemigroup of G generated by the set

$$\{(1, n) : n \in \mathbb{N}\}.$$

None of the elements in this generating set can be decomposed in S: therefore the semigroup S is not finitely generated. However, the universal group of S is G itself, which has rank 2. [To prove that G is the universal group of S, one can either reason directly or appeal to Corollary 5.2.7.]

Definition 0.9.4. Let G be a group. A subsemigroup S of G is a *positive subsemigroup* if it generates G as a group.

The next two results show that the embedding of a semigroup in a group is reflected by an embedding of Cayley graphs. These results are particularly important in the theory of automatic semigroups embeddable in groups (see Section 2.3).

Proposition 0.9.5. *Let S be a semigroup and let G be a group into which it embeds and which it generates. Let A represent a generating set for S and thus a group generating set for G. Then $\Gamma(S, A)$ is a subgraph of $\Gamma(G, A \cup A^{-1})$.* $\boxed{0.9.5}$

Corollary 0.9.6. *Let S, G, and A be as in Proposition 0.9.5. Let $s, t \in S$. Then the distance from s to t in $\Gamma(G, A \cup A^{-1})$ is no more than the distance from s to t in $\Gamma(S, A)$.* $\boxed{0.9.6}$

0.9.1. Survey of conditions for group-embeddability

Most of this thesis deals with subsemigroups of groups: semigroups that are already known to be group-embeddable. However, there has been much research into conditions necessary or sufficient for a semigroup to be embeddable into a group. The present section gives a brief overview of this work.

Groups are cancellative: thus cancellativity is certainly a necessary condition for a semigroup to be embeddable into a group. Furthermore, all periodic cancellative semigroups are groups (Clifford & Preston 1961, Exercise 1.7:6(c)). It was asked whether all cancellative semigroups were group-embeddable. Malcev (1937) answered this question in the negative, and later established a necessary and sufficient condition for a semigroup to be group-embeddable (Malcev 1939). Malcev's condition takes the form of an infinite set of equational implications that are satisfied if and only if the semigroup is embeddable in a group. Malcev also showed that no finite subset of these implications suffices to distinguish group-embeddable semigroups (Malcev 1940). A special case of Malcev's result gives the *quotient condition*: if a semigroup S is embeddable into a group, it satisfies the quotient condition (Clifford & Preston 1967, Section 12.4):

$$(\forall a, b, c, d, e, f, g, h \in S)(ae = bf \wedge cf = de \wedge dg = ch \implies ag = bh). \qquad (3)$$

[Malcev (1937) calls (3) 'condition Z'. It is interesting to compare this condition with Example 1.4.1.] Another necessary and sufficient condition for group-embeddablility was given by Lambek (1951). Clifford & Preston (1967, Chapter 12) discuss and compare Malcev's and Lambek's results.

Ore's (1931) Theorem was originally stated as a sufficient condition for embedding rings without zero divisors into division rings. However, it readily adapts to embedding semigroups into groups. Several definitions are needed in order to state the theorem.

Definition 0.9.7. A semigroup is *right-reversible* (respectively, *left-reversible*) if any two of its principal left ideals (respectively, principal right ideals) intersect.

Therefore, a semigroup S is right-reversible if every two elements of S have a common left multiple: if, for every $s, t \in S$ there exist $p, q \in S$ such that $ps = qt$. Similarly, a semigroup is left-reversible if every two elements have a common right multiple.

Definition 0.9.8. A *group of left quotients* (respectively, *group of right quotients*) of a semigroup S is a group containing S in which every element can be expressed as $s^{-1}t$ (respectively, st^{-1}) for some elements $s, t \in S$.

Observe that a group of left or right quotients of a semigroup (even one that is group-embeddable) may not exist. However, if a group of left or right quotients of a semigroup S does exist, then by definition S embeds therein.

Theorem 0.9.9 (Ore's Theorem). *Any right-reversible or left-reversible cancellative semigroup S embeds in a group.* $\boxed{0.9.9}$

Ore's (1931) original proof for rings applies to semigroups. Rees (1948) gave an elegant proof, reproduced in Clifford & Preston (1961, Theorem 1.23). Dubreil (1943 & 1954, p. 269) strengthened Ore's result to give the following:

Theorem 0.9.10 (Dubreil's Theorem). *Let S be a cancellative semigroup. Then S is right-reversible if and only if a group of left quotients of S exists. Similarly, such a semigroup is left-reversible if and only if a group of right quotients of S exists.* $\boxed{0.9.10}$

[The theorems of Ore and Dubreil will prove their worth in Chapter 5, where they are used to prove results on Malcev presentations. This is why they are discussed in some detail above. Although further results on groups of quotients will be required, their statement is deferred until Subsection 5.2.1.]

Adjan (1966*b*, Theorem II.4) established the following sufficient condition for group-embeddability:

Definition 0.9.11. The *left graph* of the presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ is an undirected graph with vertex set A and, for every relation $(au, bv) \in \rho$ (where $a, b \in A$ and $u, v \in A^*$), an edge from a to b. The *right graph* is defined analogously, using the last letters on each side of relations $(ua, vb) \in \rho$.

Theorem 0.9.12 (Adjan's Theorem). *Let S be the semigroup defined by a presentation $\mathrm{Sg}\langle A \mid \rho\rangle$ whose left and right graphs contain no non-trivial cycles. Then S is embeddable in a group.* $\boxed{0.9.12}$

[Theorem 0.9.12 specializes to Adyan's (1960*a*) result that every one-relator cancellative semigroup embeds into a group.] An elegant geometric proof of Adjan's Theorem is due to Remmers (1980). Stallings (1987) used a graph-theoretical lemma to give a third proof.

Guba (1994*b*) and Kashintsev (1992) studied generalizations of Adjan's Theorem. Their results require the following definitions:

Definition 0.9.13. Consider a semigroup presentation $\mathrm{Sg}\langle A \mid \rho\rangle$. Let $R = \{u, v : (u, v) \in \rho\}$. An *s-piece* (relative to R) is word $w \in A^+$ such that there exist $u, v \in R$, $p, p', q, q' \in A^*$, with $p \neq q$ or $p' \neq q'$, satisfying $u = pwq$ and $v = p'wq'$.

Condition $C_s(k)$ is satisfied if no word in R can be expressed as a product of fewer than k s-pieces.

Condition $D(l)$ is satisfied if neither the left graph nor the right graph of $\mathrm{Sg}\langle A|\rho\rangle$ has a cycle of length less than l.

A semigroup belongs to the class K_k^l if it has a presentation satisfying conditions $C_s(k)$ and $D(l)$. Formally define K_1^∞ as $\bigcap_{l\in\mathbb{N}} K_1^l$.

[The condition $C_s(k)$ is based on the 'small cancellation conditions' of combinatorial group theory (see Lyndon & Schupp 1977, Chapter V). Observe that $K_k^1 \supseteq K_k^2 \supseteq K_k^3 \supseteq \ldots$ and $K_1^l \supseteq K_2^l \supseteq K_3^l \supseteq \ldots$ for all $k, l \in \mathbb{N}$. More concisely,

$$K_k^l \subseteq K_{k-1}^l \cap K_k^{l-1} \text{ for all } k, l \geq 2.$$

Guba (1994*b*) proved that $K_4^2 \cup K_3^3$ is strictly contained in K_3^2.]

Using Definition 0.9.13, Adjan's Theorem can be restated as: 'every semigroup of class K_1^∞ is group-embeddable'.

Kashintsev (1992) showed that any semigroup of class K_3^3 or K_4^2 embeds into a group, and also that for any natural number l there exists a semigroup of class K_2^l that is not group-embeddable. Answering questions posed by Kashintsev, Guba (1994*b*) proved that semigroups of class K_3^2 are group-embeddable, and strengthened Kashintsev's other result to show that the class K_2^l always contains a *cancellative* semigroup that does not embed in a group. Both Guba and Kashinstev noted that Adjan's result is therefore the 'best possible' for semigroups satisfying $D(\infty)$.

Kashintsev (2001*a*) continued working on this theme, though the results obtained and the arguments involved are perhaps less elegant than his earlier research.

0.10. MISCELLANEOUS

GATHERED in this section are a few remaining definitions and results needed elsewhere.

Definition 0.10.1. Let S be a semigroup. The semigroup with a zero adjoined S^0 is the set $S \cup \{0\}$ with multiplication in $S \subset S^0$ being as before, and $s0 = 0s = 00 = 0$ for all $s \in S$. Similarly, the semigroup with an identity adjoined S^1 has set $S \cup \{1\}$, where $s1 = 1s = s$ for all $s \in S \cup \{1\}$, multiplication in $S \subset S^1$ being unchanged.

Recall that a *group-theoretical property* is formally defined as a property possessed by the trivial group and which is preserved under isomorphism (see Robinson 1996, Section 2.3). Examples of group-theoretical properties are commutativity, freedom, nilpotency, and solvability. A group-theoretical property $\mathfrak{P}$ is *hereditary* if a subgroup of a $\mathfrak{P}$ group also has property $\mathfrak{P}$. A group is *virtually* $\mathfrak{P}$ if it contains a finite-index subgroup with property $\mathfrak{P}$.

The three main instances of virtually $\mathfrak{P}$ groups encountered in this thesis are virtually free groups, introduced in Chapter 4, and virtually nilpotent and virtually abelian groups, both found in Chapter 5.

Propositions 0.10.2 and 0.10.3 are well-known, but the difficulty of finding their formal statement in the literature and the brevity of their proofs makes their reproduction worthwhile:

Proposition 0.10.2. *Let $\mathfrak{P}$ be a hereditary group-theoretical property. Then a virtually $\mathfrak{P}$ group possesses a finite-index normal subgroup with property $\mathfrak{P}$.*

Proof of 0.10.2. Let G be a virtually $\mathfrak{P}$ group. Let H be a subgroup of G with property $\mathfrak{P}$ and with $[G : H]$ finite. Then the core of H in G,

$$H_G = \bigcap_{g \in G} g^{-1} H g \leq H,$$

is a normal subgroup of G and has index dividing $[G : H]!$ (see, for example, Robinson 1996, p. 16 and Theorem 1.6.9). Since the class of groups with property $\mathfrak{P}$ is closed under taking subgroups, H_G also has property $\mathfrak{P}$. $\boxed{0.10.2}$

Proposition 0.10.3. *Let $\mathfrak{P}$ be a hereditary group-theoretical property. Then the property of being virtually $\mathfrak{P}$ is also hereditary.*

The following lemma is obvious, but the use of the argument it contains at several different points in this thesis justifies its explicit statement.

Lemma 0.10.4. *If G is a group and E is a finite extension of G, then each subgroup H of E is a finite extension of the subgroup $G \cap H$ of G.*

Proof of 0.10.4. Let H be a subgroup of E. By the second isomorphism theorem, $H/G \cap H$ is isomorphic to GH/G, which is a subgroup of the finite factor group E/G. Therefore $[H : G \cap H]$ is finite and H is a finite extension of $G \cap H$. $\boxed{0.10.4}$

Proof of 0.10.3. Let V be a virtually $\mathfrak{P}$ group. By Proposition 0.10.2, suppose V is a finite extension of a group G with property $\mathfrak{P}$. Choose an arbitrary subgroup H of V. By Lemma 0.10.4, H is a finite extension of a subgroup of G, which, since the property $\mathfrak{P}$ is hereditary, also has property $\mathfrak{P}$. Therefore H is virtually $\mathfrak{P}$. $\boxed{0.10.3}$

CHAPTER ONE

FUNDAMENTAL THEORY OF MALCEV PRESENTATIONS

I feel like someone who wades out into the sea after being initially attracted to the water by the shallows next to the shore; and I foresee any advance only taking me into even more enormous, indeed bottomless, depths, and that this undertaking of mine, which seemed to be diminishing as I was completing the earliest sections, is now almost increasing in size.
— Livy, *Ab urbe condita libri*, ch. xxxi.1
(trans. J. C. Yardley)

1.1. INTRODUCTION

A MALCEV PRESENTATION is a presentation of a special type for a semigroup that embeds in a group. Informally, a Malcev presentation defines a semigroup by means of generators, defining relations, and the unwritten rule that the semigroup so defined must be embeddable in a group. Spehner (1977) introduced Malcev presentations, though they are based on Malcev's necessary and sufficient condition for the embeddability of a semigroup in a group (Malcev 1939). Spehner exhibited an example of a finitely generated submonoid of a free monoid that admitted a finite Malcev presentation but which was not finitely presented. He later showed that all finitely generated submonoids of free monoids have finite Malcev presentations (Spehner 1989). Until recently, Spehner's articles represented the whole of the literature on Malcev presentations.

The foundations of the theory of Malcev presentations have never appeared in detail: Spehner sketches the bare minimum and refers the reader to Malcev's work. The following sections therefore explain the basic concepts in depth.

1.2. DEFINITIONS AND PRELIMINARIES

THIS SECTION contains the foundations of the theory of Malcev presentations for semigroups, but each definition and result applies *mutatis mutandis* to monoids. To obtain the corresponding monoidal result, it usually suffices to replace 'semigroup' with 'monoid' and free semigroups A^+ with free monoids A^* for all alphabets A. At any point where this rule breaks down, a note will indicate the problem and its solution.

A loose definition of a Malcev presentation was given in Section 1.1: the semigroup defined by a Malcev presentation has the given generators, satisfies the given defining relations, and is embeddable into a group. The first task is to formalize this notion. The second is to establish syntactic rules that determine when two words represent the same element of the semigroup defined by the presentation. These rules are more complex than the corresponding ones for 'ordinary' semigroup presentations.

Definition 1.2.1. Let S be any semigroup. A congruence σ on S is called a *Malcev congruence* if the corresponding factor semigroup S/σ is embeddable in a group.

Proposition 1.2.2. *Let S be a semigroup, and let $\{\sigma_i : i \in I\}$ be a collection of Malcev congruences on S. Then $\sigma = \bigcap_{i\in I} \sigma_i$ is also a Malcev congruence on S.*

Proof of 1.2.2. For each $i \in I$, there exists a group G_i and a monomorphism from S/σ_i to G_i. These embedding mappings define a monomorphism $\psi : \prod^{\times}_{i\in I} S/\sigma_i \to \prod^{\times}_{i\in I} G_i$. Define a mapping

$$\phi : S/\sigma \to \prod_{i\in I}^{\times} S/\sigma_i, \text{ by } [s]_\sigma \mapsto \left([s]_{\sigma_i}\right)_{i\in I} \text{ for all } s \in S.$$

Since $\sigma \subseteq \sigma_i$, all elements related by σ are related by every σ_i, and so the mapping ϕ is well-defined. It is clearly a homomorphism.

Suppose that $s, t \in S$ are such that $[s]_\sigma\phi = [t]_\sigma\phi$. Then $[s]_{\sigma_i} = [t]_{\sigma_i}$ for each $i \in I$. By the definition of σ, therefore, $[s]_\sigma = [t]_\sigma$. The mapping ϕ is therefore injective, and so $\phi\psi$ is an embedding of S/σ into the group $\prod^{\times}_{i\in I} G_i$. So σ is a Malcev congruence. $\boxed{1.2.2}$

The following definition makes sense in light of Proposition 1.2.2 and the observation that there is always at least one Malcev congruence on any semigroup S, namely $S \times S$.

Definition 1.2.3. Let A^+ be the free semigroup on an alphabet A. Let $\rho \subseteq A^+ \times A^+$ be any binary relation on A^+. Denote by ρ^{M} the smallest Malcev congruence containing ρ — namely,

$$\rho^{\mathrm{M}} = \bigcap \left\{\sigma : \sigma \supseteq \rho,\ \sigma \text{ is a Malcev congruence on } A^+\right\}.$$

The relation ρ^{M} is called the *Malcev congruence generated by ρ*.

Definition 1.2.4. Let A^+ be the free semigroup on an alphabet A. Let $\rho \subseteq A^+ \times A^+$ be any binary relation on A^+. Then $\mathrm{SgM}\langle A \mid \rho\rangle$ is a *Malcev presentation* for A^+/ρ^{M}, or any semigroup to which it is isomorphic. If the alphabet A and the set of *defining relations* ρ are both finite, then the Malcev presentation $\mathrm{SgM}\langle A \mid \rho\rangle$ is said to be finite.

The notation $\mathrm{SgM}\langle A \mid \rho\rangle$ distinguishes the Malcev presentation with generators A and defining relations ρ from the ordinary semigroup presentation $\mathrm{Sg}\langle A \mid \rho\rangle$, which defines $A^+/\rho^{\#}$, and the group presentation $\mathrm{Gp}\langle A \mid \rho\rangle$. [The monoid Malcev presentation $\mathrm{MonM}\langle A \mid \rho\rangle$ defines the monoid A^*/ρ^{M}.]

The first of the two goals of this section has been achieved: the remainder of the section is devoted to establishing the syntactic rules mentioned above. The reasoning that follows is elementary but rather technical.

Fix A^+ and ρ as in Definition 1.2.4 and let $S = A^+/\rho^{\mathrm{M}}$. Let $A^{\mathsf{L}}, A^{\mathsf{R}}$ be two sets in bijection with A under the mappings $a \mapsto a^{\mathsf{L}}$, $a \mapsto a^{\mathsf{R}}$, respectively, with $A, A^{\mathsf{L}}, A^{\mathsf{R}}$ being pairwise disjoint.

Extend the mappings $a \mapsto a^{\mathsf{L}}$ and $a \mapsto a^{\mathsf{R}}$ to anti-isomorphisms from A^* to $(A^{\mathsf{L}})^*$ and $(A^{\mathsf{R}})^*$ in the obvious way: for $w = a_1 a_2 \cdots a_n \in A^*$, with $a_i \in A$, define

$$w^{\mathsf{L}} = (a_1 \cdots a_n)^{\mathsf{L}} = a_n^{\mathsf{L}} a_{n-1}^{\mathsf{L}} \cdots a_1^{\mathsf{L}} \text{ and } w^{\mathsf{R}} = (a_1 \cdots a_n)^{\mathsf{R}} = a_n^{\mathsf{R}} a_{n-1}^{\mathsf{R}} \cdots a_1^{\mathsf{R}}.$$

Let

$$\tau = \rho \cup \Big\{(aa^{\mathsf{R}}, \varepsilon), (a^{\mathsf{L}}a, \varepsilon) : a \in A\Big\}. \tag{1}$$

For each $a \in A$, the letters a^{L} and a^{R} will both represent the inverse of $\overline{a}$ in the universal group of the semigroup S. The reason for having two representatives for each such inverse is to develop the syntactic rules given in Proposition 1.2.9, which distinguish between the insertion and deletion of 'generator–inverse pairs' aa^{R} and $a^{\mathsf{L}}a$. [At this point, the reader may wish to look ahead to Proposition 1.2.9 to see the destination of the course of reasoning below.]

Proposition 1.2.5. *The Malcev congruence generated by ρ is the congruence generated by τ restricted to A^+:*

$$\tau^{\#} \cap (A^+ \times A^+) = \rho^{\mathrm{M}}.$$

Proof of 1.2.5. Let $\sigma = \tau^{\#}$. (Recall that $\tau^{\#}$ is a congruence on $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$.) Let $G = \mathrm{Mon}\langle A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}} \mid \tau\rangle = (A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\sigma$.

Lemma 1.2.6. *Let ω be any congruence on $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$ that contains $\sigma = \tau^{\#}$. Then the factor monoid $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\omega$ is a group, with $[a]_\omega^{-1} = [a^{\mathsf{L}}]_\omega = [a^{\mathsf{R}}]_\omega$. In particular, the monoid G is a group.*

Proof of 1.2.6. Let T be the factor monoid $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\omega$. The element $[\varepsilon]_\omega$ is the identity element of T. For any $a \in A$,

$$[a^{\mathsf{L}}]_\omega = [a^{\mathsf{L}}aa^{\mathsf{R}}]_\omega = [a^{\mathsf{R}}]_\omega.$$

Furthermore, for each $a \in A$,

$$[a^{\mathsf{L}}]_\omega [a]_\omega = [a^{\mathsf{L}}a]_\omega = [\varepsilon]_\omega.$$

and

$$[a]_\omega[a^{\mathsf{L}}]_\omega = [aa^{\mathsf{L}}]_\omega = [aa^{\mathsf{L}}aa^{\mathsf{R}}]_\omega = [aa^{\mathsf{R}}]_\omega = [\varepsilon]_\omega.$$

So the inverse of $[a]_\omega$ in T is $[a^{\mathsf{L}}]_\omega = [a^{\mathsf{R}}]_\omega$. A similar argument applies to $[a^{\mathsf{R}}]_\omega$. Since $A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}}$ generates T, all elements of T have inverses. Therefore T is a group. $\boxed{1.2.6}$

Since τ^{M} is defined to be the smallest congruence on $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$ containing τ such that the corresponding factor monoid embeds in a group, $\tau^{\mathrm{M}} = \tau^{\#} = \sigma$.

Lemma 1.2.7. *The Malcev congruence ρ^{M} is contained in $\sigma \cap (A^+ \times A^+)$.*

Proof of 1.2.7. Let μ_i (where $i \in I$) be the congruences on $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$ that contain τ. By Lemma 1.2.6, the factor monoid $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\mu_i$ forms a group for each $i \in I$. Therefore for each $i \in I$, $\nu_i = \mu_i \cap (A^+ \times A^+)$ is a Malcev congruence on A^+ that contains ρ, since A^+/ν_i embeds in $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\mu_i$ and $\tau \cap (A^+ \times A^+) = \rho$. Therefore

$$\begin{aligned}
\rho^{\mathrm{M}} &= \bigcap \{\mu : \mu \supseteq \rho, \mu \text{ is a Malcev congruence on } A^+\} \\
&\subseteq \bigcap_{i \in I} \nu_i \\
&= \bigcap_{i \in I} \left(\mu_i \cap (A^+ \times A^+)\right) \\
&= \Big[\bigcap_{i \in I} \mu_i\Big] \cap (A^+ \times A^+) \\
&= \tau^{\#} \cap (A^+ \times A^+) \\
&= \sigma \cap (A^+ \times A^+),
\end{aligned}$$

and the result follows. $\boxed{1.2.7}$

Define $\gamma : S \to G$ by $[w]_{\rho^{\mathrm{M}}} \to [w]_\sigma$, so that $(\rho^{\mathrm{M}})^\natural \gamma = \sigma^\natural|_{A^+}$. Lemma 1.2.7 shows that if $(w, w') \in \rho^{\mathrm{M}}$, then $(w, w') \in \sigma$: the homomorphism γ is thus well-defined. Figure 1.1 illustrates the situation.

[For any equivalence relation α on a set X, the *natural map* $\alpha^\natural : X \to X/\alpha$ sends $x \in X$ to its equivalence class $[x]_\alpha$. If X is a semigroup and α is a congruence, then $\alpha^\natural$ is a homomorphism.]

Lemma 1.2.8. *The universal group of the semigroup S is (G, γ).*

Proof of 1.2.8. The image of S under γ generates G as a group: (G, γ) is an S-group. Let (H, η) be another S-group, as per Definition 0.8.1. Therefore $\eta : S \to H$ is a homomorphism whose image generates H. Define a homomorphism $\theta : G \to H$ by

$$[a]_\sigma \mapsto a(\rho^{\mathrm{M}})^\natural\eta, \quad [a^{\mathsf{L}}]_\sigma \mapsto (a(\rho^{\mathrm{M}})^\natural\eta)^{-1}, \quad [a^{\mathsf{R}}]_\sigma \mapsto (a(\rho^{\mathrm{M}})^\natural\eta)^{-1}$$

for any $a \in A$, and extend to G in the obvious way.

$$\begin{tikzcd}
A \arrow[r, hook] \arrow[d, hook] & A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}} \arrow[d, hook] \\
A^+ \arrow[r, hook] \arrow[d, "(\rho^{\mathrm{M}})^\natural"'] & (A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^* \arrow[d, "\sigma^\natural"] \\
S = A^+/\rho^{\mathrm{M}} \arrow[r, hook, "\gamma"'] & G = (A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*/\sigma
\end{tikzcd}$$

Figure 1.1. The relationship between A^+, $(A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^+$, S, and G.

Let $u, v \in (A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$ be such that $u \; \tau \; v$. If $(u, v) \in \rho$, then $u(\rho^{\mathrm{M}})^\natural = v(\rho^{\mathrm{M}})^\natural$, so $u\theta = v\theta$. Similarly, if $(u, v) \in \tau - \rho$, then u and v differ only by $a^{\mathsf{L}}a$ or aa^{R}. These letters, however, have inverse images in H, and so $u\theta = v\theta$. The mapping θ is therefore well-defined on G.

Finally, $(\rho^{\mathrm{M}})^\natural\gamma\theta = \sigma^\natural|_{A^+}\theta$ (by the definition of γ) and $\sigma^\natural|_{A^+}\theta = (\rho^{\mathrm{M}})^\natural\eta$ (by the definition of θ). Therefore $\gamma\theta = \eta$, since $(\rho^{\mathrm{M}})^\natural$ maps onto S, which is the domain of γ and η. Hence, by Definition 0.8.1, (G, γ) is the universal group of S. $\boxed{1.2.8}$

The congruence ρ^{M} is defined so that S embeds in a group. Thus it embeds in its own universal group G: the homomorphism γ is therefore injective. Recall that $(\rho^{\mathrm{M}})^\natural\gamma = \sigma^\natural|_{A^+}$.

Let $(u, v) \in \sigma \cap (A^+ \times A^+)$. Then $u\sigma^\natural = v\sigma^\natural$, so $u(\rho^{\mathrm{M}})^\natural\gamma = v(\rho^{\mathrm{M}})^\natural\gamma$. The injectivity of γ implies that $u(\rho^{\mathrm{M}})^\natural = v(\rho^{\mathrm{M}})^\natural$, whence $(u, v) \in \rho^{\mathrm{M}}$. So ρ^{M} contains $\sigma \cap (A^+ \times A^+)$. The opposite inclusion was established by Lemma 1.2.7. Therefore:

$$\rho^{\mathrm{M}} = \sigma \cap (A^+ \times A^+) = \tau^\# \cap (A^+ \times A^+). \qquad \boxed{1.2.5}$$

Proposition 1.2.5 shows that two words $u, v \in A^+$ represent the same element of S if and only if there is a sequence

$$u = u_0 \to u_1 \to \ldots \to u_n = v,$$

with $n \geq 0$, where, for each $i \in \{0, \ldots, n-1\}$, there exist $p_i, q_i, q_i', r_i \in (A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}})^*$ such that $u_i = p_i q_i r_i$, $u_{i+1} = p_i q_i' r_i$, and $(q_i, q_i') \in \tau$ or $(q_i', q_i) \in \tau$.

In fact, one can find a more restrictive set of syntactic rules:

Proposition 1.2.9. *Two words $u, v \in A^+$ represent the same element of S if any only if there exists a sequence*

$$u = u_0 \to u_1 \to \ldots \to u_n = v,$$

with $n \geq 0$, where, for each $i \in \{0, \ldots, n-1\}$, there exist $p_i \in (A \cup A^{\mathsf{L}})^*$, $r_i \in (A \cup A^{\mathsf{R}})^*$, and $(q_i, q_i') \in \tau$ or $(q_i', q_i) \in \tau$ such that $u_i = p_i q_i r_i$, $u_{i+1} = p_i q_i' r_i$.

Proof of 1.2.9. See Lemma 12.19 of Clifford & Preston (1967). $\boxed{1.2.9}$

This restriction on the letters that can appear in p_i and r_i simply means that no changes can occur to the left of an a^{L} or to the right of an a^{R}. For the purposes of this thesis, a sequence of the form in Proposition 1.2.9 is called a *Malcev ρ-chain* from u to v, or, where there is no risk of confusion, simply a *Malcev chain*. If $(u,v) \in \rho^{\mathrm{M}}$ — and are thus linked by a Malcev ρ-chain — then (u,v) is said to be a *Malcev consequence* of ρ.

Malcev chains that obey this restriction have proved a useful tool. They were introduced by Malcev (1939) in order to prove his necessary and sufficient condition for group-embeddability. The original proof of Adjan's (1966*a*) sufficient condition for a semigroup to be embeddable in a group also makes use of them. (See Subsection 0.9.1 for discussion of Adjan's result.)

[*Caveat lector:* there is much disagreement on the correct terminology for sequences obeying the restriction described by Proposition 1.2.9. Adjan (1966*a*) — or at least the translated version of his work (Adjan 1966*b*) — uses the adjective 'proper' to describe Malcev chains obeying these rules, as does Spehner (1977, 1989). However, the explanation of Malcev's work in Clifford & Preston (1967) uses the word 'normal', reserving 'proper' for a stronger, more technical, restriction. Furthermore, Adjan and Spehner prefer 'Malcev sequence' to 'Malcev chain', though Clifford & Preston use the former term for a different purpose.

Matters are further obscured by Spehner's discussion of a third class of sequence obeying a restriction weaker than that in Proposition 1.2.9. These weaker rules govern the order in which generator–inverse pairs $a^{\mathsf{L}}a$ and aa^{R} can be inserted and deleted. Spehner's 'ordinary' Malcev chains always obey this lesser restriction.

This thesis follows Clifford & Preston in preferring 'Malcev chain', and, as the only Malcev chains explicitly used henceforth are as described in Proposition 1.2.9, adjectives 'proper' or 'normal' are suppressed throughout.]

1.3. BASIC PROPERTIES

THE ELEMENTARY PROPERTIES of Malcev presentations established below are generally easy to prove. They are included for the sake of rigour, for completeness, and because they have not yet appeared in the literature.

Proposition 1.3.1. *Let S be a semigroup that embeds into a group. If $\mathrm{SgM}\langle A \mid \rho\rangle$ is a Malcev presentation S, then the universal group of S is presented by $\mathrm{Gp}\langle A \mid \rho\rangle$. Conversely, if $\mathrm{Gp}\langle A \mid \rho\rangle$ is a presentation for the universal group of S, where A represents a generating set for S and $\rho \subseteq A^+ \times A^+$, then $\mathrm{SgM}\langle A \mid \rho\rangle$ is a Malcev presentation S.*

Proof of 1.3.1. The proof of Proposition 1.2.5 shows that the universal group G of S is defined by the monoid presentation $\mathrm{Mon}\langle A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}} \mid \tau\rangle$, where τ is given by (1). Arguing as in the proof of Lemma 1.2.6, the relations $(a^{\mathsf{L}}, a^{\mathsf{R}})$ for each $a \in A$ are consequences of τ. For each $a \in A$, using Tietze transformations, replace a^{L} and a^{R}) by the new symbol $a^{-1} \in A^{-1}$. So G is presented by

$$\mathrm{Mon}\big\langle A \cup A^{-1} \bigm| \rho \cup \{(aa^{-1},\varepsilon),(a^{-1}a,\varepsilon) : a \in A\}\big\rangle,$$

which, by Equation 0.5–(1), is the group $\mathrm{Gp}\langle A \mid \rho\rangle$.

To obtain the converse result, reverse the Tietze transformations to pass from $\mathrm{Gp}\langle A \mid \rho\rangle$ to $\mathrm{Mon}\langle A \cup A^{\mathsf{L}} \cup A^{\mathsf{R}} \mid \tau\rangle$. Every relation $(u, v) \in A^+ \times A^+$ that holds in G is a consequence of τ, and is therefore a Malcev consequence of those in ρ. Since the image of A^+ under $(\tau^\#)^\natural$ in G is [the embedded image of] S, the Malcev presentation $\mathrm{SgM}\langle A \mid \rho\rangle$ defines S. $\boxed{1.3.1}$

Corollary 1.3.2. *If a group-embeddable semigroup S has a finite Malcev presentation, then its universal group is finitely presented. Conversely, if the universal group of S is finitely presented and S itself is finitely generated, then S admits a finite Malcev presentation.*

Proof of 1.3.2. Any finite Malcev presentation for S is a finite presentation for G by Proposition 1.3.1.

To prove the second statement, let $\mathrm{Sg}\langle A \mid \rho\rangle$ be any presentation for S with A being finite. Then the universal group G of S is presented by $\mathrm{Gp}\langle A \mid \rho\rangle$, by Proposition 1.3.1. Since G is finitely presented, there is a finite subset σ of ρ such that $\mathrm{Gp}\langle A \mid \sigma\rangle$ is a presentation for G. Using Proposition 1.3.1 again, $\mathrm{SgM}\langle A \mid \sigma\rangle$ is a finite Malcev presentation for S. $\boxed{1.3.2}$

[Example 0.9.3 exhibits a group-embeddable semigroup that is not itself finitely generated but whose universal group is finitely generated. The finite generation condition in the second part of Corollary 1.3.2 is therefore not superfluous.]

The proof of Propostion 1.3.3 follows easily by appealling to the corresponding group-theoretical result. A direct proof exists, but is less concise.

Proposition 1.3.3. *Let S be a group-embeddable semigroup that admits a finite Malcev presentation, and let $\mathrm{SgM}\langle A \mid \rho\rangle$ be a Malcev presentation for S, with A being finite and ρ possibly being infinite. Then S admits a finite Malcev presentation $\mathrm{SgM}\langle A \mid \sigma\rangle$, where $\sigma \subseteq \rho$.*

Proof of 1.3.3. Let G be the universal group of S. Since S admits a finite Malcev presentation, G is finitely presented by Corollary 1.3.2. In particular, since G is presented by $\mathrm{Gp}\langle A \mid \rho\rangle$, it is presented by $\mathrm{Gp}\langle A \mid \sigma\rangle$, where σ is a finite subset of ρ. Since A represents a finite generating set for S, the semigroup S has a finite Malcev presentation $\mathrm{SgM}\langle A \mid \sigma\rangle$ by Proposition 1.3.1. $\boxed{1.3.3}$

Proposition 1.3.4. *Let S be a semigroup embeddable in a group. Any 'ordinary' presentation for S is also a Malcev presentation.*

Proof of 1.3.4. Suppose S has a presentation $\mathrm{Sg}\langle A \mid \rho\rangle$. Then $S \simeq A^+/\rho^\#$. Since S embeds in a group, $\rho^\# = \rho^{\mathrm{M}}$. So S has Malcev presentation $\mathrm{SgM}\langle A \mid \rho\rangle$. $\boxed{1.3.4}$

Corollary 1.3.5. *Let S be a semigroup embeddable in a group. If S is finitely presented, it admits a finite Malcev presentation.* $\boxed{1.3.5}$

The converse of Corollary 1.3.5 does not hold: see Example 1.4.2 below.

Proposition 1.3.6. *Let A be an alphabet and $\rho \subseteq A^+ \times A^+$. Let $S = \mathrm{SgM}\langle A \mid \rho\rangle$ and $M = \mathrm{MonM}\langle A \mid \rho\rangle$.*

i.) *The semigroup S is a monoid if and only if $S \simeq M$.*

ii.) *If S is not a monoid, then $S^1 \simeq M$.*

Proof of 1.3.6. Let $G = \mathrm{Gp}\langle A \mid \rho\rangle$. The group G is the universal group of both M and S. The subsemigroup of G generated by $\overline{A}$ is S, and the submonoid generated by $\overline{A}$ is M. Clearly, S is isomorphic to M if and only if S contains 1_G. Furthermore, if S does not contain an identity, then S^1 is the submonoid of G generated by $\overline{A}$, and so is isomorphic to M. $\boxed{1.3.6}$

1.4. COMPARISON WITH ORDINARY PRESENTATIONS

Proposition 1.3.1 asserts that a Malcev presentation for a semigroup that embeds in a group is actually a presentation for the universal group of that semigroup. However, Example 0.9.2 shows that the universal group of a subsemigroup S of a group G does not, in general, coincide with the subgroup of G generated by S. For this reason, the study of Malcev presentations is more than simply the study of presentations for subgroups. Indeed, in Section 6.2, an example will be given of a coherent group that contains finitely generated subsemigroups that do not admit finite Malcev presentations. (Recall that a group is coherent if all of its finitely generated subgroups are finitely presented.)

Example 1.4.1. Let F be the free group with basis $\{p, q, r, s, x, y\}$. Let the alphabet $A = \{a, b, c, d, e, f, g, h\}$ represent elements of F in the following way:

$$\begin{aligned} \overline{a} &= px, & \overline{e} &= x^{-1}r, \\ \overline{b} &= py, & \overline{f} &= y^{-1}r, \\ \overline{c} &= qy, & \overline{g} &= x^{-1}s, \\ \overline{d} &= qx, & \overline{h} &= y^{-1}s. \end{aligned}$$

Let S be the subsemigroup of F generated by $\overline{A}$.

The semigroup S has ordinary presentation $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where

$$\mathcal{R} = \{(ae, bf), (cf, de), (dg, ch), (ag, bh)\}.$$

To see this, proceed as follows. Identify elements of F with reduced words over $\{p, q, r, s, x, y\}$. All the relations in $\mathcal{R}$ are valid: for example,

$$\overline{ae} = pxx^{-1}r = pr = pyy^{-1}r = \overline{bf}.$$

Define a set of normal forms N to be all words over A that do not include subwords bf, de, ch, or bh. Every word in A^+ can be rewritten to one in N using the relations in $\mathcal{R}$; the process of rewriting must terminate since each step decreases the number of subwords bf, de, ch, and bh (since no word ae, cf, dg, or ag contains or overlaps such a subword). Therefore every element of S is represented by a word in N. With the aim of obtaining a contradiction, suppose that $u = u_1 \cdots u_m$ and $v = v_1 \cdots v_n$ are distinct words in N (where $u_i, v_i \in A$) with $\overline{u} = \overline{v}$. Without loss of generality, assume that $u_1 \neq v_1$. Letters p, q, r, and s cannot be cancelled in S, so (possibly interchanging u and v) either $u_1 = a$ and $v_1 = b$ or $u_1 = c$ and $v_1 = d$. Assume the former case: the other is similar. Since $\overline{u_1} = px$ and $\overline{v_1} = py$, the letter x or y must cancel. Again, assume the former; the latter is similar. So u_2 is either e or g. In the first case, this forces v_2 to be f; in the second, v_2 must be h. So the word v begins $bf \cdots$ or $bh \cdots$, which contradicts the fact that v lies in N. Therefore N is a set of unique normal forms for S and so $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$ presents S.

Let $\mathcal{Q} = \mathcal{R} - \{(ag, bh)\}$. The following Malcev chain shows that $(ag, bh) \in \mathcal{Q}^{\mathrm{M}}$:

$$ag \to aee^{\mathsf{R}}g \to bfe^{\mathsf{R}}g \to bc^{\mathsf{L}}cfe^{\mathsf{R}}g \to bc^{\mathsf{L}}dee^{\mathsf{R}}g \to bc^{\mathsf{L}}dg \to bc^{\mathsf{L}}ch \to bh.$$

So S has a Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{Q}\rangle$. However, $(ag, bh) \notin \mathcal{Q}^{\#}$. Indeed, no proper subset of $\mathcal{R}$ will suffice for an ordinary presentation for S.

Example 1.4.1 shows that a Malcev presentation may require fewer defining relations than an ordinary presentation. Spehner (1977) exhibited an example of a semigroup that admits a finite Malcev presentation but is not finitely presented. In fact, he compared the concept of Malcev presentations with 'ordinary' presentations; cancellative presentations; and left- and right-cancellative presentations. Cancellative and left- and right-cancellative presentations are defined in a similar manner to Malcev presentations: one is given generators, defining relations, and the fact that the semigroup being defined is cancellative, or left- or right-cancellative. Spehner gave several examples to show that a semigroup could admit:

- a finite Malcev presentation, but no finite cancellative presentation (Spehner 1977, Theorem 3.4);
- a finite cancellative presentation, but no finite left- or right-cancellative presentation (Spehner 1977, Theorem 3.1(ii));
- a finite left- or right-cancellative presentation, but no finite 'ordinary' presentation (Spehner 1977, Theorem 3.1(i)).

All of Spehner's examples were submonoids of a free monoid. His 1977 article was one of a series of papers he authored devoted to properties of finitely generated submonoids of free monoids. [Subsection 4.2.2 extends the results of another of these papers (Spehner 1974/75).] The following example exhibits a subsemigroup of a free semigroup that has a finite Malcev presentation but is not finitely presented. It is not one of Spehner's own examples.

Example 1.4.2. Let F be the free semigroup $\{x, y, z, t\}^+$. Suppose the alphabet

$A = \{a, b, c, d, e, f\}$ represents elements of F as follows:

$$\begin{aligned} \overline{a} &= x^2yz, & \overline{d} &= x^2y, \\ \overline{b} &= yz, & \overline{e} &= zy, \\ \overline{c} &= yt^2, & \overline{f} &= zyt^2, \end{aligned}$$

and let S be the subsemigroup of F generated by $\overline{A}$. Elementary reasoning shows that S has presentation $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where

$$\mathcal{R} = \{(ab^\alpha c, de^\alpha f) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

The elements $\overline{ab^\alpha} = x^2yz(yz)^\alpha$ and $\overline{b^\alpha c} = (yz)^\alpha yt^2$ have unique representatives over the alphabet A. Therefore no valid relations hold in S that can be applied to a proper subword of $ab^\alpha c$. Each of the words $ab^\alpha c$ must therefore appear as one side of a defining relation in a presentation for S on the generating set $\overline{A}$. The semigroup S is thus not finitely presented.

However, S has a finite Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{Q}\rangle$, where

$$\mathcal{Q} = \{(ac, df), (abc, def)\}.$$

Each defining relation in $\mathcal{R}$ is a Malcev consequence of the two defining relations in $\mathcal{Q}$ — that is, $\mathcal{R} \subseteq \mathcal{Q}^{\mathrm{M}}$. One can easily prove this by induction on α: assume that, for $\beta < \alpha$, the relations $(ab^\beta c, de^\beta f)$ are in $\mathcal{Q}^{\mathrm{M}}$. Then

$$\begin{aligned} ab^\alpha c &\to ab^{\alpha-1}cc^{\mathsf{R}}bc \\ &\to de^{\alpha-1}fc^{\mathsf{R}}bc && \text{(by the induction hypothesis)} \\ &\to de^{\alpha-1}d^{\mathsf{L}}dfc^{\mathsf{R}}bc \\ &\to de^{\alpha-1}d^{\mathsf{L}}acc^{\mathsf{R}}bc \\ &\to de^{\alpha-1}d^{\mathsf{L}}abc \\ &\to de^{\alpha-1}d^{\mathsf{L}}def && \text{(by the induction hypothesis)} \\ &\to de^\alpha f, \end{aligned}$$

and so $(ab^\alpha c, de^\alpha f) \in \mathcal{Q}^{\mathrm{M}}$.

1.4.1. Malcev presentations with one defining relation

Example 1.4.2 shows that a Malcev presentation with two defining relations may define a semigroup that is not finitely presented. This subsection is dedicated to proving Proposition 1.4.4, which asserts that a one-relator Malcev presentation always defines a finitely presented semigroup. This result is the first about Malcev presentations in this thesis that is neither trivial nor technical.

Lemma 1.4.3. *Any monoid with a presentation of the form*

$$\mathrm{Mon}\big\langle A \,\big|\, \{(w, \varepsilon) : w \in C\}\big\rangle, \tag{2}$$

where $C \subseteq A^*$ is closed under cyclic permutation of words, is group-embeddable.

Proof of 1.4.3. Let M be the monoid defined by the presentation (2). Let $a \in A$ be a letter that appear in some word in C. Since C is closed under cyclic permutations of words, there is a word $w' \in A^*$ such that $(w'a, \varepsilon)$ and (aw', ε) are defining relations in the presentation (2). Then $\overline{w'}\,\overline{a} = \overline{a}\,\overline{w'} = 1_M$. Thus $\overline{w'}$ is an inverse for $\overline{a}$. Therefore, any letter of a appearing in a word in C — that is, any letter involved in a defining relation in (2) — represents an element lying in a subgroup of M.

The monoid M is therefore [isomorphic to] a free product $B^* * G$, where G is a group and B consists of all letters of A that do not appear in any word in C. So the monoid M embeds into the group free product $\mathrm{FG}(B) * G$. $\boxed{1.4.3}$

Proposition 1.4.4. *The semigroup defined by a Malcev presentation with a single defining relation is finitely presented.*

Proof of 1.4.4. Let $\mathrm{SgM}\langle A \mid (u,v)\rangle$ be a Malcev presentation, where $u, v \in A^+$. Let S be the semigroup so defined. Let u' and v' be the words obtained by deleting any common prefix $p \in A^*$ and suffix $q \in A^*$ from u and v. There are three cases:

i.) Suppose $u' = v' = \varepsilon$. Then u and v must be identical, and therefore

$$S = \mathrm{SgM}\langle A \mid (u,v)\rangle \simeq \mathrm{SgM}\langle A \mid \emptyset\rangle \simeq \mathrm{Sg}\langle A \mid \emptyset\rangle \simeq A^+,$$

and so S is finitely presented.

ii.) Suppose u' and v' are both non-empty. The Malcev chain

$$u' \to p^{\mathsf{L}}pu'qq^{\mathsf{R}} = p^{\mathsf{L}}uq^{\mathsf{R}} \to p^{\mathsf{L}}vq^{\mathsf{R}} = p^{\mathsf{L}}pv'qq^{\mathsf{R}} \to v'$$

establishes that $(u', v') \in \{(u,v)\}^{\mathrm{M}}$. The presence of (u,v) in $\{(u',v')\}^{\mathrm{M}}$ is obvious, so

$$S = \mathrm{SgM}\langle A \mid (u,v)\rangle \simeq \mathrm{SgM}\langle A \mid (u',v')\rangle.$$

Since u' and v' have no common prefix or suffix, the left and right graphs of the presentation $\mathrm{Sg}\langle A \mid (u',v')\rangle$ are cycle-free. Applying Adjan's Theorem (see Theorem 0.9.12) shows that $\mathrm{Sg}\langle A \mid (u',v')\rangle$ is group-embeddable. Thus $\{(u',v')\}^{\mathrm{M}} = \{(u',v')\}^{\#}$ and so

$$\mathrm{Sg}\langle A \mid (u',v')\rangle \simeq \mathrm{SgM}\langle A \mid (u',v')\rangle = S.$$

Therefore S is finitely presented.

iii.) Suppose one of u' and v' is empty. Without loss of generality, assume that $v' = \varepsilon$. Then u' must represent the identity element, and thus by Proposition 1.3.6,

$$S = \mathrm{SgM}\langle A \mid (u,v)\rangle \simeq \mathrm{MonM}\langle A \mid (u,v)\rangle.$$

Reasoning as in part ii., $\mathrm{MonM}\langle A \mid (u,v)\rangle \simeq \mathrm{MonM}\langle A \mid (u',\varepsilon)\rangle$. Now let C be the set of all cyclic permutations of u':

$$C = \{u'[t]u'(t) : t = 0, \ldots, |u'| - 1\}.$$

The Malcev chain

$$u'[t]u'(t) \to (u'(t))^{\mathsf{L}}u'(t)u'[t]u'(t) = (u'(t))^{\mathsf{L}}u'u'(t) \to (u'(t))^{\mathsf{L}}u'(t) \to \varepsilon$$

shows that $(u'[t]u'(t), \varepsilon)$ is a Malcev consequence of (u', ε). So

$$S = \mathrm{MonM}\langle A \mid (u', \varepsilon)\rangle \simeq \mathrm{MonM}\langle A \mid \{(w, \varepsilon) : w \in C\}\rangle.$$

By Lemma 1.4.3, the monoid with presentation $\mathrm{Mon}\langle A \mid \{(w,\varepsilon) : w \in C\}\rangle$ is group-embeddable and so is isomorphic to the monoid with the equivalent Malcev presentation, namely S. Therefore S is finitely presented. $\boxed{1.4.4}$

1.5. THE THEORY OF MALCEV PRESENTATIONS IN 2003

PRIOR TO the recent work of Cain, Robertson & Ruškuc, the theory of Malcev presentations had perhaps been neglected. Only two papers had ever been written on the subject: the article in which they were introduced (Spehner 1977), and a later paper, also by Spehner, dedicated to proving the following result:

Theorem 1.5.1 (Spehner 1989). *Every finitely generated submonoid of a free monoid admits a finite Malcev presentation.* $\boxed{1.5.1}$

[In the terminology of Section 1.6, this result — or more precisely the subsemigroup version of this result — becomes: 'free monoids are Malcev coherent'.]

Thus stood the theory of Malcev presentations in early 2003. Since then, a great deal of progress, discussed elsewhere in this thesis, has been made.

1.6. COHERENCE AND MALCEV COHERENCE

A GROUP is *coherent* if all of its finitely generated subgroups are finitely presented. [Serre (1974) coined the term 'coherent group'.] There are few known examples of coherent groups: free groups; abelian, nilpotent, or polycyclic groups; surface groups; the fundamental groups of three-dimensional manifolds (Scott 1973); Baumslag–Solitar groups (Kropholler 1990); mapping tori of free group automorphisms (Feighn & Handel 1999); and groups of the form $\mathrm{Gp}\langle A \mid w^n\rangle$ for large n (McCammond & Wise 2005). [Baumslag (1974, Section B) asks whether all one-relator groups are coherent.] On the other hand, $\mathrm{SL}_n(\mathbb{Z})$ and $\mathrm{GL}_n(\mathbb{Q})$ are incoherent for $n \geq 4$ (Serre 1974, Lennox & Wiegold 1974). The Rips (1982) construction shows that small cancellation groups may be incoherent.

Proposition 1.6.1. *The class of coherent groups is closed under:*

i.) *forming free products.*

ii.) *constructing finite extensions.*

The two assertions in this result are well-known, but do not seem to appear explicitly anywhere in the literature. The proof is therefore included below.

Proof of 1.6.1. i.) Let $\{G_i : i \in I\}$ be a collection of coherent groups. Let H be a finitely generated subgroup of the free product $\prod^*_{i \in I} G_i$. Then, by the Kurosh Subgroup Theorem (see Lyndon & Schupp 1977, Section III.3), the subgroup H is of the form

$$F * \prod_{j \in J}^{*} H_j, \tag{3}$$

where F is a free group and each group H_j is a subgroup of a conjugate of one of the free factors G_i. Since H is finitely generated, the free group F and each H_j are finitely generated. The free group F is then clearly finitely presented; since each G_i is coherent, each H_j is finitely presented. Therefore H — being the free product (3) — is finitely presented. Thus the free product $\prod^*_{i \in I} G_i$ is coherent.

ii.) Let E be a finite extension of a coherent group G. Let K be a finitely generated subgroup of E. Then K is a finite extension of $G \cap K$ by Lemma 0.10.4. Since K is finitely generated, the Reidemeister–Schreier Theorem (see Lyndon & Schupp 1977, Section II.4) asserts that $G \cap K$ is also finitely generated. The group G is coherent, so its subgroup $G \cap K$ is finitely presented. Again by the Reidemeister–Schreier Theorem, the group K is also finitely presented. The subgroup K was arbitrary; the group E is therefore coherent. $\boxed{1.6.1}$

The class of coherent groups is *not* closed under taking direct products: the direct product of two free groups of rank 2 is not coherent (Grunewald 1978). Indeed, every finitely presented subgroup of such a direct product is a finite extension of a direct product of two free groups (Baumslag & Roseblade 1984). However, the direct product of a free group and a polycyclic group is coherent (see Theorem 7.4.1).

Extend the concept of coherence to Malcev presentations as follows:

Definition 1.6.2. A group — or more generally a group-embeddable semigroup — is *Malcev coherent* if all of its finitely generated subsemigroups have finite Malcev presentations.

[The term 'Malcev coherent' is of recent provenance: neither Spehner (1977, 1989) nor Cain, Robertson & Ruškuc (2005*a*, 2005*b*) use it.]

A large part of this thesis is dedicated to proving results on Malcev coherence. Chapters yet to come contain proofs of the Malcev coherence of virtually free groups (Theorem 4.3.1); virtually nilpotent groups (Theorem 5.3.5); free products of free monoids and abelian groups (Corollary 6.3.5); and direct products of virtually free groups and abelian groups (Theorem 7.5.1). Theorem 6.2.6 also establishes that the class of Malcev coherent groups, unlike that of coherent groups, is not closed under forming free products.

Proposition 1.6.3. *Every Malcev coherent group is coherent.*

Proof of 1.6.3. Let G be a Malcev coherent group. Let H be a finitely generated subgroup of G. Then H has a finite Malcev presentation. By Proposition 1.3.1, the universal group of H — which is H itself — is finitely presented. Since H was an arbitrary finitely generated subgroup of G, this establishes the coherence of G. $\boxed{1.6.3}$

Theorem 6.2.7 proves that the class of Malcev coherent groups is properly contained in the class of coherent groups. Proposition 1.6.1 therefore provokes the question, considered in Section 9.5, of whether the class of Malcev coherent groups is closed under finite extensions.

CHAPTER TWO

AUTOMATIC SEMIGROUPS

In the present chapter I propose to consider whether anything... can be inferred from the structure of language as to the structure of the world.
— Bertrand Russell,
An Inquiry into Meaning and Truth (1940), ch. xxv

2.1. INTRODUCTION & HISTORICAL SKETCH

AUTOMATIC GROUPS were first studied in the 1980s, starting from results of Cannon (1984) on the geometry of the Cayley graph of a group of isometries of hyperbolic space. Thurston observed that these geometric results could be expressed using two-tape finite state automata. During the following few years, the basic theory of automatic groups was established and published by Epstein, Cannon, Holt, Levy, Paterson & Thurston in *Word Processing in Groups* [Boston: Jones & Bartlett, 1992], which remains the definitive reference in the area. The subject has stayed active since, although many of the questions posed by Epstein et al. remain open.

The definition of an automatic group treats the group as being generated as a semigroup. There is therefore a natural extension of the definition to semigroups. The earliest instance in the literature of such an extension is due to Hudson (1996), who explored the connection between automatic structures and string rewriting systems. Campbell, Robertson, Ruškuc & Thomas (2001) made the first systematic study of automatic semigroups. They generalized some parts of the theory of automatic groups to semigroups, showed that other properties did not generalize, proved a number of original results, and asked some questions.

Campbell et al. discovered that many of the elegant properties of automatic groups do not hold when one passes to general automatic semigroups. As a consequence, there has been much investigation into automatism in classes of semigroups closely related to groups: for example, completely simple semigroups (Campbell, Robertson, Ruškuc & Thomas 2002); monoids (Duncan, Robertson & Ruškuc 1999); and semigroups embeddable into groups (Cain et al. 2005*a*, Cain 2005*b*). Dombi (2004), noting that the free inverse semigroup is not automatic (Cutting & Solomon

2001), developed the new notions of Schützenberger and strong Schützenberger automatism for regular semigroups.

This chapter has three main purposes:

- Firstly, to introduce the basic theory of automatic semigroups. Sections 2.2 and 2.4 give the necessary definitions and elementary results.
- Secondly, to expound the elegant geometric characterizations of automatic groups and their generalization to automatic group-embeddable semigroups (Section 2.3).
- Thirdly, to establish that all automatic semigroups embeddable into groups have finite Malcev presentations (Section 2.5). [Section 3.2 gives an algorithm that obtains a finite Malcev presentation from an automatic structure.]

[This thesis disagrees with the general trend in the literature by using the term 'automatism' rather than 'automaticity' to mean 'the condition of being automatic'. The author finds the latter term both inelegant and etymologically unsound: the root of 'automatic' is the Greek αὐτόματος, but '-ity' comes from the Latin *-itātem*. The suffix '-ism' has the Greek origin -ισμός. (*Oxford English Dictionary*, Second Edition, 1989)]

2.2. SYNCHRONOUS & ASYNCHRONOUS AUTOMATIC STRUCTURES

THERE ARE two slightly different modes of automatism, called *synchronous* and *asynchronous*. (The reason behind these names is explained below.) Most of the research carried out on automatic semigroups has concentrated on synchronous automatism. The only articles in the literature that consider asynchronous automatic semigroups are Hoffmann et al. (2002*a*) and Cain, Robertson & Ruškuc (2005*a*). Epstein et al. (1992) treat of groups that are synchronously and asynchronously automatic, but introduce them in separate chapters. This section defines both of the analogous concepts for semigroups.

Definition 2.2.1. Let S be a semigroup. A *rational structure* for S is a pair (A, L), where A is a finite alphabet representing a set of generators for S, and L is a regular language over A such that $\overline{L} = S$.

[Sakarovitch (1987) uses the term 'rational structure' for a stronger concept, whereby each element of S has exactly one representative in L: the language maps bijectively onto the semigroup.]

Let (A, L) be a rational structure for a semigroup S. For each $a \in A \cup \{\varepsilon\}$, define

$$L_a = \{(u, v) : u, v \in L, \overline{ua} = \overline{v}\}. \tag{1}$$

Let \$ be a new symbol not in A. The symbol \$ is usually called the *padding symbol*. The *padded alphabet* $A(2, \$)$ is the set $\{(a, b) : a, b \in A \cup \{\$\}\} - \{(\$, \$)\}$. Define the mapping $\delta_A : A^+ \times A^+ \to A(2,\$)^+$ by

$$(u_1 \cdots u_m, v_1 \cdots v_n) \mapsto \begin{cases} (u_1, v_1) \cdots (u_m, v_n) & \text{if } m = n, \\ (u_1, v_1) \cdots (u_n, v_n)(u_{n+1}, \$) \cdots (u_m, \$) & \text{if } m > n, \\ (u_1, v_1) \cdots (u_m, v_m)(\$, v_{m+1}) \cdots (\$, v_n) & \text{if } m < n, \end{cases}$$

where $u_i, v_i \in A$.

Definition 2.2.2. Let S be a semigroup. A *synchronous automatic structure* for S is a rational structure (A, L) such that, for each $a \in A \cup \{\varepsilon\}$, the set $L_a\delta_A$ is a regular language over $A(2,\$)$. A *synchronous automatic semigroup* is a semigroup that admits a synchronous automatic structure.

Definition 2.2.2 is essentially the definition of an automatic group of Epstein et al. (1992, Definition 2.3.1) with 'group' replaced by 'semigroup'. Therefore a group that is an automatic semigroup in the sense of Definition 2.2.2 is an automatic group in the established sense.

The following proposition is a technical result which will prove useful elsewhere.

Proposition 2.2.3 (Campbell et al. 2001, Propositions 2.2 and 3.2). *Let A be a finite alphabet and let M and N be subsets of $A^+ \times A^+$ such that $M\delta_A$ and $N\delta_A$ are regular languages over $A(2,\$)$. Then $(M \circ N)\delta_A$ — $M \circ N$ being the composition of the relations M and N — is also a regular language over $A(2,\$)$. Furthermore, from finite state automata recognizing $M\delta_A$ and $N\delta_A$, one can effectively construct an automaton recognizing $(M \circ N)\delta_A$. In particular, if (A, L) is a synchronous automatic structure, the language*

$$\begin{aligned} L_{a_1 \cdots a_n}\delta_A &= (L_{a_1} \circ L_{a_2} \circ \cdots \circ L_{a_n})\,\delta_A \\ &= \{(u, v) : u, v \in L, \overline{ua_1 \cdots a_n} = \overline{v}\}\delta_A \end{aligned}$$

is regular, where $a_i \in A$. $\boxed{2.2.3}$

[In an automatic group or monoid, one can always find a word $u_a \in L$ representing the same element as the letter $a \in A$: one simply enumerates words in L until one finds a word w such that $L_w = L_\varepsilon$. This word w must represent the identity of the group or monoid. (This process is effective by Proposition 2.2.3; by the fact that the images of L_w and L_ε under δ_A are regular languages and so their equality can be checked by Theorem A.5.5; and by the knowledge that an identity must be present and so such a word w must eventually be found.) Having found this representative for the identity, one simply uses L_a to find a word $u_a \in L$ representing $\overline{wa} = \overline{a}$. The question of whether one can find such a representative in a general automatic semigroup is open. However, knowledge of these representatives for generators appears so fundamental that if one *cannot* compute them, then the definition of an automatic semigroup is in some sense incomplete. Therefore, this thesis assumes that an explicitly given automatic structure (A, L) includes, for each

$a \in A$, a specified word $u_a \in L$ representing the generator $\overline{a}$. This assumption is especially important in Chapter 3, which studies algorithmic questions for automatic semigroups.]

The term 'synchronous' refers to the fact that a finite state automaton recognizing the language $L_a\delta_A$ may be thought of as a two-tape automaton that reads its two inputs — one for u and one for v, where $(u, v) \in L_a$ — at the same speed. Thus the two 'read heads' on the input tapes are synchronized. There is a broader concept of automatism where one allows the two heads to advance at different speeds, or to consume their input asynchronously. Recall that rational relations are those recognized by [possibly non-deterministic] asynchronous two-tape automata (see Theorem A.6.2).

Definition 2.2.4. Let S be a semigroup. An *asynchronous automatic structure* for S is a rational structure (A, L) such that, for each $a \in A \cup \{\varepsilon\}$, the relation L_a is rational. An *asynchronous automatic semigroup* is a semigroup that admits an asynchronous automatic structure.

Definition 2.2.4 was first stated by Hoffmann et al. (2002*a*), who assert that any group that is an asynchronous automatic semigroup in this sense is an asynchronous automatic group in the sense of Epstein et al. (1992, Definition 7.2.1). The fact that Hoffmann et al. do not substantiate this statement is rather puzzling: their assertion is not obviously true, and the only explanation required is a reference to the work of Shapiro (1992). The difference between the definition above and that of Epstein et al. is that the latter requires that the relations L_a are recognized by *deterministic* asynchronous automata. Shapiro shows that any group that admits an asynchronous automatic structure (A, L) in the sense of Definition 2.2.4 possesses an asynchronous automatic structure (A, K), where $K \subseteq L$, in which all the relations K_a are recognized by deterministic asynchronous automata. Subsection 2.3.2 discusses Shapiro's work further, and in particular extends it to asynchronous automatic semigroups embeddable into groups.

Theorem 2.2.5 (Hoffmann et al. 2002*a*, Proposition 6.1). *A synchronous automatic semigroup is also asynchronously automatic.* $\boxed{2.2.5}$

The classes of synchronous and asynchronous automatic semigroups do not, however, coincide: Epstein et al. (1992, Example 7.4.1) show that the Baumslag–Solitar groups

$$\mathrm{Gp}\langle x, y \mid (yx^m, x^n y)\rangle,$$

where $m, n \in \mathbb{N}$, are asynchronously automatic but not synchronously automatic for $m \neq n$.

By default, for the purposes of this thesis, 'automatic' means 'synchronously automatic'.

[An automatic or asynchronous automatic structure (A, L) uniquely determines the semigroup to which it corresponds. The elements of the semigroup are the L_ε-equivalence classes of L. The product of the classes $[u]_{L_\varepsilon}$ and $[v]_{L_\varepsilon}$ is $[w]_{L_\varepsilon}$, where $w \in L$ is any word such that $(u, w) \in L_v$.]

Definition 2.2.6. A semigroup is *locally automatic* if all of its finitely generated subsemigroups admit automatic structures. Similarly, a semigroup is *locally asynchronously automatic* if all of its finitely generated subsemigroups are asynchronously automatic.

[Descalço (2002, p. 89) uses the term 'strongly automatic' instead of 'locally automatic'. The partial homonymy between the 'strong automatism' of Descalço and the 'strong Schützenberger automatism' of Dombi (2004, Definition 3.2.5) is unfortunate. The two concepts bear no resemblance.]

2.2.1. Notions of automatism

The definition of an automatic structure for a semigroup (Definition 2.2.2) is intrinsically 'right-handed': the relations L_a describe multiplication on the right by a generator; and the mapping δ_A inserts padding symbols \$ on the right. Noting this, Hoffmann & Thomas (2003) distinguished four parallel concepts of automatism for semigroups, depending on whether multiplication is carried out on the right or on the left and whether padding symbols are inserted at the right or at the left.

Let (A, L) be a rational structure for a semigroup S. For each $a \in A \cup \{\varepsilon\}$, define 'left-handed multiplication' relations

$$_aL = \{(u, v) : u, v \in L, \overline{au} = \overline{v}\}.$$

The definition of the relations $_aL$ parallels that of the normal 'right-handed multiplication' relations (1).

Define the 'left-handed padding' map $\gamma_A : A^+ \times A^+ \to A(2, \$)^+$ by

$$(u_1 \cdots u_m, v_1 \cdots v_n) \mapsto \begin{cases} (u_1, v_1) \cdots (u_m, v_n) & \text{if } m = n, \\ (u_1, \$) \cdots (u_{m-n}, \$)(u_{m-n+1}, v_1) \cdots (u_m, v_n) & \text{if } m > n, \\ (\$, v_1) \cdots (\$, v_{n-m})(u_1, v_{n-m+1}) \cdots (u_m, v_n) & \text{if } m < n, \end{cases}$$

where $u_i, v_i \in A$.

Definition 2.2.7. Let S be a semigroup. A rational structure (A, L) is a:

- *right–right automatic structure* for S if, for each $a \in A \cup \{\varepsilon\}$, the language $L_a\delta_A$ is regular;
- *right–left automatic structure* for S if, for each $a \in A \cup \{\varepsilon\}$, the language $_aL\delta_A$ is regular;
- *left–right automatic structure* for S if, for each $a \in A \cup \{\varepsilon\}$, the language $L_a\gamma_A$ is regular;
- *left–left automatic structure* for S if, for each $a \in A \cup \{\varepsilon\}$, the language $_aL\gamma_A$ is regular.

A semigroup is P–Q *automatic* if it admits a P–Q automatic structure, where P and Q are each either 'right' or 'left'.

[Observe that the first 'right' or 'left' describes the location of the padding; the second described the multiplication. 'Right–right automatic' is the same as 'automatic' in the standard sense. One could also define the concept of 'left asynchronous automatism', where one simply requires that the relations ${}_aL$ are rational.]

Hoffmann & Thomas (2003, Section 7) gave examples to show that the four concepts of automatism are independent: a semigroup can possess any subset of the four concepts of automatism but not possess any of the remainder. There are, however, some positive results:

Proposition 2.2.8 (Hoffmann & Thomas 2003, Theorem 5.8). *If a group is automatic in any of the four senses of Definition 2.2.7, then it is automatic in each of the other senses.* $\boxed{2.2.8}$

Proposition 2.2.9 (Hoffmann & Thomas 2003, Remark 8.3). *A cancellative semigroup is left–Q automatic if and only if it is right–Q automatic, where Q is either 'right' or 'left'.* $\boxed{2.2.9}$

[Therefore, it makes sense to speak of right or left automatism for cancellative semigroups. An example of a cancellative (and indeed group-embeddable) semigroup that is right automatic but not left automatic is the Baumslag–Solitar semigroup $\mathrm{Sg}\langle x, y \mid (yx^m, x^n y)\rangle$, where $m > n$; if $m < n$, the semigroup is left automatic but not right automatic (Hoffmann 2001, Corollary 4.20).]

Results elsewhere in this chapter discuss only 'standard' (right–right) automatism. However, they all have obvious variants that deal with the other concepts of automatism discussed above. These variations are only called upon in Subsection 8.2.2.

2.3. GEOMETRIC PROPERTIES

SYNCHRONOUS AUTOMATIC STRUCTURES for groups have a very elegant geometric characterization in terms of the 'fellow traveller property'. [Asynchronous automatic structures for groups also have a geometric characterization, albeit slightly less elegant.] Campbell et al. (2001) showed that automatic structures for semigroups *possess* the fellow traveller property, but are not characterized by it, and asked whether a geometric characterization existed for automatic semigroups. Campbell et al. (2002, Proposition 1.3) proved that automatic structures for completely simple semigroups, like those for groups, are characterized by the fellow traveller property. Hoffmann (2001, Theorem 8.11) gave a characterization of automatic structures for general semigroups, but its nature is more linguistic than geometric.

This section shows that the geometric characterizations of synchronous and asynchronous automatic structures for groups (Epstein et al. 1992, Theorems 2.3.5 & 7.2.8) extend to semigroups embeddable into groups. The synchronous case is dealt with first, as the asynchronous case is more technical even for groups alone.

2.3.1. Synchronous automatic structures

Definition 2.3.1. Let S be a semigroup. Let (A, L) be a rational structure for S. Let $\lambda \in \mathbb{N}$. The paths $\widehat{u}$ and $\widehat{v}$ in $\Gamma(S, A)$, labelled by words u and v in L, are said to λ-*fellow travel* if, for all $t \in \mathbb{N} \cup \{0\}$, the distance (in the Cayley graph $\Gamma(S, A)$) from $\overline{u(t)}$ to $\overline{v(t)}$ is at most λ. [That is, if one traces along both paths at the same 'speed', the two 'current' points are at most λ apart.]

The rational structure (A, L) is said to have the *fellow traveller property* if there exists a constant $\lambda \in \mathbb{N}$ such that, for all $a \in A \cup \{\varepsilon\}$ and $(u, v)\delta_A \in L_a$, the paths that u and v label λ-fellow travel. In this case, the constant λ is called a *fellow traveller constant* for (A, L).

Theorem 2.3.2. *Let S be a semigroup that embeds in a group. Let (A, L) be a rational structure for S. Then (A, L) is an automatic structure for S if and only if it has the fellow traveller property.*

Proof of 2.3.2. Follow the proof of Epstein et al. (1992, Theorem 2.3.5), with one difference: rather than using an open ball in the Cayley graph of S, take the ball to be in the Cayley graph of a group into which S embeds (such as the universal group of S) and use Proposition 0.9.5 and Corollary 0.9.6. $\boxed{2.3.2}$

Actually, the proof that automatic structures possess the fellow traveller property shows that a slightly stronger result holds:

Proposition 2.3.3. *Let S be a semigroup with automatic structure (A, L). Let $a \in A \cup \{\varepsilon\}$ and let $(u, v) \in L_a$. Then there exists a constant $\lambda \in \mathbb{N}$, dependent only on (A, L), such that for all $t \in \mathbb{N} \cup \{0\}$, there exist $p_t, q_t \in A^*$ such that $\overline{u(t)p_t} = \overline{v(t)q_t}$ with $|p_t|, |q_t| < \lambda/2$.* $\boxed{2.3.3}$

[Whilst the constant $\lambda \in \mathbb{N}$ in Proposition 2.3.3 is a fellow traveller constant for (A, L), there may exist a fellow traveller constant that does not have the given property: the words p_t and q_t may not label the shortest path between the elements $\overline{u(t)}$ and $\overline{v(t)}$.]

Observe that Proposition 2.3.3 makes no mention of group-embeddability: it applies to all automatic semigroups.

2.3.2. Asynchronous automatic structures

As the comments following Definition 2.2.4 explain, the original definition of asynchronous automatic structures (Epstein et al. 1992, Definition 7.2.1) requires that the relations L_a be recognized by deterministic asynchronous automata. Were one only concerned with one-tape finite state automata, this would obviously be no restriction: as Theorem A.4.3 asserts, any language recognized by a finite state automaton is recognized by a deterministic one. However, the class of relations recognized by deterministic asynchronous automata is strictly contained in the class recognized by all asynchronous automata (see Section A.6). This raises the possibility that there may exist groups that are asynchronously automatic in the sense of Definition 2.2.4 but not in the sense of Epstein et al. (1992, Definition 7.2.1). Shapiro (1992) shows that no such groups exist, and in so doing simplifies the original geometric characterization of asynchronous automatic structures for groups (Epstein et al. 1992, Theorem 7.2.8).

The remainder of this section is devoted to showing that Shapiro's technique extends to asynchronous automatic structures for semigroups embeddable into groups, thus generalizing both the original and the simpler geometric characterizations. First of all, it is necessary to assemble various definitions and results extending those of Epstein et al. to semigroups embeddable into groups.

Recall from Definition A.6.4 that an asynchronous automaton $\mathcal{A}$ is called *boundedly asynchronous* if there exists a constant $k \in \mathbb{N}$ such that $\mathcal{A}$ never reads more than k letters consecutively from one tape.

Definition 2.3.4. A *boundedly asynchronous automatic structure* is one in which all the relations L_a are recognized by boundedly asynchronous automata. An asynchronous automatic structure (A, L) is *deterministic* if all the relations L_a are recognized by deterministic asynchronous automata.

Theorem 2.3.5. *Let S be a semigroup that embeds in a group. Let (A, L) be a deterministic asynchronous automatic structure for S. Then there exists a language $K \subseteq L$ such that (A, K) is a deterministic boundedly asynchronous automatic structure for S.*

Proof of 2.3.5. Reason as in the proof of Epstein et al. (1992, Theorem 7.2.4). At those points where the identity or inverses are required, work in the universal group of S. $\boxed{2.3.5}$

Definition 2.3.6. Let S be a semigroup, and (A, L) a rational structure for S. A *departure function* for (A, L) is a function $D : \mathbb{R} \to \mathbb{R}$ such that, if $w \in L$, $r, s \geq 0$, $t \geq D(r)$, and $s + t \leq |w|$, then the distance between $\overline{w(s)}$ and $\overline{w(s + t)}$ in the Cayley graph of S exceeds r.

The existence of a departure function means that every word in L labels a path in the Cayley graph that eventually departs from every finite neighbourhood. The following lemma is useful in establishing the existence of departure functions for certain rational structures.

Lemma 2.3.7. *Suppose (A, L) is a rational structure for S, where S is a semigroup that embeds into a group. Suppose that L maps finite-to-one onto S. For all $r \in \mathbb{N}$, there exist only finitely many words $y \in A^+$ such that $xyz \in L$ for some $x, z \in A^*$ and $d(\overline{x}, \overline{xy}) < r$.*

Proof of 2.3.7. The proof of Lemma 7.4 of Baumslag, Gersten, Shapiro & Short (1991) applies unchanged. $\boxed{2.3.7}$

Definition 2.3.8. Let S be a semigroup and A an alphabet representing a set of generators for S. Let $u, v \in A^+$. The *Hausdorff distance* between the two paths $\widehat{u}, \widehat{v}$ in $\Gamma(S, A)$ is

$$h = \inf\left\{ r : \hat{u} \subseteq \bigcup B_r(\overline{v(t)}) \text{ and } \hat{v} \subseteq \bigcup B_r(\overline{u(t)}) \right\}.$$

In other words, every point on $\widehat{u}$ is at most h from some point on $\widehat{v}$, and vice versa.

A rational structure (A, L) for S has the *Hausdorff closeness property* if there exists a constant $\lambda \in \mathbb{N}$ such that, for all $a \in A \cup \{\varepsilon\}$ and $(u, v) \in L_a$, the paths $\widehat{u}$ and $\widehat{v}$ in the Cayley graph of S are at most a Hausdorff distance λ from one another.

Definition 2.3.9. Suppose (A, L) a rational structure for a semigroup S. The structure (A, L) has the *asynchronous fellow traveller property* if there exists a constant $\lambda \in \mathbb{N}$ such that, for all $a \in A \cup \{\varepsilon\}$ and $(u, v) \in L_a$, there are monotone increasing functions

$$\phi, \psi : \mathbb{N} \cup \{0\} \to \mathbb{N} \cup \{0\}$$

for which $d(\overline{u(\phi(t))}, \overline{v(\psi(t))}) < \lambda$ for all $t \in \mathbb{N} \cup \{0\}$. In this case, the constant λ is called an *asynchronous fellow traveller constant* for (A, L).

Proposition 2.3.10. *Let (A, L) be a rational structure for a semigroup S, and suppose (A, L) has the asynchronous fellow traveller property. Then (A, L) has the Hausdorff closeness property.*

Proof of 2.3.10. Suppose (A, L) has the asynchronous fellow traveller property. Let $(u, v) \in L_a$ for some $a \in A \cup \{\varepsilon\}$. Let λ be the constant and ϕ and ψ be monotone functions as in Definition 2.3.9. Let $s \in \mathbb{N} \cup \{0\}$. Pick $t \in \phi^{-1}(s)$. Then

$$d(\overline{u(s)}, \overline{v(\psi(t))}) = d(\overline{u(\phi(t))}, \overline{v(\psi(t))}) < \lambda.$$

Since s was arbitrary, every point on the path $\widehat{u}$ is a distance at most λ from some point on $\widehat{v}$. Similarly, every point on the path $\widehat{v}$ is at most λ from some point on $\widehat{u}$. Since $a \in A \cup \{\varepsilon\}$ and $(u, v) \in L_a$ were arbitrary, (A, L) has the Hausdorff closeness property. $\boxed{2.3.10}$

The following theorem, when restricted to groups, is the original characterization of asynchronous automatic structures due to Epstein et al. (1992).

Theorem 2.3.11. *Let S be a semigroup that embeds in a group. Let (A, L) be a rational structure for S. Then (A, L) is a deterministic boundedly asynchronous automatic structure for S if and only if the following two conditions hold:*

i.) *There exists a departure function for (A, L).*

ii.) *The structure (A, L) has either the Hausdorff closeness property or the asynchronous fellow traveller property.*

Proof of 2.3.11. The proof is analogous to that of Theorem 7.2.8 of Epstein et al. (1992). There is one difference: rather than using an open ball in the Cayley graph of S, take the ball to be in the Cayley graph of a group into which S embeds (such as the universal group of S). $\boxed{2.3.11}$

The proof of Theorem 2.3.11 actually shows that a slightly stronger version of the 'if' part of the result holds:

Proposition 2.3.12. *Let S be a semigroup with deterministic boundedly asynchronous automatic structure (A, L). Let $a \in A \cup \{\varepsilon\}$ and let $(u, v) \in L_a$. Consider a deterministic boundedly asynchronous two-tape automaton recognizing L_a. When the automaton has read t letters of (u, v) in total, let $s_{\mathsf{L}}(t)$ be the number of letters read from the left-hand tape and $s_{\mathsf{R}}(t)$ the number read from the right-hand tape, so that $t = s_{\mathsf{L}}(t) + s_{\mathsf{R}}(t)$. There exists a constant $\lambda \in \mathbb{N}$, dependent only on (A, L), such that for all $t \in \mathbb{N} \cup \{0\}$, there exist $p_t, q_t \in A^*$ such that $\overline{u(s_{\mathsf{L}}(t))p_t} = \overline{v(s_{\mathsf{R}}(t))q_t}$ with $|p_t| + |q_t| < \lambda$.* $\boxed{2.3.12}$

[The constant $\lambda \in \mathbb{N}$ in Proposition 2.3.12 is an asynchronous fellow traveller constant for (A, L), but there may exist asynchronous fellow traveller constants that do not have the given property.]

Having assembled the necessary theory for deterministic asynchronous automatic semigroups embeddable into groups, one can now turn to Shapiro's (1992) results.

Theorem 2.3.13. *Let S be a semigroup that embeds in a group. Let (A, L) be an asynchronous automatic structure for S. Then there exists a regular language $K \subseteq L$ with $\overline{K} = S$ such that the two conditions of Theorem 2.3.11 are satisfied:*

i.) *There is a departure function for (A, K).*

ii.) *The rational structure (A, K) has the asynchronous fellow traveller property.*

That is, (A, K) forms a deterministic boundedly asynchronous automatic structure for S.

Proof of 2.3.13. Proceed as in the proof of Theorem 1 of Shapiro (1992), working in the universal group as necessary. $\boxed{2.3.13}$

Theorem 2.3.11 characterizes deterministic boundedly asynchronous automatic structures in terms of departure functions and a choice of two 'distance properties'. The presence of the word 'boundedly' and the requirement of the departure function seem to make this characterization less 'natural' than the corresponding result for synchronous automatism (Theorem 2.3.2). Shapiro, by allowing non-determinism, gave a more elegant characterization which does generalize to semigroups embeddable into groups:

Theorem 2.3.14. *Let S be a semigroup that embeds in a group. Let (A, L) be a rational structure for S. Then (A, L) is an asynchronous automatic structure for S if and only if it has the asynchronous fellow traveller property.*

Proof of 2.3.14. The construction of Shapiro (1992, Theorem 2) generalizes by taking the ball to be in the Cayley graph of a group into which S embeds, such as the universal group of S. $\boxed{2.3.14}$

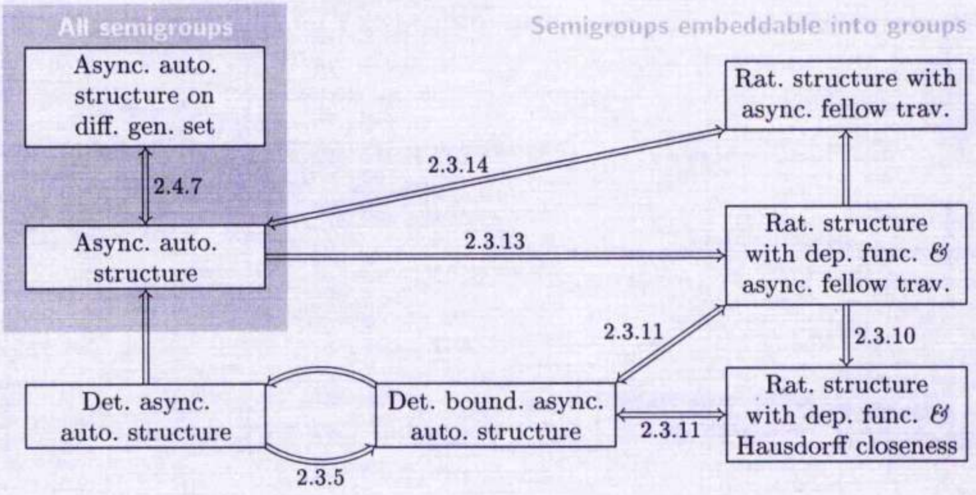

Figure 2.1. Characterization of asynchronous automatic structures for semigroups embeddable into groups. The label on an arrow refers to the result stating that implication. Unlabelled arrows are trivial implications.

2.4. BASIC RESULTS FOR GENERAL AUTOMATIC SEMIGROUPS

2.4.1. Word problem

THE RESULTS in this subsection apply to all automatic semigroups.

Theorem 2.4.1 (Campbell et al. 2001, Theorem 3.6). *Let S be a semigroup with automatic structure (A, L). There is an algorithm that takes a word $w \in A^+$ and returns a word in L representing the same element of S as w. This algorithm completes in time proportional to $|w|^2$.* $\boxed{2.4.1}$

The algorithm of Theorem 2.4.1 shows that the word problem is solvable for automatic semigroups: given two words $u, v \in A^+$, one obtains words $u', v' \in L$ representing the same elements of S as u and v. Then $\overline{u} = \overline{v}$ if and only if $(u', v')\delta_A \in L_\varepsilon$, and membership of L_ε is testable by Theorem A.5.3.

A similar result applies to asynchronous automatic semigroups:

Theorem 2.4.2 (Hoffmann et al. 2002*a*, p. 387). *Let S be a semigroup with asynchronous automatic structure (A, L). There is an algorithm that takes a word $w \in A^+$ and returns a word in L representing the same element of S as w. This algorithm completes in exponential time.* $\boxed{2.4.2}$

2.4.2. Finite Rees index subsemigroups and extensions

The following three results show how automatism is preserved under adjoining or removing elements from semigroups.

Theorem 2.4.3 (Campbell et al. 2001, Theorem 7.2). *Let S be a semigroup. The semigroup S^1 formed by adjoining a two-sided identity to S is automatic if and only* *if S is automatic. [Moreover, from an automatic structure for S one can effectively construct an automatic structure for S^1; and from an automatic structure for S^1 one can effectively construct an automatic structure for S.]* $\boxed{2.4.3}$

Theorem 2.4.4 (Campbell et al. 2001, Proposition 3.13). *Let S be a semigroup. The semigroup S^0 formed by adjoining a two-sided zero to S is automatic if and only if S is automatic.* $\boxed{2.4.4}$

Theorem 2.4.5 (Hoffmann, Thomas & Ruškuc 2002*b*, Theorem 1.1). *Let S be a semigroup and let T be a subsemigroup of finite Rees index. (That is, with $|S - T|$ being finite.) Then S is automatic if and only if T is automatic.* $\boxed{2.4.5}$

Of course, Theorems 2.4.3 and 2.4.4, though established earlier, follow as corollaries of Theorem 2.4.5.

2.4.3. Changing generators

Theorem 2.4.6 (Duncan et al. 1999, Theorem 1.1). *Let M be a monoid with automatic structure (A, L). Let B be any finite alphabet representing a set of semigroup generators for M. Then M has an automatic structure (B, K), where K is a regular language. Moreover, one can effectively construct the automatic structure (B, K) from (A, L), a set of words over B representing the generators in $\overline{A}$, and a set of words over A representing the generators $\overline{B}$.* $\boxed{2.4.6}$

Theorem 2.4.6 generalizes Theorem 2.4.1 of Epstein et al. (1992), which applies only to groups. This result does not hold for automatic semigroups generally: Campbell et al. (2001, Example 4.5) give an example of a finitely generated subsemigroup of a free semigroup that does not have an automatic structure on all generating sets. The free semigroup embeds in the free group, so this example shows that Theorem 2.4.6 does not generalize to include semigroups embeddable into groups.

Observe that the alphabets A and B of Theorem 2.4.6 represent *semigroup* generating sets for the monoid M. One could also define automatic structures for monoids using a *monoid* generating set. Otto & Sattler-Klein (1997) observe that, using this definition, a monoid can be automatic with respect to one [monoid] generating set but not with respect to another. However, as Duncan et al. (1999, Section 5) note, a monoid that is automatic with respect to this alternative definition is automatic in the sense of Definition 2.2.2.

When one passes to asynchronous automatism, the presence of an identity is no longer required:

Theorem 2.4.7. *Let S be a semigroup with asynchronous automatic structure (A, L). Let B be any finite alphabet representing a finite generating set for S. Then S has an asynchronous automatic structure (B, K), where K is a regular language.*

This result was established by Hoffmann et al. (2002*a*, Proposition 4.1) for asynchronous automatic monoids. Their proof requires minimal modification to deal with semigroups:

Proof of 2.4.7. For each $a \in A$ and each $b \in B$, choose $w_a \in B^+$ and $w_b \in A^+$ such that $\overline{a} = \overline{w_a}$ and $\overline{b} = \overline{w_b}$. Define a homomorphism $\phi : A^+ \to B^+$ by extending the maps $a \mapsto w_a$. The language $K = L\phi$ is regular and maps onto S.

The relation $R = \{(a, w_a) : a \in A\}^+$ is rational by Lemma A.3.3. Let $b \in B$ and suppose $w_b = a_1 \cdots a_n$, where $a_i \in A$. Let $u, v \in K$. Then

$$\begin{aligned}(u,v) \in K_b &\iff \overline{ub} = \overline{v} \\ &\iff \overline{u'w_b} = \overline{v'}, \text{ and } u', v' \in L, \text{ where } (u',u),(v',v) \in R \\ &\iff (u',v') \in L_{w_b}, (u,u') \in R^{-1}, (v',v) \in R \\ &\iff (u,v) \in R^{-1} \circ L_{a_1} \circ L_{a_2} \circ \ldots \circ L_{a_n} \circ R.\end{aligned}$$

Therefore

$$K_b = R^{-1} \circ L_{a_1} \circ L_{a_2} \circ \ldots \circ L_{a_n} \circ R,$$

and so K_b is rational by Theorem A.6.5. Similarly,

$$K_\varepsilon = R^{-1} \circ L_\varepsilon \circ R.$$

is rational. Therefore (B, K) is an asynchronous automatic structure for S. $\boxed{2.4.7}$

2.4.4. Unique representatives

Definition 2.4.8. A synchronous or asynchronous *automatic structure with uniqueness* for a semigroup S is an automatic structure (A, L) with the property that every element of S has a unique representative in L: the language L maps bijectively onto S.

Definition 2.4.9. Let $\prec$ be any total ordering of an alphabet A. Define the *lexicographic* ordering $\prec_{\mathrm{L}}$ on A^+ as follows:

$$\begin{aligned}u_1 \cdots u_m \prec_{\mathrm{L}} v_1 \cdots v_n \iff &\big(u(k-1) = v(k-1) \text{ and } u_k \prec v_k \text{ for some } k \leq m\big) \\ &\text{or } \big(u = v(m) \text{ and } m < n\big).\end{aligned}$$

Define the *ShortLex* or *length plus lexicographic* ordering $\prec_{\mathrm{SL}}$ on A^+ as follows: for $u, v \in A^+$,

$$u \prec_{\mathrm{SL}} v \iff (|u| < |v|) \text{ or } (|u| = |v| \text{ and } u \prec_{\mathrm{L}} v).$$

The ShortLex ordering is invariant under multiplication in A^+:

$$(\forall u, v, w \in A^+)\big(u \prec_{\mathrm{SL}} v \implies (uw \prec_{\mathrm{SL}} vw \wedge wu \prec_{\mathrm{SL}} wv)\big).$$

Theorem 2.4.10. *Let (A, L) be a synchronous automatic structure for a semigroup S. Let*

$$\mathsf{ShortLex}(L) = \big\{u \in L : (\forall v \in L)((\overline{u} = \overline{v}) \implies (u \preceq_{\mathrm{SL}} v))\big\}.$$

Then $(A, \mathsf{ShortLex}(L))$ is a synchronous automatic structure with uniqueness for S.

Proof of 2.4.10. See Epstein et al. (1992, Theorems 2.5.1) and Campbell et al. (2001, Proposition 5.4). $\boxed{2.4.10}$

Observe that the language $\mathsf{ShortLex}(L)$ is the subset of L consisting of the ShortLex-minimal representative for each element. Furthermore, given (A, L), the the automatic structure $(A, \mathsf{ShortLex}(L))$ can be effectively constructed.

Theorem 2.4.11. *Let (A, L) be an asynchronous automatic structure for a group-embeddable semigroup S. Then there is a language K contained in L such that (A, K) is an asynchronous automatic structure with uniqueness for S.*

Proof of 2.4.11. By Theorem 2.3.5, a boundedly asynchronous automatic structure exists for S. Using this new structure, the proof of Epstein et al. (1992, Theorem 7.3.2) proceeds unchanged. $\boxed{2.4.11}$

2.4.5. Constructions

Theorem 2.4.12 (Campbell, Robertson, Ruškuc & Thomas 2000). *Let S_1 and S_2 be automatic semigroups. Their direct product $S_1 \times S_2$ is automatic if and only if it is finitely generated.* $\boxed{2.4.12}$

Since the direct product of two finitely generated monoids is always finitely generated, Theorem 2.4.12 shows that the direct product of two automatic monoids is automatic. The following result, which appears here for the first time, generalizes this to asynchronous automatic monoids.

Theorem 2.4.13. *Let M_1 and M_2 be asynchronous automatic monoids. Then their direct product $M_1 \times M_2$ is asynchronously automatic.*

Proof of 2.4.13. Let M_1 and M_2 have asynchronous automatic structures (A_1, L_1) and (A_2, L_2) respectively. In the direct product $M_1 \times M_2$, view the letter $a_1 \in A_1$ as representing $(\overline{a_1}, 1_{M_2})$ and the letter $a_2 \in A_2$ as representing $(1_{M_1}, \overline{a_2})$. So $A_1 \cup A_2$ represents a generating set for $M_1 \times M_2$. Observe that the language L_1L_2 is regular and maps onto $M_1 \times M_2$. The aim is to show that $(A_1 \cup A_2, L_1L_2)$ is an asynchronous automatic structure for $M_1 \times M_2$.

Let $u_1u_2, v_1v_2 \in L_1L_2$, where $u_1, v_1 \in L_1$ and $u_2, v_2 \in L_2$. Suppose that $\overline{ua} = \overline{v}$, where $a \in A_1 \cup A_2 \cup \{\varepsilon\}$. If $a \in A_1$, then projecting to M_1 shows that $\overline{u_1a} = \overline{v_1}$ and $\overline{u_2} = \overline{v_2}$. Similarly, if $a \in A_2$, then $\overline{u_1} = \overline{v_1}$ and $\overline{u_2a} = \overline{v_2}$. If $a = \varepsilon$, then $\overline{u_1} = \overline{v_1}$ and $\overline{u_2} = \overline{v_2}$. Hence:

$$(L_1L_2)_a = \begin{cases} (L_1)_a(L_2)_\varepsilon & \text{if } a \in A_1, \\ (L_1)_\varepsilon(L_2)_a & \text{if } a \in A_2, \\ (L_1)_\varepsilon(L_2)_\varepsilon & \text{if } a = \varepsilon. \end{cases}$$

In each case, $(L_1L_2)_a$ is a rational relation. Therefore $(A_1 \cup A_2, L_1L_2)$ is an asynchronous automatic structure for the direct product $M_1 \times M_2$. $\boxed{2.4.13}$

The proofs of the group-theoretic results corresponding to Theorems 2.4.12 and 2.4.13 (Epstein et al. 1992, Theorems 4.1.1 & 7.3.5(1)) use the geometric characterization of synchronous and asynchronous automatic structures for groups. If one is only concerned with semigroups embeddable into groups, the group-theoretic arguments can be generalized using the results of Section 2.3.

Theorem 2.4.14 (Campbell et al. 2001, Theorems 6.1 and 6.2). *Let S_1 and S_2 be automatic semigroups. Then their [semigroup] free product $S_1 * S_2$ is automatic. Futhermore, if S_1 and S_2 are monoids, then their monoid free product is also automatic.* $\boxed{2.4.14}$

Theorem 2.4.15. *Let S_1 and S_2 be asynchronous automatic semigroups. Then their [semigroup] free product $S_1 * S_2$ is asynchronously automatic. Furthermore, if S_1 and S_2 are monoids, then their monoid free product is also asynchronously automatic.*

Proof of 2.4.15. Let (A_1, L_1) and (A_2, L_2) be asynchronous automatic structures for S_1 and S_2 respectively. Let

$$L = \big(L_1 \cup \{\varepsilon\}\big)\Big[L_2L_1\Big]^*\big(L_2 \cup \{\varepsilon\}\big) - \{\varepsilon\}.$$

Then L is a regular language. Clearly, $\overline{L} = S_1 * S_2$, since elements of $S_1 * S_2$ are alternating products of elements of S_1 and S_2 and words in L consist of all possible alternating products of words from L_1 and L_2. For $a_1 \in A_1$, let U_a be the set of words in L_1 representing $\overline{a_1}$, and similarly let V_a be the set of words in L_2 representing $\overline{a_2}$. Notice that each U_a and V_a is regular. It is clear that for each $a \in A_1$,

$$L_a = \big((L_1)_\varepsilon \cup (\varepsilon, \varepsilon)\big)\Big[(L_2)_\varepsilon(L_1)_\varepsilon\Big]^*\big((L_2)_\varepsilon(L_1)_a \cup (L_2)_\varepsilon\{(\varepsilon, u) : u \in U_a\}\big),$$

while for each $a \in A_2$, it is clear that

$$L_a = \big((L_1)_\varepsilon \cup (\varepsilon, \varepsilon)\big)\Big[(L_2)_\varepsilon(L_1)_\varepsilon\Big]^*\big((L_2)_a \cup \{(\varepsilon, v) : v \in V_a\}\big) - \{(\varepsilon, v) : v \in V_a\},$$

which are rational relations. Therefore $(A_1 \cup A_2, L)$ is an asynchronous automatic structure for $S_1 * S_2$.

The assertion about monoid free products is due to Hoffmann et al. (2002*a*, Theorem 5.2). $\boxed{2.4.15}$

Theorem 2.4.16 (Epstein et al. 1992, Theorems 4.1.4 and 7.3.5(2)). *Every finite extension and every finite-index subgroup of an automatic group is automatic, and every finite extension and every finite-index subgroup of an asynchronous automatic group is asynchronously automatic.* $\boxed{2.4.16}$

Theorems 2.4.14–2.4.16 have the following consequence:

Proposition 2.4.17. *The class of groups all of whose finitely generated subgroups are automatic and the class of groups all of whose finitely generated subgroups are asynchronously automatic are both closed under:*

i.) *forming free products.*

ii.) *constructing finite extensions.*

[Compare the proof of Proposition 2.4.17 with the proof of the equivalent result for the class of coherent groups (Proposition 1.6.1).]

Proof of 2.4.17. Let $\mathfrak{A}$ be the class of groups all of whose finitely generated subgroups are automatic (respectively, asynchronously automatic).

i.) Let $\{G_i : i \in I\}$ be a collection of groups in $\mathfrak{A}$. Let H be a finitely generated subgroup of the free product $\prod^*_{i\in I} G_i$. Then, by the Kurosh Subgroup Theorem (see Lyndon & Schupp 1977, Section III.3), the subgroup H is of the form

$$F * \prod_{j\in J}^{*} H_j,$$

where F is a free group and each group H_j is a subgroup of a conjugate of one of the free factors G_i. Since H is finitely generated, the free group F and each H_j are finitely generated. The free group F is manifestly automatic; since each G_i is in $\mathfrak{A}$, each H_j is automatic (respectively, asynchronously automatic). Therefore H is automatic by Theorem 2.4.14 (respectively, asynchronously automatic by Theorem 2.4.15). Since H was arbitrary, the free product $\prod^*_{i\in I} G_i$ is in $\mathfrak{A}$.

ii.) Let E be a finite extension of a group $G \in \mathfrak{A}$. Let K be a finitely generated subgroup of E. Then K is a finite extension of $G \cap K$ by Lemma 0.10.4. Since K is finitely generated, the Reidemeister–Schreier Theorem (see Lyndon & Schupp 1977, Section II.4) asserts that $G \cap K$ is also finitely generated. Since the group G is in $\mathfrak{A}$, the subgroup $G \cap K$ is automatic (respectively, asynchronously automatic). By Theorem 2.4.16, the group K is therefore also automatic (respectively, asynchronously automatic). The subgroup K was arbitrary; the group E is therefore also in $\mathfrak{A}$. $\boxed{2.4.17}$

Theorems 5.5.5 and 6.2.8 assert that Proposition 2.4.17 no longer remains true if one passes to the class of locally automatic or locally asynchronous automatic groups.

2.5. MALCEV PRESENTATIONS AND AUTOMATIC SEMIGROUPS

EVERY automatic or asynchronous automatic group admits a finite presentation (Epstein et al. 1992, Theorems 2.3.12 and 7.3.4). Automatic semigroups may not be finitely presented: see Campbell et al. (2001, Examples 3.9, 4.4 and 4.5). In particular, Campbell et al.'s Example 4.5 shows that automatic semigroups embeddable in

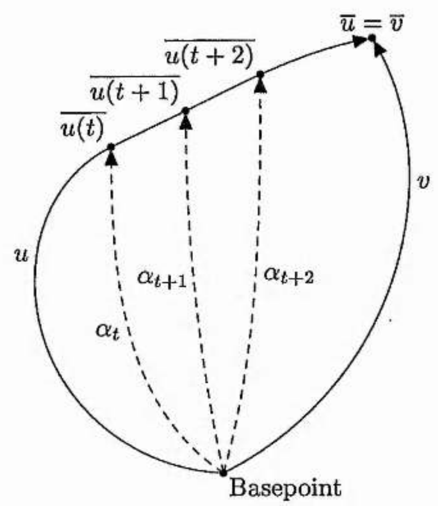

Figure 2.2. Walks in $\Gamma(S, A)$ labelled by valid relations.

groups need not be finitely presented. Campbell et al. (2002, Corollary 1.2) proved that automatic completely simple semigroups do, however, have finite presentations.

Restricting attention to subsemigroups of groups yields the following result, which first appeared as Theorem 2 of Cain, Robertson & Ruškuc (2005*a*).

Theorem 2.5.1. *Every automatic semigroup embeddable into a group admits a finite Malcev presentation.*

Proof of 2.5.1. The proof begins along similar lines to Epstein et al. (1992, Theorem 2.3.12).

Let S be an automatic semigroup that can be embedded in a group; let (A, L) be an automatic structure for S. Let λ be the constant of Proposition 2.3.3.

For $a \in A$, let $\gamma_a \in L$ represent the same element as a. Let $\mathcal{T} = \{(a, \gamma_a) : a \in A\}$. Every relation in $\mathcal{T}$ is valid in S.

Let $(u, v) \in A^+ \times A^+$ be a relation that holds in S. Then u and v label paths in the Cayley graph $\Gamma(S, A)$ from the basepoint to the same vertex. Suppose $u = u_1 \cdots u_k$ and $v = v_1 \cdots v_l$, where $u_i, v_i \in A$. Let α_i be a representative in L of $\overline{u(i)}$ for $0 \leq i \leq k$ and similarly β_j a representative of $\overline{v(j)}$ for $0 \leq j \leq l$ (see Figure 2.2). Assume without loss of generality that $\alpha_1 = \gamma_{u_1}$ and $\beta_1 = \gamma_{v_1}$. The relations

$$(u_1, \alpha_1), (v_1, \beta_1), (\alpha_i u_{i+1}, \alpha_{i+1}), (\beta_j v_{j+1}, \beta_{j+1}), \text{ and } (\alpha_k, \beta_l) \tag{2}$$

hold in S for $i = 0, \ldots, k-1$ and $j = 0, \ldots, l-1$.

Lemma 2.5.2. *The relation (u, v) is a consequence of the relations (2).*

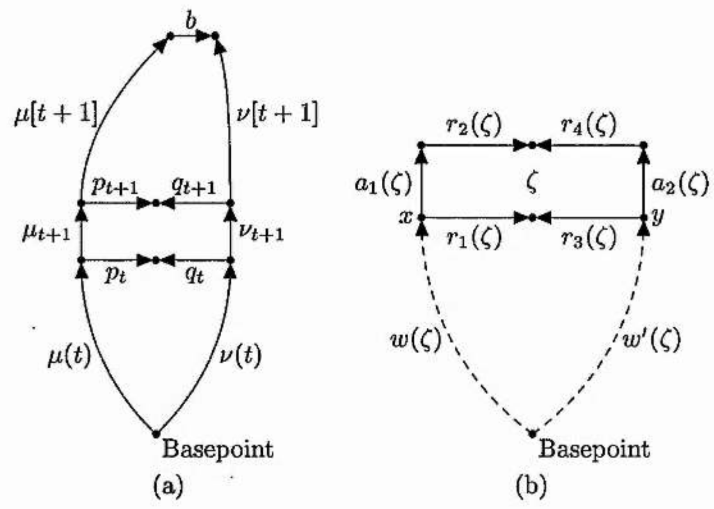

Figure 2.3. Loops in $\Gamma(S, A)$.

Proof of 2.5.2. The chain

$$\begin{aligned} u = u_1 \cdots u_k \to \alpha_1 u_2 \cdots u_k \to \alpha_2 u_3 \cdots u_k \to \ldots \to \alpha_{k-1} u_k \to \alpha_k \\ \to \beta_l \to \beta_{l-1} v_l \to \ldots \to \beta_1 v_2 \cdots v_l \to v_1 \cdots v_l = v \end{aligned}$$

shows that (u, v) is a consequence of the given relations. $\boxed{2.5.2}$

Choose and fix one of the relations (2) other than (u_1, α_1) and (v_1, β_1). To simplify notation, write it as $(\mu b, \nu)$, where $\mu, \nu \in L$ and $b \in A \cup \{\varepsilon\}$. Suppose $\mu = \mu_1 \cdots \mu_{|\mu|}$ and $\nu = \nu_1 \cdots \nu_{|\nu|}$, where $\mu_i, \nu_i \in A$. Since $(\mu, \nu) \in L_b$, for each $t \in \mathbb{N} \cup \{0\}$, there exist p_t and q_t in A^* such that $\overline{\mu(t) p_t} = \overline{\nu(t) q_t}$, with $|p_t| + |q_t|$ being bounded by the constant λ. Assume that $p_0, q_0 = \varepsilon$ and that $p_m = b$ and $q_m = \varepsilon$, where $m = \max\{|\mu|, |\nu|\}$. Geometrically, these p_t and q_t give $m - 1$ undirected loops between the paths $\widehat{\mu}$ and $\widehat{\nu}$, the total length of each loop being bounded by $2\lambda + 2$ (see Figure 2.3(a)).

Consider the set of all loops ζ in $\Gamma(S, A)$ of the form shown in Figure 2.3(b), where $|a_1(\zeta)|, |a_2(\zeta)| \leq 1$ and $|r_i(\zeta)| \leq \lambda/2$ for $i = 1, \ldots, 4$. There is no requirement that the vertices x and y be distinct: this covers the case when $t = 0$ in Figure 2.3(a).

There are only a finite number of possible labels for such loops. Let Z be a set of such loops in $\Gamma(S, A)$ containing exactly one with each possible label. Consider some loop $\zeta \in Z$ and retain the notation from Figure 2.3(b). Choose directed paths from the basepoint of $\Gamma(S, A)$ to the vertices x and y. Let $\widehat{w(\zeta)}$ and $\widehat{w'(\zeta)}$ be the words labelling these paths. The two relations

$$(w(\zeta) r_1(\zeta), w'(\zeta) r_3(\zeta)) \text{ and } (w(\zeta) a_1(\zeta) r_2(\zeta), w'(\zeta) a_2(\zeta) r_4(\zeta)) \tag{3}$$

both hold in S. Let Q be the set of relations (3) thus obtained as ζ ranges over Z. Since Z is a finite set, so is Q.

Lemma 2.5.3. *The relation $(\mu b, \nu)$ is a Malcev consequence of $\mathcal{Q}$.*

Proof of 2.5.3. Let $t \in \{0, \ldots, m-1\}$. The key is to show that there is a Malcev $\mathcal{Q}$-chain that leads from $\mu(t+1)p_{t+1}q_{t+1}^{\mathsf{R}}\nu[t+1]$ to $\mu(t)p_t q_t^{\mathsf{R}}\nu[t]$. Let ζ be the loop in Z with the same label as that in Figure 2.3(a). In this case,

$$a_1(\zeta) = \mu_{t+1}, a_2(\zeta) = \nu_{t+1}, r_1(\zeta) = p_t, r_2(\zeta) = p_{t+1}, r_3(\zeta) = q_t, r_4(\zeta) = q_{t+1};$$

and the corresponding relations (3) become

$$(w(\zeta)p_t, w'(\zeta)q_t) \text{ and } (w(\zeta)\mu_{t+1}p_{t+1}, w'(\zeta)\nu_{t+1}q_{t+1}). \tag{4}$$

The following is the desired Malcev chain:

$$\begin{aligned}
\mu(t+1)p_{t+1}q_{t+1}^{\mathsf{R}}\nu[t+1] &= \mu(t)\mu_{t+1}p_{t+1}q_{t+1}^{\mathsf{R}}\nu[t+1] \\
&\to \mu(t)w(\zeta)^{\mathsf{L}}w(\zeta)\mu_{t+1}p_{t+1}q_{t+1}^{\mathsf{R}}\nu[t+1] \\
&\to \mu(t)w(\zeta)^{\mathsf{L}}w'(\zeta)\nu_{t+1}q_{t+1}q_{t+1}^{\mathsf{R}}\nu[t+1] \\
&\to \mu(t)w(\zeta)^{\mathsf{L}}w'(\zeta)\nu_{t+1}\nu[t+1] \\
&= \mu(t)w(\zeta)^{\mathsf{L}}w'(\zeta)\nu[t] \\
&\to \mu(t)w(\zeta)^{\mathsf{L}}w'(\zeta)q_t q_t^{\mathsf{R}}\nu[t] \\
&\to \mu(t)w(\zeta)^{\mathsf{L}}w(\zeta)p_t q_t^{\mathsf{R}}\nu[t] \\
&\to \mu(t)p_t q_t^{\mathsf{R}}\nu[t].
\end{aligned}$$

Concatenating such chains for $t = 0, \ldots, m-1$ yields a Malcev chain from μs to ν, and thus $(\mu b, \nu)$ is a Malcev consequence of $\mathcal{Q}$. $\boxed{2.5.3}$

By Lemma 2.5.3, each pair in (2) is a Malcev consequence of $\mathcal{Q}$. Lemma 2.5.2 states that (u, v) is a consequence of (2) and $\mathcal{T}$. Therefore, since neither $\mathcal{T}$ nor $\mathcal{Q}$ depends on (u, v), each valid relation in S is a Malcev consequence of $\mathcal{T} \cup \mathcal{Q}$, and so $\mathrm{SgM}\langle A \mid \mathcal{T} \cup \mathcal{Q}\rangle$ is a finite Malcev presentation for S. $\boxed{2.5.1}$

Theorem 2.5.1 extends to the case of asynchronous automatism.

Theorem 2.5.4. *Every asynchronous automatic semigroup that can be embedded in a group admits a finite Malcev presentation.*

Proof of 2.5.4. This proof is similar to that of Theorem 2.5.1. First of all, by Theorem 2.3.13, assume without loss of generality that a boundedly deterministic asynchronous automatic structure exists. Following the proof for the synchronous case, show that every valid relation is a consequence of those in (2). The walks $\widehat{u}, \widehat{v}$, where $(u, v) \in L_s$ are then linked by loops in a manner similar to Figure 2.3(a), but by appealing to Proposition 2.3.12 rather than Proposition 2.3.3.

Equipped with these loops of length at most $2\lambda + 1$, one obtains a finite set of relations of the form (3). Reason as in Lemma 2.5.3 to show that the valid relation $(\mu s, \nu)$ is a Malcev consequence of the relations arising from these loops. $\boxed{2.5.4}$

The final result in this section is an immediate consequence of Theorems 2.5.1 and 2.5.4:

Theorem 2.5.5. *Any locally automatic or locally asynchronous automatic group-embeddable semigroup is Malcev coherent.* $\boxed{2.5.5}$

CHAPTER THREE

ALGORITHMS FOR AUTOMATIC SEMIGROUPS

In practice we not only want algorithms, we want algorithms that are good *in some loosely defined aesthetic sense. One criterion of goodness is the length of time taken... Other criteria are the adaptability of the algorithm... its simplicity and elegance, etc.*

— Donald E. Knuth,
Fundamental Algorithms (1997), 1.1

3.1. INTRODUCTION

AN AUTOMATIC STRUCTURE allows one to algorithmically answer various questions regarding the semigroup in question. For example, the word problem is solvable for any automatic semigroup (see Theorem 2.4.1). It is also known that one can decide whether two elements of an automatic semigroup are $\mathcal{L}$-related. However, whether two elements are $\mathcal{R}$-related is undecidable (Otto & Ruškuc 2000). [In a semigroup S, the Green's relations $\mathcal{L}$ and $\mathcal{R}$ are such that $x \mathrel{\mathcal{L}} y$ (respectively, $x \mathrel{\mathcal{R}} y$) if and only if there exists elements $p, q \in S^1$ such that $x = py$ and $y = qx$ (respectively, $x = yp$ and $y = xq$). See Howie (1995, Section 2.1) for further information.]

This chapter considers four problems for automatic semigroups that relate to group-embeddability. Section 3.2 deals with automatic semigroups that are known to be group-embeddable: an algorithm is given that takes an automatic structure for such a semigroup and yields a finite Malcev presentation for that semigroup. Section 3.3 describes an algorithm that tests the freedom of an automatic semigroup; Section 3.4 shows how to decide whether an automatic semigroup is a group. That left-cancellativity is undecidable for automatic semigroups is proven in Section 3.5.

Recall that this thesis assumes that an explicitly given automatic structure (A, L) includes, for each $a \in A$, a specified word $u_a \in L$ representing the generator $\overline{a}$. (See the remarks following Proposition 2.2.3 for the reasoning behind this assumption.)

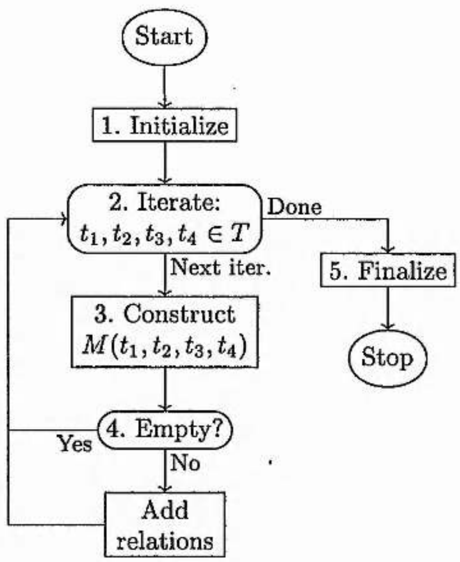

Figure 3.1. Flowchart for Algorithm 3.2.1.

3.2. ALGORITHMICALLY OBTAINING A MALCEV PRESENTATION

THEOREM 2.5.1 asserts that every automatic semigroup embeddable into a group possesses a finite Malcev presentation. It does not give a method for obtaining a finite Malcev presentation from an automatic structure. The following algorithm is such a method:

Algorithm 3.2.1. Let (A, L) be an automatic structure for a group-embeddable semigroup S, with the languages L and L_a for $a \in A \cup \{\varepsilon\}$ being explicitly described by finite state automata. Suppose λ is the constant of Proposition 2.3.3.

1. [Initialize.] Let $T = \{w \in A^* : |w| < \lambda/2 + 1\}$. Let $\mathcal{R}$ be an empty set of relations.
2. [Iterate over T.] For every $t_1, t_2, t_3, t_4 \in T$, do Steps 3–4. Then go to Step 5.
3. [Build language.] Construct the language

$$M(t_1, t_2, t_3, t_4) = (L_{t_1} \circ L_{t_3}^{-1}) \cap (L_{t_2} \circ L_{t_4}^{-1}).$$

4. [Empty?] Test the language $M(t_1, t_2, t_3, t_4)$ for emptiness. If it is non-empty, choose any $(w, w') \in M(t_1, t_2, t_3, t_4)$ and add the relations $(wt_1, w't_3)$ and $(wt_2, w't_4)$ to $\mathcal{R}$. Continue with Step 2.
5. [Finalize.] For each $a \in A$, obtain a word $\gamma_a \in L$ representing the same element of S as A. Add the relations (a, γ_a) to $\mathcal{R}$.

The semigroup S has the finite Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{R}\rangle$.

The proof of Proposition 2.3.3 shows that one may take any value for λ that is greater than twice the number of states in any of the finite state automata recognizing the various languages $L_a\delta_A$. Using these automata, one can construct an automaton recognizing the language $L_{t_i}\delta_A$ for any $t_i \in T$ (Proposition 2.2.3). One can then effectively construct an automaton recognizing $M(t_1, t_2, t_3, t_4)\delta_A$. Theorem A.5.3 shows that one can test the emptiness of a regular language. The set T is finite, so Steps 3–4 are only carried out finitely many times, and Theorem 2.4.1 shows that Step 5 is effective. Therefore the procedure described in Algorithm 3.2.1 is indeed an algorithm. The following two results show that $\mathrm{SgM}\langle A \mid \mathcal{R}\rangle$ is a finite Malcev presentation for S.

Proposition 3.2.2. *Every relation in the set $\mathcal{R}$ obtained using Algorithm 3.2.1 is a valid relation in the semigroup S.*

Proof of 3.2.2. If a relation in $\mathcal{R}$ is of the form (a, γ_a), then it is valid by definition. Therefore consider the relations found in Steps 3–4. Fix $t_1, t_2, t_3, t_4 \in T$ and suppose $(w, w') \in M(t_1, t_2, t_3, t_4)$. Then $(w, w') \in L_{t_1} \circ L_{t_3}^{-1}$ and $(w, w') \in L_{t_2} \circ L_{t_4}^{-1}$. Therefore there exist words $s, t \in L$ such that $(w, s) \in L_{t_1}$, $(w', s) \in L_{t_3}$, $(w, t) \in L_{t_2}$, and $(w', t) \in L_{t_4}$. Therefore, by definition of the languages L_{t_i}, $\overline{wt_1} = \overline{s} = \overline{w't_3}$ and $\overline{wt_2} = \overline{t} = \overline{wt_4}$. The relations $(wt_1, w't_3)$ and $(wt_2, w't_4)$ are therefore valid in S. $\boxed{3.2.2}$

Proposition 3.2.3. *Every relation of the form $(\mu b, \nu)$, where $(\mu, \nu) \in L_b$ for $b \in A \cup \{\varepsilon\}$ is a Malcev consequence of the relations $\mathcal{R}$ found using Algorithm 3.2.1.*

Proof of 3.2.3. Consider any relations 2.5–(4) and how they are used in the Malcev chain in the proof of Lemma 2.5.3. The present proof proceeds by showing that in $\mathcal{R}$ there are relations that can be used in place of 2.5–(4) in the Malcev chain.

Using the notation from the proof of Lemma 2.5.3, let $t_1 = p_t$, $t_2 = \mu_{t+1}p_{t+1}$, $t_3 = q_t$, and $t_4 = \nu_{t+1}q_{t+1}$. Observe that for $i = 1, 2, 3, 4$, the word t_i has length at most $\lambda/2+1$ and is therefore contained in T. Note also that language $M(t_1, t_2, t_3, t_4)$ is non-empty, since it contains $(\mu(t), \nu(t))$. Therefore $\mathcal{R}$ contains relations $(wt_1, w't_3)$ and $(wt_2, w't_4)$, or rather $(wp_t, w'q_t)$ and $(w\mu_{t+1}p_{t+1}, w'\nu_{t+1}q_{t+1})$. These are the two relations that can be used in the Malcev chain in the proof of Lemma 2.5.3.

Therefore every relation $(\mu b, \nu)$ is a Malcev consequence of the relations $\mathcal{R}$. $\boxed{3.2.3}$

3.3. TESTING FOR FREEDOM

IT HAS BEEN considered 'obvious' that the question of whether an automatic semigroup is a free semigroup is decidable. No algorithm to decide this property has yet been published. The present section fills this particular lacuna in the theory.

Testing whether an automatic semigroup is free is a special case of the isomorphism problem. It is unknown whether the isomorphism problem is solvable generally for automatic semigroups; the question remains open even for automatic groups. Epstein et al. (1992, Open Question 2.3.11) conjecture that the isomorphism problem for automatic groups is not solvable.

Given an automatic structure with uniqueness (A, L), define a mapping $\Phi_L : A \to \mathbb{P}(LA \cup L)$ by

$$a\Phi_L = \{ub : u \in L, b \in A \cup \{\varepsilon\}, (u, v_a)\delta_A \in L_b\},$$

where $v_a \in L$ represents $\overline{a}$. Informally, the set $a\Phi_L$ consists of the left-hand side of any 'relation stored in the automatic structure' whose right-hand side is the unique word $v_a \in L$ representing $\overline{a}$. Observe that the language $a\Phi_L$ is regular.

Algorithm 3.3.1. Let S be a semigroup with a known automatic structure.

1. [Initialize.] Adjoin an identity to S and construct the new automatic structure with uniqueness $(A \cup \{1\}, L)$ for S^1. (The new symbol 1 represents the adjoined identity.)
2. [Obviously not free?] Does there exist a letter $a \in A$ such that $a\Phi_L$ contains a word with at least two letters from A, at least one of which is a?
   No: Continue to Step 3.
   Yes: Halt — S is NOT FREE.
3. [Redundant generator?] Does there exist an $a \in A$ such that $a\Phi_L$ contains a word consisting only of letters from $(A \cup \{1\}) - \{a\}$?
   Yes: (The generator $\overline{a}$ is redundant.) Construct an automatic structure with uniqueness with respect to this new set of generators $\overline{(A \cup \{1\}) - \{a\}}$. Replace $(A \cup \{1\}, L)$ by this new structure. Go to Step 2.
   No: Continue to Step 4.
4. [All of A^*?] Does L project onto A^*?
   Yes: Halt — S is FREE.
   No: Halt — S is NOT FREE.

Justification of Algorithm 3.3.1. *Step 2.* The algorithm may reach this step from either Step 1 or Step 3. In either case, $(A \cup \{1\}, L)$ is an automatic structure with uniqueness for S^1.

Suppose that S is free with basis B. Proceed as follows to prove that, for each $a \in A$, there is no word $u \in a\Phi_L$ that contains at least two letters from A, at least one of which is a. Suppose, with the aim of obtaining a contradiction, that u is a word in $a\Phi_L$ containing a amongst at least two symbols from A, so that (u, a) is a valid relation in S. Express each letter in A in terms of B and substitute into (u, a), deleting symbols 1. This yields a relation holding between words in B^+, non-trivial since the length (over B) of the left-hand side exceeds that of the right-hand side. This contradicts S being free on B.

Therefore, if S is free on some basis, the algorithm proceeds to Step 3.

Step 3. On reaching this step, $(A \cup \{1\}, L)$ is an automatic structure with uniqueness for S^1 with the property that, for every $a \in A$, no word in $a\Phi_L$ contains at least two letters from A, at least one of which is a.

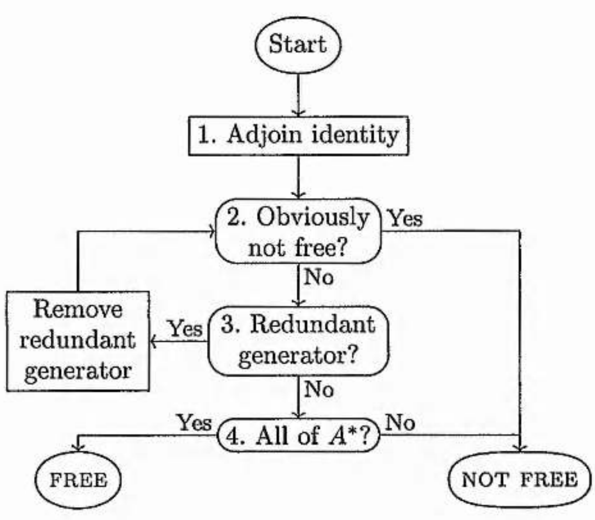

Figure 3.2. Flowchart for Algorithm 3.3.1.

Suppose that $u \in a\Phi_L$ contains only letters from $(A \cup \{1\}) - \{a\}$. Then the generator $\overline{a}$ is redundant. The construction of an automatic structure on the generating set $\overline{(A \cup \{1\}) - \{a\}}$ is effective by Theorem 2.4.6.

This step reduces the number of letters in the set A, so the algorithm cannot loop between Steps 2 and 3 indefinitely.

Step 4. When Step 4 is reached, $(A \cup \{1\}, L)$ an automatic structure with uniqueness for S^1 with the property that each word in $a\Phi_L$ consists only of words containing a single letter a and possibly symbols 1. Any words in $a\Phi_L$ not of this form would either have halted the algorithm at Step 2 or would have resulted in the elimination of a generator in Step 3 and looping back to Step 2.

Suppose, aiming for a contradiction, that $a \in A$ represents a redundant generator for S. Then there is a word $w \in (A - \{a\})^+$ such that $\overline{w} = \overline{a}$. Write $w = w'b$, where $b \in A - \{a\}$. Let $u \in L$ represent $\overline{w'}$. Then $ub \in a\Phi_L$, which contradicts the observations in the last paragraph. So there are no redundant generators in $\overline{A}$. If S is free, therefore, it is free on $\overline{A}$.

The projection map $\pi : (A \cup \{1\})^* \to A^*$ sends $w \in (A \cup \{1\})^*$ to the word obtained by deleting all symbols 1 from w. Observe that π is a homomorphism from $(A \cup \{1\})^*$ to A^* and so preserves regularity by Theorem A.5.2. Testing equality of regular languages is effective. Therefore the question of whether $L\pi = A^*$ is effectively decidable.

Let there be some word u in $A^* - L\pi$. This word must be of non-zero length, for only the words 1^k map to ε: these are the only words representing the adjoined identity, and so at least one of them is present in L. Then there is some word $w \in L$ representing $\overline{u} \in S$. Hence there is a non-trivial relation holding in S, and so S is not free.

Now let L map onto A^*. Suppose S is not free on $\overline{A}$. Then there exist words

$u, v \in A^+$ such that $\overline{u} = \overline{v}$ but $u \neq v$. Using the surjectivity of π, pick words $u', v' \in L$ such that $u'\pi = u$ and $v'\pi = v$. Since u and v are not identical, neither are u' and v'. Yet $\overline{u'} = \overline{u} = \overline{v} = \overline{v'}$, which contradicts $(A \cup \{1\}, L)$ being an automatic structure with uniqueness. Therefore S must be free. $\boxed{3.3.1}$

3.4. TESTING WHETHER AN AUTOMATIC SEMIGROUP IS A GROUP

THIS SECTION exhibits an algorithm that takes as input an automatic structure for a semigroup and determines whether that semigroup is in fact a group. The reasoning is essentially due to Silva & Steinberg (2004), but is generalized from their 'prefix-automatic' monoids to standard automatic semigroups. [The definition of prefix-automatic monoids appears to be more restrictive than that of automatic semigroups, although no example is yet known of a monoid that is automatic but not prefix-automatic (Silva & Steinberg 2004, Section 1).] Firstly, one needs an algorithm to test whether an automatic semigroup is a monoid:

Algorithm 3.4.1. Let S be a semigroup with automatic structure (A, L).

1. [Initialize.] Replace (A, L) with an automatic structure with uniqueness for S. For each $a \in A$, let $u_a \in L$ be the unique word representing $\overline{a}$.
2. [Unique left identity?] Does the set $Z = \bigcap_{a \in A} \{w : (w, u_a) \in L_a\}$ contain exactly one element e?
   Yes: Continue to Step 3.
   No: Halt — S is NOT A MONOID.
3. [Also right identity?] Is the relation L_e the diagonal relation $\{(u, u) : u \in L\}$?
   Yes: Halt — S is a MONOID (with identity $\overline{e}$).
   No: Halt — S is NOT A MONOID.

Justification of Algorithm 3.4.1. Suppose S is a monoid with identity 1_S. Let e be the unique word in L representing 1_S. Then, for each $a \in A$, $\overline{ea} = \overline{a} = \overline{u_a}$, so $(e, u_a) \in L_a$. So the word e lies in the set Z. Furthermore, since S is a monoid and L maps bijectively onto S, the set Z contains e alone. The algorithm therefore continues to Step 3. Since $\overline{e} = 1_S$, $\overline{ue} = \overline{u}$ for all $u \in L$, and so L_e is the diagonal relation and the algorithm halts and indicates that S is a monoid with identity $\overline{e}$.

Now suppose the algorithm completes and indicates that S is a monoid. The fact that the algorithm continues past Step 2 establishes the existence of a unique left identity $\overline{e}$ for S, since $\overline{ea} = \overline{a}$ for each generator $\overline{a}$ of S. and any other left identity for S would also have this property and so would lie in Z. As the algorithm gives a positive answer to Step 3, the element $\overline{e}$ must be a right identity since $\overline{ue} = \overline{u}$ for all $u \in L$ and L maps onto S. The semigroup S is therefore a monoid with identity $\overline{e}$.

Finally, observe that each step of the algorithm is effective: Step 1 is effective by Theorems 2.4.1 and 2.4.10; Step 2 by the regularity of Z (Theorem A.5.1) and that regular languages can be tested for emptiness (Theorem A.5.3) and indeed for whether they contain only one element; and Step 3 by the fact that $L_e\delta_A$ and $\{(u, u) : u \in L\}\delta$ are regular and that testing equality of regular languages is possible (Theorem A.5.5).

Algorithm 3.4.2. Let S be a semigroup with automatic structure (A, L).

1. [Monoid?] Replace (A, L) with an automatic structure with uniqueness for S and test whether S is a monoid using Algorithm 3.4.1.
   Yes: Continue to Step 2.
   No: Halt — S is NOT A GROUP.
2. [Inverses?] Let π_2 be the projection of $w \in A^* \times A^*$ to its second component. For each $a \in A$, check whether $L_a\pi_2 = L$. Do all these equalities hold?
   Yes: Halt — S is a GROUP.
   No: Halt — S is NOT A GROUP.

Justification of Algorithm 3.4.2. Suppose S is a group. Then, in particular, S is a monoid and so the algorithm continues to Step 2. Let $a \in A$. For any $v \in L$, let u represent $\overline{va^{-1}}$. Then $(u, v) \in L_a$ and $v \in L_a\pi$. Therefore $L \subseteq L_a\pi$. The opposite inclusion is obvious, so $L_a\pi = L$. Therefore the algorithm halts and indicates that S is a group.

Now suppose the algorithm indicates that S is a group. Then, since the algorithm passes Step 1, S is a monoid. It is a group if $S\overline{a} = S$ for all $a \in A$, and this holds if each word in L appears as the second component of some element of each L_a, since the inclusion of $S\overline{a}$ in S is obvious. These are the equalities described in Step 2, which must all hold: S is therefore a group.

Step 1 of the algorithm is known to be effective. Step 2 is effective by the fact that π_2 is a homomorphism and so preserves regularity (Theorem A.5.2), and equality of regular languages is testable.

3.5. UNDECIDABILITY OF LEFT-CANCELLATIVITY

A SEMIGROUP with automatic structure (A, L) is right-cancellative if, for each $a \in A$,

$$L_a \circ L_a^{-1} \subseteq L_\varepsilon.$$

Thus, since the images of the relations $L_a \circ L_a^{-1}$ and L_ε under δ_A are regular, and one can test containment of regular languages by Theorem A.5.5, it is algorithmically possible to decide whether an automatic semigroup is right-cancellative. The present section shows that there is no algorithm that takes as input an automatic structure for a semigroup and decides whether that semigroup is left-cancellative. The proof proceeds by showing that one can reduce the Modified Post's Correspondence Problem to the question of deciding left-cancellativity for automatic semigroups.

An instance of *Post's Correspondence Problem* (PCP) consists of two lists of words over an alphabet X:

$$u_1, \ldots, u_n; \quad v_1, \ldots, v_n. \tag{1}$$

A solution to this instance of PCP is a sequence $i_1, \ldots, i_k$ drawn from the set $\{1, \ldots, n\}$ such that

$$u_{i_1} u_{i_2} \cdots u_{i_k} = v_{i_1} v_{i_2} \cdots v_{i_k}. \tag{2}$$

An instance of the *Modified Post's Correspondence Problem* (MPCP) consists of two lists of words (1) over an alphabet X. A solution to such an instance of MPCP is a sequence $i_1, \ldots, i_k$ with $i_1 = 1$ drawn from the set $\{1, \ldots, n\}$ such that the equality (2) holds.

Theorem 3.5.1 (Hopcroft & Ullman 1979, Theorem 8.8). *There is no algorithm that takes an instance of* PCP *and decides whether it has a solution. Likewise, there is no algorithm that takes an instance of* MPCP *and decides whether it admits a solution.* $\boxed{3.5.1}$

The strategy is now to encode an arbitrary instance of MPCP inside an automatic semigroup in such a way that the semigroup is left-cancellative if and only if that instance of MPCP has no solution.

Pick any instance (1) of MPCP. Define a semigroup as follows. Let

$$A = X \cup \{O, U, V\} \cup B \cup B' \cup C \cup C' \cup D,$$

where

$$\begin{aligned} B &= \{b_1, \ldots, b_n\}, \\ B' &= \{b'_1, \ldots, b'_n\}, \\ C &= \{c_1, \ldots, c_n\}, \\ C' &= \{c'_1, \ldots, c'_n\}, \\ D &= \{d_0, d_1, \ldots, d_n\}, \end{aligned}$$

and the sets X, $\{O, U, V\}$, B, B', C, C', and D are all pairwise disjoint. Let

$$\begin{aligned} \mathcal{R} = \{&(Ob_1, u_1 U d_0), (Oc_1, v_1 V d_0)\} && (3) \\ \cup \{&(Ub_i, u_i U d_i), (V c_i, v_i V d_i) : i = 1, \ldots, n\} && (4) \\ \cup \{&(Ub'_i, u_i d_i), (V c'_i, v_i d_i) : i = 1, \ldots, n\} && (5) \\ \cup \{&(d_j b_i, b_i d_j), (d_j b'_i, b'_i d_j), (d_j c_i, c_i d_j), (d_j c'_i, c'_i d_j) && (6) \\ &: i = 1, \ldots, n;\ j = 0, \ldots, n\}. \end{aligned}$$

Let S be the semigroup with presentation $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$. The aim is to show that

i.) The semigroup S is automatic (Lemma 3.5.3).

ii.) Every generator in $\overline{A - \{O\}}$ is left-cancellable: if $a \in A - \{O\}$, then $\overline{ap} = \overline{aq} \implies \overline{p} = \overline{q}$ for all $p, q \in A^+$ (Lemma 3.5.4).

iii.) The generator $\overline{O}$ is left-cancellable if and only if the instance (1) of MPCP has no solution (Lemma 3.5.5).

This will show that the left-cancellativity of S is equivalent to the the instance of MPCP (1) having no solution.

Let L be the language of all words over A that do not contain the left-hand side of an element of $\mathcal{R}$. Then L is the regular language

$$A^+ - \Big[A^*\{Ob_1, Oc_1, Ub_i, Vc_i, Ub_i', Vc_i', d_jb_i, d_jb_i', d_jc_i, d_jc_i' \\ : i = 1, \ldots, n;\ j = 0, \ldots, n\}A^*\Big].$$

The language L is the set of irreducible words for the rewriting system $(A, \mathcal{R})$. [Section 0.6 contains the necessary definitions and results regarding string-rewriting systems.]

Lemma 3.5.2. *The rewriting system $(A, \mathcal{R})$ is noetherian and confluent.*

Proof of 3.5.2. Noetherian. To show that reduction using the rewriting rules $\mathcal{R}$ must terminate, proceed as follows: for any $w \in A^+$ and $t \in \{1, \ldots, |w|\}$, define $\vartheta(w, t)$ to be the number of letters from $B \cup B' \cup C \cup C'$ lying in $w[t]$ if the t-th letter of w lies in D, and 0 otherwise. Now define

$$\Theta(w) = \sum_{t=0}^{|w|} \vartheta(w, t).$$

Let

$$\Phi(w) = \vartheta(w, 0) = \text{number of letters from } B \cup B' \cup C \cup C' \text{ in } w.$$

Define a partial order on A^* as follows: for $w, w' \in A^*$,

$$w \ll w' \iff \big(\Phi(w) < \Phi(w')\big) \text{ or } \big(\Phi(w) = \Phi(w') \text{ and } \Theta(w) < \Theta(w')\big).$$

Now, reduction using rules of types (3), (4), and (5) strictly decreases the value of $\Phi(w)$. A reduction step $w' \Rightarrow w$ using a rule of type (6) implies that $\Phi(w') = \Phi(w)$ and $\Theta(w') > \Theta(w)$: if the reduction step involves interchanging the t-th and $(t+1)$-th letters of w', then $\vartheta(w', t) = \vartheta(w, t+1) + 1$ and $\vartheta(w', t+1) = 0 = \vartheta(w, t)$. So reduction always strictly $\ll$-decreases a word: thus the process of reduction must terminate. The rewriting system $(A, \mathcal{R})$ is therefore noetherian.

Confluent. As there are no overlaps between left-hand sides of rules in $\mathcal{R}$, the rewriting system is confluent by Proposition 0.6.2. $\boxed{3.5.2}$

Theorem 0.6.3 therefore shows that the language of irreducible words L is a set of unique normal forms for S. Identify S with this set of normal forms, so that $\overline{w} = \mathrm{NF}(w)$ for all words $w \in A^+$.

Lemma 3.5.3. *The semigroup S admits (A, L) as an automatic structure.*

Proof of 3.5.3. That L maps onto S has already been established. So let $u, v \in L$ and $a \in A$ and suppose that $\overline{ua} = \overline{v}$. The word v is in normal form, so $\mathrm{NF}(ua) = v$. Since u is also in normal form, one of the following two possibilities holds:

- $a \in X \cup \{O, U, V\} \cup D$. Since no left-hand side of a rule in $\mathcal{R}$ ends with the letter a, the word ua is in normal form and so $ua = v$.
- $a \in B \cup B' \cup C \cup C'$. Let $u = u'u''$, where u'' is the longest suffix of u lying in D^*. Then $\mathrm{NF}(ua) = \mathrm{NF}(u'u''a) = \mathrm{NF}(u'au'')$. Noting that u' does not

end in a letter from D, if further reduction takes place, it must begin with an application of a rule of type (3), (4), or (5). This shows that $\mathrm{NF}(ua) = \mathrm{NF}(u'(|u'|-1)rTdu'')$, where rTd is the right-hand side of a rewriting rule with $r \in X^*$, $T \in \{U, V, \varepsilon\}$, $d \in D$. The word $u'(|u'|-1)rTdu''$ is in normal form, since $u'(|u'|-1)$ is in normal form, letters from X, such as those in r, do not appear on the left-hand side of any rule, T only appears on a left-hand side when followed by a letter from $B \cup B' \cup C \cup C'$, and no rule can be applied to $du'' \in D^*$. So either $\mathrm{NF}(ua) = u'au''$ or $NF(ua) = u'(|u'|-1)rTdu''$.

Therefore, it is clear that a finite state automaton can keep track of these differences and so recognize the language $L_a\delta_A$. Thus (A, L) is an automatic structure for the semigroup S. $\boxed{3.5.3}$

Lemma 3.5.4. *In the semigroup S, all generators except $\overline{O}$ left-cancel. That is, for $a \in A - \{O\}$*

$$(\forall p, q \in S)(\overline{a}p = \overline{a}q \implies p = q).$$

Proof of 3.5.4. Let $a \in A - \{O\}$ and let p and q be elements of S, viewed as normal form words in L. Distinguish the following cases:

i.) $a \in X \cup B \cup B' \cup C \cup C'$. No left-hand side of a relation in $\mathcal{R}$ begins with a letter a, so the words ap and aq are already in normal form. Therefore $ap = aq$ and so $p = q$.

ii.) $a \in D$. A letter from D only appears on the left-hand side of a rewriting rule when it is followed by a letter of $B \cup B' \cup C \cup C'$. By induction on t, one can see that $\mathrm{NF}(p(t)ap[t]) = \mathrm{NF}(p(t+1)ap[t+1])$ for $t = 0, 1, \ldots$ whenever all letters of $p(t+1)$ lie in $B \cup B' \cup C \cup C'$. Furthermore, if $p[t]$ does not begin with a letter from $B \cup B' \cup C \cup C'$, the word $p(t)ap[t]$ is in normal form. So $\mathrm{NF}(ap) = p'ap''$, where p' is the longest prefix of p over the alphabet $B \cup B' \cup C \cup C'$. Similarly, $\mathrm{NF}(aq) = q'aq''$, where q' is the longest prefix of q over $B \cup B' \cup C \cup C'$. Since $\overline{ap} = \overline{aq}$, $p'ap'' = q'aq''$, and so $p' = q'$ and $p'' = q''$, whence $p = q$.

iii.) $a = U$. (This case is rather more complicated than i. and ii.) The first step is to show that one can restrict to prefixes of p and q of a fairly simple form. First of all, eliminate a trivial possibility: if Up is in normal form, then Uq must be also, whence $p = q$. Therefore assume that neither ap nor aq are in normal form.

Let $p = p'p''$ and $q' = q'q''$, where p' and q' are the longest prefixes of p and q over the alphabet $B \cup B' \cup C \cup C'$. Assume without loss of generality that $|p''| \leq |q''|$. Consider the first reduction steps of Up and Uq. These must use rules of type (4) or (5), which produce letters d and e of D which are moved to the end of p' and q' by reduction rules of type (6). Subsequent reduction cannot affect dp'' or eq'', since p'' and q'' are already in normal form and d and e only appear on the left-hand side of a rewriting rule when followed by a letter of $B \cup B' \cup C \cup C'$. Therefore, since $\mathrm{NF}(Up) = \mathrm{NF}(Uq)$, and noting the assumption that $|p''| \leq |q''|$, the suffix p'' must appear in q''. Suppose $q'' = \tilde{q}p''$ for some word $\tilde{q}$.

The reasoning thus far shows that $\mathrm{NF}(Up') = \mathrm{NF}(Uq'\tilde{q})$, and that $\tilde{q}$ is not affected by reducing $Uq'\tilde{q}$ to normal form. Suppose $\tilde{q}$ contains a letter x from $X \cup \{O, U, V\}$. Then $\mathrm{NF}(Uq'\tilde{q})$ contains the letter x to the right of the letter $e \in D$ produced by the first reduction step. However, reduction of Up' cannot yield a letter from $X \cup \{O, U, V\}$ to the right of a letter from D. Therefore $\tilde{q}$ is either empty or contains only letters from $B \cup B' \cup C \cup C' \cup D$. Furthermore, as q'' begins with a letter not in $B \cup B' \cup C \cup C'$, the word $\tilde{q}$ — if it is non-empty — must begin with a letter of D. Since normal form words never have a letter of D immediately to the left of one from $B \cup B' \cup C \cup C'$, the word $\tilde{q}$ must lie in D^*.

Each application of a rule of type (3), (4), or (5) introduces a letter from D which is moved to the right of letters from $B \cup B' \cup C \cup C'$ using type (6) rules. Suppose $\tilde{q} = d_{j_l} \cdots d_{j_1}$, where $d_{j_h} \in D$. Then the first l reduction steps of Up' of types (3), (4), or (5) must produce these symbols: either

$$\mathrm{NF}(Up') = \mathrm{NF}\big(u_{j_1} \cdots u_{j_l} Up'[l] d_{j_l} \cdots d_{j_1}\big)$$

or

$$\mathrm{NF}(Up') = \mathrm{NF}\big(u_{j_1} \cdots u_{j_l} p'[l] d_{j_l} \cdots d_{j_1}\big)$$

depending on whether the reduction step that produces d_{j_l} is of type (4) or type (5). On the other hand,

$$\mathrm{NF}(Uq'\tilde{q}) = \mathrm{NF}\big(Uq' d_{j_l} \cdots d_{j_1}\big).$$

Now, since $p'[l], q' \in (B \cup B' \cup C \cup C')^*$, and the letters from D yielded by any further reduction steps must match, the prefixes of $p'[l]$ and q' involved in reduction using rules of type (4) or (5) must be the same length. Suppose this prefix of $p'[l]$ is $b_{i_1} \cdots b_{i_k}$, where $b_{i_h} \in B \cup B'$. (These letters must be drawn from $B \cup B'$ since letters from $C \cup C'$ do not appear alongside the letter U on the left-hand side of any relation in $\mathcal{R}$.) This word $b_{i_1} \cdots b_{i_k}$ must also be a prefix of q' to yield matching letters of D. Therefore

$$\mathrm{NF}(Up') = u_{j_1} \cdots u_{j_l} u_{i_1} \cdots u_{i_k} Up'[l+k] d_{i_k} \cdots d_{i_1} d_{j_l} \cdots d_{j_1}.$$

[The letter U is not present on the right-hand side if $b_{i_k} \in B'$.] However,

$$\mathrm{NF}(Uq'\tilde{q}) = u_{i_1} \cdots u_{i_k} Uq'[k] d_{i_k} \cdots d_{i_1} d_{j_l} \cdots d_{j_1}.$$

[Again, the letter U is not present on the right-hand side if $b_{i_k} \in B'$.] Since $p'[l+k]$ and $q'[k]$ are words over $B \cup B' \cup C \cup C'$, this forces $l = 0$. So

$$\mathrm{NF}(Up') = u_{i_1} \cdots u_{i_k} Up'[k] d_{i_k} \cdots d_{i_1}$$

and

$$\mathrm{NF}(Uq'\tilde{q}) = u_{i_1} \cdots u_{i_k} Uq'[k] d_{i_k} \cdots d_{i_1}.$$

[The same caveat applies to the presence of U as before.] Therefore $p'[k] = q'[k]$. The first k letters of both p' and q' are $b_{i_1} \cdots b_{i_k}$, so $p' = q'$. Since $|\tilde{q}| = l = 0$, the words p'' and q'' are equal. Thus $p = p'p'' = q'q'' = q$.

iv.) $a = V$. The reasoning exactly parallels that of case iii. $\boxed{3.5.4}$

Lemma 3.5.4 shows that the left-cancellativity of S depends solely on the behaviour of the generator $\overline{O}$.

Lemma 3.5.5. *In the semigroup S, the generator $\overline{O}$ left-cancels if and only if the instance of* MPCP (1) *does not have a solution.*

Proof of 3.5.5. Let p and q be elements of S viewed as normal form words, and assume that $\overline{Op} = \mathrm{NF}(Op) = \mathrm{NF}(Oq) = \overline{Oq}$. By reasoning as in case iii. of the proof of Lemma 3.5.4, assume that p lies in $(B \cup B' \cup C \cup C')^*$ and q in $(B \cup B' \cup C \cup C')^* D^*$.

If one of the words Op and Oq is in normal form, the other must be also, and $p = q$. So suppose some reduction occurs. The first letters of p and q must be drawn from $\{b_1, c_1\}$, since reduction must start with applications of rewriting rules of type (3). If these first letters are both b_1 or both c_1, these first reduction steps both produce letters U or letters V and the reasoning reduces to cases iii. and iv. of Lemma 3.5.4, showing that $p = q$.

Therefore O can be left-cancelled except possibly when p and q begin with different letters from $\{b_1, c_1\}$.

Suppose that O does not left-cancel. Choose normal form words $p \in (B \cup B' \cup C \cup C')^*$ and $q \in (B \cup B' \cup C \cup C')^* D^*$ such that $\mathrm{NF}(Op) = \mathrm{NF}(Oq)$ but $p \neq q$. Let $q = q'\tilde{q}$, where $\tilde{q}$ is the maximal suffix of q lying in D^*. By the observation in the last paragraph, assume that p begins with b_1 and q with c_1; the other case is symmetrical. This implies that

$$\mathrm{NF}(Op) = \mathrm{NF}(u_1 U p[1] d_0) \text{ and } \mathrm{NF}(Oq) = \mathrm{NF}(v_1 V q'[1] d_0 \tilde{q}).$$

Now, if $|\tilde{q}| > 0$, the symbol d_0 present in $\mathrm{NF}(Oq)$ must be matched in $\mathrm{NF}(Op)$ by a symbol produced by a subsequent reduction step. However, this is impossible, since symbols d_0 are only produced by reduction rules of type (3). Thus $|\tilde{q}| = 0$ and $q \in (B \cup B' \cup C \cup C')^*$.

Since suffixes over D in $\mathrm{NF}(Op)$ and $\mathrm{NF}(Oq)$ must match, p must begin with a string $b_1 b_{i_2} \cdots b_{i_k}$ over $B \cup B'$ and q with a string $c_1 c_{i_2} \cdots c_{i_k}$ over $C \cup C'$. Furthermore, since the letter U cannot appear in $\mathrm{NF}(Op)$ and the letter V cannot appear in $\mathrm{NF}(Oq)$, all letters b_{i_h} lie in B except $b_{i_k} \in B'$ and all letters c_{i_h} lie in C except $c_{i_k} \in C'$. This gives

$$\mathrm{NF}(Op) = u_1 u_{i_2} \cdots u_{i_k} p[k] d_{i_k} \cdots d_{i_2} d_0$$

and

$$\mathrm{NF}(Oq) = v_1 v_{i_2} \cdots v_{i_k} q[k] d_{i_k} \cdots d_{i_2} d_0.$$

Therefore the prefixes $u_1u_{i_2}\cdots u_{i_k}$ and $v_1v_{i_2}\cdots v_{i_k}$, which are words over X, must be identical. So, letting $i_1 = 1$, the sequence $i_1, i_2, \ldots, i_k$ is a solution to the instance of MPCP (1).

Now suppose that the instance of MPCP (1) has a solution $i_1, \ldots, i_k$ with $i_1 = 1$. Let $p = b_1b_{i_2}\cdots b_{i_{k-1}}b'_{i_k}$ and let $q = c_1c_{i_2}\cdots c_{i_{k-1}}c'_{i_k}$. Then

$$\mathrm{NF}(Op) = u_1u_{i_2}\cdots u_{i_k}d_{i_k}\cdots d_{i_2}d_0$$

and

$$\mathrm{NF}(Oq) = v_1v_{i_2}\cdots v_{i_k}d_{i_k}\cdots d_{i_2}d_0.$$

Since $i_1, \ldots, i_k$ is a solution to (1), $\overline{Op} = \mathrm{NF}(Op) = \mathrm{NF}(Oq) = \overline{Oq}$. Yet the elements $\overline{p} = \mathrm{NF}(p) = p$ and $\overline{q} = \mathrm{NF}(q) = q$ are unequal. Therefore the generator $\overline{O}$ does not left-cancel. $\boxed{3.5.5}$

Since the set $\overline{A}$ generates S, the upshot of Lemmata 3.5.4 and 3.5.5 is that S is left-cancellative if and only if the instance of MPCP (1) has no solution. The semigroup S is automatic by Lemma 3.5.3. Since there is no algorithm that determines whether an instance of MPCP has a solution (Theorem 3.5.1), the undecidability of left-cancellativity for automatic semigroups is established:

Theorem 3.5.6. *There is no algorithm that takes as input an automatic structure for a semigroup and decides whether that semigroup is left-cancellative.* $\boxed{3.5.6}$

[The encoding of MPCP into the left-cancellativity problem for an automatic semigroup is similar in spirit to the reduction of PCP to deciding the ambiguity of a particular context-free grammar; see Subsection A.8.1 and Hopcroft & Ullman (1979, Theorem 8.9).]

It is easy to see that the semigroup S is always right-cancellative by analyzing the rewriting that can occur upon right-multiplication by a single generator. [The possibile rewriting after right-multiplication by a generator is much more limited than that for left-multiplication. The necessary reasoning is therefore much simpler than the proofs of Lemmata 3.5.4 and 3.5.5.] Therefore the [two-sided] cancellativity of S depends only on whether the generator $\overline{O}$ left-cancels, which in turn depends only on whether the instance of MPCP (1) has a solution (Lemma 3.5.5).

Theorem 3.5.7. *There is no algorithm that takes as input an automatic structure for a semigroup and decides whether that semigroup is cancellative.* $\boxed{3.5.7}$

The semigroup S is not group-embeddable, even when it is cancellative. To see this, observe that if S *were* group-embeddable, the relations $\mathcal{R}^{\#}$ and $\mathcal{R}^{\mathrm{M}}$ would coincide. However, for any $i \in \{1, \ldots, n\}$, the relation $\mathcal{R}^{\mathrm{M}}$ includes the pair (b'_iU, b_i), as is shown by the following Malcev chain:

$$\begin{aligned} b_i'U &\to b_i'u_i^{\mathsf{L}}u_iUd_id_i^{\mathsf{R}} \\ &\to b_i'u_i^{\mathsf{L}}Ub_id_i^{\mathsf{R}} \\ &\to b_i'u_i^{\mathsf{L}}Ub_i'(b_i')^{\mathsf{R}}b_id_i^{\mathsf{R}} \\ &\to b_i'u_i^{\mathsf{L}}u_id_i(b_i')^{\mathsf{R}}b_id_i^{\mathsf{R}} \\ &\to b_i'd_i(b_i')^{\mathsf{R}}b_id_i^{\mathsf{R}} \\ &\to d_ib_i'(b_i')^{\mathsf{R}}b_id_i^{\mathsf{R}} \\ &\to d_ib_id_i^{\mathsf{R}} \\ &\to b_id_id_i^{\mathsf{R}} \\ &\to b_i. \end{aligned}$$

Clearly, the pair $(b_i'U, b_i)$ is not in $\mathcal{R}^{\#}$, since no relation in $\mathcal{R}$ can be applied to b_i. Therefore the following question remains open:

Open Problem 3.5.8. Is there an algorithm that takes an automatic structure for a semigroup and decides whether that semigroup is group-embeddable?

CHAPTER FOUR

SUBSEMIGROUPS OF VIRTUALLY FREE GROUPS

The system of words was eccentric. At times it proceeded in a single direction, at other times it went backwards, at still others in a circle...

— Umberto Eco,
The Name of the Rose (1980)
(trans. W. Weaver)

4.1. INTRODUCTION

FREE GROUPS have many pleasant properties, the most notable being the Nielsen–Schreier Theorem, which asserts that every subgroup of a free group is again a free group. An immediately consequence of the Nielsen–Schreier Theorem is that free groups are coherent (see Section 1.6). Free groups are also automatic (Epstein et al. 1992, Example 2.1.3). See Lyndon & Schupp (1977, Chapter I) for a formal definition and basic properties of free groups.

Denote the free group with basis X by $\mathrm{FG}(X)$. The free group $\mathrm{FG}(X)$ is presented by $\mathrm{Gp}\langle X \mid \emptyset\rangle$. Every element of the group $\mathrm{FG}(X)$ is represented by a unique word over $X \cup X^{-1}$ that contains no subword xx^{-1} or $x^{-1}x$ for any $x \in X$. Such words are known as *reduced words*. Identify the free group with the set of reduced words under the operation of 'concatenation plus cancellation': to multiply two reduced words, concatenate them and eliminate any subwords xx^{-1} or $x^{-1}x$ until a new reduced word is obtained. For example, the process of multiplying xyx and $x^{-1}y^{-1}x$ in $\mathrm{FG}(x,y)$ is as follows:

$$xyx \cdot x^{-1}y^{-1}x = xyxx^{-1}y^{-1}x = xyy^{-1}x = x^2.$$

The free group $\mathrm{FG}(X)$ contains the free monoid X^* and thus the free semigroup X^+. In X^* and X^+, multiplication is simply concatenation. A subsemigroup of the free group may be more complicated than any of the free semigroup, because of the more complex multiplication. Much of this chapter is devoted to the generalization to virtually free groups of results dealing with subsemigroups of free semigroups: see Theorems 4.2.4, 4.3.1, and 4.4.1.

Let F be a virtually free group. By Proposition 0.10.2 and the Nielsen–Schreier Theorem, F possesses a free normal subgroup N of finite index n. The virtually free group F therefore admits a presentation

$$\mathrm{Gp}\langle X, D \mid (d_i x_h d_i^{-1}, w_{i,h}), (d_i d_j, z_{i,j} d_{\mu(i,j)})\rangle, \tag{1}$$

where h ranges over an index set I and i, j over $\{1, \ldots, n\}$; and where $X = \{x_h : h \in I\}$, $D = \{d_1, \ldots, d_n\}$, $w_{i,h}, z_{i,j}$ are words over $X \cup X^{-1}$. The set X is a basis for the free normal subgroup N, and

$$\mathrm{Gp}\langle d_1, \ldots, d_n \mid (d_i d_j, d_{\mu(i,j)})\rangle$$

is the 'multiplication table' presentation for the finite group F/N. So, in (1), each d_i represents an element from the corresponding coset of N. Assume without loss that d_1 represents the element 1_F, drawn from the coset N itself. Observe that any word over $X \cup X^{-1} \cup D$ can be rewritten using the relations in (1) to a normal form wd, where w is a reduced word on $X \cup X^{-1}$ and $d \in D$. Note in particular that the normal form word representing 1_F is d_1. [Analysis of multiplication using normal form words is deferred until Theorem 4.4.1.]

4.1.1. Context-free word problem

Let A be a finite alphabet representing a semigroup generating set for a group G. The *word problem* for G (with respect to A) is the language of words over A that represent the identity of G:

$$\{w \in A^* : \overline{w} = 1_G\}.$$

Furthermore, if the word problem for G with respect to A lies in a family of languages $\mathfrak{F}$, and if $\mathfrak{F}$ is closed under forming inverse homomorphic images, then the word problem for G with respect to any other finite generating set also lies in $\mathfrak{F}$ (see, for example, Herbst & Thomas 1993, Corollary 2.2).

Much research has been carried out on classifying groups whose word problem lies in a particular family of languages. The earliest such classification is due to Anīsīmov (1971), who showed that the groups with regular word problem were the finite groups. Muller & Schupp (1983) proved that all finitely generated virtually free groups have deterministic context-free word problem, and that all groups that have context-free word problem *and are accessible* are virtually free. (Accessibility is a technical condition whose precise definition is not required here.) Dunwoody (1985) established the accessibility of all finitely presented groups. This, together with Anīsīmov's (1972) result that groups with context-free word problem are finitely presented, proves that a group has context-free word problem if and only if it is virtually free. [Dunwoody (1993) later constructed an example of an inaccessible finitely generated group.]

Muller & Schupp (1983, Lemma 3) show how to construct a deterministic pushdown automaton over $X \cup X^{-1} \cup D$ recognizing the word problem for a finitely generated virtually free group F. The automaton simulates multiplication of reduced words over $X \cup X^{-1}$ using its stack and multiplication in F/N using its states. Thus the automaton stores the element wd represented by the string of generators read thus far. It ultimately accepts a word if it leads to the element d_1 being stored. [For a detailed definition of this PDA, see Muller & Schupp.]

4.2. ALGORITHMICALLY TESTING FOR FREEDOM

4.2.1. Survey of algorithms for subsemigroups of free semigroups

LET C be a finite set of words over an alphabet X. Several conditions have been established that allow one to determine when the subsemigroup S of X^+ generated by C is free with basis C.

The earliest such condition was given by Sardinas & Patterson (1953), who stated their algorithm in terms of coding theory. A finite set C of words over the alphabet X is a code if any string formed by concatenating words from C can be uniquely decomposed back into those words. Clearly, C is a code if and only if the subsemigroup S is free with basis C. [Bandyopadhyay (1963) justifies the Sardinas–Patterson algorithm using elementary methods. Another proof is due to Riley (1967); a third to de Luca (1976). Levenšteĭn (1961a) also studied this problem from the perspective of coding theory.]

Knuth (2002, p. 2) relates how R. W. Floyd rediscovered the Sardinas–Patterson algorithm whilst investigating the problem of testing ambiguity of context-free grammars (see Section A.8). Although ambiguity is undecidable in general (Hopcroft & Ullman 1979, Theorem 8.9), Floyd gave a method to test the ambiguity of a context-free grammar Γ whose productions are of the form

$$N \to x_1 \mid \ldots \mid x_n \mid Nx_1 \mid \ldots \mid Nx_n,$$

where N is the only non-terminal and $x_1, \ldots, x_n$ are non-empty words over the terminal alphabet X. The grammar Γ is unambiguous if and only if the the semigroup generated by $\{x_1, \ldots, x_n\}$ is free on that set.

Cohn's (1962) condition asserts that the subsemigroup S is free (on some generating set) if

$$(\forall a, a', b, b' \in S)(ab' = ba' \implies (\exists x \in S^1)(a = bx \vee b = ax)).$$

This result does not form an algorithmic test for freedom.

Blum (1965a) gave another algorithm for testing for freedom, and also noted (Blum 1965b) that when C has size 2, the semigroup S is not free on C if and only if it is commutative. [Blum's (1965b) result can be proved rapidly as follows. Suppose S is not free on C. Then the subgroup of the free group $\mathrm{FG}(X)$ generated by C is also not free on C. The Nielsen–Schreier Theorem asserts that this subgroup is free of rank at most 2 (Lyndon & Schupp 1977, Proposition I.2.6). Since C is not a basis, the subgroup cannot be free of rank 2 (Lyndon & Schupp 1977, Proposition I.2.7): it is therefore free of rank 1 and thus abelian. Therefore the subsemigroup S — which is contained in this subgroup — is commutative.]

Spehner (1974/75) gave an algorithm that tests whether a finitely generated submonoid of a free monoid is free and — should it prove to be non-free — yields a presentation for the submonoid. [This presentation may be infinite even if the submonoid is finitely presented. Spehner's algorithm is explained in Lallement (1979).]

4.2.2. From free semigroups to virtually free groups

The remainder of the present section is devoted to establishing the existence of an algorithm that takes as input a finite subset of a virtually free group and tests whether the subsemigroup generated by that subset is free on some generating set. The reasoning in this subsection is an extended version of Sections 2 and 3 of Cain, Robertson & Ruškuc (2005b).

This algorithm generalizes those discussed in Subsection 4.2.1, which test only subsemigroups of free semigroups. Comparing Spehner's (1974/75) algorithm to the one below is particularly interesting. Spehner's arguments use directed graphs that resemble finite state automata; the arguments below rely heavily on the theory of context-free languages and pushdown automata. (Section 4.3 contains a similar observations with regard to arguments on Malcev coherence.)

[Ševrin (1960a) studied the freedom of subsemigroups of the free product of a free group and a free monoid. However, his condition does not yield an algorithmic test for freedom.]

Let F be a finitely generated virtually free group, and assume that a presentation for F of the form (1) is known. Let A be a finite alphabet representing a set of elements of F, with each element of $\overline{A}$ specified as a normal form word over $X \cup X^{-1} \cup D$. Let S be the subsemigroup of F generated by $\overline{A}$. Define the language

$$L(A) = \{uv^{-1} : u \in A^+,\ v \in A^*,\ \overline{u} = \overline{v},\ u \text{ and } v \text{ have no common suffix}\}.$$

Lemma 4.2.1. *The language $L(A)$ is context-free, and one can effectively construct a pushdown automaton recognizing it.*

Proof of 4.2.1. Let the language W be the word problem of F with respect to $X \cup X^{-1} \cup D$. Construct the pushdown automaton recognizing W as per Muller & Schupp (1983, Lemma 3). Let $\phi : (A \cup A^{-1})^* \to (X \cup X^{-1} \cup D)^*$ be the homomorphism that extends the mapping $a \mapsto \overline{a}$. By Theorem A.8.6, the language $W \subseteq (X \cup X^{-1} \cup D)^*$ is context-free; its inverse image $W\phi^{-1}$ is also context-free, and Theorem A.8.6 asserts that a pushdown automaton recognizing $W\phi^{-1}$ can be constructed. Observe that

$$W\phi^{-1} = \{w \in (A \cup A^{-1})^* : \overline{w} = 1_F\}. \tag{2}$$

The language

$$R = A^+(A^{-1})^* - \Big[\bigcup_{a \in A} A^* a a^{-1}(A^{-1})^*\Big]$$

is regular. Therefore, by Theorem A.8.5, $W\phi^{-1} \cap R$ is context-free, and one can effectively construct a PDA recognizing this language. Furthermore,

$$\begin{aligned} W\phi^{-1} \cap R &= \{w \in (A \cup A^{-1})^* : \overline{w} = 1_F\} \cap R \\ &= \{uv^{-1} : u \in A^+,\ v \in A^*,\ \overline{uv^{-1}} = 1_F, \\ &\qquad\qquad uv^{-1} \text{ does not include a subword } aa^{-1}\} \\ &= \{uv^{-1} : u \in A^+,\ v \in A^*,\ \overline{u} = \overline{v},\ u \text{ and } v \text{ have no common suffix}\} \\ &= L(A). \end{aligned}$$

This completes the proof. $\boxed{4.2.1}$

Proposition 4.2.2. *There is an algorithm that tests whether the subsemigroup S is free on $\overline{A}$.*

Proof of 4.2.2. Let S be the subsemigroup generated by $\overline{A}$. Suppose that S is not free on $\overline{A}$. Then some non-trivial relation holds: there exist $u, v \in A^+$ with $u \neq v$ such that $\overline{u} = \overline{v}$. Without loss of generality, assume $|u| \geq |v|$. Suppose $u = u's$, $v = v's$, where $s \in A^*$ is the common suffix of u and v of maximum length, and $u' \in A^+, v' \in A^*$. (Since this is a non-trivial relation, at most one of u' and v' can be the empty word ε and the length assumption shows that u' is not empty.) Then $\overline{u'} = \overline{v'}$, since S is a subsemigroup of a group and therefore cancellative, and $u'(v')^{-1}$ cannot contain a subword aa^{-1}. Therefore the language $L(A)$ is non-empty, since it contains $u'(v')^{-1}$.

Conversely, suppose $L(A)$ is non-empty. Let $uv^{-1} \in L(A)$. Then, since uv^{-1} does not contain aa^{-1} as a subword, u and v are not identical. If v is empty, replace u by ua and v by va for some $a \in A$. Then $\overline{uv^{-1}} = 1_F$, so $\overline{u} = \overline{v}$ and the non-trivial relation (u, v) holds in S. Therefore S is not free on $\overline{A}$.

Testing the emptiness of $L(A)$ is therefore equivalent to testing the freedom of S on $\overline{A}$, and there is an algorithm that checks whether the language defined by a given context-free grammar is empty (Theorem A.8.3). $\boxed{4.2.2}$

In a free semigroup C^+, the basis C is contained in every generating set. Therefore C^+ is only free on C. Thus, if the subsemigroup S is free on some generating set Y, then Y must be a subset of $\overline{A}$. There are only a finite number of subsets Y of $\overline{A}$. If it is possible to determine which of these subsets generate S, then the algorithm of Proposition 4.2.2 can be applied to each one. The subsemigroup S will then be free if and only if some subset Y of $\overline{A}$ generates S and the algorithm finds that the subsemigroup S is free on Y.

Let $B \subseteq A$ and let S' be the subsemigroup generated by $\overline{B}$. Clearly, S contains S'. The question of deciding whether $S = S'$ therefore reduces to that of determining whether $\overline{A - B} \subseteq S'$.

Lemma 4.2.3. *The element $\overline{a}$, where $a \in A - B$, lies in the subsemigroup S' if and only if the language $L(A) \cap a(B^{-1})^+$ is non-empty.*

Proof of 4.2.3. Suppose $\overline{a} \in S'$. Express a as $\overline{b_1 \cdots b_l}$, where $b_i \in B$. So there exists $w \in B^+$ such that $\overline{a} = \overline{w}$. Thus $\overline{aw^{-1}} = 1_F$. Since $a \notin B$, the word aw^{-1} does not contain aa^{-1} as a subword. So aw^{-1} is a member of the language $L(A) \cap a(B^{-1})^+$, which is therefore non-empty.

Conversely, if $L(A) \cap a(B^{-1})^+$ is non-empty, then there exists $w \in B^+$ such that $\overline{a} = \overline{w}$, whence $\overline{a} \in S'$. $\boxed{4.2.3}$

The language $L(A) \cap a(B^{-1})^+$ is context-free, and one can effectively construct a PDA recognizing it (Theorem A.8.5). Therefore the emptiness of $L(A) \cap a(B^{-1})^+$, and so the question of whether $S = S'$, is decidable. The discussion following Proposition 4.2.2 shows that:

Theorem 4.2.4. *There is an algorithm that takes as input:*

i.) *a finitely generated virtually free group F, specified by a presentation of the form (1),*

ii.) *a finite subset of F, specified as normal form words,*

and decides whether the subsemigroup of F generated by that subset is free. $\boxed{4.2.4}$

4.3. MALCEV COHERENCE

AS OBSERVED in Section 1.6, virtually free groups are coherent: all of their finitely generated subgroups are finitely presented. Indeed, every subgroup of a virtually free group is virtually free by Proposition 0.10.3. Example 1.4.2 shows, however, that virtually free groups — even free semigroups — contain finitely generated subsemigroups that are not finitely presented. Spehner (1989) proved that submonoids of a free monoid admit finite Malcev presentations. This section is dedicated to generalizing Spehner's result to subsemigroups of virtually free groups. [The reasoning below first appeared in Section 4 of Cain et al. (2005*b*).]

Theorem 4.3.1. *Virtually free groups are Malcev coherent. Moreover, there is an algorithm that takes as input:*

i.) *a finitely generated virtually free group F, specified by a presentation of the form (1),*

ii.) *a finite subset of F, specified as normal form words,*

and returns a finite Malcev presentation for the subsemigroup of F generated by that subset.

Proof of 4.3.1. Let F be a virtually free group; let A be a finite alphabet representing a subset of F; let S be the subsemigroup of F generated by $\overline{A}$. Since one can replace F by the [virtually free] subgroup generated by S, assume without loss of generality that F is finitely generated. Let W be the word problem of F with respect to $X \cup X^{-1} \cup D$ and let $\phi : A^+ \to F$ extend the mapping $a \mapsto \overline{a}$. Let

$$J(A) = W\phi^{-1} \cap A^+(A^{-1})^+ = \{uv^{-1} : u, v \in A^+,\ \overline{u} = \overline{v}\}.$$

Equation (2) and Theorem A.8.5 together show that $J(A)$ is a context-free language over $A \cup A^{-1}$ and that one can effectively construct a pushdown automaton recognizing $J(A)$.

Observe that S has an ordinary presentation $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where

$$\mathcal{R} = \{(u, v) : u \in A^+,\ v \in A^+,\ uv^{-1} \in J(A)\}.$$

The set $\mathcal{R}$ is the kernel of the representation mapping $w \mapsto \overline{w}$ and thus consists of all relations holding in S. The strategy is to define an ordering on $\mathcal{R}$ and show that all except a finite number of elements of $\mathcal{R}$ are Malcev consequences of preceding elements in that order.

Let $\Gamma = (N, A \cup A^{-1}, P, O)$ be a context-free grammar that generates $J(A)$. For the purposes of this proof, a path in a derivation tree from O to a leaf node is referred to as a *derivation path*. (Refer to Subsection A.8.1 for a discussion of derivation trees.)

Let T be a derivation tree of a word in $J(A)$. Define

$$n(T) = \text{number of internal vertices of } T.$$

(Recall that the internal vertices of T are labelled by non-terminals.) For $w \in J(A)$, define

$$n(w) = \min\{n(T) : T \text{ is a derivation tree for } w\}.$$

One could now define an order on $\mathcal{R}$ directly. However, defining the ordering on $J(A)$, then naturally mirroring it in $\mathcal{R}$ will prove advantageous later.

Define the partial order $\prec$ on $J(A)$ by

$$w_1 \prec w_2 \iff n(w_1) < n(w_2).$$

for $w_1, w_2 \in J(A)$. Mimic this ordering in $\mathcal{R}$ by defining

$$(u_1, v_1) \prec (u_2, v_2) \iff u_1 v_1^{-1} \prec u_2 v_2^{-1}$$

for $(u_1, v_1), (u_2, v_2) \in \mathcal{R}$.

Let $K(A)$ be the subset of $J(A)$ consisting of all words that have a derivation tree in which no derivation path contains the same non-terminal more than twice. The set of such derivation trees is finite, since the length of each of their derivation paths is bounded by $2|N|$. Therefore $K(A)$ is finite. Let

$$\mathcal{Q} = \{(u, v) \in \mathcal{R} : uv^{-1} \in K(A)\}.$$

The binary relation $\mathcal{Q}$ is clearly finite. This set will form the finite set mentioned above: all relations in $\mathcal{R} - \mathcal{Q}$ will be shown to be a Malcev consequences of $\prec$-preceding elements of $\mathcal{R}$.

Let $(u, v) \in \mathcal{R} - \mathcal{Q}$. Consider the derivation of $uv^{-1} \in J(A) - K(A)$ in Γ. Let T be a derivation tree for uv^{-1} with $n(T) = n(uv^{-1})$ — that is, with T having the minimum number of internal vertices of all derivation trees for uv^{-1}. At least one

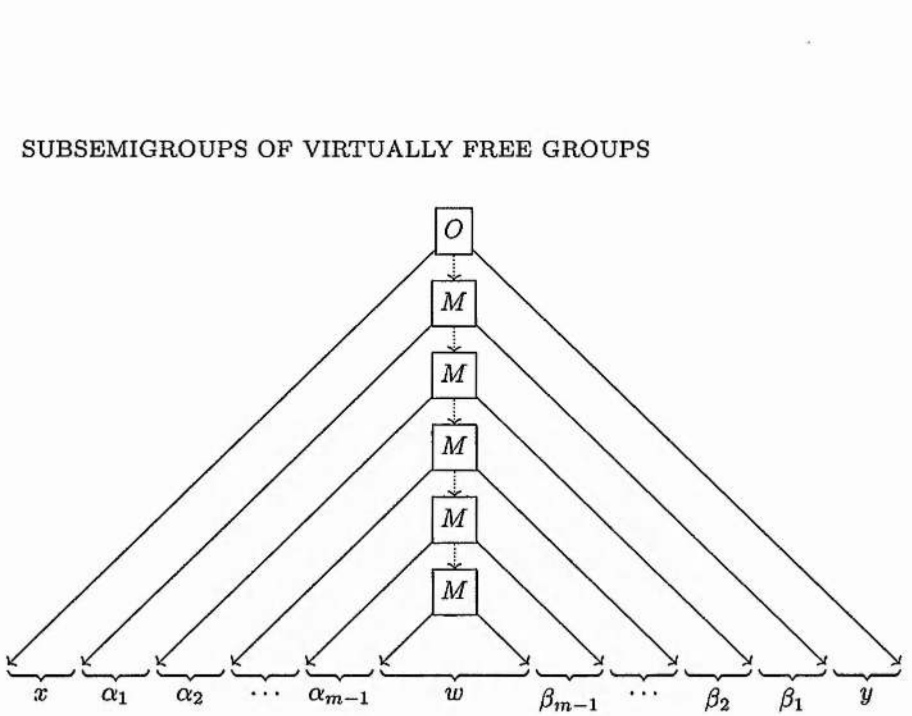

Figure 4.1. Schematic of the derivation tree for a word in $J(A)$. The derivations are stated explicitly in (3).

derivation path in T must have three or more internal vertices labelled by the same non-terminal. Distinguish such a derivation path with $m \geq 3$ appearances of this non-terminal M. Suppose

$$uv^{-1} = x\alpha_1 \cdots \alpha_{m-1} w \beta_{m-1} \cdots \beta_1 y,$$

where x, y, w and the α_i and β_i (for $i \in \{1, \ldots, m-1\}$) are all words in $(A \cup A^{-1})^*$ such that

$$O \overset{*}{\Rightarrow} xMy,\ M \overset{*}{\Rightarrow} \alpha_i M \beta_i \text{ for each } i \in \{1, \ldots, m-1\},\ M \overset{*}{\Rightarrow} w. \tag{3}$$

(See Figure 4.1 for a schematic illustration.)

The *u-v^{-1} boundary* of the word $uv^{-1} \in J(A)$ is the point between the last letter of the word $u \in A^+$ and the first letter of $v^{-1} \in (A^{-1})^+$.

Lemma 4.3.2. *The u-v^{-1} boundary is in either x, w, or y (possibly at the end of x or w or the start of w or y).*

Proof of 4.3.2. Suppose the u-v^{-1} boundary is in α_i for some $i \in \{1, \ldots, m-1\}$ (not at the start of α_1 or the end of α_{m-1}). Then there exist $s, t \in A^+$ such that $\alpha_1 \cdots \alpha_{m-1} = st^{-1}$. By the derivation

$$\begin{aligned} O &\overset{*}{\Rightarrow} xMy \\ &\overset{*}{\Rightarrow} x(\alpha_1 \cdots \alpha_{m-1}) M (\beta_{m-1} \cdots \beta_1) y \\ &\overset{*}{\Rightarrow} x(\alpha_1 \cdots \alpha_{m-1})^2 M (\beta_{m-1} \cdots \beta_1)^2 y \\ &\overset{*}{\Rightarrow} x(\alpha_1 \cdots \alpha_{m-1})^2 w (\beta_{m-1} \cdots \beta_1)^2 y, \end{aligned}$$

one sees that

$$J(A) \ni x(\alpha_1 \cdots \alpha_{m-1})^2 w(\beta_{m-1} \cdots \beta_1)^2 y = xst^{-1}st^{-1}w\beta_{m-1} \cdots \beta_1\beta_{m-1} \cdots \beta_1 y,$$

which is a contradiction, because this word is not in $A^+(A^{-1})^+$. A similar contradiction arises should the u–v^{-1} boundary be in some β_i. $\boxed{4.3.2}$

The relation $(u, v) \in \mathcal{R} - \mathcal{Q}$ therefore takes one of the following three forms:

i.) $(x\alpha_1 \cdots \alpha_{m-1}w', y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1}w'')$, where $w = w'(w'')^{-1}$, if the u–v^{-1} boundary is in w;

ii.) $(x\alpha_1 \cdots \alpha_{m-1}w\beta_{m-1} \cdots \beta_1 y', y'')$, where $y = y'(y'')^{-1}$, if the u–v^{-1} boundary is in y;

iii.) $(x', y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1}w^{-1}\alpha_{m-1}^{-1} \cdots \alpha_1^{-1}x'')$, where $x = x'(x'')^{-1}$, if the u–v^{-1} boundary is in x.

The second and third cases are almost symmetrical — only the possibility that v (and so also y'') could be the empty word makes the second case more general than the third. It therefore suffices to prove that, in cases i. and ii., (u, v) is a Malcev consequence of $\prec$-preceding elements.

Observe that in Γ, since $m > 2$,

$$O \overset{*}{\Rightarrow} xMy,\ M \overset{*}{\Rightarrow} \alpha_1 \cdots \alpha_{m-2}M\beta_{m-2} \cdots \beta_1,\ M \overset{*}{\Rightarrow} \alpha_{m-1}M\beta_{m-1},\ M \overset{*}{\Rightarrow} w$$

and therefore

$$\begin{aligned} &x\alpha_1 \cdots \alpha_{m-2}w\beta_{m-2} \cdots \beta_1 y, \\ &\qquad x\alpha_{m-1}w\beta_{m-1}y,\ xwy \in L(X). \end{aligned} \tag{4}$$

Furthermore, derivation trees with fewer than $n(uv^{-1}) = n(T)$ internal vertices exist for each of these words, as the following three derivations show:

$$\begin{aligned} &O \overset{*}{\Rightarrow} xMy \overset{*}{\Rightarrow} x\alpha_1 \cdots \alpha_{m-2}M\beta_{m-2} \cdots \beta_1 y \overset{*}{\Rightarrow} x\alpha_1 \cdots \alpha_{m-2}w\beta_{m-2} \cdots \beta_1 y, \\ &O \overset{*}{\Rightarrow} xMy \overset{*}{\Rightarrow} x\alpha_{m-1}M\beta_{m-1}y \overset{*}{\Rightarrow} x\alpha_{m-1}w\beta_{m-1}y, \\ &O \overset{*}{\Rightarrow} xMy \overset{*}{\Rightarrow} xwy. \end{aligned}$$

The words (4) therefore precede uv^{-1} in the $\prec$-ordering on $J(A)$.

i.) Firstly, observe that (4) implies that

$$\begin{aligned} &(x\alpha_1 \cdots \alpha_{m-2}w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}w''), \\ &\qquad (x\alpha_{m-1}w', y^{-1}\beta_{m-1}^{-1}w''), (xw', y^{-1}w'') \in \mathcal{R}, \end{aligned}$$

and that they each precede (u, v) in the $\prec$-ordering on $\mathcal{R}$. The following Malcev chain shows that (u, v) is a Malcev consequence of the given three relations:

$$\begin{aligned}
u &= x\alpha_1 \cdots \alpha_{m-1}w' \\
&\to x\alpha_1 \cdots \alpha_{m-2}w'(w')^{\mathsf{R}}\alpha_{m-1}w' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}w''(w')^{\mathsf{R}}\alpha_{m-1}w' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}y^{-1}w''(w')^{\mathsf{R}}\alpha_{m-1}w' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}xw'(w')^{\mathsf{R}}\alpha_{m-1}w' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}x\alpha_{m-1}w' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}y^{-1}\beta_{m-1}^{-1}w'' \\
&\to y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}\beta_{m-1}^{-1}w'' \\
&= v.
\end{aligned}$$

ii.) In this case, (4) means that

$$\begin{gathered}
(x\alpha_1 \cdots \alpha_{m-2}w\beta_{m-2} \cdots \beta_1 y', y''), \\
(x\alpha_{m-1}w\beta_{m-1}y', y''), (xwy', y'') \in \mathcal{R}
\end{gathered}$$

and these precede (u, v) in the $\prec$-ordering on $\mathcal{R}$. Once again, the following Malcev chain shows that (u, v) is a Malcev consequence of these relations:

$$\begin{aligned}
u &= x\alpha_1 \cdots \alpha_{m-1}w\beta_{m-1} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}x^{\mathsf{L}}x\alpha_{m-1}w\beta_{m-1} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}x^{\mathsf{L}}x\alpha_{m-1}w\beta_{m-1}y'(y')^{\mathsf{R}}\beta_{m-2} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}x^{\mathsf{L}}y''(y')^{\mathsf{R}}\beta_{m-2} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}x^{\mathsf{L}}xwy'(y')^{\mathsf{R}}\beta_{m-2} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}x^{\mathsf{L}}xw\beta_{m-2} \cdots \beta_1 y' \\
&\to x\alpha_1 \cdots \alpha_{m-2}w\beta_{m-2} \cdots \beta_1 y' \\
&\to y'' \\
&= v.
\end{aligned}$$

Therefore, (u, v) is a Malcev consequence of $\prec$-preceding elements, and this applies to all elements of $\mathcal{R} - \mathcal{Q}$. Hence the Malcev congruence generated by $\mathcal{Q}$ contains $\mathcal{R}$, and so $\mathrm{SgM}\langle A \mid \mathcal{Q}\rangle$ is a finite Malcev presentation for S.

Furthermore, since one can construct a CFG Γ from a pushdown automaton recognizing $J(A)$, and enumerate every derivation tree in Γ with at most two repetitions of any non-terminal in each derivation path, one can effectively construct the set $\mathcal{Q}$. Therefore, a finite Malcev presentation for the subsemigroup S can be found algorithmically from a finite generating set for S. $\boxed{4.3.1}$

4.3.1. Some further observations

Spehner (1989, Remark 2.10) notes a certain parallel between his proof that finitely generated submonoids of a free monoid have finite Malcev presentations and the proof of the Ehrenfeucht conjecture for regular languages given by Culik & Salomaa (1978). The proof of Theorem 4.3.1 resembles the proof of Albert, Culik & Karhumäki (1982) of the Ehrenfeucht conjecture in the case of context-free languages. In particular, the idea of examining repeated non-terminals in a derivation tree is drawn from that source.

[The Ehrenfeucht conjecture asserts that any language L over a finite alphabet A has a finite subset X such that, for any alphabet B and homomorphisms $\phi, \psi : A^+ \to B^+$, the mappings ϕ and ψ agree on L if and only if they agree on X. The conjecture has been proven independently by Albert & Lawrence (1985) and Guba (1986).]

A natural question is whether Theorem 4.3.1 generalizes to groups with context-free *co*-word problem, introduced by Holt, Rees, Röver & Thomas (2004). [The *co-word problem* of a group is the complement of the word problem — the language of all words representing elements other than the identity.] The closure of the class of deterministic context-free languages under complement (see Theorem A.8.7) implies that every virtually free group has context-free co-word problem. However, this latter class is closed under finite direct products (Holt et al. 2004, Proposition 6), and therefore contains the direct product of two free groups of rank 2, which, as was noted in Section 1.6, is not even coherent.

As the class of context-sensitive languages contains the class of context-free languages (Proposition A.9.2) and is closed under complementation (Proposition A.9.3), every group with context-free co-word problem has context-sensitive word problem. Therefore the observation of the last paragraph also shows that Theorem 4.3.1 does not generalize to the class of groups with context-sensitive word problem.

4.4. AUTOMATISM OF SUBSEMIGROUPS

FINITELY GENERATED SUBSEMIGROUPS of free semigroups were proven to be automatic by Campbell et al. (2001, Theorem 8.1). Their proof relies on the fact that multiplication in the free semigroup X^+ is simply concatenation of words. This enforces an approximate correspondence between lengths of words over an alphabet A and the lengths of words they represent in X^+. This correspondence is made more exact by adjoining an identity and 'padding' words in the language A^+ using symbols representing the identity.

This section is devoted to proving the analogous result for finitely generated subsemigroups of virtually free groups. The strategy is loosely based on that of Campbell et al., but is necessarily more complex. The main obstacle is finding a regular language L with an approximate correspondence between the lengths of a word in L and the length of the element it represents in the virtually free group. Having found such a language, one pads it in a manner similar to that of Campbell et al., before proving the conditions necessary to use Theorem 2.3.2.

Theorem 4.4.1. *Virtually free groups are locally automatic.*

[The proof below first appeared in Section 6 of Cain et al. (2005*a*).]

Proof of 4.4.1. Let F be a virtually free group presented by (1). Identify elements of F with their normal form representatives. Thus the length $|g|$ of an element $g \in F$ is the length of its normal form representative. Use $\cdot$ to denote concatenation of reduced words on $X \cup X^{-1} \cup D$.

Let A be a finite alphabet representing a subset of F, and let S be the subsemigroup of F generated by $\overline{A}$. The first step is to prove:

Proposition 4.4.2. *The semigroup S — viewed as a set of normal form representatives — is a regular language over $X \cup X^{-1} \cup D$.*

The proof begins by constructing a finite state automaton over $X \cup X^{-1} \cup D$ that recognizes S. To immediately define the automaton would be to sacrifice clarity to brevity: some motivational remarks will aid understanding of the ideas behind the proof.

Consider a simple example from a free group: forget about the letters d, so that multiplication is simply concatenation and free reduction. Let B be a finite alphabet representing the following elements of the free group $\mathrm{FG}(x, y, z)$:

$$\overline{B} = \{xy^2, y^{-1}z, z^{-1}y^{-1}zyx, x^{-1}y\}$$

Now, xzy^2 is in the subsemigroup T generated by $\overline{B}$, since

$$xy^2 \cdot y^{-1}z \cdot z^{-1}y^{-1}zyx \cdot x^{-1}y = xzy^2.$$

Analyze this factorization as follows:

$$x\,\overbrace{yy \cdot y^{-1}z \cdot z^{-1}y^{-1}}^{=\varepsilon}\,zy\,\overbrace{x \cdot x^{-1}}^{=\varepsilon}\,y = xzy^2.$$

A pair of letters may end up being consecutive in the word $xzyy$ in one of two ways: either, like zy, they are adjacent in the same generator, or, like xz, they are not adjacent in the same generator, but are united by cancellation.

One may use this observation to construct a non-deterministic finite state automaton recognizing every reduced word that lies in T. Here is a conceptual idea of how the automaton works: Let p be a 'first letter' of a word in T. This first letter must come from some generator in $\overline{B}$, say $s \cdot p \cdot t$. Since p can be a 'first letter', there exists a string $u \in B^*$ such that $\overline{u} = s^{-1}$. (This includes the possibility that $s = \varepsilon$ and $|u| = 0$.) The automaton reads the letter p and moves to a state (t, p). Suppose the next letter of the word is q. Either q is the next letter of $s \cdot p \cdot t$ — that is, the first letter of t — or is brought next to p by cancellation. In the former case, the automaton moves to the state $(t[1], q)$ after reading q. In the latter, q arises from another generator $s' \cdot q \cdot t'$, there exists a string $v \in B^*$ with $tvs' = \varepsilon$ and

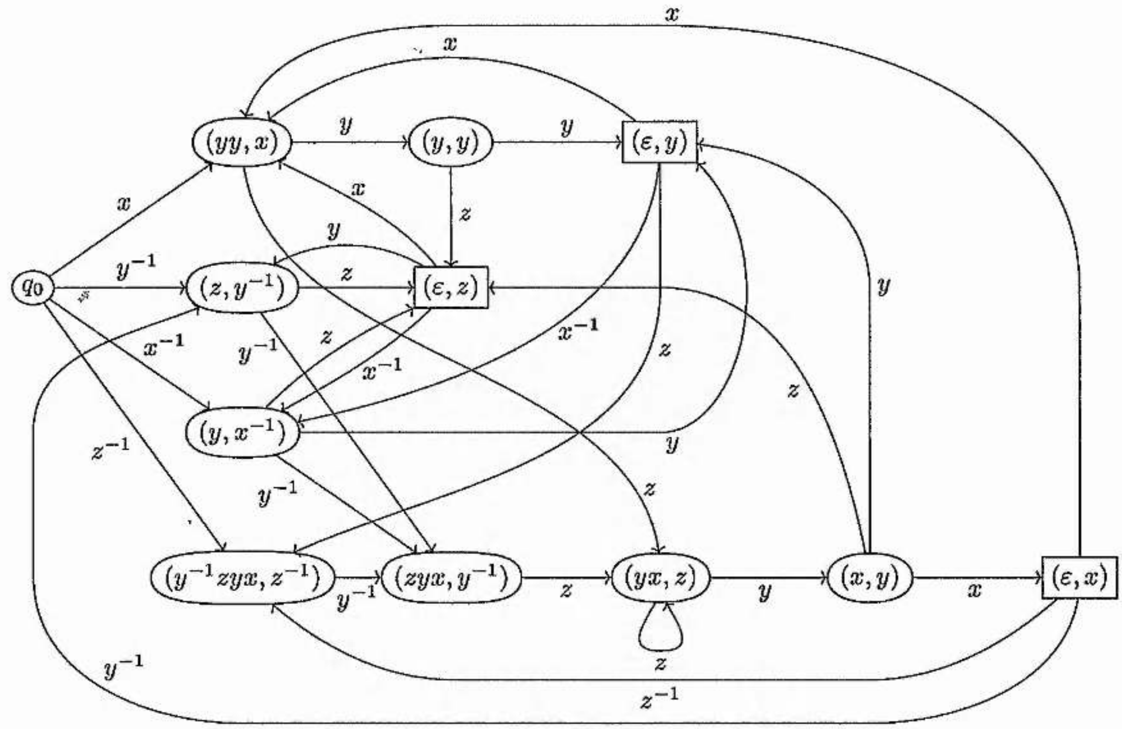

Figure 4.2. Finite state automaton recognizing reduced words in the subsemigroup of $\mathrm{FG}(x, y, z)$ generated by $\{xy^2, y^{-1}z, z^{-1}y^{-1}zyx, x^{-1}y\}$. The start state is q_0; accept states have rectangular outlines.

the automaton reads q and moves to the state (t', q). So in reading the word $xzyy$ above, the transitions of the automaton are as follows:

$$q_0 \xrightarrow{x} (yy, x) \xrightarrow{z} (yx, z) \xrightarrow{y} (x, y) \xrightarrow{y} (\varepsilon, y).$$

The first transition exists because x is the first letter of xyy; the second because $yy \cdot y^{-1}z \cdot z^{-1}y^{-1} = \varepsilon$, the remaining suffix of xyy being yy and $z^{-1}y^{-1}$ being the prefix of $z^{-1}y^{-1}zyx$ before z; the third because y follows z in the generator $z^{-1}y^{-1}zyx$; the fourth because $x \cdot x^{-1} = \varepsilon$, x being the remainder of $z^{-1}y^{-1}zyx$ after y and x^{-1} the prefix of $x^{-1}y$ before y. The state (ε, y) is an accept state; more generally any state $(r, \cdot)$ is an accept state if there exists $w \in B^*$ with $r\overline{w} = \varepsilon$. Obviously, such an automaton will be non-deterministic in general because there may be many possible factorizations of the same word into a set of generators, and because the fact that a particular string of generators is not a factorization may not be apparent until a long prefix of the word has been read.

Figure 4.2 shows the automaton for the subsemigroup T of $\mathrm{FG}(x, y, z)$.

The automaton defined below for subsemigroups of *virtually* free groups is more complex, as it must deal with the more complex multiplication of normal form words. Let w_1d_1 and w_2d_2 be elements of the virtually free group G, with w_1, w_2

being reduced words on $X \cup X^{-1}$ and $d_1, d_2 \in D$. To multiply these elements, concatenate $w_1 d_2$ and $w_1 d_2$ to obtain $w_1 d_1 \cdot w_2 d_2$; move d_1 to the right using the defining relations of the presentation (1) to obtain $w_1 \cdot w_3 d_3$, where w_3 is a reduced word on $X \cup X^{-1}$ and $d_3 \in D$; and freely reduce $w_1 \cdot w_3$ to obtain a reduced word w_4. Thus $w_1 d_1 w_2 d_2 = w_4 d_3$.

Proof of 4.4.2. Let

$$T = \{w : w \text{ is a suffix of the normal form of } d\overline{a} \in G, d \in D, a \in A\},$$

and let

$$Q = (T \times (X \cup X^{-1} \cup \{\varepsilon\})) \cup \{q_\infty\}.$$

Let $\mathcal{A} = (Q, X \cup X^{-1} \cup D, \eta, (\varepsilon, \varepsilon), \{q_\infty\})$ be a finite state automaton, where the transition function $\eta : Q \times (X \cup X^{-1} \cup D) \to Q$ consists of the following transitions and is elsewhere undefined:

i.) $((s, y), z) \mapsto (t, z)$ for $z \in X \cup X^{-1}$ if there exists $w \in A^+$ such that $s\overline{w} = z \cdot t$, and $y^{-1} \neq z$.

ii.) $((s, y), z) \mapsto (t, z)$ for $z \in X \cup X^{-1}$ if $s = z \cdot t$, $|t| \geq 1$, and $y^{-1} \neq z$.

iii.) $((s, y), d) \mapsto q_\infty$ for $d \in D$ if there exists $w \in A^+$ such that $s\overline{w} = d$.

iv.) $((s, y), d) \mapsto q_\infty$ for $d \in D$ if $s = d$.

Lemma 4.4.3. $S \subseteq L(\mathcal{A})$.

Proof of 4.4.3. Let $f \in S$ and let $u \in A^+$ with $u = u_1 \cdots u_n$, $u_i \in A$, and $\overline{u} = f$. Let $0 = i_0 < i_1 \leq i_2 \leq \cdots \leq i_{|f|} = n$ be such that $u(i_j)$ is the shortest prefix of u such that $f(j)$ is a prefix of $\overline{u(i_j)}$ that remains unaffected by subsequent right-multiplication by $\overline{u_{i_{j+1}}}, \ldots, \overline{u_n}$ in forming $\overline{u}$. [Observe that $i_{|f|} = n$ since any part of u after $u(i_{|f|})$ would affect at least the element of D at the end of $\overline{u(i_{|f|})}$. Even if the letter after $u(i_{|f|})$ represented the identity, the d at the end would be involved in rewriting using the defining relations in (1), although it would simply be rewritten back to d.]

Assume the automaton has read the prefix $f(j)$, ending in y, and has reached a state (s, y), where s is the suffix of $\overline{u(i_j)}$ after $f(j)$.

There are now four possibilities:

i.) $i_{j+1} > i_j$ and $z \in X \cup X^{-1}$. A transition of type i. exists, with $w \in A^+$ in its definition being such that $u(i_j)w = u(i_{j+1})$. The state reached by this transition is (t, z) with t being such that $\overline{u(i_{j+1})} = f(j+1) \cdot t$.

ii.) $i_{j+1} = i_j$ and $z \in X \cup X^{-1}$. A type ii. transition exists and leads to a state (t, z) with $s = z \cdot t$ and so t is such that $\overline{u(i_{j+1})} = f(j+1) \cdot t$.

iii.) $i_{j+1} > i_j$ and $z \in D$. A type iii. transition exists, with $w \in A^+$ in its definition being such that $u(i_j)w = u(i_{j+1})$. The state reached by this transition is q_∞.

iv.) $i_{j+1} = i_j$ and $z \in D$. A transition of type iv. exists and leads to q_∞.

Induction on j shows that $\mathcal{A}$ accepts f, since entering state q_∞ occurs only on reading $z \in D$, which can only happen at the end of f. $\boxed{4.4.3}$

Lemma 4.4.4. $L(\mathcal{A}) \subseteq S$.

Proof of 4.4.4. Let $f \in L(\mathcal{A})$. Only transitions labelled by D reach q_∞, and the definitions of transitions labelled by $X \cup X^{-1}$ mean that f contains no adjacent letters that are mutually inverse. Therefore f is certainly in normal form.

Choose and fix a walk γ in $\mathcal{A}$ that is labelled by f. Let

$$\gamma = \gamma_1 \cdots \gamma_n,$$

where each γ_i is a single transition. Let s_i be the first component of the state to which γ_i leads, with s_n formally defined as ε.

For transition γ_i of type i. or iii., define $\gamma_i\rho$ to be w, where $w \in A^+$ is as in the definition of this transition. If γ_i is of type ii. or iv., define $\gamma_i\rho = \varepsilon$.

For some i, assume that

$$\overline{\gamma_1\rho \cdots \gamma_i\rho} = f(i) \cdot s_i. \tag{5}$$

Suppose γ_i is labelled by z, so that $f(i+1) = f(i) \cdot z$. If γ_{i+1} is of type ii. or iv., then $s_i = z \cdot s_{i+1}$ and $\gamma_{i+1}\rho = \varepsilon$, and so:

$$\begin{aligned}\overline{\gamma_1\rho \cdots \gamma_i\rho\gamma_{i+1}\rho} &= \overline{\gamma_1\rho \cdots \gamma_i\rho} \\ &= f(i) \cdot s_i \\ &= f(i) \cdot z \cdot s_{i+1} \\ &= f(i+1) \cdot s_{i+1}.\end{aligned}$$

If, on the other hand, γ_{i+1} is of type i. or iii., then $\gamma_{i+1}\rho = w$, where w is such that $s_i\overline{w} = z \cdot s_{i+1}$, whence:

$$\begin{aligned}\overline{\gamma_1\rho \cdots \gamma_i\rho\gamma_{i+1}\rho} &= \overline{\gamma_1\rho \cdots \gamma_i\rho w} \\ &= f(i) \cdot s_i\overline{w} \\ &= f(i) \cdot z \cdot s_{i+1} \\ &= f(i+1) \cdot s_{i+1}.\end{aligned}$$

In either case, (5) holds for $i+1$. Observe that (5) holds trivially for $i = 0$. Induction now shows that it holds for all i. In particular, it holds for $i = n$. Therefore $\overline{\gamma_1\rho \cdots \gamma_n\rho} = f$, and so $f \in S$. $\boxed{4.4.4}$

Lemmas 4.4.3 and 4.4.4 together prove Proposition 4.4.2. $\boxed{4.4.2}$

If S is not already a monoid, adjoin an identity to obtain a monoid $M = S^1$, otherwise let $M = S$. Let 1 be a new symbol not in A that represents the identity. The strategy is now to construct from $\mathcal{A}$ a generalized finite state automaton $\mathcal{W}$ that recognizes a language over $A \cup \{1\}$ that maps onto S. (For the case $M = S^1$, the language of normal forms will be $L(\mathcal{W}) \cup \{1\}$.)

Let $p \xrightarrow{y} q$ be a transition in $\mathcal{A}$. Suppose it is of type i. or iii. Let w be as in the definition of this transition, with $|w|$ being minimal. Define $(p, y, q)\rho = w$. For all other transitions $p \xrightarrow{y} q$, define $(p, y, q)\rho = \varepsilon$. (This definition of ρ echoes that in the proof of Lemma 4.4.4.) Let $m = \max\{|w| : w \in \operatorname{im} \rho\}$.

The construction of $\mathcal{W}$ is as follows. Retain the state set and the start and accept states from $\mathcal{A}$. Replace the label on each transition $p \xrightarrow{y} q$ of the automaton $\mathcal{A}$ by $(p, y, q)\rho 1^{m-|(p,y,q)\rho|}$, so that each edge in $\mathcal{W}$ has a label of length m.

The following two lemmata relate a word recognized by $\mathcal{A}$ to the word labelling the corresponding path in $\mathcal{W}$.

Lemma 4.4.5. *If u labels a walk in $\mathcal{W}$ from the start state to an accept state and s labels the corresponding walk in $\mathcal{A}$, then $\overline{u} = s$ and $m|s| = |u|$.*

Proof of 4.4.5. This lemma follows from the proof of Lemma 4.4.4, noting the coincidence of the definition of ρ in that proof and in the construction of $\mathcal{W}$. $\boxed{4.4.5}$

Lemma 4.4.6. *Let $u \in L(\mathcal{W})$ and let $t \in \mathbb{N} \cup \{0\}$ with $m \mid t$. Then the first t/m letters of $\overline{u}$ form a word w_u over $X \cup X^{-1} \cup D$ that label the walk in $\mathcal{A}$ corresponding to the walk labelled by $u(t)$ in $\mathcal{W}$. Furthermore, if s_u is the first component of the state to which this walk leads (in either $\mathcal{A}$ or $\mathcal{W}$), then $\overline{u(t)} = w_u \cdot s_u$.*

Proof of 4.4.6. Observe that, since $m \mid t$, upon reading the prefix $u(t)$, the generalized finite state automaton $\mathcal{A}$ does indeed enter a state. Reasoning parallel to that in the proof of Lemma 4.4.4 shows that $\overline{u(t)} = w_u \cdot s_u$. $\boxed{4.4.6}$

Observe that Lemma 4.4.5 and the fact that $\mathcal{A}$ recognizes S together imply that $L(\mathcal{W})$ maps onto S. The remainder of the proof consists of showing that if $M = S$; then $(A \cup \{1\}, L(\mathcal{W}))$ is an automatic structure for M; and that if $M = S^1$, then $(A \cup \{1\}, L(\mathcal{W}) \cup \{1\})$ is an automatic structure for M.

First case. Suppose $M = S$. Let $u, v \in L(\mathcal{W})$ be such that $\overline{ua} = \overline{v}$ for $a \in A \cup \{1\} \cup \{\varepsilon\}$. In order to invoke Theorem 2.3.2, it is necessary to show that for all $t \in \mathbb{N} \cup \{0\}$, $\overline{u(t)}$ and $\overline{v(t)}$ are within a bounded distance of one another. One may assume without loss that $t \leq \max\{|u|, |v|\}$.

Since cancellation can proceed at most $l = \max\{|d\overline{a}| : d \in D, a \in A\}$ leftwards from the end of an element of S, $\overline{u}$ and $\overline{v}$ must be equal except for suffixes of length at most l.

Let $t \leq \max\{|u|, |v|\}$ and suppose that $m \mid t$.

i.) If t is such that $|\overline{u}| - t/m$ is greater than l, apply Lemma 4.4.6 to u and v to obtain w_u, s_u, w_v and s_v. The prefix w_u, being of length $t/m < |\overline{u}| - l$, is confined to the part of $\overline{u}$ that is unaffected by cancellation in forming the product $\overline{u}\,\overline{a}$, and so w_u and w_v are identical. Therefore $\overline{u(t)}^{-1}\overline{v(t)} = s_u^{-1} s_v$, and since there are only a finite number of choices for s_u and s_v, $\overline{u(t)}$ and $\overline{v(t)}$ are within a bounded distance of one another.

ii.) If t is such that $|\overline{u}| - t/m$ does not exceed l, then $|u| = m|\overline{u}| \leq ml + t$, so $|u| - |u(t)| \leq ml$. The distance between $\overline{u(t)}$ and $\overline{v}$ is at therefore at most $ml + 1$.

Consider the maximum amount of cancellation that can occur in forming the product $\overline{ua}$ to see that $|\overline{v}| \geq |\overline{u}| - l$. So $m|\overline{v}| \geq m|\overline{u}| - ml$, whence $|v| \geq |u| - ml$ by Lemma 4.4.5. So $t \leq \max\{|u|, |v|\} \leq |v| + ml$, or $|v| \geq t - ml$. On the other hand, considering the case when no cancellation occurs in forming the product $\overline{ua}$ shows that $|\overline{v}| \leq |\overline{u}| + l$. So $|v| = m|\overline{v}| \leq m|\overline{u}| + ml \leq 2ml + t$. Therefore

$$-ml \leq |v| - t \leq 2ml,$$

or $\big||v| - t\big| \leq 2ml$. So the distance from $\overline{v}$ to $\overline{v(t)}$ is at most $2ml$. So the total distance from $\overline{u(t)}$ to $\overline{v(t)}$ is at most $3ml + 1$.

Clearly, if $m \nmid t$, then a distance at most $2m$ is added to these bounds.

Therefore $(A \cup \{1\}, L(\mathcal{W}))$ possesses the fellow traveller property. The conditions of Theorem 2.3.2 are thus satisfied and so $(A \cup \{1\}, L(\mathcal{W}))$ is an automatic structure for M.

Second case. Now suppose that $M \neq S$ and $a \in A \cup \{1\} \cup \{\varepsilon\}$. The reasoning above remains valid except when u or v is 1. Suppose $a = \varepsilon$ or $a = 1$. Clearly, $u = 1$ if and only if $v = 1$, since otherwise there would be a word in $L(\mathcal{W}) \subseteq A^+$ representing the identity of M, contradicting the definition of M. Trivially, therefore, $\overline{u(t)}$ and $\overline{v(t)}$ are within a bounded distance of one another for all t.

If $a \in A$, then v cannot be 1, since otherwise ua would represent the identity, contradicting the definition of M. So assume $u = 1$, so that $\overline{v} = \overline{a}$. Then v can have at length at most $m|\overline{a}|$, by the definition of $\mathcal{W}$. Therefore, for all t, the distance between $\overline{u(t)}$ and $\overline{v(t)}$ is bounded.

Therefore $(A \cup \{1\}, L(\mathcal{W}) \cup \{1\})$ satisfies the conditions of Theorem 2.3.2 and so is an automatic structure for M. The monoid M is therefore automatic. Now apply Theorem 2.4.3 to show that S is automatic. $\boxed{4.4.1}$

Theorem 4.4.7. *There is an algorithm that takes as input:*

i.) *a finitely generated virtually free group F, specified by a presentation of the form* (1),

ii.) *a finite subset of F, specified as normal form words,*

and returns an automatic structure for the subsemigroup of F generated by that subset.

Proof of 4.4.7. Adopt the notation from the proof of Theorem 4.4.1. The set T can be constructed from $\overline{A}$ and the presentation (1), so the state set Q can also be effectively constructed. To complete construction of the automaton $\mathcal{A}$ from Proposition 4.4.2, one needs to compute the set of transitions.

Clearly, the presence of a transition of type ii. or iv. can be decided simply by inspection of the state set.

Let $A' = A \cup \{\alpha\} \cup \{\omega\}$, where α and ω are new symbols not in A. To decide whether there should be a transition of type i. from (s, y) to (t, z) labelled by $z \in Y \cup Y^{-1}$ with $y^{-1} \neq z$, proceed as follows: let $\overline{\alpha} = s$ and $\overline{\omega} = z \cdot t$. Construct the context-free language $L(A')$ as described in Lemma 4.2.1. Construct the language

$$L(A') \cap \alpha A^+ \omega^{-1},$$

which is also context-free, and check whether it is empty. If it contains $\alpha w \omega^{-1}$, then $s\overline{w} = \overline{\alpha w} = \overline{\omega} = z \cdot t$, and there should be an edge $(s, y) \xrightarrow{z} (t, z)$.

Similarly, to decide whether there should be a transition of type iii. from (s, y) to q_∞ labelled by $d \in D$, let $\overline{\alpha} = s$ and $\overline{\omega} = d$ and again check the emptiness of $L(A') \cap \alpha A^+ \omega^{-1}$.

Thus one can effectively construct the automaton $\mathcal{A}$. It is then clear that the construction of $\mathcal{W}$ proceeds by inspection of $\mathcal{A}$.

To discover whether S is a monoid, one simply checks whether $d_1 \in D$ — the normal form word representing the identity of F (see Section 4.1) — lies in $L(\mathcal{A})$.

As observed in the proof of Theorem 4.4.1, if $(u, v) \in L_a$ for $a \in A \cup \{\varepsilon\}$ and t is such that $|\overline{u}| - t/m$ and $|\overline{v}| - t/m$ exceed l, then $\overline{u(t)}^{-1}\overline{v(t)} = s_u^{-1} s_v$, where s_u and s_v are in T. (The constants l and m were defined in the proof of Theorem 4.4.1.) Let k_1 be the maximum length (in F) of all possible $s_u^{-1} s_v$. For other values of t, the elements $\overline{u(t)}$ and $\overline{v(t)}$ are at most $k_2 = (3ml + 1)\max\{|\overline{a}| : a \in A\}$ apart in F by Proposition 0.7.4. So $\overline{u(t)}^{-1}\overline{v(t)}$ always lies in a ball of radius $\max\{k_1, k_2\}$ around the identity in $\Gamma(F, X \cup X^{-1} \cup D)$. Applying the construction of Theorem 2.3.2 using this ball yields a finite state automaton recognizing the language L_a.

Finally, to obtain, for some $a \in A$, a representative $u_a \in L$ such that $\overline{u_a} = \overline{a}$, simply trace a path in $\mathcal{A}$ labelled by $\overline{a}$ and take the word labelling the corresponding path in $\mathcal{W}$. $\boxed{4.4.7}$

4.5. CONSEQUENCES OF AUTOMATISM

4.5.1. Finite Malcev presentations

THEOREM 2.5.1 asserts that every automatic semigroup embeddable into a group admits a finite Malcev presentation. Therefore, the fact that finitely generated subsemigroups of virtually free groups have finite Malcev presentations is an immediate corollary of Theorem 4.4.1. Furthermore, Theorem 4.4.7 shows that one can effectively construct an automatic structure for such a subsemigroup, starting from a presentation (1) for the virtually free group and a finite generating set for the subsemigroup specified as normal form words. Algorithm 3.2.1 yields a Malcev presentation given the automatic structure. This reasoning forms an alternative proof for Theorem 4.3.1.

4.5.2. Testing for freedom

Section 3.3 describes an algorithm that takes as input an automatic structure for a semigroup and decides whether that semigroup is free. As mentioned in the preceding subsection, one can effectively compute an automatic structure for a subsemigroup of a finitely generated virtually free group. Algorithm 3.3.1 can then be run with this automatic structure as input. This yields a method of testing for freedom alternative to Theorem 4.2.4.

4.5.3. Spehner's theorem

The machinery of asynchronous automatism allows one to rapidly prove the semigroup version of Spehner's (1989) result that every finitely generated submonoid of a free monoid has a finite Malcev presentation (see Theorem 1.5.1):

Theorem 4.5.1. *Every finitely generated subsemigroup of a free semigroup is asynchronously automatic and so has a finite Malcev presentation.*

Proof of 4.5.1. Let X^+ be a free semigroup with basis X and let A be a finite alphabet representing a set of generators for a subsemigroup S of X^+. Let $L = A^+$. The aim is to show that (A, L) is an asynchronous automatic structure for S by showing that it satisfies the hypotheses of Theorem 2.3.11. Notice firstly that, by definition, $\overline{L} = S$.

Let $u, v \in L$ with $(u, v) \in L_a$ for some $a \in A \cup \{\varepsilon\}$. Let $t \in \mathbb{N} \cup \{0\}$. The element $\overline{u(t)}$ of S is a word over X. Let $v(s)$ be the shortest prefix of u such that $\overline{u(t)}$ is a prefix of $\overline{v(s)}$. Then $\overline{v(s)}w = \overline{u(t)}$, where w is a (possibly empty) suffix of a word in $\overline{A}$. The distance from $\overline{u(t)}$ and $\overline{v(s)}$ in $\Gamma(S, A)$ is thus bounded by a quantity dependent only on $\overline{A}$. Since t was arbitrary, every point on $\widehat{u}$ is a bounded distance from some point on $\widehat{v}$. The same reasoning applies with u and v interchanged. Since u and v were arbitrary, (A, L) has the Hausdorff closeness property.

Since every element of $\overline{A}$ has non-zero length as a word over X, any word $w \in S$ has a finite number of representatives in $L = A^+$. Therefore L maps finite-to-one onto S. By Lemma 2.3.7, a departure function for (A, L) exists.

Thus, by Theorem 2.3.11, (A, L) is an asynchronous automatic structure for S. Finally, by Theorem 2.5.4, S admits a finite Malcev presentation. $\boxed{4.5.1}$

4.6. QUESTIONS ON FINITE PRESENTABILITY

MARKOV (1971/72, THEOREM III) exhibits an algorithm that takes as input a finite set of words C over an alphabet X and determines whether the subsemigroup of X^+ generated by C is finitely presented. Given the above extension of various results regarding free semigroups to free and virtually free groups, the following question arises:

Open Problem 4.6.1. Is there an algorithm that takes a finite subset of a [virtually] free group and determines whether the subsemigroup generated by that subset is finitely presented?

Budkina & Markov (1973*a*) describe all subsemigroups of rank 3 of a free semigroup and showed in particular that all such subsemigroups are finitely presented. [Spehner (1981) builds upon the work of Budkina & Markov to show that all such subsemigroups have a presentation with at most two defining relations unless they are commutative.] One naturally asks whether this extends to subsemigroups of a free or virtually free group:

Open Problem 4.6.2. Is every three-generated subsemigroup of a [virtually] free group finitely presented?

CHAPTER FIVE

GROUPS SATISFYING SEMIGROUP LAWS

Law is order, and good law is good order.
— Aristotle, *Politics*, bk. VII
(trans. B. Jowett)

5.1. INTRODUCTION

INFORMALLY, a semigroup S obeys a law $u = v$ (where u and v are words over an alphabet X) if every possible substitution of elements of S for letters of X in u and v gives equals elements of S. For example, commutative semigroups satisfy the law $xy = yx$. [Semigroup laws are defined formally in Section 5.2.]

The present chapter studies subsemigroups of groups that obey non-tautological semigroup laws. Nilpotent, virtually nilpotent, abelian, and virtually abelian groups are considered, and, in particular, whether their subsemgroups admit finite Malcev presentations or automatic structures.

5.2. SEMIGROUP LAWS

THE FOLLOWING DEFINITION formalizes the intuitive definition of a semigroup law given above:

Definition 5.2.1. A *semigroup law* consists of two words u, v over an alphabet X. Such a law is simply denoted $u = v$. A semigroup S *satisfies* or *obeys* the law $u = v$ if $u\vartheta = v\vartheta$ for every homomorphism $\vartheta : X^+ \to S$. A semigroup law is *tautological* if it is of the form $u = u$; it is otherwise *non-tautological*.

If a semigroup obeys a particular semigroup law, then all of its subsemigroups also obey that law.

Proposition 5.2.2 (Lewin & Lewin 1969, Section IV). *Let G be a group, and let E be a finite extension of G, with $[E : G] = n$. Suppose G satisfies the semigroup law $u = v$. Then E satisfies the law obtained by replacing each letter x by x^n in both u and v.* $\boxed{5.2.2}$

Malcev (1953) classified semigroups embeddable into nilpotent groups as being cancellative semigroups that satisfy a particular non-tautological semigroup law. Neumann & Taylor (1963) independently discovered a similar classification that involves a slightly different law, although Neumann (1964) later acknowledged Malcev's precedence. The Malcev and Neumann–Taylor classifications are discussed further in Section 5.3.

5.2.1. Semigroup laws and group-embeddability

Recall the definitions of the right- and left-reversibility of a semigroup and those of groups of left and right quotients (Subsection 0.9.1): a semigroup is right-reversible (respectively, left-reversible) if any two elements have a common left (respectively, right) multiple; a group G is a group of left (respectively, right) quotients of a semigroup S if $G = S^{-1}S$ (respectively, SS^{-1}). Dubreil's Theorem 0.9.10 asserts the existence of the group of left (respectively, right) quotients of a cancellative right-reversible (respectively, left-reversible). Some additional results on these subjects will be needed in forthcoming sections:

Proposition 5.2.3 (Clifford & Preston 1961, Theorem 1.25). *Let S be a right-reversible (respectively, left-reversible) cancellative semigroup, and let G and H be groups of left quotients (respectively, right quotients) of S. Then there is an isomorphism from G to H fixing all elements of S.* $\boxed{5.2.3}$

It is therefore sensible to discuss *the* group of left or right quotients of a right- or left-reversible semigroup S.

Proposition 5.2.4 (Clifford & Preston 1961, Proof of Theorem 1.24). *Let G be a group and let S be a right-reversible (respectively, left-reversible) subsemigroup of G. Let H be the subgroup of G generated by S. Then H is [isomorphic to] the group of left quotients (respectively, right quotients) of S.* $\boxed{5.2.4}$

Proposition 5.2.4 implies that the only group that a right- or left-reversible cancellative semigroup S embeds into and generates is the group of left or right quotients of S. This observation yields the following lemma:

Lemma 5.2.5. *Let S be a right-reversible (respectively, left-reversible) cancellative semigroup. Let G_1 be the universal group of S. Let G_2 be the group of left (respectively, right) quotients of S. Let G_3 be the subgroup generated by S inside some group into which S embeds. Then G_1, G_2, and G_3 are all isomorphic.* $\boxed{5.2.5}$

The next result links the study of right- and left-reversible semigroups with that of semigroups that obey non-tautological laws:

Lemma 5.2.6. *If S is a cancellative semigroup that satisfies a non-tautological law, then S is both right- and left-reversible.*

The proof below is essentially that of Neumann & Taylor (1963, Corollary 2), although Lemma 5.2.6 is phrased slightly differently.

Proof of 5.2.6. Let $u = v$ be a non-tautological law satisfied by S, with $|u|$, the length of u, being minimal. If $|u| = |v| = 1$, then the law is of the form $x = y$ (where x and y are variables), and S must be the trivial group. If $|u| = 1$, $|v| > 1$, then put all variables in u, v equal to one another and use the cancellative property to deduce that S is again a group (of finite exponent). Groups are manifestly right- and left-reversible.

If $|u|, |v| > 1$, write $u = xu'$ and $v = yv'$, so that $|u| = 1 + |u'|$ and $|v| = 1 + |v'|$. Observe that if x and y were the same variable, cancellation would show $u' = v'$ to be a law, which would contradict the minimality of $|u|$. Substitute any two distinct elements of S for x and y (and arbitrary elements for any other letters in u and v) to show that they have a common right multiple. Therefore the semigroup S is left-reversible. Reason similarly to show the right-reversibility of S. $\boxed{5.2.6}$

Lemmata 5.2.6 and 5.2.5 together imply the following result:

Corollary 5.2.7. *Let G be a group. Let S be a subsemigroup of G. Suppose that S satisfies a non-tautological semigroup law. Then the subgroup of G generated by S is isomorphic to the universal group of S.* $\boxed{5.2.7}$

Remember that Corollary 5.2.7 does not hold for arbitrary subsemigroups of groups (see Example 0.9.2).

Theorem 5.2.8. *A coherent group that satisfies a non-tautological semigroup law is Malcev coherent.*

Proof of 5.2.8. Let G be a coherent group that obeys a non-tautological semigroup law. Let S be a finitely generated subsemigroup of G. Let H be the subgroup of G generated by S. The group H is finitely generated and thus, by the coherence of G, is finitely presented. Corollary 5.2.7 shows that H coincides with the universal group of S: the universal group of S is finitely presented. Therefore S admits a finite Malcev presentation by Corollary 1.3.2. Thus G is Malcev coherent. $\boxed{5.2.8}$

5.3. NILPOTENT AND VIRTUALLY NILPOTENT GROUPS

RECALL the definition of a nilpotent group:

Definition 5.3.1. A group G is *nilpotent* if it possesses a series of normal subgroups

$$\{1_G\} = H_0 \leq H_1 \leq \ldots \leq H_n = G,$$

such that, for each $i \in \{1, \ldots, n\}$, the factor group H_i/H_{i-1} lies in the centre of G/H_{i-1}. The shortest length n of such a series is the *nilpotency class* of the group G.

The only nilpotent group of class 0 is the trivial group; the groups of nilpotency class 1 are the abelian groups. Therefore nilpotent groups are a generalization of abelian groups, or abelian groups are a special case of nilpotent groups. Like the class of abelian groups, the class of nilpotent groups is closed under taking subgroups (Robinson 1996, Theorem 5.1.4). In addition, every finitely generated nilpotent group is finitely presented. Consequently, nilpotent groups are coherent.

The goal of this section is to show that every finitely generated subsemigroup of a virtually nilpotent group admits a finite Malcev presentation. [This result first appeared — albeit with a rather different exposition — as Theorem 1 of Cain et al. (2005a).] Although every finitely generated subsemigroup of an abelian group has a finite 'ordinary' presentation by Rédei's Theorem (see Theorem 5.4.1), the same is not generally true of nilpotent groups:

Example 5.3.2. Let H be the Heisenberg group:

$$H = \mathrm{Gp}\langle p, q, r \mid (qp, pqr), (rq, qr), (pr, rp)\rangle.$$

The group H is nilpotent — indeed, it is the free nilpotent group of class 2 and rank 2.

Every element of H can be represented by a unique word of the form $p^\alpha q^\beta r^\gamma$, where $\alpha, \beta, \gamma \in \mathbb{Z}$; see, for example, Johnson (1997, Section 5.2). Clearly, any word in $\{p, q, r\}^+$ can be transformed to one in this normal form using only the relations (qp, pqr), (rq, qr), and (pr, rp). The positive subsemigroup T of H consisting of elements represented by words in $\{p, q, r\}^+$ is therefore presented by

$$T = \mathrm{Sg}\langle p, q, r \mid (qp, pqr), (rq, qr), (pr, rp)\rangle. \tag{1}$$

Identify H — and so T — with words in normal form.

Let $A = \{a, b\}$ be an alphabet representing elements of T in the following way:

$$\overline{a} = p,\ \overline{b} = q.$$

Let S be the subsemigroup of T generated by $\overline{A}$. Observe that for all $n \in \mathbb{N}$,

$$\overline{ab^{2n}a} = pq^{2n}p = p^2q^{2n}r^{2n} = q^np^2q^n = \overline{b^na^2b^n}.$$

Therefore $(ab^{2n}a, b^na^2b^n)$ is a valid relation in S for all $n \in \mathbb{N}$. [Here is an alternative line of reasoning: the Neumann–Taylor law $xyzyx = yxzxy$ (that is, $q_c(x,y,z) = q_c(y,x,z)$ with $c = 2$) is obeyed by $\mathrm{Sg}\langle \overline{A} \cup \{1_H\}\rangle$ (see Theorem 5.3.4 below). Substituting $\overline{a}$ for x, $\overline{b^n}$ for y, and 1_H for z yields $\overline{ab^{2n}a} = \overline{b^na^2b^n}$.]

An easy argument shows that for any proper subword w of $ab^{2n}a$, the element $\overline{w}$ can be factorized into elements of $\overline{A}$ in only one way. For example, the subword ab^{2n} represents pq^{2n}. Since the relations in (1) do not change the numbers of letters p or q, the occurrence of p can only arise from a single letter a. Similarly, the q^{2n} requires $2n$ letters b. Therefore, any word representing pq^{2n} must be a rearrangement of ab^{2n}. However, any letters b to the left of the a would lead at least one r being present in the normal form. A similar arguments applies to the subword $b^{2n}a$. Therefore no valid relations can be applied to a proper subword of $ab^{2n}a$, and hence any presentation for S on this set of generators must have $ab^{2n}a$ as one side of a defining relation for each $n \in \mathbb{N}$. Thus S is not finitely presented.

Before proceeding with the body of this section, make one final observation: 'virtually nilpotent' and 'nilpotent-by-finite' are synonymous: Proposition 0.10.2 implies that a group having a nilpotent subgroup of finite index possesses a normal such subgroup.

Section 5.2 remarked that Malcev (1953) and Neumann & Taylor (1963) characterized semigroups embeddable into nilpotent groups in terms of cancellativity and satisfaction of a semigroup law. Detailed discussion of their results is now appropriate.

Both the Malcev and Neumann–Taylor classifications depend on sequences of words over an alphabet $\{x, y, z_1, z_2, \ldots\}$. For the Malcev classification, define

$$X_0 = x, Y_0 = y, \text{ and } X_{i+1} = X_i z_{i-1} Y_i,\ Y_{i+1} = Y_i z_{i-1} X_i \text{ for } i \in \mathbb{N},$$

and for the Neumann–Taylor classification, let

$$q_1(x,y,z) = xy, \text{ and } q_{i+1}(x,y,z) = q_i(x,y,z) z_i q_i(y,x,z) \text{ for } i \in \mathbb{N}.$$

Theorem 5.3.3 (Malcev 1953). *A semigroup S embeds in a nilpotent group of class c if and only if S is cancellative and satisfies the semigroup law $X_c = Y_c$.* $\boxed{5.3.3}$

Theorem 5.3.4 (Neumann & Taylor 1963). *A semigroup S embeds in a nilpotent group of class c if and only if S is cancellative and satisfies the semigroup law $q_c(x,y,z) = q_c(y,x,z)$.* $\boxed{5.3.4}$

Theorems 5.3.3 and 5.3.4 both imply that the semigroups embeddable into abelian groups are precisely the commutative cancellative semigroups. [This is obvious for the Neumann–Taylor result; Malcev's law $xzy = yzx$ is also clearly satisfied by every commutative semigroup.] An alternative proof of the group-embeddability of commutative cancellative semigroups proceeds by using the same ordered pair construction as one uses to embed an integral domain into its field of fractions (see, for example, Fraleigh 1998, Section 5.4). As Clifford & Preston (1961, Section 1.10) observe, the reasoning is actually simpler because the operation of addition is not a concern.

Armed with Theorems 5.3.3 and 5.3.4 and the tools of Subsection 5.2.1, the Malcev coherence of virtually nilpotent groups now follows easily:

Theorem 5.3.5. *Virtually nilpotent groups are Malcev coherent.*

Proof of 5.3.5. Let G be a virtually nilpotent group. Then G is a finite extension of a nilpotent group N, which, by Theorem 5.3.3 or 5.3.4, satisfies a non-tautological law. By Proposition 5.2.2, G satisfies a non-tautological law. Furthermore, G is a coherent group by Proposition 1.6.1 and the coherence of nilpotent groups. The group G therefore satisfies the hypotheses of Theorem 5.2.8 and so is Malcev coherent. $\boxed{5.3.5}$

Theorem 5.3.5 is simply an application of Theorem 5.2.8. Noting this, one immediately asks the following questions:

Open Problem 5.3.6. What other coherent groups satisfy a non-tautological semigroup law?

[Olshanskii & Storozhev (1996) proved the existence of finitely generated groups that are not nilpotent-by-finite but which satisfy non-tautological semigroup laws.]

Open Problem 5.3.7. What other coherent groups have the property that the universal group of a subsemigroup coincides with the subgroup it generates?

The class of polycyclic groups properly contains the class of finitely generated nilpotent groups whilst retaining many of the pleasant properties of the latter. In particular, polycyclic groups are coherent. However, as is proven in Section 7.3, there exist polycyclic groups that are not Malcev coherent.

5.4. ABELIAN GROUPS

5.4.1. Generators and presentations for subsemigroups

ALL SUBGROUPS of finitely generated abelian groups are finitely generated. However, the same is not true of their *subsemigroups*: for example, let S be the subsemigroup of the free abelian group of rank 2 generated by

$$X = \{(1, n) : n \in \mathbb{N}\}.$$

Then S is not finitely generated, for no member of X is decomposable in S: none can be expressed as products of other elements of S.

However, subsemigroups of abelian groups — and commutative semigroups generally — do have the following elegant property:

Theorem 5.4.1 (Rédei's Theorem). *Every finitely generated commutative semigroup is finitely presented.*

Proof of 5.4.1. See Rédei (1963) for the original proof. Freyd (1968) deduces this result as a consequence of the Hilbert Basis Theorem. An elementary proof is due to Grillet (1993). $\boxed{5.4.1}$

Therefore every finitely generated commutative semigroup that embeds in a group trivially has a finite Malcev presentation (Corollary 1.3.5). It therefore follows immediately from Rédei's Theorem that abelian groups are Malcev coherent.

By Theorem 5.3.5, virtually abelian groups are also Malcev coherent. However, unlike abelian groups, they can contain finitely generated subsemigroups that do not admit finite ordinary presentations (see Example 5.5.1 and Proposition 5.5.7).

5.4.2. Automatism of subsemigroups

All finitely generated abelian groups are automatic (Epstein et al. 1992, Section 4.1). Campbell et al. (2001, Question 4.6) asked whether all finitely generated commutative semigroups were automatic; Hoffmann & Thomas (2002) answered this question negatively by proving that the semigroup

$$\begin{aligned}\mathrm{Sg}\langle a,b,x,y \mid\ & (a^2x,bx),(ay,b^2y),\\ & (ab,ba),(ax,xa),(ay,ya),(bx,xb),(by,yb),(xy,yx)\rangle \qquad (2)\end{aligned}$$

does not admit an automatic structure.

However, finitely generated subsemigroups of abelian groups — which, as observed in Section 5.3, are precisely the finitely generated commutative cancellative semigroups — are automatic:

Theorem 5.4.2. *Let A be a finite alphabet representing a generating set for a subsemigroup S of an abelian group. Then S admits an automatic structure (A, K).*

Proof of 5.4.2. Suppose $A = \{a_1, \ldots, a_n\}$. Represent elements of S using tuples: identify the tuple $(\alpha_1, \ldots, \alpha_n)$ with the element $\overline{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}$. Define the ShortLex ordering on these tuples by

$$\begin{aligned}(\alpha_1,\ldots,\alpha_n) \prec_{\mathrm{SL}} (\beta_1,\ldots,\beta_n) \iff & \sum_{i=1}^n \alpha_i < \sum_{i=1}^n \beta_i, \text{ or}\\ & \Bigg[\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i\\ & \qquad \text{and } (\alpha_1,\ldots,\alpha_n) \sqsubset_{\mathrm{L}} (\beta_1,\ldots,\beta_n)\Bigg],\end{aligned}$$

where $\sqsubset_{\mathrm{L}}$ is the lexicographical order of tuples: $(\alpha_1,\ldots,\alpha_n) \sqsubset_{\mathrm{L}} (\beta_1,\ldots,\beta_n)$ if the leftmost non-zero coördinate of $(\beta_1-\alpha_1,\ldots,\beta_n-\alpha_n)$ is positive.

Rédei's Theorem 5.4.1 asserts that S is finitely presented. An approach to this theorem found in Rosales & García-Sánchez (1999, Chapter 5) (which is a modification of Grillet's (1993) proof) shows that the semigroup S is isomorphic to

$$[(\mathbb{N}\cup\{0\})^n - \{(0,\ldots,0)\}] / \{(u_1,v_1),\ldots,(u_n,v_n)\}^{\#},$$

where $u_i \prec_{\mathrm{SL}} v_i$, and such that the ShortLex-minimal representative of $\overline{w} \in (\mathbb{N} \cup \{0\})^n - \{(0,\ldots,0)\}$ can be found by repeatedly replacing w by $w - v_i + u_i$ whenever every coördinate of $w - v_i$ is non-negative. (Addition is performed componentwise on tuples.)

Since the ShortLex order is compatible with the operation (that is, for all $x \in S$, $u \prec_{\mathrm{SL}} v \implies u + x \prec_{\mathrm{SL}} v + x$), the set of ShortLex-minimal elements is simply

$$M = \{w \in (\mathbb{N}\cup\{0\})^n - \{(0,\ldots,0)\} : w - v_i \text{ is not in } (\mathbb{N}\cup\{0\})^n \text{ for any } i\}.$$

Let

$$K = \{a_1^{\alpha_1}\cdots a_n^{\alpha_n} : (\alpha_1,\ldots,\alpha_n) \in M\}.$$

Since the number of v_i is finite, a finite state automaton can check whether a word $a_1^{\alpha_1}\cdots a_n^{\alpha_n}$ lies in K. Therefore K is regular.

[Henceforth the proof follows that of Theorem 4.3.1 of Epstein et al. (1992).]

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$ be tuples. If $\beta_i \leq \alpha_i$ for every i, then α is said to be *contained* in β; this is denoted $\alpha \ll \beta$. [Notice that M is closed under $\ll$.] If $\alpha \ll \beta$, then the complement of α in β is the tuple $\beta - \alpha$. If α and β are distinct tuples representing the same element of S, then $\alpha \sim \beta$ is a *relation*. A relation is *minimal* if it is minimal with respect to the containment order

$$(\alpha \sim \beta) \ll (\alpha' \sim \beta') \iff (\alpha \ll \beta) \wedge (\alpha' \ll \beta').$$

Let λ be the tuple $(\ldots, 0, 1, 0, \ldots)$ corresponding to some letter a_i. Let α and β be tuples in M such that $(\alpha + \lambda) \sim \beta$. Let $\gamma \sim \delta$ be a minimal relation contained in $(\alpha + \lambda) \sim \beta$. Let γ' be the complement of γ in $(\alpha + \lambda)$ and δ' be the complement of δ in β. As the semigroup S is cancellative, $\gamma' \sim \delta'$ is a relation. At least one of γ and γ' is contained in α. If $\gamma \ll \alpha$, then γ and δ are both in M (by the closure of M under $\ll$), which is a contradiction since M maps bijectively onto S. So $\gamma' \ll \alpha$. Therefore γ' and δ' both lie in M, and so $\gamma' = \delta'$. So α and β differ only by the minimal relation $\gamma \sim \delta$. The number of such relations is bounded by Lemma 4.3.2 of Epstein et al. (1992). So the paths labelled by the words in K corresponding to α and β fellow travel.

Therefore, since the tuples α and β were arbitrary, (A, K) has the fellow traveller property. Theorem 2.3.2 thus implies that (A, K) is an automatic structure for the semigroup S. $\boxed{5.4.2}$

The upshot of Theorem 5.4.2 is the following:

Corollary 5.4.3. *Abelian groups are locally automatic.* $\boxed{5.4.3}$

5.5. VIRTUALLY ABELIAN GROUPS

IN LIGHT OF the local automatism of abelian groups (Corollary 5.4.3); that their finitely generated subsemigroups are finitely presented (Rédei's Theorem 5.4.1); and the Malcev coherence of virtually abelian groups (Theorem 5.3.5); one naturally asks whether local automatism and the finite presentability of all finitely generated subsemigroups extends from abelian to virtually abelian groups. The present section is devoted to a single example, due to Cain (2005*b*), that establishes that neither of these extensions holds.

Example 5.5.1. Let $\mathcal{S}_8$ be the symmetric group on eight elements. Let $\mathbb{Z}^8$ be the direct product of eight copies of the integers under addition. View elements of $\mathbb{Z}^8$ as octuples of integers. Let G be the semidirect product $\mathcal{S}_8 \ltimes \mathbb{Z}^8$, where $\mathcal{S}_8$ acts (on the right) by permuting the coördinates of elements of $\mathbb{Z}^8$. (Index the $\mathbb{Z}$-coördinates from 1 at the left to 8 at the right.) The abelian normal subgroup $\mathbb{Z}^8$ of G has index $8!$, so G is a virtually abelian group.

Let $A = \{a, b, c, d, e, f, g, h\}$ be an alphabet representing elements of G in the following way:

$$\begin{aligned}
&\overline{a} = [(1\ 3),\ (0,1,1,0,0,0,1,0)], && \\
&\overline{b} = [\mathrm{id},\ (0,0,1,0,0,0,0,0)], && \overline{f} = [(1\ 5)(2\ 6),\ (0,0,0,0,1,1,2,0)], \\
&\overline{c} = [(1\ 3)(2\ 4),\ (1,0,0,0,0,0,1,1)], && \overline{g} = [\mathrm{id},\ (0,0,0,0,1,1,0,0)], \\
&\overline{d} = [\mathrm{id},\ (0,0,0,1,0,0,0,0)], && \overline{h} = [(1\ 5)(2\ 6),\ (1,1,0,0,0,0,0,2)]. \\
&\overline{e} = [(2\ 4),\ (0,1,0,0,0,0,0,1)], &&
\end{aligned}$$

Let S be the subsemigroup of G generated by $\overline{A}$.

This example first appeared in Cain (2005*b*), although the proof therein corresponding to Proposition 5.5.2 shows only that the semigroup S is not *synchronously* automatic.

Proposition 5.5.2. *The semigroup S is not asynchronously automatic.*

Proof of 5.5.2. Let $A' = \{a, c, e, f, h\}$. Observe that only generators in $\overline{A'}$ possess non-identity $\mathcal{S}_8$-components, and that these are precisely the generators that have non-zero seventh or eighth $\mathbb{Z}$-coördinates. Note further that the seventh and eighth $\mathbb{Z}$-coördinates are not affected by any of the $\mathcal{S}_8$-components in $\overline{A}$, and that there are no negative integers amongst the $\mathbb{Z}$-coördinates.

Now, for any $\alpha \in \mathbb{N} \cup \{0\}$,

$$\begin{aligned}
\overline{ab^\alpha cd^\alpha e} &= [(1\ 3),\ (0,1,1,0,0,0,1,0)][\mathrm{id},\ (0,0,\alpha,0,0,0,0,0)]\overline{cd^\alpha e} \\
&= [(1\ 3),\ (0,1,\alpha+1,0,0,0,1,0)][(1\ 3)(2\ 4),\ (1,0,0,0,0,0,1,1)]\overline{d^\alpha e} \\
&= [(2\ 4),\ (\alpha+2,0,0,1,0,0,2,1)][\mathrm{id},\ (0,0,0,\alpha,0,0,0,0)]\overline{e} \\
&= [(2\ 4),\ (\alpha+2,0,0,\alpha+1,0,0,2,1)][(2\ 4),\ (0,1,0,0,0,0,0,1)] \\
&= [\mathrm{id},\ (\alpha+2,\alpha+2,0,0,0,0,2,2)],
\end{aligned}$$

and

$$\begin{aligned}
\overline{fg^\alpha h} &= [(1\ 5)(2\ 6),\ (0,0,0,0,1,1,2,0)][\mathrm{id},\ (0,0,0,0,\alpha,\alpha,0,0)]\overline{h} \\
&= [(1\ 5)(2\ 6),\ (0,0,0,0,\alpha+1,\alpha+1,2,0)][(1\ 5)(2\ 6),\ (1,1,0,0,0,0,0,2)] \\
&= [\mathrm{id},\ (\alpha+2,\alpha+2,0,0,0,0,2,2)].
\end{aligned}$$

So $\overline{ab^\alpha cd^\alpha e} = \overline{fg^\alpha h}$ for all $\alpha \in \mathbb{N} \cup \{0\}$.

Lemma 5.5.3. *For each $\alpha \in \mathbb{N} \cup \{0\}$, the elements of S represented by $ab^\alpha cd^\alpha$ and fg^α are represented by those words alone.*

Proof of 5.5.3. Suppose w represents

$$s = \overline{fg^\alpha} = [(1\ 5)(2\ 6),\ (0,0,0,0,\alpha+1,\alpha+1,2,0)].$$

Since the seventh and eighth $\mathbb{Z}$-coördinates are 2 and 0, the only letters from A' in w are either two letters a or one letter f. The first option is impossible, since the $\mathcal{S}_8$-component of $\overline{w}$ would then be the identity permutation. Since the third and fourth $\mathbb{Z}$-coördinates of s are zero, and these coördinates are unaffected by the $\mathcal{S}_8$-component of $\overline{f}$, the rest of w must consist of letters g. So w is a rearrangement of fg^β for some β. Since the first two $\mathbb{Z}$-coördinates of s are 0, no letters g can precede the letter f. So $w = fg^\beta$. Considering the fifth and sixth $\mathbb{Z}$-coördinates of s shows that $\beta = \alpha$. Therefore fg^α is the unique word over A representating s.

Now suppose that v represents

$$t = \overline{ab^\alpha cd^\alpha} = [(2\ 4),\ (\alpha+2, 0, 0, \alpha+1, 0, 0, 2, 1)].$$

The first task is to determine what letters from A' appear in v. The letter h is ruled out by the last $\mathbb{Z}$-coördinate of t being 1. If an f is present, the only other letter from A' must be e, since the last two $\mathbb{Z}$-coördinates of t are 2 and 1. However, this gives the wrong $\mathcal{S}_8$-component. The other possibilities are a, a, and e; or a and c. Suppose the former. If the letter e is the last of these three letters, then $\overline{v}$ has a non-zero second $\mathbb{Z}$-coördinate. If one of the letters a is the last of the three, then $\overline{v}$ has non-zero second and third $\mathbb{Z}$-coördinate. So the letters from A' must be a and c.

Since the fifth and sixth $\mathbb{Z}$-coördinates of t are 0, and these coördinates are unaffected by the $\mathcal{S}_8$ components of $\overline{a}$ or $\overline{c}$, no letters g can be present in v. So v is a rearrangement of $acb^\beta d^\gamma$ for some $\beta, \gamma \in \mathbb{N} \cup \{0\}$. The letter a must precede the letter c, for otherwise $\overline{v}$ would have non-zero second and third $\mathbb{Z}$-coördinates. Similarly, the third $\mathbb{Z}$-coördinate of s being zero forces the letters b to lie between the letter a and the letter c (since the $\mathcal{S}_8$-components of $\overline{a}$ and $\overline{c}$ together send the third $\mathbb{Z}$-coördinate to itself). The letters d must lie to the right of the letter c, since otherwise $\overline{v}$ would have non-zero second $\mathbb{Z}$-coördinate. So $w = ab^\beta cd^\gamma$. The values of the first and fourth $\mathbb{Z}$-coördinates of t together force $\beta = \gamma = \alpha$. So $ab^\alpha cd^\alpha$ is the unique word over A representing t. $\boxed{5.5.3}$

Suppose S is asynchronously automatic. Then, by Theorem 2.4.7, S has an asynchronous automatic structure (A, L). By definition, the relations L_e and L_h are both rational. Theorem A.6.5 shows that the relation

$$\begin{aligned} L_e \circ L_h^{-1} &= \{(u, w) : u, w \in L, \overline{ue} = \overline{w}\} \circ \{(w, v) : w, v \in L, \overline{vh} = \overline{w}\} \\ &= \{(u, v) : u, v \in L, \overline{ue} = \overline{vh}\} \end{aligned}$$

is also rational. Let N be the number of states in an asynchronous automaton $\mathcal{A}$ recognizing $L_e \circ L_h^{-1}$.

For each $\alpha \in \mathbb{N} \cup \{0\}$, let u_α and v_α be representatives in L of the elements $\overline{ab^\alpha cd^\alpha}$ and $\overline{fg^\alpha}$, respectively. Since these elements have unique representatives over A by Lemma 5.5.3, it is clear that $u_\alpha = ab^\alpha cd^\alpha$ and $v_\alpha = fg^\alpha$. By its definition, the relation $L_e \circ L_h^{-1}$ contains (u_α, v_α) for all $\alpha \in \mathbb{N} \cup \{0\}$.

Fix $\alpha > N$. Consider the automaton $\mathcal{A}$ reading (u_α, v_α), and the states it enters immediately after reading each of the letters b from the word u_α. As the number of letters b exceeds N, the automaton enters the same state after reading two different letters b. Let ab^β and ab^γ, where $\beta < \gamma$, be the prefixes of u_α up to and including these two different letters b. Let v' be the subword of v_α read by $\mathcal{A}$ between having read ab^β from u_α and reaching ab^γ. The subword v' is either fg^η or g^η for some $\eta \in \mathbb{N} \cup \{0\}$. (The former possibility arises because $\mathcal{A}$ may not read any letters from v_α whilst reading ab^β from u_α.) As the automaton $\mathcal{A}$ is in the same state immediately before and after reading $(b^{\gamma-\beta}, v')$, it must accept the pair formed by reading $(b^{2(\gamma-\beta)}, (v')^2)$ instead of $(b^{\gamma-\beta}, v')$.

Suppose that $v' = fg^\eta$. Then $\mathcal{A}$ accepts

$$(ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha, fg^\eta fg^\alpha).$$

So, by the definition of the relation $L_e \circ L_h^{-1}$,

$$\overline{ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha e} = \overline{fg^\eta fg^\alpha h}.$$

This is a contradiction, since the $\mathcal{S}_8$-component of the left-hand side is the identity permutation and that of the right-hand side is $(1\ 5)(2\ 6)$.

Therefore suppose that $v' = g^\eta$. Then $\mathcal{A}$ accepts

$$(ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha, fg^{\alpha+\eta}),$$

whence, again by the definition of $L_e \circ L_h^{-1}$,

$$\overline{ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha e} = \overline{fg^\eta g^\alpha h}.$$

This too is a contradiction, since $\gamma - \beta$ is at least 1, but

$$\overline{ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha e} = [\mathrm{id},\ (\alpha - \beta + \gamma + 2, \alpha + 2, 0, 0, 0, 0, 2, 2)],$$

while

$$\overline{fg^\eta g^\alpha h} = [\mathrm{id},\ (\alpha + \eta + 2, \alpha + \eta + 2, 0, 0, 0, 0, 2, 2)].$$

Therefore S is not asynchronously automatic. $\boxed{5.5.2}$

The consequences of Proposition 5.5.2 go beyond showing that virtually abelian groups are not in general locally automatic. Several authors asked whether the automatism of the universal group of a finitely generated group-embeddable semigroup implies the automatism of the semigroup itself (see Cain et al. 2005*a*, Question 7.4; Hoffmann 2001, Question 4.22; Kambites 2003, Question 6.8).

Examples in favour of this implication include: free groups and semigroups; braid groups and semigroups (Epstein et al. 1992, Chapter 9); abelian groups and their subsemigroups (Theorem 5.4.2).

However, Proposition 5.5.2 shows that the implication does not hold: Let H be the subgroup of G generated by S. The group H is a subgroup of the virtually abelian group G and is therefore itself virtually abelian by Proposition 0.10.3. It is finitely generated, and so is automatic by Section 4.1 of Epstein et al. (1992).

Furthermore, as it contains $\mathbb{Z}^8$ as an index 8! normal subgroup, the group G satisfies the non-tautological semigroup law $x^{8!}y^{8!} = y^{8!}x^{8!}$ by Proposition 5.2.2. Corollary 5.2.7 applies to show that the subgroup H coincides with the universal group of S. Therefore S is a finitely generated non-automatic semigroup that is embeddable in a group but whose universal group is automatic.

Theorem 5.5.4. *The automatism of the universal group of a finitely generated group-embeddable semigroup does not in general imply the automatism of the semigroup itself.* $\boxed{5.5.4}$

[A similar question asked whether the automatism of a group implied the automatism of its positive subsemigroups. (Recall that a positive subsemigroup is one generated by a group generating set.) Proposition 5.5.2 of course implies that the answer to this second question is also 'no', but this negative answer was established earlier than the answer to the original question by an example of Cain et al. (2005*a*, Section 7). This earlier example uses the theory of Malcev presentations and is found in Section 6.2.]

The following result is a consequence of Proposition 5.5.2 and Corollary 5.4.3:

Theorem 5.5.5. *The class of locally automatic groups is not closed under constructing finite extensions. Indeed, a finite extension of a locally automatic group may not be locally asynchronously automatic.* $\boxed{5.5.5}$

Contrasting Theorem 5.5.5 with the situation for groups with all finitely generated subgroups automatic described in Proposition 2.4.17 yields the next result:

Theorem 5.5.6. *The class of locally automatic groups is properly contained in the class of groups all of whose finitely generated subgroups are automatic.* $\boxed{5.5.6}$

The last consequence drawn from Proposition 5.5.2 is that the property of having all finitely generated subsemigroups finitely presented does not extend from abelian to virtually abelian groups.

Proposition 5.5.7. *The semigroup S is not finitely presented.*

Proof of 5.5.7. For all $\alpha \in \mathbb{N} \cup \{0\}$, the relation $(ab^\alpha cd^\alpha e, fg^\alpha h)$ is valid in S. Lemma 5.5.3 shows that $\overline{fg^\alpha}$ is represented by the word fg^α alone. Similar reasoning shows that, for each $\alpha \in \mathbb{N} \cup \{0\}$, the element of S represented by $g^\alpha h$ is represented by that word alone.

Therefore no non-trivial relation in S can be applied to a proper subword of $fg^\alpha h$, and so in any presentation for S on the generating set $\overline{A}$, each word $fg^\alpha h$ must appear as one side of a defining relation. Therefore S is not finitely presented. $\boxed{5.5.7}$

Open Problem 5.5.8. Does there exist a *finitely presented* group-embeddable non-automatic semigroup whose universal group is automatic?

CHAPTER SIX

FREE PRODUCTS & THEIR SUBSEMIGROUPS

But when the words free, *and* liberty, *are applied to any thing but* bodies, *they are abused; for that which is not subject to motion, is not subject to impediment...*
— Thomas Hobbes,
Leviathan (1651), ch. xxi

6.1. INTRODUCTION

THE CLASS OF COHERENT GROUPS is closed under free products (Proposition 1.6.1). It is therefore natural to examine whether the same is true for the class of Malcev coherent groups.

[Throughout this chapter, 'free product' refers to the group or monoid free product, not to the semigroup free product, which is a different construction (Howie 1995, Section 8.2). Refer to Lyndon & Schupp (1977, Section IV.1) for background reading on free products.]

Section 6.2 considers a particular finitely generated subsemigroup of the free product of a free group and an abelian group. Three noteworthy theorems follow from the failure of this semigroup to admit a finite Malcev presentation. Section 6.3 shows that by replacing the free group with a free monoid, one recovers Malcev coherence via automatism.

6.2. FREE PRODUCT OF FREE GROUPS AND ABELIAN GROUPS

FREE GROUPS and abelian groups are coherent; the free product of a free group and an abelian group is therefore coherent by Proposition 1.6.1. Free groups are Malcev coherent by Theorem 4.3.1; abelian groups by Theorem 5.3.5. One therefore naturally asks whether the free product of a free group and an abelian group is Malcev coherent, and more generally whether the class of Malcev coherent groups is closed under free products. The following example, which first appeared in Section 5 of Cain et al. (2005*a*), provides a negative answer to both questions:

Example 6.2.1. Let F be the free product $\mathrm{FG}(x, y, z, s, t) * (\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z})$. Identify elements of F with alternating products of elements of $\mathrm{FG}(x, y, z, s, t)$ (viewed as reduced words) and of $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ (viewed as triples of integers).

Let $A = \{a, b, c, d, e, f, g, h, i, j\}$ be an alphabet, and let this alphabet represent elements of F in the following way:

$$\begin{aligned} \overline{a} &= x^2 y, & \overline{f} &= x^2 s, \\ \overline{b} &= y^{-1}(1,0,1)y, & \overline{g} &= s^{-1}(1,0,0)s, \\ \overline{c} &= y^{-1}z, & \overline{h} &= s^{-1}t, \\ \overline{d} &= z^{-1}(0,1,0)z, & \overline{i} &= t^{-1}(0,1,1)t, \\ \overline{e} &= z^{-1}x^2, & \overline{j} &= t^{-1}x^2. \end{aligned}$$

Let S be the subsemigroup of F generated by $\overline{A}$.

Lemma 6.2.2. *The semigroup S is presented by $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where*

$$\mathcal{R} = \{(ab^\alpha cd^\alpha e, fg^\alpha hi^\alpha j) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

Proof of 6.2.2. Let $\alpha \in \mathbb{N} \cup \{0\}$. Then

$$\begin{aligned} \overline{ab^\alpha cd^\alpha e} &= x^2 y (y^{-1}(1,0,1)y)^\alpha y^{-1} z (z^{-1}(0,1,0)z)^\alpha z^{-1} x^2 \\ &= x^2 y y^{-1}(\alpha,0,\alpha) y y^{-1} z z^{-1}(0,\alpha,0) z z^{-1} x^2 \\ &= x^2 (\alpha,0,\alpha)(0,\alpha,0) x^2 \\ &= x^2 (\alpha,\alpha,\alpha) x^2 \\ &= x^2 (\alpha,0,0)(0,\alpha,\alpha) x^2 \\ &= x^2 s s^{-1}(\alpha,0,0) s s^{-1} t t^{-1}(0,\alpha,\alpha) t t^{-1} x^2 \\ &= x^2 s (s^{-1}(1,0,0)s)^\alpha s^{-1} t (t^{-1}(0,1,1)t)^\alpha t^{-1} x^2 \\ &= \overline{fg^\alpha hi^\alpha j}. \end{aligned}$$

So all of the relations in $\mathcal{R}$ hold in S.

Define a set of normal forms N to be the set of all words in A^+ that do not contain $fg^\alpha hi^\alpha j$ for any $\alpha \in \mathbb{N} \cup \{0\}$. Every element of S is represented by at least one element of N since a word over A^+ can be rewritten to one in N using the relations $\mathcal{R}$: such rewriting cannot continue indefinitely since each step decreases the number of letters from $\{f, g, h, i, j\}$ present.

Let $u = u_1 \cdots u_p$, $v = v_1 \cdots v_q$ ($u_i, v_i \in A$ for all i) be distinct words in the set of normal forms N, and suppose they represent the same element of S. Without loss of generality, suppose that $u_1 \neq v_1$ and that u precedes v in the lexicographic ordering based on $a \prec b \prec c \prec \ldots \prec j$.

Consider which letters u_1 and v_1 may be. Observe that the element x^{-1} appears in no word in $\overline{A}$, nor does a negative number appear in any coordinate of a tuple. Letters x and tuples therefore cannot be cancelled. The possibilities for u_1 and v_1 marked 1 in Table 6.1 are immediately excluded for this reason.

		v_1								
		b	c	d	e	f	g	h	i	j
u_1	a	1	2	1	1		1	2	1	1
	b		3	1	1	1	1	2	1	1
	c			2	2	2	2	2	2	2
	d				1	1	1	2	1	1
	e					1	1	2	1	1
	f						1	2	1	1
	g							2	1	1
	h								2	2
	i									1

Table 6.1. Possibilities for u_1 and v_1 in the proof of Lemma 6.2.2.

If $u_1 = c$, then the only way the occurrence of z can be cancelled from $\overline{c}$ is if $u_2 = d$ or $u_2 = e$, which leaves the y^{-1} unaffected. Reasoning parallel to this, together with the observation in the last paragraph, excludes the possibilities marked 2 in Table 6.1.

Suppose that $u_1 = b$ and $v_1 = c$. Then the occurrence of y must be cancelled from $\overline{b}$ by a string of β letters b and one c. So

$$\overline{u} = y^{-1}(\beta+1, 0, \beta+1)z\overline{u_{\beta+3}\cdots u_p}$$
$$\overline{v} = y^{-1}z\overline{v_2\cdots v_q}.$$

However, the occurrence of z must now be cancelled from $\overline{v}$. Yet, using a string of letters d or e can only replace z with a tuple with 0 in the first component, followed by x or an uncancelled z. This is a contradiction, which eliminates the possibility marked 3 from Table 6.1.

Therefore $u_1 = a$ and $v_1 = f$. Either $u_2 = b$ or $v_2 = g$ in order to cancel the occurrence of y or s. Suppose the former; the latter case is analogous. Suppose that $u_2 = \ldots = u_{\alpha+1} = b$ and $u_{\alpha+2} \neq b$ for some α. Then

$$\overline{u} = x^2(\alpha, 0, \alpha)y\overline{u_{\alpha+2}\cdots u_p}.$$

To match the first component of this tuple, $v_2 = \ldots = v_{\alpha+1} = g$ and $v_{\alpha+2} \neq g$. So

$$\overline{v} = x^2(\alpha, 0, 0)s\overline{v_{\alpha+2}\cdots v_q}.$$

Since $u_{\alpha+2} \neq b$ and $v_{\alpha+2} \neq g$, either $u_{\alpha+2} = c$ or $v_{\alpha+2} = h$. Once again, suppose the first case; the other follows in the same way. So

$$\overline{u} = x^2(\alpha, 0, \alpha)z\overline{u_{\alpha+3}\cdots u_p}.$$

Repeating the reasoning above, with reference to the second component of the tuple rather than the first, shows that $v_{\alpha+2} = h$ and that $u_{\alpha+3}$ is the first of a string of β

letters d and $v_{\alpha+3}$ a string of letters i. This shows that

$$\overline{u} = x^2(\alpha, \beta, \alpha)\overline{z\cdots u_p},$$
$$\overline{v} = x^2(\alpha, \beta, \beta)\overline{t\cdots v_q}.$$

The only way the letters z and t can be cancelled is if the next letters of u and v are e and j, respectively. In order to match the third component of the tuple, therefore, $\alpha = \beta$. So v begins $fg^\alpha hi^\alpha j$, which contradicts the fact that N contains v.

The set N is therefore a set of unique normal forms for S, and so $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$ does indeed present S. $\boxed{6.2.2}$

The following result will be needed shortly:

Theorem 6.2.3 (Baumslag 1962). *Let J and L be finitely presented groups. Then $J *_K L$, their free product with amalgamated subgroup K, is finitely presented if and only if K is finitely generated.* $\boxed{6.2.3}$

Proposition 6.2.4. *The semigroup S does not admit a finite Malcev presentation.*

Proof of 6.2.4. By Lemma 6.2.2, S is presented by $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where $A = \{a, \ldots, j\}$ and

$$\mathcal{R} = \{(ab^\alpha cd^\alpha e, fg^\alpha hi^\alpha j) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

Proposition 1.3.1 shows that the universal group of S is $U = \mathrm{Gp}\langle A \mid \mathcal{R}\rangle$. The group U is [isomorphic to] the amalgamated free product

$$\mathrm{FG}(a, b, c, d, e) *_K \mathrm{FG}(f, g, h, i, j),$$

where $K \simeq \mathrm{Gp}\langle ab^\alpha cd^\alpha e, \alpha \in \mathbb{N} \cup \{0\}\rangle \simeq \mathrm{Gp}\langle fg^\alpha hi^\alpha j, \alpha \in \mathbb{N} \cup \{0\}\rangle$.

Lemma 6.2.5. *The amalgamated subgroup K is not finitely generated.*

Proof of 6.2.5. One proceeds by showing that the set

$$B = \{ab^\alpha cd^\alpha e : \alpha \in \mathbb{N} \cup \{0\}\}$$

is Nielsen-reduced (see Lyndon & Schupp 1977, p. 6) and therefore a basis for $K \leq \mathrm{FG}(a, b, c, d, e)$ (Lyndon & Schupp 1977, Proposition I.2.5).

If $\alpha > \beta$, then $ab^\alpha cd^\alpha e(ab^\beta cd^\beta e)^{-1} = ab^\alpha cd^{\alpha-\beta}c^{-1}b^{-\beta}a^{-1}$, with $\alpha - \beta > 0$. If $\alpha \neq \beta \neq \gamma$, then at least the middle c remains uncancelled in the product $ab^\alpha cd^\alpha e(ab^\beta cd^\beta e)^{-1}ab^\gamma cd^\gamma e$. Therefore B is Nielsen-reduced.

Therefore B is a basis for the amalgamated subgroup K, and so K cannot be finitely generated (Lyndon & Schupp 1977, Proposition I.2.7). $\boxed{6.2.5}$

By Theorem 6.2.3, the group U cannot be finitely presented. Therefore, by Corollary 1.3.2, the semigroup S does not admit a finite Malcev presentation. $\boxed{6.2.4}$

Example 6.2.1 yields several important results. Observing that free groups and abelian groups are Malcev coherent by Theorems 4.3.1 and 5.3.5 produces the first result:

Theorem 6.2.6. *The class of Malcev coherent groups is not closed under forming free products.* $\boxed{6.2.6}$

Proposition 1.6.1 asserts that the class of coherent groups *is* closed under forming free products. Consequently, the free product of a free group and an abelian group is coherent. Therefore F is an example of a coherent group that is not Malcev coherent. Conversely, Malcev coherent groups are always coherent (Proposition 1.6.3).

Theorem 6.2.7. *The class of Malcev coherent groups is properly contained in the class of coherent groups.* $\boxed{6.2.7}$

Applying Theorems 2.5.1 and 2.5.4 shows that the semigroup S is neither automatic nor asynchronously automatic. Yet S is a subsemigroup of F, a free product of a free group and an abelian group, which are locally automatic by Theorems 4.4.1 and 5.4.3.

Theorem 6.2.8. *The class of locally automatic groups is not closed under forming free products. Indeed, the free product of two locally automatic groups may not even be locally asynchronously automatic.* $\boxed{6.2.8}$

[Descalço (2002, Question 6.7) asks whether the class of locally automatic semigroups is closed under forming *semigroup* free products. Theorem 6.2.8, which deals with *group* free products, therefore does not answer Descalço's question.]

Finally, let H be the subgroup of F generated by S. Proposition 2.4.17 shows that H is automatic. Therefore S is an example of a positive subsemigroup of an automatic group that is not itself automatic.

Example 6.2.1 was the first known example of such a positive subsemigroup. However, the question of whether the automatism of the universal group of a semigroup that embeds in a group implies the automatism of the semigroup remained open until Cain (2005*b*) constructed the semigroup in Example 5.5.1. Although related to the problem regarding positive subsemigroups answered above, one could not hope to use an argument analogous to the one above to find a counterexample. If the universal group of a particular group-embeddable semigroup were automatic, it would be finitely presented (Epstein et al. 1992, Theorem 2.3.12). Corollary 1.3.2 would then imply that the semigroup had a finite Malcev presentation.

The free product of a non-abelian free group and the free abelian group of rank 3 is not Malcev coherent. The free product of a non-abelian free group with the free abelian group of rank 1 is again a free group and therefore Malcev coherent. The rank 2 case remains undecided:

Open Problem 6.2.9. Is the free product of a non-abelian free group and the free abelian group of rank 2 Malcev coherent?

[Notice that the free product of a free group and $\mathbb{Z} \times \mathbb{Z}$ is a one-relator group; see Open Problem 8.3.5.]

6.3. FREE PRODUCTS OF FREE MONOIDS AND ABELIAN GROUPS

EXAMPLE 6.2.1 shows that the free product of a free group and an abelian group is not in general Malcev coherent. Contrast this with the following result:

Theorem 6.3.1. *A [monoid] free product of a free monoid and an abelian group is locally automatic.*

[Cain et al. (2005*a*, Theorem 6) proved that every finitely generated subsemigroup of the free product of a free monoid and an abelian group is asynchronously automatic. The proof of Theorem 6.3.1 owes much to that earlier result, but is rather more complex.]

Proof of 6.3.1. Let $F = X^* * H$, where X is finite and H is an abelian group. View elements of F as alternating products of elements of H and *letters* of X. (Successive letters of X may be separated by $1_F \in H$.) Let A be a finite alphabet representing a subset of F. Add a symbol 1 to A with $\overline{1} = 1_F$. Let $A = A' \cup A''$, where $\overline{A''} \subseteq H$ and $\overline{A'} \subseteq F - H$. (Observe that $1 \in A''$.) Let M be the semigroup generated by $\overline{A}$.

Each element of M can be written in the form

$$z_0 x_1 z_1 \cdots x_n z_n, \tag{1}$$

where each z_i lies in H and each x_i is in X. For the purposes of this proof, the element z_0 is called an *H-prefix*; z_n is called an *H-suffix*; all other z_i are called *H-subwords*. Let S be the set of H-suffixes of elements of $\overline{A'}$; P the set of H-prefixes of elements of $\overline{A'}$; and Z the set of H-subwords elements of $\overline{A'}$. Formally add 1_H to each of P, S, and Z. An H-subword of an element of M may arise in two ways:

i.) as an H-subword of an element of $\overline{A'}$;

ii.) as a product of an H-suffix of an element of $\overline{A'}$, an element $\overline{w}$ where $w \in (A'')^*$, and an H-prefix of $\overline{A'}$.

Let (A'', K) be an automatic structure with uniqueness for $\mathrm{Sg}\langle\overline{A''}\rangle \subseteq H$. (Such an automatic structure exists by Theorem 5.4.2.)

Lemma 6.3.2. *Let $p, r \in S$ and $q, s \in P$. Then the language*

$$\begin{aligned} K_{pqs^{-1}r^{-1}}\delta_A &= \{(u,v) : u, v \in K, \overline{u}pq = \overline{v}rs\}\delta_A \\ &= \{(u,v) : u, v \in K, p\overline{u}q = r\overline{v}s\}\delta_A \end{aligned}$$

is regular. Furthermore, if $(u,v) \in K_{pqs^{-1}r^{-1}}$, then the paths $\widehat{u}$ and $\widehat{v}$ fellow-travel.

[This lemma, although easily proved, is not trivial: the elements p, q, r, and s may not lie in $\mathrm{Sg}\langle A''\rangle$, and thus the relations $K_x = \{(u,v) \in K \times K : \overline{u}x = \overline{v}\}$ (where $x \in \{p,q,r,s\}$) may be empty even if $K_{pqs^{-1}r^{-1}}$ is not.]

Proof of 6.3.2. If there are no words $w, w' \in K$ with $\overline{w}pq = \overline{w'}rs$, then $K_{pqs^{-1}r^{-1}}\delta_A$ is empty and so regular. Otherwise choose any words $w, w' \in A^+$ with $\overline{w}pq = \overline{w'}rs$. Proposition 2.2.3 shows that the language $(K_{w'} \circ K_w^{-1})\delta_A$ is regular. If $(u,v) \in$

$K_{w'} \circ K_w^{-1}$, then multiple applications of Proposition 2.3.3 (for each letter of w and w') show that $\widehat{u}$ and $\widehat{v}$ fellow-travel. Then

$$\begin{aligned} K_{pqs^{-1}r^{-1}} &= \{(u,v) : u, v \in K, \overline{u}pq = \overline{v}rs\} \\ &= \{(u,v) : u, v \in K, \overline{u}\,\overline{w'} = \overline{v}\,\overline{w}\} && \text{since } H \text{ is abelian} \\ &= K_{w'} \circ K_w^{-1}. \end{aligned}$$

So $K_{pqs^{-1}r^{-1}}$ and $K_{w'} \circ K_w^{-1}$ coincide. $\boxed{6.3.2}$

Constructing a language of normal forms for M requires some delicate manoeuvring in order to ensure that if u and v are normal form words with $\overline{ua} = \overline{v}$ for $a \in A$, then u and v 'keep pace' with one another. The main obstacle is the two different ways in which an H-subword may arise. In comparison, letters of X are easily dealt with, since they do not cancel.

The first step is to use K to construct languages $K(p,q)$ (where $p \in S$ and $q \in P$) with the property that if $u \in K(p,q)$ and $v \in K(r,s)$ are such that $p\overline{u}q = r\overline{v}s$, then $|u| = |v|$.

For every $p \in S$ and $q \in P$, define the language

$$\begin{aligned} K(p,q) = \Big\{u1^k : u \in K \wedge (\exists r \in S, s \in P)(\exists v \in K) \\ \Big[(|v| = |u| + k) \wedge ((u,v) \in K_{pq(rs)^{-1}}) \\ \wedge \big((\forall r' \in S, s' \in P)(\forall v' \in K) \\ ((u,v') \in K_{pq(r's')^{-1}} \implies |v'| \leq |v|)\big)\Big]\Big\}. \end{aligned}$$

The language $K(p,q)$ is defined so that each word $u \in K$ is padded (by appending symbols 1) to the length of the longest word v such that $p\overline{u}q = r\overline{v}s$ for some $r \in S$ and $s \in P$.

In the definition of $K(p,q)$, the value k has only a finite range: two words u and v in K such that $p\overline{u}q = r\overline{v}s$ for $p, r \in S$ and $q, s \in P$ cannot differ in length too much without violating the uniqueness condition. Furthermore, the existential quantifiers over S and P could be replaced by a finite union, the universal quantifiers over S and P by a finite intersection, and all other quantifiers are over regular languages. The conditions $|v| = |u| + k$ and $|v'| \leq |v|$ can be checked inside a finite state automaton. Therefore the language $K(p,q)$ is regular by Theorem A.5.6.

The second step is to modify these languages $K(p,q)$ to deal with the case when a word $w \in K(p,q)$ is such that $p\overline{w}q$ is an H-subword of an element of $\overline{A'}$ — the language of normal forms must cope with the circumstance that $\overline{ua} = \overline{v}$, where u and v are normal form words, $a \in A$, and in $\overline{u}$, a particular H-subword arises from an H-subword of a generator $\overline{a}$ in $\overline{A'}$, whilst in $\overline{v}$, the corresponding H-subword arises from some $p\overline{w}q$.

Let

$$m = \max\big\{|u| : u \in K(p,q),\ p\overline{u}q = z \text{ for some } z \in Z,\ p \in S,\ q \in P\big\}.$$

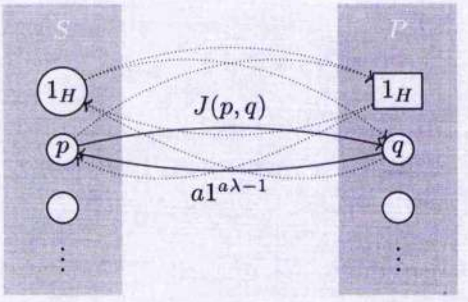

Figure 6.1. The automaton $\mathcal{A}$, which recognizes normal form words for the submonoid M. Its start state is $1_H \in S$ and its unique accept state is $1_H \in P$.

This maximum exists because K maps bijectively to $\mathrm{Sg}\langle \overline{A''} \rangle$. To every word $u \in K(p,q)$ with $|u| \leq m$, add a suffix string $1^{m-|u|}$ to yield a new language $J(p,q)$ with no word having length less than m. This new language is still regular.

The construction of the languages $J(p,q)$ guarantees two facts about any word $u \in J(p,q)$:

i.) The length of u is at least m.

ii.) If $v \in J(r,s)$ for $r \in S$ and $s \in P$ is such that $p\overline{u}q = r\overline{v}s$, then the lengths of u and v are equal.

Define $\lambda : A' \to \mathbb{N}$ as follows. Suppose, for $a \in A'$, that $\overline{a} = z_0 x_1 z_1 \cdots x_n z_n$, where each z_i lies in H and each x_i is a letter of X. Then

$$a\lambda = n + (n-1)m.$$

Interpret this definition of $a\lambda$ as follows: each of the n letters x_i contributes 1 to the total, and each of the H-subwords z_i, which are $n-1$ in number, contributes m. Define λ on the languages $J(p,q)$ by $w\lambda = |w|$.

Construct a generalized finite state automaton $\mathcal{A}$ as follows. Let its state set be the disjoint union of S and P. (Notice particularly that $1_H \in P$ and $1_H \in S$ are distinct states.) The start state is $1_H \in S$ and the unique accept state is $1_H \in P$. For each $a \in A'$, add an edge from the H-prefix of $\overline{a}$ to the H-suffix of $\overline{a}$ labelled by $a1^{a\lambda-1}$; add an edge from each $p \in S$ to each $q \in S$ labelled by [a regular expression defining] $J(p,q)$. (See Figure 6.1.)

The aim is now to show that $(A, L(\mathcal{A}))$ is an automatic structure for M. However, some further information about $L(\mathcal{A})$ must be gathered before embarking on a proof.

Observe that every word in $L(\mathcal{A})$ alternates between letters of A' padded with symbols 1 and words in some language $J(p,q)$. Extend λ to such alternating words by inductively defining

$$(wu)\lambda = w\lambda + u\lambda,$$

where $u = 1$ (with $1\lambda = 0$), $u = a$ or $u \in J(p,q)$ for some $p \in S$, $q \in P$.

Let w be such an alternating word. Suppose $\overline{w} = z_0x_1z_1\cdots x_nz_n$, where each z_i lies in H and each x_i is a letter of X. Each letter x_i — arising from $\overline{a}$ for some $a \in A'$ — contributes 1 towards the value of $w\lambda$. If z_i lies in the set of H-subwords Z, it contributes m towards $w\lambda$, regardless of whether it arises from $\overline{a}$ for some $a \in A'$ or from $p\overline{u}q$ for some $p \in S$, $q \in P$, and $u \in J(p,q)$. If $z_i \notin Z$, then it must arise as $p\overline{u}q$ for some $u \in J(p,q)$ and therefore contributes $|u|$ towards $w\lambda$. However, by observation ii. above, all words u such that $p\overline{u}q = z_i$ for some $p \in S$ and $q \in P$ have the same length. Therefore, the image of w under λ depends on $\overline{w}$, not on w itself.

Furthermore, since λ is a measure of length on the various languages $J(p,q)$ and each $a \in A'$ is padded out to length $a\lambda$, the length of $w \in L(\mathcal{A})$ is $w\lambda$. Slightly more general reasoning along these lines gives the following result:

Lemma 6.3.3. *If $\mathcal{A}$ reads a word u and arrives in a state, $|u| = u\lambda$.* $\boxed{6.3.3}$

Lemma 6.3.4. *The language $L(\mathcal{A})$ maps onto M.*

Proof of 6.3.4. The identity element of M is represented by one word in every language $J(p,q)$. In particular, therefore, it is represented by a word in $J(1_F, 1_F)$ that is accepted by $\mathcal{A}$.

Let $w \in A^+$, and that $\overline{w} \neq 1_F$. Begin by deleting all symbols 1.

For each subword aub, with $a, b \in A'$ and $u \in (A'')^*$, replace u by the element of $J(p,q)$ representing $\overline{u}$, where p is the H-suffix of $\overline{a}$ and q is the H-prefix of $\overline{b}$. For a subword over A'' at the start of w, take $p = 1_F$; at the end of w, take $q = 1_F$.

Finally, replace each letter $a \in A'$ by $a1^{a\lambda-1}$.

This yields a word accepted by $\mathcal{A}$. None of these transformations alters the element represented by w, so this shows that $L(\mathcal{A})$ maps onto M. $\boxed{6.3.4}$

Let $u, v \in L(\mathcal{A})$ and suppose that $\overline{u} = \overline{v}$.

Suppose that after reading $u(t)$ the automaton $\mathcal{A}$ arrives in a state in S. Then $\overline{u(t)} = z_0x_1z_1\cdots x_nz_n$, where $z_i \in H$ and $x_i \in X$, and $z_n \in S$. (Indeed, z_n is the state $\mathcal{A}$ reaches.) Notice that $t = |u(t)| = (u(t))\lambda$.

Let $v(s)$ be the shortest prefix of v such that $z_0x_1z_1\cdots x_n$ is a prefix of $\overline{v(s)}$. Then the s-th letter of v is drawn from A'. Since these letters lie at the start of labels on edges originating in P, $\mathcal{A}$ can read $v(s-1)$ and arrive in a state in P. Therefore $(v(s-1))\lambda = |v(s-1)| = s-1$ by Lemma 6.3.3. Furthermore, $\mathcal{A}$ will reach a state in S after reading $v(s)$ and l additional symbols 1, where $l < \max\{a\lambda : a \in A'\}$. Therefore, using Lemma 6.3.3 again,

$$s - 1 = (v(s-1))\lambda \leq \underbrace{(u(t))\lambda}_{=t} \leq (v(s))\lambda = (v(s+l))\lambda = s + l.$$

So $|s - t| < \max\{a\lambda : a \in A'\}$.

Now, $\overline{v(s)} = z_0x_1z_1\cdots x_n\underline{z_n\cdots x_{n+h}y}$, where $h \geq 0$ and $z_n\cdots x_{n+h}y$ is a suffix of an element of $\overline{A'}$. Therefore, $\overline{u(t)v(s)^{-1}} = z_n(z_n\cdots x_{n+h}y)^{-1}$. Since both z_n and $z_n\cdots x_{n+h}y$ are drawn from finite sets, this bounds the distance between $\overline{u(t)}$ and

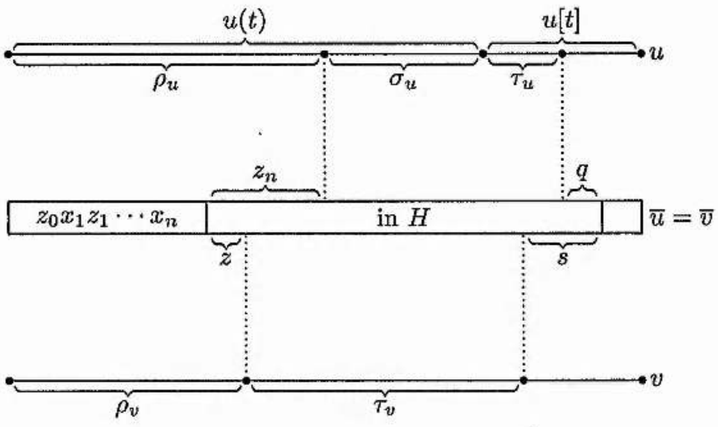

Figure 6.2. Words in L and corresponding elements of $M \subseteq X^* * H$.

$\overline{v(s)}$. Since the difference between t and s is also bounded, the distance between the elements $\overline{u(t)}$ and $\overline{v(t)}$ is bounded.

If $u(t)$ takes $\mathcal{A}$ to part-way along an edge leading to a state in S, at most $2\max\{a\lambda : a \in A'\}$ is added to this bound.

The remaining case is when $u(t)$ concludes with part of a word in $J(p,q)$. Let $u(t) = \rho_u\sigma_u$, where σ_u is the longest suffix of $u(t)$ over A''. Suppose $\overline{\rho_u} = z_0x_1z_1\cdots x_nz_n$ and pick ρ_v to be the shortest prefix of v such that $z_0x_1z_1\cdots x_n$ is a prefix of $\overline{\rho_v}$. Reason as before to see that $||\rho_v| - |\rho_u|| < \max\{a\lambda : a \in A'\}$. Let τ_u be the longest prefix of $u[t]$ over A''. Let τ_v be the longest prefix of $v[|\rho_v|]$ over A''. Let q and s be the H-prefixes of the elements represented by the letters of u and v immediately after τ_u and τ_v.

Suppose $\overline{\rho_v} = z_0x_1z_1\cdots x_mz_m$ where $m > n$. Then $z_n\overline{\sigma_u\tau_u}q$ matches an H-subword of $\overline{\rho_v}$. There are therefore only finitely many possibilities for $\sigma_u\tau_u$, which therefore bounds the distance between $\overline{\rho_u}$ and $\overline{\rho_v}$.

Now suppose $\overline{\rho_v} = z_0x_1z_1\cdots x_nz$. (See Figure 6.2.) Then $z_n\overline{\sigma_u\tau_u}q = z\overline{\tau_v}s$, and so $(\sigma_u\tau_u, \tau_v)$ — minus the appended padding symbols 1 — lies in $K_{z_nqs^{-1}z^{-1}}$. Therefore the paths labelled by $\sigma_u\tau_u$ and τ_v fellow travel. That is, $\overline{(\sigma_u\tau_u)(t-|\rho_u|)}$ is within a bounded distance of $\overline{\tau_v(t-|\rho_v|)}$. The established boundedness of the distance between $\overline{\rho_u}$ and $\overline{\rho_v}$ and of the difference in the lengths of ρ_u and ρ_v, together with the fact that $\overline{\tau_u}$ and $\overline{\tau_v}$ commute with z_n and z, show that $\overline{u(t)}$ and $\overline{v(t)}$ are a bounded distance apart.

This reasoning shows that if $\overline{u} = \overline{v}$, then u and v fellow travel. Now suppose that $\overline{ua} = \overline{v}$ for some $a \in A$. Treat two cases:

i.) Suppose $a \in A'$. Let $u = u'u''$, where u'' is the longest suffix of u over A''. By previous reasoning, the paths $\widehat{u'}$ and $\widehat{v}$ must fellow-travel. Let p be the H-suffix of $\overline{u'}$ and let q be the H-prefix of $\overline{a}$. The element $p\overline{u''}q$ corresponds to an H-subword of $\overline{v}$. The word u'' is a word of K with some padding added. The structure (A'', K) is an automatic structure with uniqueness for the semigroup

$\mathrm{Sg}\langle\overline{A''}\rangle$. Therefore, regardless of how the H-subword $p\overline{u''}q$ arises in $\overline{v}$, the fellow traveller property holds throughout u''.

ii.) If $a \in A''$, then $\overline{u}$ and $\overline{v}$ differ only in their H-suffixes. Let $u = u'u''$ and $v = v'v''$, where u'' and v'' are the longest suffixes of u and v over A''. By previous reasoning, the paths labelled by u' and v' must fellow-travel. Let p and r be the H-suffixes of $\overline{u'}$ and $\overline{v'}$ respectively. Then $u'' \in J(p, 1_F)$ and $v'' \in J(r, 1_F)$ and $p\overline{u''a} = r\overline{v''}$. By the construction of the various languages $J(p,q)$ from K, and the fact that (A'', K) is an automatic structure for $\mathrm{Sg}\langle\overline{A''}\rangle$, the paths labelled by u'' and v'' fellow travel. So, since H is abelian, the paths $\widehat{u}$ and $\widehat{v}$ also satisfy the fellow traveller property.

Theorem 2.3.2 shows that $(A, L(\mathcal{A}))$ is an automatic structure for the monoid M. If the subsemigroup generated by the *original* set $\overline{A}$ is not a monoid, remove the adjoined identity $\overline{1}$ and apply Theorem 2.4.3. Since the original subset $\overline{A}$ was arbitrary, the free product F is locally automatic. $\boxed{6.3.1}$

Theorem 2.5.5 yields an immediate corollary of Theorem 6.3.1:

Corollary 6.3.5. *The [monoid] free product of a free monoid and an abelian group is Malcev coherent.* $\boxed{6.3.5}$

Open Problem 6.3.6. Is the free product of a free monoid with a virtually abelian, nilpotent, or virtually nilpotent group Malcev coherent?

As finitely generated subsemigroups of virtually abelian groups may not be asynchronously automatic (see Example 5.5.1), and as nilpotent groups themselves are asynchronously automatic only when they are virtually abelian (Epstein et al. 1992, Theorem 8.2.8), one cannot establish the Malcev coherence of these free products by means of automatism.

A more general question is the following:

Open Problem 6.3.7. Is the class of Malcev coherent monoids closed under taking free products with a free monoid?

CHAPTER SEVEN

SUBSEMIGROUPS OF DIRECT PRODUCTS

Iago: ... *take note, take note, O world,*
To be direct and honest is not safe!
— William Shakespeare,
Othello, Moor of Venice (1603–4), iii.3

7.1. INTRODUCTION

THE DIRECT PRODUCT of two coherent groups is not in general coherent: if F_2 is the free group of rank 2, then $F_2 \times F_2$ is not coherent (Grunewald 1978). Therefore the direct product of two Malcev coherent groups may not even be coherent. However, the direct product of a free group and a polycyclic group is coherent: a proof of this, adapted from reasoning of Miller (2002), is given in Section 7.4. [Polycyclic groups are of course themselves coherent (see Section 7.3).]

One naturally asks whether the direct product of a free group and a polycyclic group is Malcev coherent. The present chapter answers this question negatively: indeed, Section 7.3 shows that polycyclic groups themselves are not in general Malcev coherent. The proof of this result relies on the fact, established in Section 7.2, that the direct product of two free semigroups of rank at least 2 is not Malcev coherent. Section 7.5 proves a positive result: any direct product of a virtually free group and an abelian group is Malcev coherent. It also exhibits an example of a non-automatic finitely generated subsemigroup of the direct product of a free semigroup and the natural numbers. Section 7.6 suggests some open problems on the Malcev coherence of direct products. [This chapter is a revised and extended version of Cain (2005*c*).]

7.2. DIRECT PRODUCTS OF FREE SEMIGROUPS

ALTHOUGH the direct product of two free non-abelian groups is not coherent, this does not immediately preclude the Malcev coherence of the direct product of two free semigroups. However, the following example exhibits a finitely generated subsemigroup of such a direct product that does not admit a finite Malcev presentation.

Example 7.2.1. Let $A = \{a, b, c, d, e, f, g, h, i, j\}$ be an alphabet representing elements of $\{x, y, p, q, r, s\}^+ \times \{x, y, p, q, r, s\}^+$ as follows:

$$\begin{aligned}
\overline{a} &= (x^2pqrs, x), & \overline{f} &= (x^2pq, x),\\
\overline{b} &= (pqrspqrs, p), & \overline{g} &= (rspq, p),\\
\overline{c} &= (pqr, q), & \overline{h} &= (rsp, q),\\
\overline{d} &= (spqr, r), & \overline{i} &= (qrspqrsp, r),\\
\overline{e} &= (sy^2, s), & \overline{j} &= (qrsy^2, s).
\end{aligned}$$

Let S be the semigroup generated by $\overline{A}$.

Proposition 7.2.2. *The semigroup S is presented by* $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, *where*

$$\mathcal{R} = \{(ab^\alpha cd^\alpha e, fg^\alpha hi^\alpha j) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

It is therefore isomorphic to the semigroup in Example 6.2.1, and so does not admit a finite Malcev presentation.

Proof of 7.2.2. Every relation in $\mathcal{R}$ holds in S:

$$\begin{aligned}
\overline{ab^\alpha cd^\alpha e} &= (x^2pqrs(pqrspqrs)^\alpha pqr(spqr)^\alpha sy^2, xp^\alpha qr^\alpha s)\\
&= (x^2(pqrs)^{3\alpha+2}y^2, xp^\alpha qr^\alpha s)\\
&= (x^2pq(rspq)^\alpha rsp(qrspqrsp)^\alpha qrsy^2, xp^\alpha qr^\alpha s)\\
&= \overline{fg^\alpha hi^\alpha j}
\end{aligned}$$

for all $\alpha \in \mathbb{N} \cup \{0\}$.

Define a set N of normal forms to be the set of all words in A^+ that do not contain $fg^\alpha hi^\alpha j$ for any $\alpha \in \mathbb{N} \cup \{0\}$. Every element of S is represented by at least one element of N since a word over A^+ can be rewritten to one in N using the relations $\mathcal{R}$: such rewriting cannot continue indefinitely since each step decreases the number of letters from $\{f, g, h, i, j\}$ present.

Let $u = u_1 \cdots u_p$ and $v = v_1 \cdots v_q$ ($u_i, v_i \in A$ for all i) be distinct words in the set of normal forms N, and suppose they represent the same element of S. Without loss of generality, suppose that $u_1 \neq v_1$ and that u precedes v in the lexicographic ordering based on $a \prec b \prec c \prec \ldots \prec j$. Consider the second component of $\overline{u_1}$ and $\overline{v_1}$ to see that

$$(u_1, v_1) \in \{(a, f), (b, g), (c, h), (d, i), (e, j)\}.$$

Examining the first component forces $u_1 = a$ and $v_1 = f$. So

$$\begin{aligned}
\overline{u} &= (x^2pqrs\cdots, x\cdots),\\
\overline{v} &= (x^2pq\cdots, x\cdots).
\end{aligned}$$

The letter v_2 is thus either g or h, and consideration of second components then forces $u_2 = b$ or $u_2 = c$, respectively. Without loss of generality, assume that the next

$\alpha \in \mathbb{N} \cup \{0\}$ letters of v are letters g followed by a different letter: $v_2 \cdots v_{\alpha+1} = g^\alpha$ and $v_{\alpha+2} \neq g$. So:

$$\begin{aligned}\overline{u} &= (x^2pqrs\cdots, x\cdots),\\ \overline{v} &= (x^2pq(rspq)^\alpha\cdots, xp^\alpha\cdots).\end{aligned}$$

In order to match the p^α in the second component, $u_2 \cdots u_{\alpha+1} \in \{b, g\}^*$.

Now, if $v_{\alpha+2}$ were a, b, c, d, e, f, i, or j, the first component of $\overline{v}$ would include a subword qx, qp, qs, or q^2. The only way such a subword can arise in $\overline{u}$ would be if u included wa, wb, wc, wd, we, wf, wi, or wj, where w is a word ending in f or g and the first component of $\overline{w}$ is $(pqrs)^\mu pq$. It is easy to see that no such word w exists. Therefore $v_{\alpha+2} = h$. So:

$$\begin{aligned}\overline{u} &= (x^2pqrs\cdots, x\cdots),\\ \overline{v} &= (x^2pq(rspq)^\alpha rsp\cdots, xp^\alpha q\cdots).\end{aligned}$$

If $v_{\alpha+3}$ were a, b, c, d, e, f, g, or h, the first component of $\overline{v}$ would include a subword px, p^2, ps, or pr. The only way such a subword can arise in $\overline{u}$ would be if u included wa, wb, wc, wd, we, wf, wg, or wh, where w is a word ending in h or i and the first component of $\overline{w}$ is $(pqrs)^\mu p$. Again, it is easy to see that no such word w exists. Therefore $v_{\alpha+3} = i$ or $v_{\alpha+3} = j$. Without loss of generality, assume that the next $\gamma \in \mathbb{N} \cup \{0\}$ letters of v are letters i, followed by a different letter: $u_{\alpha+3} \cdots v_{\alpha+\gamma+2} = i^\gamma$ and $v_{\alpha+\gamma+3} \neq i$. So:

$$\begin{aligned}\overline{u} &= (x^2pqrs\cdots, x\cdots),\\ \overline{v} &= (x^2pq(rspq)^\alpha rsp(qrspqrsp)^\gamma\cdots, xp^\alpha qr^\gamma\cdots).\end{aligned}$$

Reasoning similar to that in the last paragraph establishes that $v_{\alpha+\gamma+3} = j$. So

$$\begin{aligned}\overline{u} &= (x^2pqrs\cdots, x\cdots),\\ \overline{v} &= (x^2pq(rspq)^\alpha rsp(qrspqrsp)^\gamma qrsy^2\cdots, xp^\alpha qr^\gamma s\cdots).\end{aligned}$$

In order to match the first component of $\overline{v}$, the string u must begin $ab^\beta cd^\delta e$, where the y^2 arising from e matches the y^2 in the first component of $\overline{v}$. In order for the second components to match, $\alpha = \beta$ and $\gamma = \delta$. Comparing first components shows that

$$x^2pqrs(pqrspqrs)^\beta pqr(spqr)^\delta sy^2 = x^2pq(rspq)^\alpha rsp(qrspqrsp)^\gamma qrsy^2,$$

or rather, $x^2(pqrs)^{2\beta+\delta+2}y^2 = x^2(pqrs)^{\alpha+2\gamma+2}y^2$. So $2\beta + \delta = \alpha + 2\gamma$. Since $\alpha = \beta$ and $\gamma = \delta$, this forces $\alpha = \beta = \gamma = \delta$.

So v begins $fg^\alpha hi^\alpha j\cdots$, which is a contradiction. Therefore S is presented by $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$. $\boxed{7.2.2}$

Theorem 7.2.3. *The direct product of two free semigroups of rank at least 2 is not Malcev coherent.*

Proof of 7.2.3. Let D be the direct product of two free semigroups of rank at least 2. The free semigroup of rank 2 contains isomorphic copies of free semigroups of every rank; D therefore contains a subsemigroup isomorphic to the direct product $\{x,y,p,q,r,s\}^+ \times \{x,y,p,q,r,s\}^+$. The semigroup D thus contains the finitely generated subsemigroup S of Example 7.2.1, which does not admit a finite Malcev presentation. Therefore D is not Malcev coherent. $\boxed{7.2.3}$

7.3. POLYCYCLIC GROUPS AND THEIR DIRECT PRODUCTS

EVERY FINITELY GENERATED NILPOTENT GROUP is polycyclic (Proposition 7.3.2). Whilst the class of polycyclic groups is larger than the class of finitely generated nilpotent groups, the former class retains many of the pleasant properties of the latter (see Sims 1994, Section 9.3). In particular, polycyclic groups are coherent (Proposition 7.3.7). This section is dedicated to proving that polycyclic groups are not in general Malcev coherent.

Definition 7.3.1. A group G is *polycyclic* if it possesses a series of subgroups

$$\{1_G\} = H_0 \trianglelefteq H_1 \trianglelefteq \ldots \trianglelefteq H_n = G,$$

such that, for each $i \in \{1,\ldots,n\}$, the factor group H_i/H_{i-1} is cyclic.

The few necessary facts about polycyclic groups are collected here:

Proposition 7.3.2 (Sims 1994, Proposition 9.3.4). *Every finitely generated nilpotent group is polycyclic.* $\boxed{7.3.2}$

Proposition 7.3.3 (Sims 1994, Corollary 9.3.8). *Every subgroup of a polycyclic group is finitely generated.* $\boxed{7.3.3}$

Proposition 7.3.4 (Sims 1994, Proposition 9.3.3). *Every extension of a polycyclic group by a polycyclic group if itself polycyclic. That is, if E is an extension of G, and G and E/G are both polycyclic, then E is polycyclic. In particular, the direct product of two polycyclic groups is polycyclic.* $\boxed{7.3.4}$

Proposition 7.3.5 (Sims 1994, Section 9.4). *Polycyclic groups are finitely presented.* $\boxed{7.3.5}$

Proposition 7.3.6 (Sims 1994, Proposition 9.3.7). *Every subgroup of a polycyclic group is polycyclic.* $\boxed{7.3.6}$

An immediate consequence of Propositions 7.3.5 and 7.3.6 is the following:

Proposition 7.3.7. *Polycyclic groups are coherent.* $\boxed{7.3.7}$

Finally, the following result of Rosenblatt is needed:

Theorem 7.3.8 (Rosenblatt 1974, Theorem 4.12). *Let G be a polycyclic group. Then exactly one of the following two statements is true:*

i.) *The group G is virtually nilpotent.*

ii.) *The group G contains a free subsemigroup of rank 2.* $\boxed{7.3.8}$

In light of Proposition 7.3.7 and Theorem 5.3.5, one naturally considers the Malcev coherence of polycyclic groups.

Theorem 7.3.9. *The direct product of two polycyclic groups that are not virtually nilpotent is not Malcev coherent.*

Proof of 7.3.9. Let G and H be polycyclic groups that are not virtually nilpotent. Let $P = G \times H$. Theorem 7.3.8 shows that G and H both contain a free subsemigroup of rank 2. Therefore P contains the direct product of two free semigroups of rank 2, which is not Malcev coherent by Theorem 7.2.3. Ergo, P itself is not Malcev coherent. $\boxed{7.3.9}$

Corollary 7.3.10. *Polycyclic groups are not in general Malcev coherent.*

Proof of 7.3.10. Let G be a polycyclic group that is not virtually nilpotent. Let $P = G \times G$. By Proposition 7.3.4, P is also a polycyclic group; by Theorem 7.3.9 P is not Malcev coherent. $\boxed{7.3.10}$

7.4. DIRECT PRODUCTS OF FREE AND POLYCYCLIC GROUPS

Theorem 7.4.1. *Every direct product of a free group and a polycyclic group is coherent.*

The following proof is a modification of the proof of Miller (2002, Theorem 1). It requires the following two theorems.

Theorem 7.4.2 (Hall 1949). *If G is a finitely generated subgroup of a free group F, then G is a free factor of a finite-index subgroup of F.* $\boxed{7.4.2}$

Theorem 7.4.3. *Let G be group with a finite presentation $\mathrm{Gp}\langle A \mid \rho\rangle$. Let P and Q be subgroups of G and $\psi : P \to Q$ an isomorphism. The HNN extension of G with associated subgroups P and Q and stable letter t,*

$$\mathrm{Gp}\langle A, t \mid \rho, (t^{-1}pt, p\psi) \text{ for all } p \in P\rangle,$$

is finitely presented if and only if P and Q are finitely generated. $\boxed{7.4.3}$

[Theorem 7.4.3 is an HNN extension version of Theorem 6.2.3, which was proved by Baumslag (1962). Although Baumslag did not publish the HNN extension version, it is usually attributed to him.]

Proof of 7.4.1. Let F be a free group and let P be a polycyclic group. Let X be a finite subset of $F \times P$ and suppose X generates the subgroup G. Let $K = G \cap F$ and $L = G \cap P$. Let $\pi_F : G \to F$ and $\pi_P : G \to P$ be the projection mappings to F and P respectively.

By restricting, if necessary, to (im π_F) $\times$ (im π_P), assume without loss of generality that these projection mappings are surjective. As $X\pi_F$ generates im π_F, the group F is a finitely generated free group by the Nielsen–Schreier Theorem (Lyndon & Schupp 1977, Proposition I.2.6). Similarly, the group P is a polycyclic group by Proposition 7.3.6.

Observe that $K = G \cap F = \operatorname{Ker} \pi_P$ and $L = G \cap P = \operatorname{Ker} \pi_F$ are both normal subgroups of G.

i.) If $K = \{1_G\}$, then π_P is injective and so is an isomorphism. The group G is therefore a polycyclic group and so finitely presented by Proposition 7.3.5.

ii.) If $L = \{1_G\}$, then π_F is an isomorphism and G is a free group and thus finitely presented.

iii.) Suppose K and L are both non-trivial. Then L, being a subgroup of a polycyclic group, is finitely generated (see Proposition 7.3.3). Let Y be a set representing L and let $\mathcal{R}$ be the kernel of the representation mapping from Y to L. Let t represent an element of $K - \{1_F\}$. Theorem 7.4.2 shows that there is a finite-index subgroup M of F which has $\operatorname{Gp}\langle \overline{t} \rangle$ as a free factor — in other words, admitting a basis $\{\overline{t}, s_1, \ldots, s_n\}$. For each i, let a_i be such that $\overline{a_i}\pi_F = s_i$. So each $\overline{a_i}$ is a lift of s_i to G. Let G_0 be the inverse image of M under π_F. Then G_0 is a subgroup of G of finite index and is presented by

$$G_0 = \operatorname{Gp}\langle Y, t, a_1, \ldots, a_n \mid \mathcal{R}, (t^{-1}bt, b), (a_i^{-1}ba_i, b\phi_i) \text{ for all } i \text{ and } b \in Y \rangle,$$

where $b\phi_i \in Y$ represents the image of $\overline{b}$ under conjugation by $\overline{a_i} \in G$.

Since L is finitely generated, multiple applications of Theorem 7.4.3 show that G_0 is finitely presented. The Reidemeister–Schreier Theorem (see Lyndon & Schupp 1977, Section II.4) applies to show that G is finitely presented.

In each case, the arbitrary finitely generated subgroup G of $F \times P$ is finitely presented. Therefore $F \times P$ is coherent. $\boxed{7.4.1}$

Corollary 7.4.4. *Every direct product of a free group and a nilpotent group is coherent.*

[This corollary is not quite a special case of Theorem 7.4.1: polycyclic groups are always finitely generated; nilpotent groups may not be.]

Proof of 7.4.4. Let F be a free group and N a nilpotent group. Let X be a finite subset of $F \times N$ and suppose X generates the subgroup G. Let $\pi_N : G \to N$ be the projection mapping to N. Then G is a finitely generated subgroup of $F \times \text{im } \pi_N$. The group im π_N is nilpotent and finitely generated by $X\pi_N$. By Proposition 7.3.2, im π_N is a polycyclic group. Therefore, by Theorem 7.4.1, $F \times \text{im } \pi_N$, the group G is finitely presented. Since X was arbitrary, $F \times N$ is coherent. $\boxed{7.4.4}$

7.5. DIRECT PRODUCTS OF VIRTUALLY FREE AND ABELIAN GROUPS

Theorem 7.5.1. *Every direct product of a virtually free group and an abelian group is Malcev coherent.*

Proof of 7.5.1. Let F be a virtually free group and let H be an abelian group. Let A be a finite alphabet representing elements of $G = F \times H$. Let $\rho : A^* \to G$ be the standard representation mapping. Let $S = \mathrm{Sg}\langle A\rho\rangle$. The semigroup S is obviously presented by $\mathrm{Sg}\langle A \mid \ker \rho\rangle$.

Let $\pi_F : G \to F$ and $\pi_H : G \to H$ be the projection mappings to F and H, respectively. Define $\rho_F : A^* \to F$ and $\rho_H : A^* \to H$ by $\rho\pi_F$ and $\rho\pi_H$, respectively. Notice that $\ker \rho = \ker \rho_F \cap \ker \rho_H$.

The strategy of the proof is based on the observation that any relation $(u, v) \in \ker \rho$ can be decomposed as

$$(u, v) = (c_1, d_1)(c_2, d_2) \cdots (c_k, d_k), \qquad (1)$$

where $u = c_1c_2 \cdots c_k$, $v = d_1d_2 \cdots d_k$, and each (c_i, d_i) is in $\ker \rho_F$ but not necessarily in $\ker \rho_H$. The first stage of the proof involves showing that every relation in $\ker \rho$ is a Malcev consequence of relations that have a decomposition (1) where each pair (c_i, d_i) is drawn from a particular finite set. However, the set of such relations is manifestly infinite. The second — rather technical — stage involves showing that all these relations are Malcev consequences of those in a different, but still infinite, set. The reader — although perhaps beginning to empathize with Sisyphus — should be reassured by the fairly simple structure of this new set of relations. The third stage is an easy proof that a finite subset of these relations suffices for a Malcev presentation.

Preliminaries. Let $S_F = S\pi_F \subseteq F$. Notice that S_F is not in general a subset of S and that $A\rho_F$ is a finite generating set for S_F. The subsemigroup S_F has a finite Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{R}\rangle$, where $\mathcal{R}$ is the finite set of relations found using the algorithm of Theorem 4.3.1. Suppose that $\mathcal{R}$ is symmetrized, so that $(u, v) \in \mathcal{R}$ implies that $(v, u) \in \mathcal{R}$. Suppose that

$$\mathcal{R} = \{(u_1, v_1), (v_1, u_1), \ldots, (u_n, v_n), (v_n, u_n)\},$$

where $u_i, v_i \in A^+$. The set of relations $\mathcal{R}$ is contained in $\ker \rho_F$. Fix these pairs (u_i, v_i) throughout the proof.

Define $\delta : A^+ \times A^+ \to H$ by $(u, v)\delta = (u\rho_H) - (v\rho_H)$. Let $D = \mathcal{R}\delta \subseteq H$. Observe that $(u, v)\delta = -(v, u)\delta$. Therefore, since $\mathcal{R}$ is symmetrized, $D = -D$. Throughout this proof, δ is used as a measure of the 'difference' in the H-components of the elements represented by the two sides of a relation in $\ker \rho_F$.

Each $(u, v) \in \ker \rho$ can be decomposed (possibly in many ways) as a product

$$(u, v) = (c_1, d_1)(c_2, d_2) \cdots (c_k, d_k),$$

where $u = c_1c_2 \cdots c_k$, $v = d_1d_2 \cdots d_k$, and $(c_i, d_i) \in \ker \rho_F$.

Recall from the proof of Theorem 4.3.1 that for $(s, t) \in \ker \rho_F$, $n(s, t)$ is the minimum number of internal vertices in a derivation tree for the word st^{-1}.

Define, for each decomposition $(c_1, d_1)(c_2, d_2) \cdots (c_k, d_k)$,

$$n'((c_1, d_1)(c_2, d_2) \cdots (c_k, d_k)) = \max\{n(c_i, d_i) : i = 1, \ldots, k\}.$$

and

$$p'((c_1, d_1)(c_2, d_2) \cdots (c_k, d_k)) = \left|\{i : n(c_i, d_i) = n'((c_1, d_1)(c_2, d_2) \cdots (c_k, d_k))\}\right|.$$

So n' is the maximum n-value of any of the (c_i, d_i), and p' is the number of times this maximum is achieved.

Fix a *canonical decomposition* of each relation $(u, v) \in \ker \rho$ by selecting the decompositions that minimize n' and from these selecting one that minimizes p'. Define

$$n''(u, v) = n'((c_1, d_1)(c_2, d_2) \cdots (c_k, d_k))$$

and

$$p''(u, v) = p'((c_1, d_1)(c_2, d_2) \cdots (c_k, d_k)),$$

where $(c_1, d_1)(c_2, d_2) \cdots (c_k, d_k)$ is the canonical decomposition of (u, v).

First stage. Let $\mathcal{S}$ be the subset of $\ker \rho$ consisting of those relations whose canonical decompositions are formed by concatenating elements of $\mathcal{R}$. Let

$$\mathcal{Q} = \{(u_i w v_i, v_i w u_i) : w \in A^* \text{ and } i = 1, \ldots, n\}.$$

Notice that $\mathcal{Q} \subseteq \mathcal{R}^{\#}$, and furthermore that if $(p, q) \in \mathcal{R}^+$ and $w \in A^*$, then (pwq, qwp) is a consequence of $\mathcal{Q}$. (The set $\mathcal{R}^+$ consists of all relations formed by concatenating elements of $\mathcal{R}$.)

Define an ordering $\ll$ of the set $\ker \rho$ as follows:

$$\begin{aligned}(u, v) \ll (u', v') \iff\ & n''(u, v) < n''(u', v') \text{ or} \\ & \bigl(n''(u, v) = n''(u', v') \text{ and } p''(u, v) < p''(u', v')\bigr).\end{aligned}$$

The line of reasoning in this first stage owes much to the proof of Theorem 4.3.1. To show that each (u, v) is a Malcev consequence of $\ll$-preceding elements of $\ker \rho$, one follows the basic outline of that earlier proof to obtain $\ll$-preceding elements of $\ker \rho_F$; one 'compensates' for the fact that these relations may not lie in $\ker \rho_H$ by inserting pairs $u_i u_i^{\mathsf{R}}$, $v_i v_i^{\mathsf{R}}$, $u_i^{\mathsf{L}} u_i$, or $v_i^{\mathsf{L}} v_i$; and one uses these newly-found relations and those in $\mathcal{Q}$ in a Malcev chain yielding (u, v).

Lemma 7.5.2. *Let $(u, v) \in \ker \rho - \mathcal{S}$. Then (u, v) is a Malcev consequence of relations in $\mathcal{Q}$ and $\ll$-preceding elements of* $\ker \rho$.

The following technical result will be needed in the proof of Lemma 7.5.2. Informally, it is this result that allows the 'compensation' mentioned above. Recall that $D = \mathcal{R}\delta$.

Lemma 7.5.3. *Let $(u, v) \in \ker \rho_F$. Then $(u, v)\delta$ is a positive (that is, semigroup) sum of elements of D.*

Proof of 7.5.3. This result is obviously true for elements of $\mathcal{R} \subseteq \ker \rho_F$. Suppose $(u, v) \in \ker \rho_F - \mathcal{R}$. As in the proof of theorem Theorem 4.3.1, proceed by induction on $n(u, v)$. There are three cases, two of which are parallel.

i.) Suppose

$$u = x\alpha_1 \cdots \alpha_{m-1} w' \text{ and } v = y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1} w'',$$

where $x, \alpha_i, w', w'', \beta_i, y$ are as in the proof of Theorem 4.3.1. Then the relations

$$(x\alpha_1 \cdots \alpha_{m-2} w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w''),$$
$$(x\alpha_{m-1} w', y^{-1}\beta_{m-1}^{-1} w''), \text{ and } (xw', y^{-1}w'') \quad (2)$$

are in $\ker \rho_F$ and have n-values less than $n(u, v)$. Assume that δ applied to each of these relations (2) gives a positive sum of elements of D. Now,

$$\begin{aligned}
(u, v)\delta &= (u\rho_H) - (v\rho_H) \\
&= (x\alpha_1 \cdots \alpha_{m-1} w')\rho_H - (y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1} w'')\rho_H \\
&= (x\alpha_1 \cdots \alpha_{m-2} w')\rho_H + (\alpha_{m-1})\rho_H \\
&\qquad - (y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w'')\rho_H - (\beta_{m-1})\rho_H \\
&= (x\alpha_1 \cdots \alpha_{m-2} w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w'')\delta + (\alpha_{m-1})\rho_H - (\beta_{m-1})\rho_H \\
&= (x\alpha_1 \cdots \alpha_{m-2} w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w'')\delta + (\alpha_{m-1})\rho_H - (\beta_{m-1})\rho_H \\
&\qquad + (x)\rho_H + (w')\rho_H - (y^{-1})\rho_H - (w'')\rho_H \\
&\qquad\qquad - (x)\rho_H - (w')\rho_H + (y^{-1})\rho_H + (w'')\rho_H \\
&= (x\alpha_1 \cdots \alpha_{m-2} w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w'')\delta \\
&\qquad + (x\alpha_{m-1} w')\rho_H - (y^{-1}\beta_{m-1} w'')\rho_H \\
&\qquad\qquad - (xw')\rho_H + (y^{-1}w'')\rho_H \\
&= (x\alpha_1 \cdots \alpha_{m-2} w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1} w'')\delta \\
&\qquad + (x\alpha_{m-1} w', y^{-1}\beta_{m-1} w'')\delta - (xw', y^{-1}w'')\delta.
\end{aligned}$$

Since $D = -D$, the assumption shows that $(u, v)\delta$ is a positive sum of elements of D.

ii.) Suppose

$$u = x\alpha_1 \cdots \alpha_{m-1} w \beta_{m-1} \cdots \beta_1 y' \text{ and } v = y''$$

where $x, \alpha_i, w, \beta_i, y', y''$ are as in the proof of Theorem 4.3.1. Then the relations

$$(x\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y', y''), (x\alpha_{m-1} w \beta_{m-1} y', y''), \text{ and } (xwy', y'') \quad (3)$$

are in $\ker \rho_F$ and have n-values less than $n(u, v)$. Again assume that δ applied to each of these relations (3) gives a positive sum of elements of D. Then

$$\begin{aligned}
(u,v)\delta &= (u\rho_H) - (v\rho_H) \\
&= (x\alpha_1\cdots\alpha_{m-1}w\beta_{m-1}\cdots\beta_1 y')\rho_H - (y'')\rho_H \\
&= (x\alpha_1\cdots\alpha_{m-2}w\beta_{m-2}\cdots\beta_1 y')\rho_H - (y'')\rho_H \\
&\qquad + (\alpha_{m-1})\rho_H + (\beta_{m-1})\rho_H \\
&= (x\alpha_1\cdots\alpha_{m-2}w\beta_{m-2}\cdots\beta_1 y', y'')\delta + (\alpha_{m-1})\rho_H + (\beta_{m-1})\rho_H \\
&\qquad + (xwy')\rho_H - (y'')\rho_H - (xwy')\rho_H + (y'')\rho_H \\
&= (x\alpha_1\cdots\alpha_{m-2}w\beta_{m-2}\cdots\beta_1 y', y'')\delta \\
&\qquad + (x\alpha_{m-1}w\beta_{m-1}y')\rho_H - (y'')\rho_H - (xwy')\rho_H + (y'')\rho_H \\
&= (x\alpha_1\cdots\alpha_{m-2}w\beta_{m-2}\cdots\beta_1 y', y'')\delta \\
&\qquad + (x\alpha_{m-1}w\beta_{m-1}y', y'')\delta - (xwy', y'')\delta.
\end{aligned}$$

From $D = -D$ and the assumption, $(u,v)\delta$ is a positive sum of elements of D.

Therefore, by induction on $n(u,v)$, the image under δ of each $(u,v) \in \ker\,\rho_F$ can be expressed as a positive sum of elements of D $\boxed{7.5.3}$

Proof of 7.5.2. Let $(c_1,d_1)\cdots(c_k,d_k)$ be the canonical decomposition of (u,v). Since $(u,v) \in \ker\,\rho - \mathcal{S}$, there exists $(c_j,d_j) \in \ker\,\rho_F - \mathcal{R}$. Reasoning as in the proof of Theorem 4.3.1, there are three cases, two of which are parallel. For brevity, let $s = c_1\cdots c_{j-1}$, $t = c_{j+1}\cdots c_k$, $s' = d_1\cdots d_{j-1}$, and $t' = d_{j+1}\cdots d_k$.

i.) Suppose

$$c_j = x\alpha_1\cdots\alpha_{m-1}w' \text{ and } d_j = y^{-1}\beta_1^{-1}\cdots\beta_{m-1}^{-1}w''$$

where $x, \alpha_i, w', w'', \beta_i, y$ as in the proof of Theorem 4.3.1. Then the relations

$$(x\alpha_1\cdots\alpha_{m-2}w', y^{-1}\beta_1^{-1}\cdots\beta_{m-2}^{-1}w''),$$
$$(x\alpha_{m-1}w', y^{-1}\beta_{m-1}^{-1}w''), \text{ and } (xw', y^{-1}w'')$$

are in $\ker\,\rho_F$. The relations

$$(sx\alpha_1\cdots\alpha_{m-2}w't, s'y^{-1}\beta_1^{-1}\cdots\beta_{m-2}^{-1}w''t'),$$
$$(sx\alpha_{m-1}w't, s'y^{-1}\beta_{m-1}^{-1}w''t'), \text{ and } (sxw', s'y^{-1}w''t')$$

are thus also in $\ker\,\rho_F$. Now, since $(u,v)\delta = 0_H$,

$$\begin{aligned}
(sx\alpha_1\cdots\alpha_{m-2}w't, s'y^{-1}\beta_1^{-1}\cdots\beta_{m-2}^{-1}w''t')\delta &= -(\alpha_{m-1})\rho_H + (\beta_{m-1}^{-1})\rho_H, \\
(sx\alpha_{m-1}w't, s'y^{-1}\beta_{m-1}^{-1}w''t')\delta &= -(\alpha_1\cdots\alpha_{m-2})\rho_H + (\beta_{m-2}^{-1}\cdots\beta_1^{-1})\rho_H, \\
(sxw't, s'y^{-1}w''t')\delta &= -(\alpha_1\cdots\alpha_{m-1})\rho_H + (\beta_{m-1}^{-1}\cdots\beta_1^{-1})\rho_H.
\end{aligned}$$

Lemma 7.5.3 asserts that there are positive sums of elements of D that equal the left-hand sides of the above equations. That is, one can choose elements of $\mathcal{R}$, concatenate them, and get a relation in $\ker\,\rho_F$ whose image under δ takes one of these three values. By switching the two sides of such a relation, one can invert the image under δ. Therefore choose (p,q) and (p',q') in $\mathcal{R}^+$ such that $(p,q)\delta = (\alpha_{m-1})\rho_H - (\beta_{m-1}^{-1})\rho_H$ and $(p',q')\delta = (\alpha_1 \cdots \alpha_{m-2})\rho_H - (\beta_{m-2}^{-1} \cdots \beta_1^{-1})\rho_H$. Observe that the relations (pwq, qwp) and $(p'wq', q'wp')$ are consequences of $\mathcal{Q}$ for any word $w \in A^*$, and that

$$(p'p, q'q)\delta = (p,q)\delta + (p',q')\delta = (\alpha_1 \cdots \alpha_{m-1})\rho_H - (\beta_{m-1}^{-1} \cdots \beta_1^{-1})\rho_H.$$

Using these relations (p,q) and (p',q') as 'compensation', one obtains the relations

$$\begin{aligned}&(psx\alpha_1 \cdots \alpha_{m-2}w't, qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}w''t'),\\ &\qquad (sx\alpha_{m-1}w'tp', s'y^{-1}\beta_{m-1}^{-1}w''t'q'), \text{ and } (q'qsxw', p'ps'y^{-1}w''t'), \quad (4)\end{aligned}$$

which lie in ker ρ since their images under δ are 0_H. By the choice of (p,q) and (p',q'), the relations (4) precede (u,v) in the $\ll$-ordering: for example, the decomposition of the first relation

$$(p,q)(s,s')(x\alpha_1 \cdots \alpha_{m-2}w', y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}w'')(t,t')$$

has a lesser value of n'', or the same n''-value and a smaller p''-value, than the canonical decomposition of (u,v), so certainly the canonical decomposition of the first relation must have the same property.

The following Malcev chain shows that (u,v) is a Malcev consequence of the relations (4) and those in $\mathcal{Q}$:

$$\begin{aligned}
&sx\alpha_1 \cdots \alpha_{m-1}w't \\
&\to p^{\mathsf{L}}psx\alpha_1 \cdots \alpha_{m-2}w'tt^{\mathsf{R}}(w')^{\mathsf{R}}\alpha_{m-1}w't \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}w''t't^{\mathsf{R}}(w')^{\mathsf{R}}\alpha_{m-1}w't && \text{by (4)} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}(q')^{\mathsf{L}}q'qs'y^{-1}w''t't^{\mathsf{R}}(w')^{\mathsf{R}}\alpha_{m-1}w't \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}(q')^{\mathsf{L}}p'psxw'tt^{\mathsf{R}}(w')^{\mathsf{R}}\alpha_{m-1}w't && \text{by (4)} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}(q')^{\mathsf{L}}p'psx\alpha_{m-1}w't \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}(q')^{\mathsf{L}}p'psx\alpha_{m-1}w'tq'(q')^{\mathsf{R}} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}(q')^{\mathsf{L}}q'psx\alpha_{m-1}w'tp'(q')^{\mathsf{R}} && \text{by } \mathcal{Q}^{\#} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}psx\alpha_{m-1}w'tp'(q')^{\mathsf{R}} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}ps'y^{-1}\beta_{m-1}^{-1}w''t'q'(q')^{\mathsf{R}} && \text{by (4)} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}ps'y^{-1}\beta_{m-1}^{-1}w''t' \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}ps'y^{-1}\beta_{m-1}^{-1}w''t'qq^{\mathsf{R}} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-2}^{-1}(y^{-1})^{\mathsf{L}}(s')^{\mathsf{L}}q^{\mathsf{L}}qs'y^{-1}\beta_{m-1}^{-1}w''t'pq^{\mathsf{R}} && \text{by } \mathcal{Q}^{\#} \\
&\to p^{\mathsf{L}}qs'y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1}w''t'pq^{\mathsf{R}} \\
&\to p^{\mathsf{L}}ps'y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1}w''t'qq^{\mathsf{R}} && \text{by } \mathcal{Q}^{\#} \\
&\to s'y^{-1}\beta_1^{-1} \cdots \beta_{m-1}^{-1}w''t'.
\end{aligned}$$

ii.) Suppose

$$c_j = x\alpha_1 \cdots \alpha_{m-1} w \beta_{m-1} \cdots \beta_1 y' \text{ and } d_j = y''$$

where $x, \alpha_i, w, \beta_i, y', y''$ are as in the proof of Theorem 4.3.1. Then the relations

$$(x\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y', y''), (x\alpha_{m-1} w \beta_{m-1} y', y''), \text{ and } (xwy', y'')$$

are in ker ρ_F. Therefore the relations:

$$(sx\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y't, s'y''t'),$$
$$(sx\alpha_{m-1} w \beta_{m-1} y't, s'y''t'), \text{ and } (sxwy't, s'y''t')$$

are also in ker ρ_F. Since $(u, v)\delta = 0_H$,

$$(sx\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y't, s'y''t')\delta = -(\alpha_{m-1})\rho_H - (\beta_{m-1})\rho_H,$$
$$(sx\alpha_{m-1} w \beta_{m-1} y't, s'y''t')\delta = -(\alpha_1 \cdots \alpha_{m-2})\rho_H - (\beta_{m-2} \cdots \beta_1)\rho_H,$$
$$(sxwy't, s'y''t')\delta = -(\alpha_1 \cdots \alpha_{m-1})\rho_H - (\beta_{m-1} \cdots \beta_1)\rho_H.$$

Choose 'compensation' relations (p, q) and (p', q') in $\mathcal{R}^+$ such that $(p, q)\delta = (\alpha_{m-1})\rho_H + (\beta_{m-1})\rho_H$ and $(p', q')\delta = (\alpha_1 \cdots \alpha_{m-2})\rho_H + (\beta_{m-2}^{-1} \cdots \beta_1)\rho_H$. The relations

$$(sx\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y'tp, s'y''t'q),$$
$$(p'sx\alpha_{m-1} w \beta_{m-1} y't, q's'y''t'), \text{ and } (pp'sxwy't, qq's'y''t') \quad (5)$$

are in ker ρ and precede (u, v) in the $\ll$-ordering. The following Malcev chain shows that the relation (u, v) is a Malcev consequence of the relations (5) and those in $\mathcal{Q}$:

$$\begin{aligned}
& sx\alpha_1 \cdots \alpha_{m-1} w \beta_{m-1} \cdots \beta_1 y't \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} p' sx\alpha_{m-1} w \beta_{m-1} y' t t^{\mathsf{R}} (y')^{\mathsf{R}} \beta_{m-2} \cdots \beta_1 y't \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} q' s' y'' t' t^{\mathsf{R}} (y')^{\mathsf{R}} \beta_{m-2} \cdots \beta_1 y't && \text{by } (5) \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} q^{\mathsf{L}} q q' s' y'' t' t^{\mathsf{R}} (y')^{\mathsf{R}} \beta_{m-2} \cdots \beta_1 y't \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} q^{\mathsf{L}} p p' sxwy' t t^{\mathsf{R}} (y')^{\mathsf{R}} \beta_{m-2} \cdots \beta_1 y't && \text{by } (5) \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} q^{\mathsf{L}} p p' sxw \beta_{m-2} \cdots \beta_1 y' t q q^{\mathsf{R}} \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} x^{\mathsf{L}} s^{\mathsf{L}} (p')^{\mathsf{L}} q^{\mathsf{L}} q p' sxw \beta_{m-2} \cdots \beta_1 y' t p q^{\mathsf{R}} && \text{by } \mathcal{Q}^{\#} \\
\to\ & sx\alpha_1 \cdots \alpha_{m-2} w \beta_{m-2} \cdots \beta_1 y' t p q^{\mathsf{R}} \\
\to\ & s' y'' t' q q^{\mathsf{R}} && \text{by } (5) \\
\to\ & s' y'' t'
\end{aligned}$$

In either case, $(u, v) \in \ker \rho - \mathcal{S}$ is a Malcev consequence of $\ll$-preceding elements of ker ρ plus relations from $\mathcal{Q}$. $\boxed{7.5.2}$

Second stage. The reasoning thus far has reduced the presentation $\mathrm{Sg}\langle A \mid \ker \rho\rangle$ for S to the Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{S} \cup \mathcal{Q}\rangle$. However, the set of relations $\mathcal{S} \cup \mathcal{Q}$ is infinite. The next stage is to show that all relations in $\mathcal{S} \cup \mathcal{Q}$ are Malcev consequences of those in a still infinite — but simpler — set $\mathcal{T}$.

This section is rather technical, so a few motivational remarks will be made immediately after some definitions required later.

Let $K = (\mathbb{N} \cup \{0\})^{2n}$ and $N = \mathbb{Z}^n$. Define $\delta' : K \to H$ by

$$(a_1, a_1', \ldots, a_n, a_n') \mapsto \sum_{i=1}^{n} \left[a_i(u_i, v_i)\delta + a_i'(v_i, u_i)\delta\right].$$

(Recall that $\mathcal{R} = \{(u_1, v_1), (v_1, u_1), \ldots, (u_n, v_n), (v_n, u_n)\}$.)

Define $\sigma : K \to N$ by

$$(a_1, a_1', \ldots, a_n, a_n') \mapsto (a_1 - a_1', \ldots, a_n - a_n').$$

Let $\delta'' : N \to H$ be given by

$$(b_1, \ldots, b_n) \mapsto \sum_{i=1}^{n} b_i(u_i, v_i)\delta.$$

Recall that every relation in $\mathcal{S}$ is a concatenation of elements of $\mathcal{R}$. It is obvious, therefore, that $\mathcal{S} \subseteq \mathcal{R}^{\#}$. However, although all elements of $\mathcal{R}$ lie in $\ker \rho_F$, they may have non-zero image under δ. (Recall that $(u, v)\delta = u\rho_H - v\rho_H$.) Suppose $(u, v) = (c_1, d_1) \cdots (c_k, d_k)$, with $(c_j, d_j) \in \mathcal{R}$ and that this decomposition contains a_i instances of (u_i, v_i) and a_i' instances of (v_i, u_i) for each i. Record this fact using a tuple $T = (a_1, a_1', a_2, a_2', \ldots, a_n, a_n') \in K$. Notice that the image of T under δ' coincides with the image of (u, v) under δ. So the tuple corresponding to any relation in $\ker \rho$ must also have image 0_H under δ'. Now, as the contributions of each a_i and a_i' to the sum (5) are mutually inverse, one may pass to a tuple $(a_1 - a_1', \ldots, a_n - a_n')$ in N (using the mapping σ) and still be able to obtain the image of T under δ' using the mapping δ''. (Lemma 7.5.4 formalizes this notion.) The kernel of δ'' (in the group-theoretical sense) is a subgroup of the finitely generated abelian group N and is therefore itself finitely generated. The strategy is to pick a finite [semigroup] generating set for this kernel, pull this set back to a set of tuples Y in K, and thus to find a particular set $\mathcal{T}$ consisting of relations formed by concatenating relations from $\mathcal{R}$ and trivial relations (a, a) such that the number of (u_i, v_i) and (v_i, u_i) in a particular pair is described by a tuple in Y. (Thus each relation in $\mathcal{T}$ has image 0_H under δ.) The aim is to express the tuple T as a positive (semigroup) sum of tuples $\sum_{j \in J} y_j$ with $y_j \in Y$. The definition of the relations in $\mathcal{T}$ then allows the construction of a Malcev chain from u to v in which there is a one-to-one correspondence between the steps of the chain and the y_j in the sum of tuples.

The details of the reasoning are, however, quite delicate: a number of technical difficulties arise in pulling back the generators of the kernel of δ'' to tuples in K. Lemmata 7.5.4–7.5.6 show how to surmount these problems.

Lemma 7.5.4. $\sigma\delta'' = \delta'$.

Proof of 7.5.4. Let $(a_1, a_1', \ldots, a_n, a_n') \in K$. Then

$$\begin{aligned}(a_1, a_1', \ldots, a_n, a_n')\sigma\delta'' &= (a_1 - a_1', \ldots, a_n - a_n')\delta'' \\ &= \sum_{i=1}^{n}(a_i - a_i')(u_i, v_i)\delta \\ &= \sum_{i=1}^{n}[a_i(u_i, v_i)\delta - a_i'(u_i, v_i)\delta] \\ &= \sum_{i=1}^{n}[a_i(u_i, v_i)\delta + a_i'(v_i, u_i)\delta] \\ &= (a_1, a_1', \ldots, a_n, a_n')\delta'.\end{aligned}$$

Therefore $\sigma\delta'' = \delta'$. $\boxed{7.5.4}$

Now, Ker δ'' is a subgroup of $N = \mathbb{Z}^n$ and so is finitely generated. Let X be a finite semigroup generating set for Ker δ''. [This proof adheres to a notational distinction between ker ϕ, which denotes a congruence on the domain of a homomorphism ϕ, and Ker ψ, which is a normal subgroup of the domain of a group homomorphism ψ.]

Let $\tau : N \to K$ be defined by

$$(b_1, \ldots, b_n) \mapsto (\max\{b_1, 0\}, \max\{-b_1, 0\}, \ldots, \max\{b_n, 0\}, \max\{-b_n, 0\}).$$

Observe that $\tau\sigma = \mathrm{id}_N$. This observation and Lemma 7.5.5 will together show that τ is 'almost' an inverse of σ. Use τ to pull back X into K as follows: Let $Y' = X\tau$. Let $Y'' \subseteq K$ be the set

$$\{(1, 1, 0, 0, \ldots, 0, 0), (0, 0, 1, 1, \ldots, 0, 0), \ldots, (0, 0, 0, 0, \ldots, 1, 1)\},$$

and let $Y = Y' \cup Y''$.

Unfortunately, the composition $\sigma\tau$ is not the identity mapping. However, it is 'close enough' for the purposes of this proof, in a sense made precise by the following lemma:

Lemma 7.5.5. *For* $(a_1, a_1', \ldots, a_n, a_n') \in K$,

$$(a_1, a_1', \ldots, a_n, a_n') - (a_1, a_1', \ldots, a_n, a_n')\sigma\tau \in \mathrm{Mon}\langle Y''\rangle.$$

Proof of 7.5.5. Let $(\ldots, a_i, a_i', \ldots) \in K$. Then

$$\begin{aligned}(\ldots, a_i, a_i', \ldots)\sigma\tau &= (\ldots, a_i - a_i', \ldots)\tau \\ &= (\ldots, \max\{a_i - a_i', 0\}, \max\{a_i' - a_i, 0\}, \ldots).\end{aligned}$$

If $a_i \geq a_i'$, then this gives $(\ldots, a_i - a_i', 0, \ldots)$, and

$$(\ldots, a_i, a_i', \ldots) - (\ldots, a_i - a_i', 0, \ldots) = a_i'(\ldots, 1, 1, \ldots) + \ldots.$$

If $a_i \leq a_i'$, then this gives $(\ldots, 0, a_i' - a_i, \ldots)$, and

$$(\ldots, a_i, a_i', \ldots) - (\ldots, 0, a_i' - a_i, \ldots) = a_i(\ldots, 1, 1, \ldots) + \ldots.$$

Reasoning thus for each i gives the result. $\boxed{7.5.5}$

Similarly, τ is 'close' to being a homomorphism:

Lemma 7.5.6. *For $(b_1, \ldots, b_n)$ and $(c_1, \ldots, c_n)$ in N,*

$$(b_1, \ldots, b_n)\tau + (c_1, \ldots, c_n)\tau - (b_1 + c_1, \ldots, b_n + c_n)\tau \in \mathrm{Mon}\langle Y'' \rangle .$$

Proof of 7.5.6. Let $(\ldots, b_i, \ldots)$ and $(\ldots, c_i, \ldots)$ be members of N. Then

$$\begin{aligned}
&(\ldots, b_i, \ldots)\tau + (\ldots, c_i, \ldots)\tau \\
&= (\ldots, \max\{b_i, 0\}, \max\{-b_i, 0\}, \ldots) + (\ldots, \max\{c_i, 0\}, \max\{-c_i, 0\}, \ldots) \\
&= (\ldots, \max\{b_i, 0\} + \max\{c_i, 0\}, \max\{-b_i, 0\} + \max\{-c_i, 0\}, \ldots).
\end{aligned}$$

Consider the following four cases:

i.) $b_i, c_i \geq 0$. This gives $(\ldots, b_i, \ldots)\tau + (\ldots, c_i, \ldots)\tau = (\ldots, b_i + c_i, 0, \ldots)$, and

$$\begin{aligned}
&(\ldots, b_i + c_i, 0, \ldots) - (\ldots, b_i + c_i, \ldots)\tau \\
&= (\ldots, b_i + c_i, 0, \ldots) - (\ldots, b_i + c_i, 0, \ldots) \\
&= (\ldots, 0, 0, \ldots).
\end{aligned}$$

ii.) $b_i, c_i \leq 0$. This gives $(\ldots, b_i, \ldots)\tau + (\ldots, c_i, \ldots)\tau = (\ldots, 0, -b_i - c_i, \ldots)$, and

$$\begin{aligned}
&(\ldots, 0, -b_i - c_i, 0, \ldots) - (\ldots, b_i + c_i, \ldots)\tau \\
&= (\ldots, 0, -b_i - c_i, 0, \ldots) - (\ldots, 0, -b_i - c_i, \ldots) \\
&= (\ldots, 0, 0, \ldots).
\end{aligned}$$

iii.) $b_i \geq 0$, $c_i \leq 0$. Now split into two sub-cases:

a.) $|b_i| \geq |c_i|$. This gives $(\ldots, b_i, \ldots)\tau + (\ldots, c_i, \ldots)\tau = (\ldots, b_i, -c_i, \ldots)$, and

$$\begin{aligned}
&(\ldots, b_i, -c_i, 0, \ldots) - (\ldots, b_i + c_i, \ldots)\tau \\
&= (\ldots, b_i, -c_i, 0, \ldots) - (\ldots, b_i + c_i, 0, \ldots) \\
&= (\ldots, b_i - b_i - c_i, -c_i \ldots) \\
&= (\ldots, -c_i, -c_i, \ldots).
\end{aligned}$$

(Observe that $-c_i \geq 0$.)

b.) $|b_i| \leq |c_i|$. This gives $(\ldots, b_i, \ldots)\tau + (\ldots, c_i, \ldots)\tau = (\ldots, b_i, -c_i, \ldots)$, and

$$\begin{aligned}&(\ldots, b_i, -c_i, 0, \ldots) - (\ldots, b_i + c_i, \ldots)\tau \\ &= (\ldots, b_i, -c_i, 0, \ldots) - (\ldots, 0, -(b_i + c_i), \ldots) \\ &= (\ldots, b_i, -c_i + b_i + c_i \ldots) \\ &= (\ldots, b_i, b_i, \ldots).\end{aligned}$$

iv.) $b_i \leq 0$, $c_i \geq 0$. This reasoning parallels that of case iii.

Apply the four cases above to each i to complete the proof. $\boxed{7.5.6}$

Finally, although Ker δ' is not guaranteed to be a subset of Mon$\langle Y\rangle$, any element of Ker δ' differs from some element of Mon$\langle Y\rangle$ only by an element of Mon$\langle Y''\rangle$:

Lemma 7.5.7. *If $(a_1, a'_1, \ldots, a_n, a'_n)\delta' = 0_H$, then there exists $y \in$ Mon$\langle Y''\rangle$ such that $(a_1, a'_1, \ldots, a_n, a'_n) + y \in$ Mon$\langle Y\rangle$.*

Proof of 7.5.7. Let $(a_1, a'_1, \ldots, a_n, a'_n)\delta' = 0_H$. Applying Lemma 7.5.4 shows that $(a_1, a'_1, \ldots, a_n, a'_n)\sigma \in$ Ker δ''. So

$$\begin{aligned}(a_1, a'_1, \ldots, a_n, a'_n)\sigma &= \sum_{j\in J} x_j && \text{for some } x_j \in X, \\ (a_1, a'_1, \ldots, a_n, a'_n)\sigma\tau &= \Big(\sum_{j\in J} x_j\Big)\tau, \\ (a_1, a'_1, \ldots, a_n, a'_n)\sigma\tau + y &= \sum_{j\in J}(x_j\tau) && \text{where } y \in \mathrm{Mon}\langle Y''\rangle \text{, by Lemma 7.5.6,} \\ (a_1, a'_1, \ldots, a_n, a'_n) + y &= \sum_{j\in J}(x_j\tau) + z && \text{where } z \in \mathrm{Mon}\langle Y''\rangle \text{, by Lemma 7.5.5,} \\ &\in \mathrm{Mon}\langle Y\rangle,\end{aligned}$$

and this completes the proof. $\boxed{7.5.7}$

Let $\mathcal{T}$ consist of all relations in $(\mathcal{R} \cup \{(a, a) : a \in A\})^+$, containing, for some $(a_1, a'_1, \ldots, a_n, a'_n) \in Y$, exactly a_i instances of (u_i, v_i) and a'_i instances of (v_i, u_i), for each i. Observe that $\mathcal{T}$ contains $\mathcal{Q}$ since Y contains Y''. (Notice that the definition of $\mathcal{T}$ makes no mention of *canonical* decompositions.) When a relation in $\mathcal{T}$ is applied to a word over A, the a_i instances of u_i change to v_i, and the a'_i instances of v_i change to u_i, and no other letters alter. Call these unchanged intermediate letters 'padding'.

Suppose $(u, v) \in \mathcal{T}$. Then obviously $(u, v) \in \ker \rho_F$. Also, $(u, v) \in \ker \rho_H$, since

$$(u, v)\delta = (a_1, a'_1, \ldots, a_n, a'_n)\delta' = 0_H,$$

where $(a_1, a'_1, \ldots, a_n, a'_n) \in Y$. So $\mathcal{T} \subseteq \ker \rho$.

Define $\chi : \mathcal{S} \to K$ as follows: $w \in \mathcal{S}$ maps to $(a_1, a_1', \ldots, a_n, a_n')$, where a_i is the number of (u_i, v_i) in the canonical decomposition of w, and a_i' is the number of (v_i, u_i).

Lemma 7.5.8. *Every relation in $\mathcal{S}$ is a Malcev consequence of those in $\mathcal{T}$.*

Proof of 7.5.8. Let $(u, v) \in \mathcal{S}$. Then $(u, v) \in \ker \rho_F \cap \ker \rho_H$, so $(u, v)\chi \in \operatorname{Ker} \delta'$. Let $y \in \operatorname{Mon}\langle Y'' \rangle$ be such that $(u, v)\chi + y \in \operatorname{Mon}\langle Y \rangle$. Suppose $y = (b_1, b_1, \ldots, b_n, b_n)$. Let

$$(s,t) = (u_1, v_1)^{b_1}(v_1, u_1)^{b_1} \cdots (u_n, v_n)^{b_n}(v_n, u_n)^{b_n}. \tag{6}$$

Observe that $(s, t) \in \mathcal{Q}^{\#}$.

Suppose that $(u, v)\chi + y = \sum_{j=1}^{p} y_j$, where $y_j \in Y$. Construct a Malcev $\mathcal{T}$-chain from u to v as follows. First of all, insert t and transform it to s using relations from $\mathcal{Q} \subseteq \mathcal{T}$:

$$u \to utt^{\mathsf{R}} \to ust^{\mathsf{R}}. \tag{7}$$

The component a_i of the tuple $(u, v)\chi + y$ describes the number of (u_i, v_i) that appear in the canonical decomposition of (u, v) concatenated with the decomposition (6) of (s, t). A similar statement applies to a_i' and (v_i, u_i). Put another way, the components a_i and a_i' of the tuple $(u, v)\chi + y$ describes the number of subwords u_i and v_i of us that must be changed to v_i and u_i, respectively, in order to transform us to vt. As $(u, v)\chi + y = \sum_{j=1}^{p} y_j$, the sum of the various components of the y_j also gives the number of subwords of each type that must be changed.

Construct a Malcev chain from us to vt by defining the j-th step in the chain (where $j = 1, \ldots, p$) as follows. Suppose $y_j = (a_1, a_1', \ldots, a_n, a_n')$ and that the first $j-1$ steps have transformed us to w. For each i, find a_i subwords u_i and a_i' subwords v_i of w that have not been changed in the chain thus far. The j-th step consists of changing those words u_i to v_i and v_i to u_i. By the definition of $\mathcal{T}$, a relation exists that permits this step. Furthermore, by the comments in the last paragraph, the word left after the p-th step is vt.

Concatenate the Malcev chain (7) with the one just constructed and append

$$vtt^{\mathsf{R}} \to v$$

to obtain a Malcev $\mathcal{T}$-chain from u to v. $\boxed{7.5.8}$

Third stage. The proof thus far has shown that S has a Malcev presentation $\operatorname{SgM}\langle A \mid \mathcal{T} \rangle$. The set $\mathcal{T}$ is still infinite. This third and final stage shows that a finite subset $\mathcal{U}$ of $\mathcal{T}$ will suffice in a Malcev presentation for S.

Let $\mathcal{U}$ be the subset of $\mathcal{T}$ where each string of padding is either empty or one letter long — elements of $\mathcal{R}$ are either adjacent or separated by a single (a, a) for some $a \in A$. Observe that the set $\mathcal{U}$ is finite because Y — which dictates how many elements of $\mathcal{R}$ can appear — is finite.

Lemma 7.5.9. *Every relation in $\mathcal{T}$ is a Malcev consequence of those in $\mathcal{U}$.*

Proof of 7.5.9. Let $(\alpha w\beta, \gamma w\zeta)$ be a relation in $\mathcal{T}$, with (w, w) being padding. Suppose $(w, w) = (a, a)(w', w')$ for $a \in A$. The relations

$$(\alpha a\beta, \gamma a\zeta), (\alpha w'\beta, \gamma w'\zeta), (\alpha\beta, \gamma\zeta) \tag{8}$$

are also in $\mathcal{T}$.

The following Malcev chain shows that $(\alpha w\beta, \gamma w\zeta)$ is a Malcev consequence of the relations (8):

$$\begin{aligned}
&\alpha w\beta \\
&= \alpha a w'\beta \\
&\to \alpha a\beta\beta^{\mathsf{R}} w'\beta \\
&\to \gamma a\zeta\beta^{\mathsf{R}} w'\beta \\
&\to \gamma a\gamma^{\mathsf{L}}\gamma\zeta\beta^{\mathsf{R}} w'\beta \\
&\to \gamma a\gamma^{\mathsf{L}}\alpha\beta\beta^{\mathsf{R}} w'\beta \\
&\to \gamma a\gamma^{\mathsf{L}}\alpha w'\beta \\
&\to \gamma a\gamma^{\mathsf{L}}\gamma w'\zeta && \text{(by induction on } |w|) \\
&\to \gamma a w'\zeta \\
&= \gamma w\zeta.
\end{aligned}$$

Apply such reasoning to every padding string in the relation to show that it is a Malcev consequence of $\mathcal{U}$. $\boxed{7.5.9}$

Conclusion. By Lemmata 7.5.2, 7.5.8, and 7.5.9,

$$\ker\rho = \mathcal{S}^{\mathrm{M}} = \mathcal{T}^{\mathrm{M}} = \mathcal{U}^{\mathrm{M}}.$$

Therefore S admits the finite Malcev presentation $\mathrm{SgM}\langle A \mid \mathcal{U}\rangle$. Since S was an arbitrary finitely generated subsemigroup G, the group G — which was an arbitrary direct product of a virtually free group and an abelian group — is Malcev coherent. $\boxed{7.5.1}$

The following is an explicit example of a non-automatic finitely generated subsemigroup of the direct product of a free semigroup and the natural numbers. Although not of the utmost interest here, except perhaps to show that one could not use automatism to prove Theorem 7.5.1, it is called upon in Subsection 8.2.3.

Example 7.5.10. Let $A = \{a, b, c, d, e, f, g, h\}$ be an alphabet representing elements of $\{x, y, p, q, r\}^+ \times \mathbb{N}$ as follows:

$$\begin{aligned}
\overline{a} &= (x^2p, 0), \\
\overline{b} &= (qrp, 1), & \overline{f} &= (x^2pq, 0), \\
\overline{c} &= (qr, 0), & \overline{g} &= (rpqrpq, 1), \\
\overline{d} &= (pqr, 0), & \overline{h} &= (rpy^2, 0). \\
\overline{e} &= (py^2, 0),
\end{aligned}$$

Let S be the semigroup generated by $\overline{A}$.

Proposition 7.5.11. *The semigroup S is not asynchronously automatic.*

[The following proof is very similar to that of Proposition 5.5.2.]

Proof of 7.5.11. First of all, notice that for each $\alpha \in \mathbb{N}$,

$$\overline{ab^\alpha cd^\alpha e} = (x^2(pqr)^{2\alpha+1}qy^2, \alpha) = \overline{fg^\alpha h}.$$

Elementary reasoning shows that for each $\alpha \in \mathbb{N} \cup \{0\}$ the elements $(x^2(pqr)^{2\alpha+1}, \alpha)$ and $\overline{fg^\alpha} = (x^2(pqr)^{2\alpha}pq, \alpha)$ have unique representives $ab^\alpha cd^\alpha$ and fg^α over the alphabet A.

Suppose that S is asynchronously automatic. Then it admits an asynchronous automatic structure (A, L), and, in particular, the relation $L_e \circ L_h^{-1}$ is rational. Let N be the number of states in an asynchronous automaton recognizing $L_e \circ L_h^{-1}$.

The language L must contain the words $ab^\alpha cd^\alpha$ and fg^α for each $\alpha \in \mathbb{N} \cup \{0\}$, and so $L_e \circ L_h^{-1}$ contains the pair $(ab^\alpha cd^\alpha, fg^\alpha)$.

Fix $\alpha > N$. Proceed as in the proof of Proposition 5.5.2 to show that one of the following two cases holds:

i.) For some $\beta, \gamma, \eta \in \mathbb{N} \cup \{0\}$ with $\beta < \gamma$,

$$(ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha, fg^\eta fg^\alpha) \in L_e \circ L_h^{-1},$$

whence

$$\overline{ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha e} = \overline{fg^\eta fg^\alpha h}.$$

This is a contradiction, since the left-hand side is

$$(x^2(pqr)^{2\alpha+\gamma-\beta+1}py^2, \alpha+\gamma-\beta)$$

and the right-hand side is

$$(x^2(pqr)^{2\eta}pqx^2(pqr)^{2\alpha+1}rpy^2, \alpha+\eta).$$

ii.) For some $\beta, \gamma, \eta \in \mathbb{N} \cup \{0\}$ with $\beta < \gamma$,

$$(ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha, fg^{\alpha+\eta}) \in L_e \circ L_h^{-1}.$$

Therefore

$$\overline{ab^\beta b^{2(\gamma-\beta)} b^{\alpha-\gamma} cd^\alpha e} = \overline{fg^{\alpha+\eta} h}.$$

The left-hand side is

$$(x^2(pqr)^{2\alpha+\gamma-\beta+1}py^2, \alpha+\gamma-\beta)$$

and the right-hand side is

$$(x^2(pqr)^{2(\alpha+\eta)+1}py^2, \eta+\alpha).$$

For the $\mathbb{N}$-components to match, $\eta = \gamma - \beta$. However, for the free semigroup components to match, $2\alpha+\gamma-\beta = 2(\alpha+\eta)+1$, which implies that $2\eta = \gamma-\beta$. So $\gamma - \beta = 0$, which contradicts $\beta < \gamma$.

Therefore S cannot be asynchronously automatic. $\boxed{7.5.11}$

7.6. FROM ABELIAN GROUPS TO NILPOTENT GROUPS

THEOREM 7.4.1 asserts that the direct product of a free group and a polycyclic group is coherent. Thus, by Proposition 1.6.1, the direct product of a virtually free group and a virtually polycyclic group is also coherent. This immediately provokes questions about how far Theorem 7.5.1 extends: can one replace the abelian group by a virtually abelian, a nilpotent, or a virtually nilpotent group?

Open Problem 7.6.1. Is every direct product of a [virtually] free group and a virtually abelian group Malcev coherent?

Open Problem 7.6.2. Is every direct product of a [virtually] free group and a [virtually] nilpotent group Malcev coherent?

Of course, Corollary 7.3.10 implies that Theorem 7.5.1 does not extend in general to the direct product of a free group and a polycyclic group.

CHAPTER EIGHT

ONE-RELATOR SEMIGROUPS & GROUPS

. . . one true inference invariably suggests others.
(spoken by Sherlock Holmes)
— Arthur Conan Doyle, *Silver Blaze* (1892)

8.1. INTRODUCTION

ONE-RELATOR GROUPS have long been and continue to be a popular subject of study. By comparison, the field of one-relator semigroups shows little activity, except perhaps regarding the question of whether the word problem for one-relator semigroups is soluble. [Magnus (1932) proved that one-relator groups have soluble word problem.] The present chapter studies subsemigroups of one-relator groups and one-relator cancellative semigroups. [One-relator cancellative semigroups are always group-embeddable by Adjan's Theorem 0.9.12 and so form positive subsemigroups of one-relator groups.]

The majority of the chapter is devoted to Baumslag–Solitar semigroups, which are positive subsemigroups of the noted Baumslag–Solitar groups. Subsections 8.2.1 and 8.2.2 establish that finitely generated subsemigroups of 'almost all' Baumslag–Solitar semigroups are either automatic or left-automatic, and, consequently, have finite Malcev presentations. Subsection 8.2.3 proves that automatism does not extend to subsemigroups of the remaining Baumslag–Solitar semigroups. [Subsections 8.2.1–8.2.3 are based on Cain (2005*a*).]

The final part of the chapter, Section 8.3, speculates on the Malcev coherence of other one-relator groups and cancellative semigroups.

8.2. BAUMSLAG–SOLITAR SEMIGROUPS

BAUMSLAG & SOLITAR (1962) introduced groups with presentations of the form

$$\mathrm{Gp}\langle x, y \mid (yx^m, x^n y)\rangle, \tag{1}$$

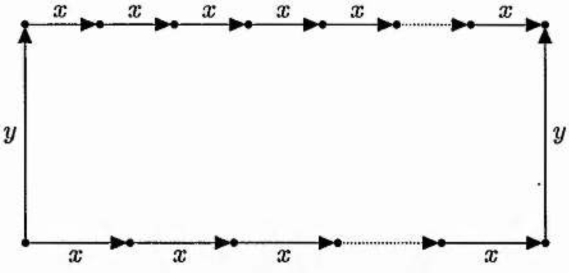

Figure 8.1. Fragment of the Cayley graph of the Baumslag–Solitar semigroup $\mathrm{BSS}(m, n)$.

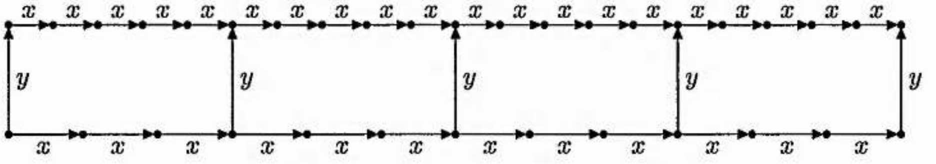

Figure 8.2. Fragment of the Cayley graph of the Baumslag–Solitar semigroup $\mathrm{BSS}(5, 3)$.

where m and n are natural numbers, in order to answer certain questions about Hopfian and non-Hopfian groups. [A group is *Hopfian* if it is not isomorphic to any of its proper factor groups (Lyndon & Schupp 1977, p. 14).]

Denote by $\mathrm{BSG}(m, n)$ the particular Baumslag–Solitar group presented by (1). Each group $\mathrm{BSG}(m, n)$ is an HNN extension of $\mathrm{Gp}\langle \overline{x} \rangle \simeq \mathbb{Z}$ with stable letter y and associated [cyclic] subgroups $\mathrm{Gp}\langle \overline{x^m} \rangle$ and $\mathrm{Gp}\langle \overline{x^n} \rangle$. Epstein et al. (1992, Example 7.4.1) showed that $\mathrm{BSG}(m, n)$ is always asynchronously automatic, but is automatic if and only if $m = n$.

Analogously, the Baumslag–Solitar semigroups are those semigroups

$$\mathrm{BSS}(m, n) = \mathrm{Sg}\langle x, y \mid (yx^m, x^n y) \rangle,$$

where $m, n \in \mathbb{N}$. In contrast to the result for Baumslag–Solitar groups, Hoffmann (2001, Lemma 4.18) proved that $\mathrm{BSS}(m, n)$ is automatic for $m > n$. [In his study of decision and separability problems for Baumslag–Solitar semigroups, Jackson (2002) suggests that Baumslag–Solitar semigroups can have radically different properties from Baumslag–Solitar groups.] This section studies the automatism of subsemigroups of Baumslag–Solitar semigroups. The first task is to gather some information about the Cayley graphs of these semigroups.

To construct the Cayley graph, start with a single cell describing the relation $(yx^m, x^n y)$, as shown in Figure 8.1. Join copies of this cell along the edges labelled y as shown in Figure 8.2. Starting from the basepoint ω, add an infinite horizontal row R of edges, each labelled by x. Add n copies of the row of cells shown in Figure 8.2 to this row, identifying the basepoint of the $(k-1)$-th such row of cells with the k-th vertex from the left of R. Viewed side-on, the fragment of the graph constructed thus far is a 'fan' with n spokes. Now iterate this construction, taking R to be the

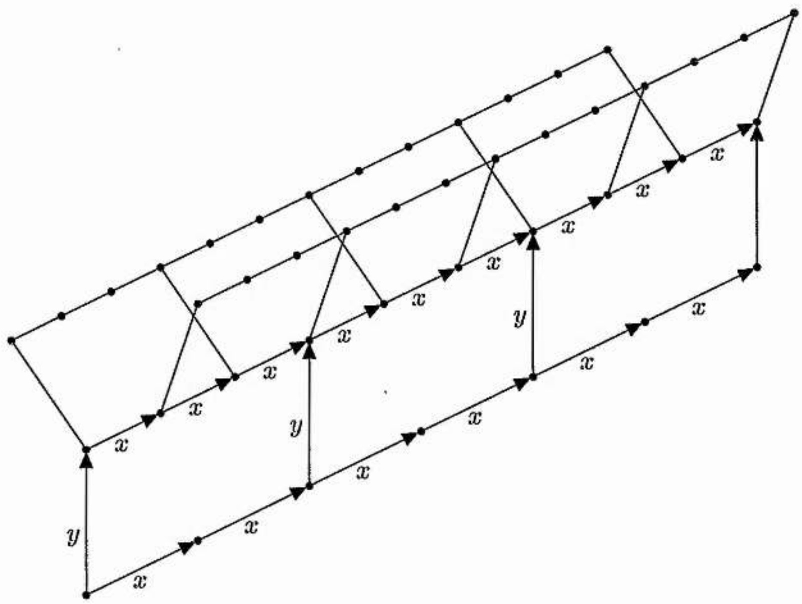

Figure 8.3. Fragment of the Cayley graph of the Baumslag–Solitar semigroup BSS(3, 2).

row of edges labelled x at the top of each row of cells just added. [Figure 8.3 shows this step in the construction for $m = 3$ and $n = 2$.] The result is the Cayley graph of BSS(m, n). Viewed side-on, this graph is an infinite tree. Select an infinite path climbing this tree and take the subset of the Cayley graph that projects to this path. Call such a subset a *branch* of the Cayley graph. [The construction of the Cayley graph just described is similar to that for the Baumslag–Solitar group BSG(m, n) given by Epstein et al. (1992, Section 7.4).]

A branch of the Cayley graph of BSS(m, n) may be embedded into the Euclidean plane as shown in Figure 8.4. Notice that all cells describing a relation $(yx^m, x^n y)$ are similar squares, being scaled by n/m as one climbs from one row to the next. Define a concept of 'horizontal distance' within rows of x-edges by taking the distance along this x-row between vertices and extending to edges by linear interpolation.

Each element of the Baumslag–Solitar semigroup BSS(m, n) has a normal form

$$yx^{k_1}yx^{k_2}\cdots yx^{k_j}yx^l,$$

where each k_i is less than n; such a normal form can be obtained from any word over x and y by using the defining relation to move letters x as far to the right as possible. Identify elements of BSS(m, n) with these normal forms.

8.2.1. Automatic subsemigroups

Let A be a finite alphabet representing a subset of BSS(m, n). Let S be the subsemigroup generated by $\overline{A}$. In the Cayley graph $\Gamma(S, A)$, one may imagine a word

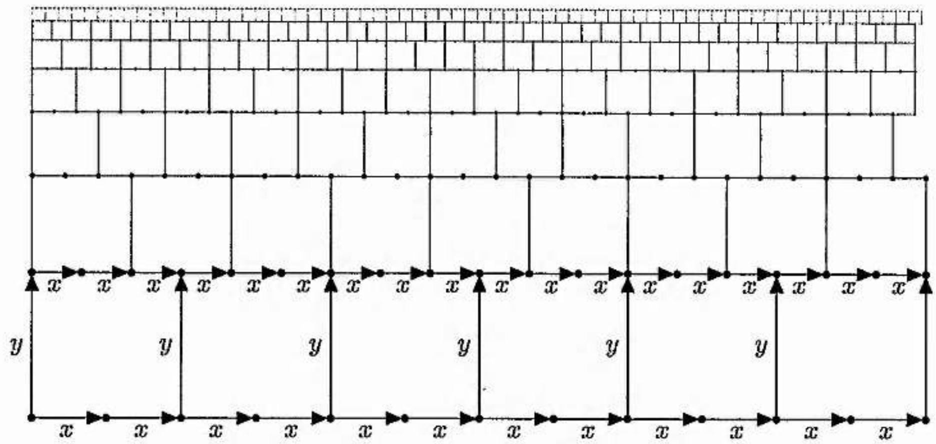

Figure 8.4. Fragment of a branch the Cayley graph of the Baumslag–Solitar semigroup BSS(3, 2) embedded into the Euclidean plane.

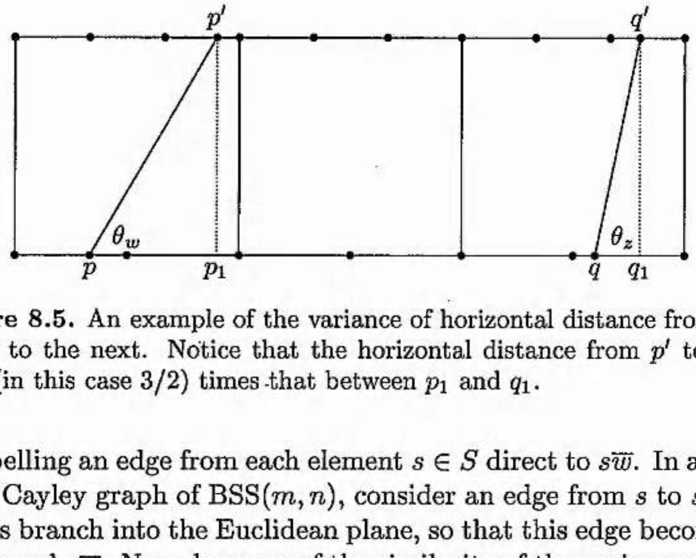

Figure 8.5. An example of the variance of horizontal distance from one x-row to the next. Notice that the horizontal distance from p' to q' is m/n (in this case $3/2$) times that between p_1 and q_1.

$w \in A^+$ as labelling an edge from each element $s \in S$ direct to $s\overline{w}$. In an appropriate branch of the Cayley graph of $\mathrm{BSS}(m, n)$, consider an edge from s to $s\overline{w}$ labelled by w. Embed this branch into the Euclidean plane, so that this edge becomes a straight line between s and $s\overline{w}$. Now, because of the similarity of the various cells mentioned above, the angle θ_w between this edge and the horizontal axis is independent of s. Notice further that this angle must lie between 0 and $\pi/2$.

Pick any two points p and q (not necessarily vertices) on an x-row. Choose any two words w and z such that θ_w and θ_z are non-zero. Consider the intersection of two lines through p and q with angles θ_w and θ_z, respectively. These lines intersect the x-row immediately above that containing p and q at points p' and q', respectively. The horizontal distance from p' to q' is given by

$$\frac{m}{n}\left|(\text{distance between } p \text{ and } q) - \frac{1}{\tan\theta_w} + \frac{1}{\tan\theta_z}\right|,$$

as can be seen from Figure 8.5. [The absolute value is needed in case the two lines cross over. The 'vertical' distance between the two x-rows is 1. Although all concepts of distance discussed here relate to the Cayley graph, one must momentarily appeal to the Euclidean plane for trigonometric purposes.] Suppose $m > n$ so that $m/n > 1$. Then, regardless of θ_w and θ_z, if the distance between p and q is larger than a certain critical value, the distance between p' and q' is larger still. This observation will prove crucial at a later stage. [This is a purely geometric remark. The words w and z may not label lines passing through a particular choice of points p, p', q and q'.]

Let A' be that subset of A whose letters represent elements of $\mathrm{BSS}(m,n)$ of the form x^k for some $k \in \mathbb{N}$ (that is, letters a such that $\theta_a = 0$). Let $A'' = A - A'$. For each $a \in A''$; let β_a be the number of symbols y in $\overline{a}$. (For all $a \in A'$, $\beta_a = 0$.)

As a consequence of the defining relation $(yx^m, x^n y)$, for any $a \in A'$,

$$y^k\overline{a^{mk}} = \overline{a^{nk}}y^k,$$

whence, for $b \in A''$ and $k = \beta_b$,

$$\overline{ba^{mk}} = \overline{a^{nk}b}.$$

Moreover, $\overline{a_1a_2} = \overline{a_2a_1}$, where $a_1, a_2 \in A'$. The upshot of this is that every element of S has a representative in the set

$$(A')^*K^* - \{\varepsilon\}, \tag{2}$$

where K is the finite set

$$\{ba_1^{\alpha_1}\cdots a_l^{\alpha_l} : b \in A'', 0 \le \alpha_i < m^{\beta_b}\}$$

and $A' = \{a_1, \ldots, a_l\}$. [In a way, this is the reverse of the set of normal forms for $\mathrm{BSS}(m,n)$ — letters representing powers of x are now moved as far *left* as possible.]

Let 1 be a new symbol representing the adjoined identity of S^1. Let k be the maximum length of any element of K. Define

$$K' = \{bw1^{\beta_b(k+1)-|w|-1} : b \in A'', bw \in K\}.$$

The set K' consists of elements bw of K padded with symbols 1 to a length that is a constant multiple of β_b. [The large constant multiple k is necessary to ensure that the exponent on 1 is always positive.] Define

$$J = \{a1^{\gamma_a - 1} : a \in A'\},$$

where $\overline{a} = x^{\gamma_a}$ for each $a \in A'$.

Let $L = J^*(K')^* - \{\varepsilon\} \cup \{1\}$. The language L differs from the set (2) only by padding using symbols 1 and the addition of the word 1 to represent the adjoined identity. Therefore, since the set (2) maps onto S, the language L maps onto S^1. The aim is now to show that $(A \cup \{1\}, L)$ is an automatic structure for S^1.

Suppose $u, v \in L$ and $a \in A \cup \{\varepsilon\}$ with $\overline{ua} = \overline{v}$. Let $u = u'u''$ and $v = v'v''$, where $u', v' \in J^*$ and $u'', v'' \in (K')^*$. The paths $\widehat{ua}$ and $\widehat{v}$ run from ω to a common vertex. As these paths never run 'downwards' through the Cayley graph, they lie in a common branch containing $\overline{ua} = \overline{v}$. Isolate such a branch and embed it into the Euclidean plane. The parts of the two paths labelled by u' and v' run along the lowest x-row.

Let $t \in \mathbb{N} \cup \{0\}$. Suppose firstly that $t \leq \min\{|u'|, |v'|\}$. As any word w in J has length equal to the number of letters x in $\overline{w}$, the same holds true for any word in J^*. Any prefix of a word in J^* is at most $M = \max\{|w| : w \in J\}$ letters short of a member of J^*; the number of letters x in $\overline{u'(t)}$ differs by at most M from t. Similar reasoning applies to v'; the distance between $\overline{u'(t)}$ and $\overline{v'(t)}$ is therefore bounded.

The words $u''a$ and v'' label subpaths from $\overline{u'}$ and $\overline{v'}$ to $\overline{ua} = \overline{v}$. Imagine these paths as made up of 'segments' $w \in K'$, with each w labelling an edge that runs directly from s to $s\overline{w}$. Consider the intersection of the paths with a given x-row, at points p and q. Let the intersections with the next x-row be p' and q'. Let w and z be the labels on the segments that run between p and p' and q and q'. (These segments may of course start *below* p and q and end *above* p' and q'.) As was observed above, if the horizontal distance between p and q exceeds a certain critical value, then the distance between p' and q' is larger still, regardless of w and z. Therefore, since the paths labelled by $u''a$ and v'' must eventually meet, the horizontal distance between their intersections with each x-row cannot exceed the maximum critical value obtained as w and z range over the finite set K'.

In particular, the points $\overline{u'}$ and $\overline{v'}$ can only be a bounded distance apart. Therefore $|u'|$ and $|v'|$ can only differ by a bounded amount.

An argument similar to that for J^* shows that if w is a prefix of a word in $(K')^*$, then the length of w differs from a constant multiple of the number of letters y in $\overline{w}$ by only a bounded amount.

Suppose now that $t \geq \min\{|u'|, |v'|\}$. Consider the elements $\overline{u(t)}$ and $\overline{v(t)}$. By altering t by a bounded amount, assume $v(t) \in v'(K')^*$. By the observation in the last paragraph, the [new] elements $\overline{u(t)}$ and $\overline{v(t)}$ lie on x-rows that are only a bounded number of elements y apart. Therefore $\overline{u(t)}$ is a bounded distance from the intersection p of the subpath labelled by u'' with the x-row containing $\overline{v(t)}$. The horizontal distance between p and $\overline{v(t)}$ cannot exceed the critical value discussed above. Therefore the distance between $\overline{u(t)}$ and $\overline{v(t)}$ in the Cayley graph of $\mathrm{BSS}(m,n)$ is bounded. Restoring the original value for t does not alter this fact. The paths $\widehat{u}$ and $\widehat{v}$ therefore fellow travel.

By Theorem 2.3.2, $(A \cup \{1\}, L)$ is an automatic structure for S^1. The semigroup S is therefore automatic by Theorem 2.4.3. Since S was an arbitrary finitely generated subsemigroup of $\mathrm{BSS}(m,n)$, this proves the following result.

Theorem 8.2.1. *Every Baumslag–Solitar semigroup* $\mathrm{BSS}(m,n)$, *where* $m > n$, *is locally automatic: each of its finitely generated subsemigroups is automatic.* $\boxed{8.2.1}$

8.2.2. Left-automatic subsemigroups

Theorem 8.2.1 has the following left-handed version (see Subsection 2.2.1 for the definition of left automatism):

Theorem 8.2.2. *Every finitely generated subsemigroup of a Baumslag–Solitar semigroup* $\mathrm{BSS}(m,n)$*, where* $m < n$*, is left automatic.*

Proof of 8.2.2. The reasoning for this proof mirrors that for Theorem 8.2.1. Firstly, notice that a branch of the Cayley graph of $\mathrm{BSS}(m,n)$ with $m < n$ resembles Figure 8.4 mirrored in a horizontal plane: horizontal distance increases by n/m as one moves *downwards* through the graph.

Retain notation from the proof of Theorem 8.2.1. One can show that every element of S has a representative in

$$K^*(A')^* - \{\varepsilon\},$$

where

$$K = \{a_1^{\alpha_1} \cdots a_l^{\alpha_l} b : b \in A'', 0 \le \alpha_i < n^{\beta_b}\}.$$

By padding elements of A' and K as before, using a symbol 1 representing an adjoined identity, one obtains the language of normal forms $L = J^*(K')^*$.

A path labelled by a word $u \in L$ 'climbs' to the x-row of the graph containing $\overline{u}$ in segments from K' is the same way that those in the earlier proof climbed *from* the bottom x-row. It then travels along that x-row to the vertex $\overline{u}$.

To show that if $u, v \in L$ are such that $\overline{au} = \overline{v}$, then the paths $\widehat{u}$ and $\widehat{v}$ fellow-travel, proceed as follows. Argue in the same way as before to show that $\widehat{u}$ and $\widehat{v}$ 'keep pace' during the climb to the relevant x-row and during travelling along that row. The earlier reasoning about horizontal distances holds true because of the increase in horizontal distance by a factor of n/m as one moves down a single row. Therefore, as the paths have a common origin at the basepoint, the distance between their intersections with a given x-row cannot exceed a certain value.

Applying the left-handed version of Theorem 2.3.2 shows that (A, L) is a left-automatic structure for S. $\boxed{8.2.2}$

8.2.3. Non-automatic subsemigroups

Theorems 8.2.1 and 8.2.2 show that all finitely generated subsemigroups of the Baumslag–Solitar semigroup $\mathrm{BSS}(m,n)$ are automatic if $m > n$ and left-automatic if $m < n$. This leaves the case when $m = n$. The present subsection shows that $\mathrm{BSS}(m,m)$, which has presentation $\mathrm{Sg}\langle x, y \mid (yx^m, x^m y)\rangle$, is not locally automatic unless $m = 1$. If $m \ge 2$, it contains finitely generated subsemigroups that are not even asynchronously automatic; a consequence of the following result:

Proposition 8.2.3. *The Baumslag–Solitar semigroup* $\mathrm{BSS}(m,m)$ *contains the semigroup* $\{p_1, \ldots, p_m\}^* \times (\mathbb{N} \cup \{0\}) - \{(\varepsilon, 0)\}$*: the direct product of the free monoid on* m *letters and the natural numbers (including zero) with the identity* $(\varepsilon, 0)$ *removed.*

Proof of 8.2.3. Let $A = \{p_1, \ldots, p_m, r\}$ be an alphabet representing elements of $\mathrm{BSS}(m,m)$ as follows

$$\overline{p_i} = x^{i-1} y \text{ for each } i, \text{ and } \overline{r} = x^m.$$

Let S be the subsemigroup of $\mathrm{BSS}(m, m)$ generated by $\overline{A}$. The aim is to show that S is presented by

$$\mathrm{Sg}\langle A \mid (p_i r, r p_i) \text{ for all } i\rangle.$$

To prove this, note firstly that every relation $(p_i r, r p_i)$ holds in S. Define a set of normal forms $N = \{p_1, \ldots, p_m\}^* r^* - \{\varepsilon\}$. Every element of S has a normal form: letters r can be moved to the left of all letters p_i using the defining relations. Consider any element of $s \in S$. Suppose first that s contains some letter y. If s begins $x^{i-1}y \cdots$, then any word in N representing it must begin $p_i \cdots$. On the other hand, if s contains no letters y, then $s = x^{m\alpha}$ for some $\alpha \in \mathbb{N}$ and the normal form word representing it is r^α. In the first case, one can cancel the $x^{i-1}y$ and iterate this reasoning to obtain the entire normal form word representing s. Thus N is a set of unique normal forms for S, and so S has the given presentation.

Thus S is isomorphic to $\{p_1, \ldots, p_m\}^* \times (\mathbb{N} \cup \{0\}) - \{(\varepsilon, 0)\}$. $\boxed{8.2.3}$

Since the free semigroup of rank 2 contains a copy of the free semigroup of any rank, the following corollary is immediate:

Corollary 8.2.4. *The Baumslag–Solitar semigroup* $\mathrm{BSS}(m, m)$, *where* $m \geq 2$, *contains the semigroup* $\{p_1, \ldots, p_k\}^* \times (\mathbb{N} \cup \{0\}) - \{(\varepsilon, 0)\}$ *for any* $k \in \mathbb{N}$. $\boxed{8.2.4}$

Therefore, excepting the case when $m = 1$, the Baumslag–Solitar semigroup $\mathrm{BSS}(m, m)$ always contains the semigroup $\{x, y, p, q, r\}^* \times (\mathbb{N} \cup \{0\}) - \{(\varepsilon, 0)\}$. This semigroup is known to contain finitely generated subsemigroups that are not asynchronously automatic; see Example 7.5.10.

Proposition 8.2.5. *The Baumslag–Solitar semigroup* $\mathrm{BSS}(m, m)$, *where* $m \geq 2$, *is neither locally automatic nor locally asynchronously automatic.* $\boxed{8.2.5}$

Of course, $\mathrm{BSS}(1, 1)$ is simply

$$\mathrm{Sg}\langle x, y \mid (yx, xy)\rangle \simeq \mathbb{N} \times \mathbb{N} \subseteq \mathbb{Z} \times \mathbb{Z},$$

and is therefore locally automatic by Theorem 5.4.2.

8.3. MALCEV COHERENCE

APPLYING the right- and left-handed versions of Theorem 2.5.1 to Theorems 8.2.1 and 8.2.2 yields the following:

Corollary 8.3.1. *The Baumslag–Solitar semigroups* $\mathrm{BSS}(m, n)$, *where* $m \neq n$, *are Malcev coherent.* $\boxed{8.3.1}$

This leaves unanswered the following questions:

Open Problem 8.3.2. Are the Baumslag–Solitar semigroups $\mathrm{BSS}(m, m)$ (where $m \geq 2$) Malcev coherent?

Open Problem 8.3.3. Are Baumslag–Solitar groups Malcev coherent?

[The Baumslag–Solitar groups are known to be coherent (Kropholler 1990).]

Baumslag (1974, Section B) asks whether all one-relator groups are coherent. Some progress has been made on this front; see Karrass & Solitar (1970) and McCammond & Wise (2005). It is therefore natural, although perhaps precipitate, to pose the following question:

Open Problem 8.3.4. Are all one-relator groups Malcev coherent?

A positive answer to this question would also provide a positive answer to the question of whether the free product of a free group and $\mathbb{Z} \times \mathbb{Z}$ is Malcev coherent (Open Problem 6.2.9), since

$$\mathrm{FG}(X) * (\mathbb{Z} \times \mathbb{Z}) \simeq \mathrm{Gp}\langle X, z_1, z_2 \mid (z_1 z_2, z_2 z_1)\rangle.$$

A restricted version of Open Problem 8.3.4 that may be easier to answer is the following:

Open Problem 8.3.5. Are all one-relator cancellative semigroups Malcev coherent?

CHAPTER NINE

EXTENSIONS, SUBSEMIGROUPS & INDICES

... what a large and immense field doth extension alone afford the mathematicians?

— John Locke,
An Essay Concerning Human Understanding (1690),
bk. i, ch. vii.10

9.1. INTRODUCTION

THE REIDEMEISTER–SCHREIER THEOREM asserts that the finite generation and finite presentability of a group is preserved under constructing finite extensions and passing to finite-index subgroups (see Lyndon & Schupp 1977, Section II.4). Much work has been carried out on concepts of index for semigroups that preserve properties such as finite generation and finite presentability. The earliest such semigroup index to be defined was the Rees index, introduced by Jura (1978) and studied in Section 9.2. The Rees index, however, is not a generalization of the group index. An index that does specialize to the group index is the syntactic index of Ruškuc & Thomas (1998), considered in Section 9.3. Section 9.4 speculates on other concepts of index. Finally, Section 9.5 considers the question of whether the class of Malcev coherent groups is closed under finite extensions.

9.2. REES INDEX

JURA (1978, DEFINITION 1) was the first to define the Rees index of a subsemigroup of a semigroup, although he simply referred to it as the 'index':

Definition 9.2.1. Let S and T be semigroups with T being contained in S. The *Rees index* of T in S is $|S - T| + 1$. If this Rees index is finite, then semigroup S is a *small extension* of T, and the semigroup T is a *large subsemigroup* of S.

[Should the subsemigroup T be an ideal, the Rees index is the order of the Rees quotient semigroup S/T; hence the name 'Rees index'. Some authors, however, consider the Rees index to be the cardinality of $S - T$. Of course, this disagreement has no effect on finiteness.] Many properties of semigroups are known to be preserved under constructing small extensions or passing to large subsemigroups. Finite generation is such a property:

Proposition 9.2.2 (Campbell, Robertson, Ruškuc & Thomas 1995). *Let S be a semigroup and let T be a large subsemigroup of S. Then S is finitely generated if and only if T is finitely generated.* $\boxed{9.2.2}$

Automatism is also preserved under passing to small extensions and large subsemigroups (see Theorem 2.4.5), as is finite presentability:

Theorem 9.2.3 (Ruškuc 1998, Theorem 1.3). *Let T be a large subsemigroup of a semigroup S. Then S is finitely presented if and only if T is finitely presented.* $\boxed{9.2.3}$

[The proof that finite presentability is preserved under constructing small extensions is concise. The proof of its preservation under forming large subsemigroups is long and technical. Gray & Ruškuc (2005, Section 5) point out and repair a gap in the latter's original proof.]

Anyone familiar with Malcev presentations will, upon reading this result of Ruškuc, ask whether the property of admitting a finite Malcev presentation is preserved under taking large subsemigroups or small extensions. The purpose of the present section is to prove that such preservation does indeed occur. Although one could prove the 'small extension' case by following the reasoning of Ruškuc but using Malcev presentations rather than 'ordinary' presentations, both cases can be deduced from the following result:

Theorem 9.2.4. *Let S be a semigroup that embeds in a group. Let T be a subsemigroup of S. Suppose that $|T| > |S - T|$. [This includes the possibility that T is infinite and $S - T$ finite.] Then the universal groups of S and T are isomorphic.*

Proof of 9.2.4. Let $\mathrm{Sg}\langle T \mid \tau\rangle$ and $\mathrm{Sg}\langle S \mid \sigma\rangle$ be the Cayley table presentations for T and S, respectively. Let G be the universal group of S, and view S and T as subsemigroups of G. By Proposition 0.8.2, G is presented by $\mathrm{Gp}\langle S \mid \sigma\rangle$. Use the sets S and T as both symbols in presentations and as elements of G, and suspend (for the duration of this proof) the notational distinction between a symbol and the element it represents. The key to the proof is the following lemma:

Lemma 9.2.5. *For each $s \in S - T$, there exist elements u_s, v_s, w_s, and x_s of T such that $s = u_s v_s^{-1}$ and $s = w_s^{-1} x_s$ in G.*

Proof of 9.2.5. Let $k = |S - T|$. Let $k < l \leq |T|$. Pick distinct elements $t_1, t_2, \ldots, t_l$ of T. Suppose the elements $st_1, st_2, \ldots, st_l$ are not all distinct. Then for some i, j with $i \neq j$, $st_i = st_j$, which means that $t_i = t_j$, contradicting the choice of $t_1, t_2, \ldots, t_l$. Therefore the elements $st_1, st_2, \ldots, st_l$ are all distinct, and so at least one of them lies in T since $l > |S - T|$. Let h be such that $st_h \in T$. Let $u_s = st_h$ and $v_s = t_h$. Then $u_s, v_s \in T$ and $s = u_s v_s^{-1}$ in G. Similar reasoning yields w_s and x_s. $\boxed{9.2.5}$

Lemma 9.2.5 shows that T generates G as a group, since the subgroup of G generated by T contains S, and S is certainly a group generating set for G. The strategy of the remainder of the proof is to show that G has a presentation $\mathrm{Gp}\langle T \mid \rho\rangle$ such that all of the defining relations in ρ are between positive words and are valid in T. All the relations in ρ must then be consequences of those in τ. Therefore $\mathrm{Gp}\langle T \mid \tau\rangle$ will present G and so G will be isomorphic to the universal group of T.

Partition σ as $\tau \cup \omega$, where

$$\omega = \{(pq, r) : p \in S - T \text{ or } q \in S - T\} \subseteq SS \times S,$$

and let

$$\mathcal{R} = \{(sv_s, u_s), (w_s s, x_s) : s \in S - T\},$$

where u_s, v_s, w_s, and x_s are as in Lemma 9.2.5. Notice that $\mathcal{R} \subseteq \omega$.

Create a new set of defining relations ω'. For each relation $(pq, r) \in \omega$, add relations to ω' in accordance with the appropriate case below:

i.) $p \in S-T$, $q \in T$, $r \in T$. Use $\mathcal{R}$ to see that the relation $(w_p^{-1}x_p q, r)$ is valid in G. Therefore add the relation $(x_p q, w_p r)$ — which is valid in T — to ω'. Observe that (pq, r) is a consequence of this new relation and $(w_p p, x_p) \in \mathcal{R}$.

ii.) $p \in T$, $q \in S - T$, $r \in T$. The relation $(pu_q v_q^{-1}, r)$ is valid in G. Add (pu_q, rv_q) to ω' and observe that the original relation is once again a consequence of the new one and $(qv_q, u_q) \in \mathcal{R}$.

iii.) $p \in S - T$, $q \in S - T$, $r \in T$. The relation $(w_p^{-1}x_p u_q v_q^{-1}, r)$ is valid; add $(x_p u_q, w_p r v_q)$ to ω'. Once more the original relation is a consequence of the new one and $(w_p p, x_p), (qv_q, u_q) \in \mathcal{R}$.

iv.) $p \in S - T$, $q \in T$, $r \in S - T$. The relation $(w_p^{-1}x_p q, u_r v_r^{-1})$ is valid; add $(x_p q v_r, w_p u_r)$ to ω'. The original relation is a consequence of the new one and $(w_p p, x_p), (rv_r, u_r) \in \mathcal{R}$.

v.) $p \in T$, $q \in S - T$, $r \in S - T$. The relation $(pu_q v_q^{-1}, w_r^{-1}x_r)$ is valid; add $(w_r p u_q, x_r v_q)$ to ω'. The original relation is a consequence of the new one and $(w_r r, x_r), (qv_q, u_q) \in \mathcal{R}$.

vi.) $p \in S - T$, $q \in S - T$, $r \in S - T$. Now, $pq = r$ in S, so $pq = u_r v_r^{-1}$ in G. Therefore $pqv_r = u_r$. Now consider two sub-cases:

  a) $qv_r = s \in S - T$. Then $ps = u_r$, and so (pq, r) is a consequence in G of (rv_r, u_r), (qv_r, s) and (ps, u_r). The set $\mathcal{R}$ contains the first of these three relations. The second and third are in ω and are of types iv. and i. above.

  b) $qv_r = t \in T$. Then $pt = u_r$, and so (pq, r) is a consequence of (rv_r, u_r), (qv_r, t) and (pt, u_r). Again, the set $\mathcal{R}$ contains the first of these three relations. The second and third are both in ω and of type i. above.

In either sub-case, (pq, r) is a consequence of $\mathcal{R}$ and relations in ω of other types. Therefore do not add any relations to ω'.

Now, $G = \mathrm{Gp}\langle S \mid \sigma\rangle = \mathrm{Gp}\langle S \mid \sigma \cup \mathcal{R}\rangle = \mathrm{Gp}\langle S \mid \tau \cup \omega \cup \mathcal{R}\rangle$. Each relation in ω' is a consequence of those in $\mathcal{R}$ and those in ω. On the other hand, each relation in ω

is a consequence of those in $\mathcal{R}$ and those in ω'. So the group G is also presented by $\mathrm{Gp}\langle S \mid \tau \cup \omega' \cup \mathcal{R}\rangle$.

Partition $\mathcal{R}$ as $\mathcal{R}' \cup \mathcal{R}''$, where

$$\mathcal{R}' = \{(sv_s, u_s) : s \in S\},$$
$$\mathcal{R}'' = \{(w_s s, x_s) : s \in S\}.$$

Every relation in

$$\mathcal{Q} = \{(x_s v_s, w_s u_s) : s \in S\}$$

is a consequence of those in $\mathcal{R}$; every relation in $\mathcal{R}''$ is a consequence of those in $\mathcal{R}' \cup \mathcal{Q}$. Therefore the group G is presented by $\mathrm{Gp}\langle S \mid \tau \cup \omega' \cup \mathcal{R}' \cup \mathcal{Q}\rangle$.

Finally, eliminate the generators contained in $S - T$ and the relations in $\mathcal{R}'$. This gives a presentation $\mathrm{Gp}\langle T \mid \tau \cup \omega' \cup \mathcal{Q}\rangle$ for G. Observing that $\tau \cup \omega' \cup \mathcal{Q}$ consists of positive relations between elements of T completes the proof. $\boxed{9.2.4}$

Theorem 9.2.6. *Let S be a semigroup that embeds in a group. Let T be a subsemigroup of S of finite Rees index. Then S has a finite Malcev presentation if and only if T has a finite Malcev presentation.*

Proof of 9.2.6. If T is finite, then S is also finite and therefore S and T trivially both admit finite Malcev presentations. Therefore assume T is infinite, in which case $|S - T| < |T|$. Theorem 9.2.4 applies to show that the universal groups of S and T are isomorphic.

By Proposition 9.2.2, S is finitely generated if and only if T is finitely generated. Let S admit a finite Malcev presentation. Then S is finitely generated and its universal group is finitely presented by Corollary 1.3.2. Therefore T is finitely generated and its universal group — isomorphic to that of S — is finitely presented. Corollary 1.3.2 applies again to show that T admits a finite Malcev presentation. Similar reasoning shows that S has a finite Malcev presentation if T does. $\boxed{9.2.6}$

9.3. SYNTACTIC INDEX

Ruškuc & Thomas (1998) defined the syntactic index of a subsemigroup:

Definition 9.3.1. Let S be a semigroup and T a subsemigroup of S. Let σ be the relation $(T \times T) \cup ((S - T) \times (S - T))$. Let σ_{R} and σ_{L} be, respectively, the largest right congruence and largest left congruence contained in σ. The *right syntactic index* of T in S, denoted $[S : T]_{\mathrm{R}}$, is the number of σ_{R}-classes in S. Similarly, the *left syntactic index* $[S : T]_{\mathrm{L}}$ of T in S is the number of σ_{L}-classes in S.

In other words, σ_{R} and σ_{L} are the largest right congruence and largest left congruence on S that respect T — that is, for which T is a union of congruence classes. [In formal language theory, the syntactic congruence of a language L over an alphabet A is the unique largest congruence on A^* that respects L (see Definition A.5.7). This is the origin of the name of the 'syntactic indices'.]

The right syntactic index of T in S is finite if and only if the left syntactic index of T in S is finite (Ruškuc & Thomas 1998, Theorem 3.2(iii)). It is therefore sensible to simply state that a subsemigroup is of finite syntactic index. If T is a finite Rees index subsemigroup of S, then T has finite syntactic index in S (Ruškuc & Thomas 1998, Corollary 4.4).

Definition 9.3.2. Let S be a semigroup and T a subsemigroup of S of finite syntactic index. Then S is a *syntactically small extension* of S and T is a *syntactically large subsemigroup* of S.

The syntactic indices have an important advantage over the Rees index: they are generalizations of the group index. Ruškuc & Thomas (1998, Theorem 3.2(iii)) show that if G and H are groups, then

$$[G : H] = [G : H]_{\mathrm{R}} = [G : H]_{\mathrm{L}}.$$

Ruškuc & Thomas (1998, Theorem 3.5) point out that any property of semigroups either fails to be inherited by syntactically large subsemigroups or by syntactically small extensions. However, their proof relies on a semigroup with a zero adjoined, which does not eliminate the possibility of restricting to semigroups embeddable into groups and obtaining positive results. As the syntactic indices generalize the group index, one might hope that group-embeddable semigroups would prove fertile in this regard.

This section investigates syntactically small extensions and syntactically large subsemigroups in the context of group-embeddable semigroups. In particular, study is made of whether the the following properties are inherited by those extensions and subsemigroups: being a group, finiteness, finite generation, finite presentability, admitting a finite Malcev presentation, and Malcev coherence. [All of these properties are preserved under forming finite extensions and taking finite-index subgroups of groups except Malcev coherence, the question of whose inheritance by finite extensions is open (Open Problem 9.5.1).]

As one can see from Table 9.1, which provides a summary, the results are generally negative. It therefore seems that the syntactic indices are 'poor relations' of the group index.

Proposition 9.3.3. *A syntactically large subsemigroup of a group is itself a group.*

Proof of 9.3.3. Let G be a group and let T be a subsemigroup of G of finite syntactic index. Let ρ be the largest right congruence contained in $(T \times T) \cup ((G-T) \times (G-T))$. Suppose the number of ρ-congruence classes is finite.

Pick $t \in T$. Consider the powers t^α for $\alpha \in \mathbb{N}$. Since there are only finitely many congruence classes, there exist natural numbers α and β with $\alpha < \beta$ such that $t^\alpha \mathrel{\rho} t^\beta$. Since G is a group, $t^{-\alpha} \in G$. As ρ is a right congruence on G, $t^\alpha t^{-\alpha} \mathrel{\rho} t^\beta t^{-\alpha}$, so $1_G \mathrel{\rho} t^{\beta-\alpha}$. Since $\beta - \alpha$ is greater than zero, $1_G \in T$. Furthermore, $t^{-1} = 1_G t^{-1} \mathrel{\rho} t^{\beta-\alpha} t^{-1} = t^{\beta-\alpha-1}$. Therefore, since $\beta - \alpha - 1$ is at least zero, $t^{-1} \in T$. Since $t \in T$ was arbitrary, this means that T is a subgroup of G. $\boxed{9.3.3}$

Table 9.1. Summary, for group-embeddable semigroups, of inheritance of certain properties by syntactically large subsemigroups and syntactically small extensions.

	INHERITED BY SYNTACTICALLY	
PROPERTY	LARGE SUBSEMIGROUPS	SMALL EXTENSIONS
Being a group	Y (Pr. 9.3.3)	N (Ex. 9.3.4)
Finite	Y (Trivial)	N (Ex. 9.3.4)
Finitely generated	N (Ex. 9.3.5)	N (Ex. 9.3.5)
Finitely presented*	N (Ex. 9.3.7)	N (Ex. 9.3.7)
Finite Malcev presentation*	N (Ex. 9.3.8)	N (Ex. 9.3.8)
Malcev coherent	Y (Trivial)	N (Ex. 9.3.8)
Automatic*	? (Op. 9.3.9)	N (Ex. 9.3.8)
Locally automatic	Y (Trivial)	N (Ex. 9.3.8)

*Assuming finite generation.

Note. Y = Yes, N = No, ? = Undecided.

The converse of the preceding result does not hold: the following example shows that a small extension of a group — even an extension known to be group-embeddable — may not be a group.

Example 9.3.4. Let $S = \mathbb{Z}_2 \times (\mathbb{N} \cup \{0\})$. Let $T = \{(m, 0) : m \in \mathbb{Z}_2\}$. Let ρ be the relation $(T \times T) \cup ((S - T) \times (S - T))$. Observe that

$$TT \subseteq T, \ \ T(S-T) \subseteq S - T, \ \text{ and } (S-T)S \subseteq S - T.$$

Therefore, if $(t_1, t_2) \in T \times T$ and $t \in T$, then $(t_1 t, t_2 t) \in T \times T$. If $(t_1, t_2) \in T \times T$ and $s \in S - T$, then $(t_1 s, t_2 s) \in (S-T) \times (S-T)$. Finally, if $(s_1, s_2) \in (S-T) \times (S-T)$ and $s \in S$, then $(s_1 s, s_2 s)$ is again in $(S - T) \times (S - T)$. Therefore ρ is a right congruence on S and $[S : T]_{\mathrm{R}} = 2$. However, T clearly is a group and S clearly is not.

Notice that Example 9.3.4 also shows that a syntactically small extension of a finite group may not be finite.

Example 9.3.5. Consider the following subsets of the group $\mathbb{Z}^2$:

$$\begin{aligned} X &= \{(1, n) : n \in \mathbb{N} \cup \{0\}\}, \\ Y &= X \cup \{(0, 1)\}, \\ Z &= \{(1, 0)\}. \end{aligned}$$

Let S, T, and U be the subsemigroups generated by Y, X and Z, respectively:

$$\begin{aligned} S &= \{(m, n) : m \in \mathbb{N} \cup \{0\}, n \in \mathbb{N} \cup \{0\}\} - \{(0, 0)\}, \\ T &= \{(m, n) : m \in \mathbb{N}, n \in \mathbb{N} \cup \{0\}\}, \\ U &= \{(m, 0) : m \in \mathbb{N}\}. \end{aligned}$$

Then

$$S - T = \{(0, n) : n \in \mathbb{N}\},$$
$$T - U = \{(m, n) : m, n \in \mathbb{N}\}.$$

Let ρ be the relation $(T \times T) \cup ((S - T) \times (S - T))$. Let σ be the relation $(U \times U) \cup ((T - U) \times (T - U))$.

Observe that

$$TT \subseteq T, \quad T(S - T) \subseteq T, \quad \text{and } (S - T)S \subseteq S - T.$$

Therefore ρ is a congruence on S and $[S : T]_{\mathrm{R}} = 2$. Similarly, observe that

$$UU \subseteq U, \quad U(T - U) \subseteq T - U, \quad \text{and } (T - U)T \subseteq (T - U).$$

Therefore σ is a congruence on T and $[T : U]_{\mathrm{R}} = 2$

Therefore S is a syntactically small extension of T and U is a syntactically large subsemigroup of T. However, no element of the subset X is decomposable in T. Therefore T is not finitely generated. However, U is finitely generated by Z and S by $\{(1, 0), (0, 1)\}$. Ergo, finite generation is inherited neither by syntactically small extensions nor by syntactically large subsemigroups, even if one restricts to the case of group-embeddable semigroups.

Proposition 9.3.6. *A finitely generated subsemigroup of a free semigroup has finite syntactic index.*

Proof of 9.3.6. Let X be an alphabet. Let Z be a finite subset of the free semigroup X^+, and let T be the subsemigroup it generates. Let X' be the subset of X consisting of all letters that appear in at least one word in Z. The alphabet X' must be finite.

Let $S = (X')^+$ and notice that S contains $T = Z^+$, which is a regular language. The syntactic congruence ρ_T, is the largest congruence on S respecting T. There are finitely many ρ_T-classes by Theorem A.5.8.

Since $(X^+ - S)X^+$ and $(X^+ - S)(X^+ - S)$ are subsets of $(X^+ - S)$, and ρ_T is a congruence on S, the relation

$$\sigma = \rho_T \cup ((X^+ - S) \times (X^+ - S))$$

is a right congruence on X^+ and has one more class than ρ_T. Furthermore, σ is contained in $(T \times T) \cup ((X^+ - T) \times (X^+ - T))$ since ρ_T is. Since there are finitely many σ-classes, $[X^+ : T]_{\mathrm{R}}$ is finite. $\boxed{9.3.6}$

Proposition 9.3.6 parallels the fact that a non-trivial finitely generated normal subgroup of a free group has finite index (Lyndon & Schupp 1977, Proposition I.3.12).

Example 9.3.7. Let $X = \{x, y, z, t\}$. Let $S = X^+$. Let $A = \{a, b, c, d, e, f\}$ be an alphabet representing a subset of S as follows:

$$\overline{a} = x^2yz, \qquad \overline{d} = x^2y,$$
$$\overline{b} = yz, \qquad \overline{e} = zy,$$
$$\overline{c} = yx^2, \qquad \overline{f} = zyx^2.$$

Let $T = \mathrm{Sg}\langle\overline{A}\rangle$ and let $U = \mathrm{Sg}\langle\overline{a}, \overline{b}, \overline{c}\rangle$. By Proposition 9.3.6, T is a finite syntactic index subsemigroup of S.

The semigroups S and U are finitely presented; the semigroup T is presented by

$$\mathrm{Sg}\langle A \mid (ab^\alpha c, de^\alpha f) : \alpha \in \mathbb{N} \cup \{0\}\rangle,$$

but does not admit a finite presentation.

Identify T with $A^+ - A^*de^*fA^*$. Let $T_d = Ude^* = \{a, b, c\}^+de^*$. Let $T' = (T - U) - T_d$. Let

$$\sigma = (U \times U) \cup (T_d \times T_d) \cup (T' \times T').$$

Notice that $U\{a, b, c\} \subseteq U$, $U\{d\} \subseteq T_d$, $U\{e, f\} \subseteq T'$, $T_d\{a, b, c, d\} \subseteq T'$, $T_d\{e\} \subseteq T_d$, $T_d\{f\} \subseteq U$, and $T'A \subseteq T'$. So, since A generates T, the relation σ is a right congruence on T contained in $(U \times U) \cup ((T - U) \times (T - U))$. So $[T : U]_{\mathrm{R}}$ is at most 3.

So T is a syntactically small extension of the finitely presented semigroup U and a syntactically large subsemigroup of the finitely presented semigroup S, yet is itself not finitely presented.

Example 9.3.8. Let $F = \mathrm{FG}(x, y, z, s, t, p) * \mathbb{Z}^3$. Identify F with alternating products of elements of $\mathrm{FG}(x, y, z, s, t, p)$ (viewed as reduced words) and of $\mathbb{Z}^3$ (viewed as triples of integers). Let $B = \{a, b, c, d, e, f, g, h, i, j, k, l, m\}$ be an alphabet representing elements of F as follows:

$$\begin{array}{lll}
\overline{a} = x^2y, & \overline{f} = x^2s, & \\
\overline{b} = y^{-1}(1,0,1)y, & \overline{g} = s^{-1}(1,0,0)s, & \overline{k} = x^2p, \\
\overline{c} = y^{-1}z, & \overline{h} = s^{-1}t, & \overline{l} = p^{-1}(1,1,1)p, \\
\overline{d} = z^{-1}(0,1,0)z, & \overline{i} = t^{-1}(0,1,1)t, & \overline{m} = p^{-1}x^2. \\
\overline{e} = z^{-1}x^2, & \overline{j} = t^{-1}x^2, &
\end{array}$$

Let $A = B - \{k, l, m\}$ and let $C = \{a, c, e, f, h, j\}$. Let $S = \mathrm{Sg}\langle\overline{B}\rangle$; let $T = \mathrm{Sg}\langle\overline{A}\rangle$; and let $U = \mathrm{Sg}\langle\overline{C}\rangle$.

By Lemma 6.2.2, the semigroup T is presented by $\mathrm{Sg}\langle A \mid \mathcal{R}\rangle$, where

$$\mathcal{R} = \{(ab^\alpha cd^\alpha e, fg^\alpha hi^\alpha j) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

Reasoning similar to that in the proof of Lemma 6.2.2 shows that the semigroup S is presented by $\mathrm{Sg}\langle B \mid \mathcal{Q}\rangle$, where

$$\mathcal{Q} = \{(ab^\alpha cd^\alpha e, kl^\alpha m), (fg^\alpha hi^\alpha j, kl^\alpha m) : \alpha \in \mathbb{N} \cup \{0\}\}.$$

Let $S_k = \{t\overline{kl^\alpha} : t \in T, \alpha \in \mathbb{N} \cup \{0\}\}$. Let $S' = (S - T) - S_k$. Let

$$\sigma = (T \times T) \cup (S_k \times S_k) \cup (S' \times S').$$

Now, $T\overline{A} \subseteq T$, $T\{\overline{k}\} \subseteq S_k$, $T\{\overline{l}, \overline{m}\} \subseteq S'$, whilst $S_k\{\overline{l}\} \subseteq S_k$, $S_k\{\overline{m}\} \subseteq T$, and $S_k(\overline{B - \{l,m\}}) \subseteq S'$. Finally, $S'\overline{B} \subseteq S'$. Therefore, since $\overline{B}$ generates S, the relation σ is a right congruence on S. Thus $[S : T]_{\mathrm{R}}$ is at most 3.

Let

$$\tau = (U \times U) \cup ((T - U) \times (T - U)).$$

The various tuples of natural numbers cannot cancel, so $U(\overline{B - C}) \subseteq (T - U)$ and $(T - U)\overline{B} \subseteq (T - U)$. That $U\overline{C} \subseteq T$ is obvious. Therefore τ is a right congruence and $[T : U]_{\mathrm{R}} = 2$.

Proposition 6.2.4 establishes that T does not admit a finite Malcev presentation. However, S admits the finite Malcev presentation $\mathrm{SgM}\langle B \mid \mathcal{S}\rangle$, where

$$\mathcal{S} = \{(ab^\alpha cd^\alpha e, kl^\alpha m), (fg^\alpha hi^\alpha j, kl^\alpha m) : \alpha = 0, 1, 2\}.$$

To prove this, proceed by inducation on α. Let α be at least 3, and assume $(ab^\beta cd^\beta e, kl^\beta m) \in \mathcal{S}^{\mathrm{M}}$, for all $\beta < \alpha$. The following Malcev chain shows that $(ab^\alpha cd^\alpha e, kl^\alpha m)$ is also in $\mathcal{S}^{\mathrm{M}}$:

$$\begin{aligned}
& ab^\alpha cd^\alpha e \\
&\to ab^{\alpha-2}a^{\mathsf{L}}ab^2cd^2ee^{\mathsf{R}}d^{\alpha-2}e \\
&\to ab^{\alpha-2}a^{\mathsf{L}}kl^2me^{\mathsf{R}}d^{\alpha-2}e && \text{by induction, with } \beta = 2 \\
&\to ab^{\alpha-2}a^{\mathsf{L}}klmm^{\mathsf{R}}lme^{\mathsf{R}}d^{\alpha-2}e \\
&\to ab^{\alpha-2}a^{\mathsf{L}}abcdem^{\mathsf{R}}lme^{\mathsf{R}}d^{\alpha-2}e && \text{by induction, with } \beta = 1 \\
&\to ab^{\alpha-1}cdem^{\mathsf{R}}lme^{\mathsf{R}}d^{\alpha-2}e \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}a^{\mathsf{L}}acem^{\mathsf{R}}lme^{\mathsf{R}}d^{\alpha-2}e \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}a^{\mathsf{L}}kmm^{\mathsf{R}}lme^{\mathsf{R}}d^{\alpha-2}e && \text{by induction, with } \beta = 0 \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}a^{\mathsf{L}}klme^{\mathsf{R}}d^{\alpha-2}e \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}a^{\mathsf{L}}abcdee^{\mathsf{R}}d^{\alpha-2}e && \text{by induction, with } \beta = 1 \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}bcd^{\alpha-1}e \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}(b^{\alpha-2})^{\mathsf{L}}a^{\mathsf{L}}ab^{\alpha-1}cd^{\alpha-1}e \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}(b^{\alpha-2})^{\mathsf{L}}a^{\mathsf{L}}kl^{\alpha-1}m && \text{by induction, with } \beta = \alpha - 1 \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}(b^{\alpha-2})^{\mathsf{L}}a^{\mathsf{L}}kl^{\alpha-2}mm^{\mathsf{R}}lm \\
&\to ab^{\alpha-1}cdc^{\mathsf{L}}(b^{\alpha-2})^{\mathsf{L}}a^{\mathsf{L}}ab^{\alpha-2}cd^{\alpha-2}em^{\mathsf{R}}lm && \text{by induction, with } \beta = \alpha - 2 \\
&\to ab^{\alpha-1}cd^{\alpha-1}em^{\mathsf{R}}lm \\
&\to kl^{\alpha-1}mm^{\mathsf{R}}lm && \text{by induction, with } \beta = \alpha - 1 \\
&\to kl^\alpha m.
\end{aligned}$$

Similarly, each relation $(fg^{\alpha}hi^{\alpha}j, kl^{\alpha}m)$ lies in $\mathcal{S}^{\mathrm{M}}$.

As U is a finitely generated subsemigroup of a free group, it admits a finite Malcev presentation and is Malcev coherent by Theorem 4.3.1. Furthermore, U is automatic and locally automatic by Theorem 4.4.1.

The semigroup S contains the syntactically large subsemigroup T, which in turn is a syntactically small extension of the semigroup U. Therefore syntactically small extensions do not in general inherit the property of admitting a finite Malcev presentation, nor Malcev coherence, automatism, or local automatism. Furthermore, the properties of having a finite Malcev presentation and of being automatic may not be inherited by a syntactically large subsemigroup.

Open Problem 9.3.9. Let S be an automatic semigroup, and let T be a syntactically large subsemigroup of S. Must T be automatic?

[If S were a group, then T would also be a group by Proposition 9.3.3 and so would be automatic by Theorem 2.4.16.]

9.4. OTHER INDICES

The proof of the result of Ruškuc (1998) that finite presentability is preserved under passing to subsemigroups of finite Rees index is long (eleven pages) and technical. In contrast, proving Theorem 9.2.6 is almost 'too easy'. One might informally view the ease of proof as a consequence of the fact that if T is a finite Rees index subsemigroup of a group-embeddable semigroup S, then a very strong restriction is imposed on the structure of T relative to S. It is this restriction that forces the universal groups of T and S to coincide, as per Theorem 9.2.4.

On the other hand, the syntactic indices seem to allow *too* much flexibility. Perhaps, then, there is a concept of index for subsemigroups of groups that lies 'between' the Rees and syntactic indices and shares the most desirable properties of both: like the syntactic indices, it could generalize the group index; like the Rees index, taking finite-index subsemigroups and constructing finite-index extensions would preserve certain pleasant properties. In particular, one would hope to preserve the property of having a finite Malcev presentation.

Open Problem 9.4.1. For group-embeddable semigroups, is there a 'better' concept of index than the Rees and syntactic indices?

9.5. EXTENSIONS OF MALCEV COHERENT GROUPS

Proposition 1.6.1 asserts that the class of coherent groups is closed under constructing finite extensions. One therefore immediately asks whether the same holds true for Malcev coherence:

Open Problem 9.5.1. Is the class of Malcev coherent groups closed under forming finite extensions?

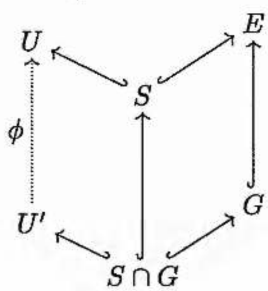

Figure 9.1. Diagram illustrating a strategy one might hope to use to prove that the class of Malcev coherent groups is closed under forming finite extensions.

[The author opines that Open Problem 9.5.1 is the most important unanswered question in the theory of Malcev presentations.]

Solving Open Problem 9.5.1 seems to be difficult. Three species of Malcev coherent groups are known: virtually free groups, virtually nilpotent groups, and direct products of virtually free groups and abelian groups. Of these three classes, the first two are by definition closed under constructing finite extensions. Therefore, only the direct product of a virtually free group and an abelian group and a finite extension of same *may* provide a counterexample to negatively answer Open Problem 9.5.1.

On the other hand, suppose the class of Malcev coherent groups is indeed closed under forming finite extensions. One might naïvely hope to prove this using a strategy similar to that used for Proposition 1.6.1:

i.) Take a finite extension E of a Malcev coherent group G. Choose an arbitrary finitely generated subsemigroup S of E.

ii.) Use the Malcev coherence of G to obtain a finite Malcev presentation of $S \cap G$, which is simply a presentation for the universal group U' of $S \cap G$.

iii.) Lift this presentation to a finite presentation of the universal group U of S using an extension ϕ of the embedding mapping $S \cap G \hookrightarrow S$. (Figure 9.1 illustrates this idea.)

However, there are two major problems with this outline. Firstly, despite the finite generation of S, the subsemigroup $S \cap G$ may not be finitely generated, so one cannot call upon the Malcev coherence of G to obtain a finite presentation for U' (see Example 9.5.2). Secondly, even if $S \cap G$ is finitely generated, so that U' admits a finite presentation, the mapping ϕ may not be injective (see Example 9.5.3).

Example 9.5.2. Let $\mathbb{Z}_2 = \{0, 1\}$ be the cyclic group of order 2. Let $G = \mathbb{Z}^3$ be the direct product of three copies of $\mathbb{Z}$. Let $E = \mathbb{Z}_2 \ltimes G$, where $1 \in \mathbb{Z}_2$ acts on G (from the right) by interchanging the leftmost two coördinates of an element of G.

Let $A = \{a, b\}$ represent elements of E as follows:

$$\overline{a} = [1, (0, 0, 1)], \ \overline{b} = [0, (1, 0, 0)].$$

Let $S = \mathrm{Sg}\langle \overline{A} \rangle$. Let T be the intersection of S and G.

Each element of T with a non-zero second integer coördinate has at least 2 in the rightmost integer coördinate: at least one generator $\overline{a}$ would be required to obtain a non-zero second coördinate; another is required to have identity $\mathbb{Z}_2$-component. Each of these generators contributes 1 to the rightmost coördinate.

However, T contains an element

$$\overline{ab^{\alpha}a} = [0, (0, \alpha, 2)]$$

for each $\alpha \in \mathbb{N}$. By the observation in the last paragraph, none of these elements decomposes into a product of other elements of T. Therefore $T = S \cap G$ cannot be finitely generated.

Example 9.5.3. Let $\mathbb{Z}_2 = \{0, 1\}$ be the cyclic group of order 2. Let $G = \mathrm{FG}(x, y, z)$. Let $E = \mathbb{Z}_2 \ltimes G$, where $1 \in \mathbb{Z}_2$ acts on $G = \mathrm{FG}(x, y, z)$ (from the right) by the automorphism induced by $x \mapsto y$ and $y \mapsto x$. Then E is a finite extension of the free group $\mathrm{FG}(x, y, z)$.

Let $A = \{a, b\}$ represent elements of E as follows:

$$\overline{a} = [1, xz], \ \overline{b} = [1, yz].$$

Let $S = \mathrm{Sg}\langle\overline{A}\rangle$. Let U' be the universal group of $S \cap G$ and U the universal group of S. Clearly, $S \cap G$ consists of those elements of S represented by words over A of even length, for it is these elements that have 0 as their $\mathbb{Z}_2$-component. Let

$$\begin{aligned} \overline{c} &= \overline{a^2} = [0, yzxz], & \overline{e} &= \overline{ba} = [0, xzxz], \\ \overline{d} &= \overline{ab} = [0, yzyz], & \overline{f} &= \overline{b^2} = [0, xzyz]. \end{aligned}$$

The subsemigroup $S \cap G$ is generated by $B = \{\overline{c}, \overline{d}, \overline{e}, \overline{f}\}$. Inspecting the free group components of these elements shows that $S \cap G$ is a free semigroup with basis $\overline{B}$. The universal group U' of $S \cap G$ is therefore the free group of rank 4. However, its image in U under the extension of the inclusion mapping $S \cap G \hookrightarrow S$ is not free, because $(ce^{-1}f, d)$ is a consequence of relations holding in S:

$$\overline{ce^{-1}f} = \overline{a^2(ba)^{-1}b^2} = \overline{a^2a^{-1}b^{-1}b^2} = \overline{ab} = \overline{d}.$$

PRÉCIS

Summary is almost exact (6)
— *The Times* Cryptic Crossword

THROUGHOUT this thesis, and especially in Chapters 4–8, subsemigroups of several classes of groups and semigroups are studied from the perspective of Malcev presentations and automatism. Certain theorems for these classes parallel one another. There are also a number of examples that establish, for example, that a particular class is not closed under a given construction. However, these results and examples are scattered throughout the thesis, depending on the context in which they arise. The purpose of this Précis is to gather and summarize these examples and results for the purposes of reference and comparison. Three tables contain the various data:

- **Table P.1** summarizes the closure of certain classes of groups under free product, finite extension, and direct product.
- **Table P.2** compares the Malcev coherence of groups satisfying various properties $\mathfrak{P}$, virtually $\mathfrak{P}$ groups, and direct products of these groups with free groups.
- **Table P.3** describes the Malcev coherence and local automatism of groups and semigroups of various classes.

Table P.1. Closure of various classes of groups under certain constructions.

	CLOSED UNDER		
CLASS OF GROUPS	FREE PRODUCT	FINITE EXTENSION	DIRECT PRODUCT
Coherent	Y (Th. 1.6.1)	Y (Th. 1.6.1)	N (§7.1)
Malcev coherent	N (Th. 6.2.6)	? (Op. 9.5.1)	N (§7.1)
All f.g. subgroups auto.	Y (Th. 2.4.17)	Y (Th. 2.4.17)	N (§7.1)
Locally automatic	N (Th. 6.2.8)	N (Th. 5.5.5)	N (§7.1)

Note. Y = Yes, N = No, ? = Undecided.

Table P.2. Malcev coherence of groups of certain classes.

	MALCEV COHERENT			
$\mathfrak{P}$	$\mathfrak{P}$	VIRT. $\mathfrak{P}$	FREE $\times$ $\mathfrak{P}$	FREE $\times$ VIRT. $\mathfrak{P}$
Free	Y (Th. 4.3.1)	Y (Th. 4.3.1)	N (Incoherent)	N (Incoherent)
Abelian	Y (Th. 5.3.5)	Y (Th. 5.3.5)	Y (Th. 7.5.1)	? (Op. 7.6.1)
Nilpotent	Y (Th. 5.3.5)	Y (Th. 5.3.5)	? (Op. 7.6.2)	? (Op. 7.6.2)
Polycyclic	N (Co. 7.3.10)	N (Co. 7.3.10)	N (Co. 7.3.10)	N (Co. 7.3.10)
One-relator	? (Op. 8.3.4)	? (Op. 8.3.4)	N (Incoherent)	N (Incoherent)

Notes.

1. Y = Yes, N = No, ? = Undecided.
2. It is unknown whether all [virtually] one-relator groups are coherent. Except for the classes so marked, groups of the other classes in this table are known to be coherent.
3. An entry 'N' in this table does not preclude a particular group of the class in question from being Malcev coherent. It merely asserts that not all such groups have that property.
4. The term 'free' could be replaced by 'virtually free' in the headings of the rightmost two columns of this table without altering its validity. However, it is conceivable that — for example — direct products of free groups and nilpotent groups are Malcev coherent but that there exists a direct product of a *virtually* free group and a nilpotent group that is not Malcev coherent. This situation, however, cannot arise if Open Problem 9.5.1 has a positive answer.

Table P.3. Malcev coherence and local automatism of various semigroups.

SEMIGROUP	MALCEV COHER.	LOCALLY AUTO.
Virtually free group	Y (Th. 4.3.1)	Y (Th. 4.4.1)
Virt. abelian/nilpotent/polycyclic group:		
Abelian group	Y (Th. 5.3.5)	Y (Th. 5.4.2)
Virtually abelian group	Y (Th. 5.3.5)	N (Ex. 5.5.1)
Nilpotent/virt. nilpotent group	Y (Th. 5.3.5)	N (Not auto.)
Polycyclic group	N (Co. 7.3.10)	N (Not auto.)
Direct product:		
Free semigroup × Free semigroup	N (Ex. 7.2.1)	N (Ex. 7.2.1)
Free semigroup × Natural numbers	Y (Th. 7.5.1)	N (Ex. 7.5.10)
Virt. free group × Abelian group	Y (Th. 7.5.1)	N (Ex. 7.5.10)
Virt. free group × Virt. abel. group	? (Op. 7.6.1)	N (Ex. 5.5.1)
Virt. free group × Nil./virt. nil. group	? (Op. 7.6.2)	N (Not auto.)
Free product:		
Free group ∗ Abelian group	N (Ex. 6.2.1)	N (Ex. 6.2.1)
Free monoid ∗ Abelian group	Y (Co. 6.3.5)	Y (Th. 6.3.1)
Free monoid ∗ Virt. abelian. group	? (Op. 6.3.6)	N (Ex. 5.5.1)
Free monoid ∗ Nil./virt. nil. group	? (Op. 6.3.6)	N (Not auto.)
Free monoid ∗ Malcev coher. monoid	? (Op. 6.3.7)	N (Not auto.)
Baumslag–Solitar semigroup/group:		
$\mathrm{BSS}(m,n)$, $m > n$	Y (Co. 8.3.1)	Y (Th. 8.2.1)
$\mathrm{BSS}(m,n)$, $m < n$	Y (Co. 8.3.1)	L (Th. 8.2.2)
$\mathrm{BSS}(m,m)$, $m > 1$	? (Op. 8.3.2)	N (Pr. 8.2.5)
$\mathrm{BSG}(m,n)$, $m \neq n$	? (Op. 8.3.3)	N (Not auto.)
$\mathrm{BSG}(m,m)$, $m > 1$	? (Op. 8.3.3).	N (Pr. 8.2.5)
One-relator cancellative semigroup	? (Op. 8.3.5)	N (Pr. 8.2.5)
One-relator group	? (Op. 8.3.4)	N (Pr. 8.2.5)

Notes.

1. Y = Yes, N = No, L = All finitely generated subsemigroups are left-automatic, ? = Undecided.
2. An entry 'N' in this table does not preclude a particular semigroup of the class in question from being Malcev coherent or locally automatic. It merely asserts that not all such semigroups have that property.

APPENDIX A

FORMAL LANGUAGES & AUTOMATA

The proverbial German 'verb-at-the-end', about which droll tales of absentminded professors who would begin a sentence, ramble on for an entire lecture, then finish up by rattling off a string of verbs by which their audience, for whom the stack had long since lost its coherence, would be totally nonplussed, are told, is an excellent example of linguistic pushing and popping.

— Douglas R. Hofstadter,
Gödel, Escher, Bach (1979), ch. v

A.1. INTRODUCTION

PURELY LINGUISTIC results required in this thesis are gathered together in this appendix. Most of the material herein may be found in the standard texts in formal language theory. The likeliest exceptions are the treatment of rational relations and asynchronous automata in Section A.6, and the discussion of regular predicates in Section A.5. For proofs of the main results quoted below, the reader is referred to Hopcroft *&* Ullman, *Introduction to Automata Theory, Languages, and Computation* [Reading: Addison–Wesley, (1979)], which is probably unsurpassed as a general reference to the subject. Alternatively, see Harrison (1978), Howie (1991), or Hopcroft, Ullman *&* Motwani (2001).

A.2. OPERATIONS ON LANGUAGES

RECALL that an alphabet is an abstract set of symbols; a word is a string of those symbols; and a language L over an alphabet A is simply a set of words over A. [Section 0.2 formally defines these concepts.]

The *concatenation* of two languages L_1 and L_2, denoted L_1L_2, is the language

$$\{u_1u_2 : u_1 \in L_1, u_2 \in L_2\}.$$

For $n \in \mathbb{N}$, let $L^n = (L^{n-1})L$ and $L^0 = \{\varepsilon\}$.

The *Kleene closure* of a language L over an alphabet A, denoted L^*, is the language

$$L^* = \bigcup_{n \in \mathbb{N} \cup \{0\}} L^n.$$

The unary operation returning the Kleene closure of a language is called the *Kleene star* operation. Observe that the notation for the Kleene star is compatible with the notation A^* for the free monoid over A. Similarly, one defines

$$L^+ = \bigcup_{n \in \mathbb{N}} L^n,$$

which agrees with the notation used for the free semigroup A^+ over A.

A.3. REGULAR LANGUAGES AND RATIONAL SUBSETS

THE MOST PLEASANT LANGUAGES — from a computational perspective — are the regular languages. This section defines regular languages by means of regular expressions, following closely the style of definition in Section 1.1 of Epstein et al. (1992).

Definition A.3.1. Let A be an alphabet. A *regular expression* over A is a string of a particular type over the extended alphabet $A \cup \{\ (\ ,\ )\ ,\ ^*\ ,\ \vee\ ,\ \varepsilon\ \}$. Any regular expression r defines a particular language, denoted by $L(r)$. If $w \in L(r)$, then the word w is said to *match* the regular expression r. The set of regular expressions over A is defined recursively as follows:

i.) The string ε is a regular expression over A, and $L(\varepsilon) = \{\varepsilon_A\}$. [Observe that '$\varepsilon$' is a symbol *in* the extended alphabet — it is not the empty word over that alphabet. It is, however, matched by the empty word over the alphabet A.]

ii.) The string $()$ is a regular expression over A, and $L(()) = \emptyset$.

iii.) For all letters $a \in A$, the string a is a regular expression over A, and $L(a) = \{a\}$.

iv.) If r is a regular expression over A, then so is (r), and $L((r)) = L(r)$.

v.) If r is a regular expression over A, then so is $(r)^*$, and $L((r)^*) = (L(r))^*$.

vi.) If r_1 and r_2 are regular expressions over A, then so is $(r_1)(r_2)$, and $L((r_1)(r_2)) = L(r_1)L(r_2)$.

vii.) If r_1 and r_2 are regular expressions over A, then so is $(r_1) \vee (r_2)$, and $L((r_1) \vee (r_2)) = L(r_1) \cup L(r_2)$.

[Rules viii–x. are not normally included in the definition of regular expressions. That they do not alter the class of languages defined by rules i–vii. is a consequence of Theorem A.5.1. They simply serve to abbreviate more complex regular expressions. Formally, these extra rules require further augmentation of the extended alphabet defined above.]

viii.) If S is a finite set of words over A, then the string S is a regular expression and $L(S) = S$.

ix.) If r_1 is a regular expression, then so is $(r_1)^+$, and $L((r_1)^+) = (L(r_1))^+$.

x.) If r_1 and r_2 are regular expressions, then so is $(r_1) - (r_2)$, and $L((r_1) - (r_2)) = L(r_1) - L(r_2)$.

A language defined by a regular expression is called a *regular language.*

To avoid a surfeit of parentheses, one adopts the convention that the Kleene star (rule v.) and the operation $^+$ (rule viii.) have higher precedence than concatenation (rule vi.), which in turn has higher precedence than $\vee$ (rule vii.). This permits the omission of parentheses in those rules. Furthermore, the body of the thesis makes no notational distinction between a regular expression and the language it defines: for example, if r is a regular expression, then '$w \in r$' means 'w matches r'.

The following definition is slightly incongruous here. Although it is mentioned in Section A.4, its main purpose is to allow comparison of regular languages with rational relations, the subject of Section A.6.

Definition A.3.2. Let M be a monoid. The class of *rational subsets* of M is the smallest subset of the power set of M satisfying the following conditions:

i.) All finite subsets of M are rational subsets.

ii.) If S is a rational subset of M, then so is the submonoid generated by S, which is denoted S^*.

iii.) If S and T are rational subsets of M, then so are $S \cup T$ and ST.

Lemma A.3.3. *Let M be a monoid. If S is a rational subset of M, then so is the semigroup S^+ that S generates.*

Proof of A.3.3. By part ii. of Definition A.3.2, the set S^* is rational. By part iii, so is the set $S(S^*) = S^+$. $\boxed{\text{A.3.3}}$

A.4. FINITE STATE AUTOMATA

INFORMALLY, a finite state automaton is a 'machine' with a finite number of internal states that reads a string of input symbols from a tape and either accepts or rejects that input string. Suppose that a finite state automaton is in internal state q and is reading the symbol a from its input tape. Then the automaton changes to a new input state q' dependent on q and a and advances to read the next input symbol from the tape. In certain cases, the new state q' depends only on q and a; in others, the automaton may have an element of choice. It may even be possible for the automaton to change spontaneously to a new state without advancing the input tape. (One supposes the automaton reads the empty word ε.) If the automaton reads and accepts an input string w — or if it *can* do so by making the right choices — then the automaton is said to accept w. The set of all such accepted words w is called the language recognized by the automaton.

The following definition formalizes the notions introduced in the last paragraph.

Definition A.4.1. A *finite state automaton* (FSA) $\mathcal{A}$ over a finite alphabet A is a quintuple (Q, A, δ, q_0, Y), where Q is a finite set of *states*; $\delta : Q \times (A \cup \{\varepsilon\}) \to \mathbb{P}(Q)$ is a function, known as the *transition function*; $q_0 \in Q$ is called the *initial state*; and $Y \subseteq Q$ is the set of *accept states*.

A word $w \in A^*$ is *accepted* by the automaton $\mathcal{A}$ if there is a sequence of states $q_0, \ldots, q_m$ and letters $a_1, \ldots, a_m \in A \cup \{\varepsilon\}$ such that q_m is an accept state, $w = a_1 a_2 \cdots a_m$ and, for each $i = 1, 2, \ldots, m$, the state q_i lies in $(q_{i-1}, a_i)\delta$.

The language $L(\mathcal{A})$ recognized by the automaton $\mathcal{A}$ is the set of words $w \in A^*$ accepted by $\mathcal{A}$.

The automaton $\mathcal{A}$ of Definition A.4.1 is best visualized as a directed graph. The vertex set is Q and for all $q, q' \in Q$ there is an edge from q to q' labelled by a whenever q' lies in $(q, a)\delta$. The label on a trail γ is the concatenation (or multiplication in A^*) of the labels on the edges making up γ. A word $w \in A^*$ is accepted by $\mathcal{A}$ if and only if there is a trail in the graph labelled by w leading from q_0 to an accept state. As a consequence of this graphical representation, it makes sense to think of δ not as a function, but as a set of edges called *transitions*: for each $q' \in (q, a)\delta$, there is a transition from q to q' labelled by a. Such a transition is denoted (q, a, q') or $q \xrightarrow{a} q'$.

Definition A.4.2. A finite state automaton $\mathcal{A} = (Q, A, \delta, q_0, Y)$ is called *deterministic* if, for each $q \in Q$ and $a \in A$, the set $(q, a)\delta$ contains at most one element, and that the set $(q, \varepsilon)\delta$ is empty for each $q \in Q$. [Graphically, this means that each vertex q is the source of at most one edge labelled with the letter a and that ε labels no edge in the automaton.]

In a deterministic finite state automaton, any word labels at most one path starting at any given state. A word is accepted by the automaton if it labels a path starting at the start state and if that necessarily unique path leads to an accept state.

Theorem A.4.3 (Hopcroft & Ullman 1979, Theorem 2.1). *Given any [possibly non-deterministic] finite state automaton, one can effectively construct a deterministic* FSA *recognizing the same language. Thus the class of languages recognized by deterministic finite state automata coincides with the class of languages recognized by all finite state automata.* $\boxed{\text{A.4.3}}$

There is an extension to the concept of an FSA called a *generalized finite state automaton*. Exactly as its name implies, generalized finite state automata are extensions of 'standard' finite state automata. Rather than having edges labelled by letters of the alphabet or by the empty word, a generalized FSA has its edges labelled by regular expressions. Suppose a generalized FSA is in a state q, and the word w remains on its input tape. The automaton can read any prefix $w(k)$ of w and move to a state q' if there is an edge from q to q' labelled by a regular expression r such that $w(k)$ matches r.

A word $w \in A^*$ is therefore recognized by a generalized FSA over A if there is a sequence of states $q_0, \ldots, q_m$, where q_0 is the initial state and q_m is an accept state, and a factorization of w as $w_1 \cdots w_m$, where $w_i \in A^*$, such that, for each $i = 1, 2, \ldots, m$, there is a transition from q_{i-1} to q_i labelled by a regular expression matched by w_i.

Theorem A.4.4 (Hopcroft & Ullman 1979, Section 2.5). *Let A be a finite alphabet and A^* be the free monoid over A. Then the following four classes coincide:*

i.) *The class of rational subsets of A^*.*

ii.) *The class of regular languages over A.*

iii.) *The class of languages recognized by finite state automata over A.*

iv.) *The class of languages recognized by generalized finite state automata over A.*

Indeed, there is an algorithm that takes a regular expression r and yields a finite state automaton recognizing $L(r)$; and an algorithm that takes a [generalized] finite state automaton $\mathcal{A}$ and yields a regular expression defining $L(\mathcal{A})$. $\boxed{\text{A.4.4}}$

As a consequence of Theorem A.4.4, the terms 'regular language', 'rational language', and — less commonly — 'recognizable language' are used interchangeably. Furthermore, one can effectively convert back and forth between a representation of a regular language as a finite state automaton and a representation of the same language as a regular expression.

For further discussion of finite state automata, refer to Chapters 2–3 of Hopcroft & Ullman (1979) or Chapter 1 of Epstein et al. (1992).

A.5. PROPERTIES OF REGULAR LANGUAGES

Theorem A.5.1 (Hopcroft & Ullman 1979, Theorems 3.1, 3.2, & 3.3). *The class of regular languages is closed under the operations of concatenation, intersection, union, complement, and set difference. Moreover, each of these operations can be effectively computed. That is, there are algorithms that take as input finite state automata $\mathcal{A}$ and $\mathcal{B}$ and yield finite state automata recognizing the languages:*

i.) $L(\mathcal{A})L(\mathcal{B})$,

ii.) $L(\mathcal{A}) \cap L(\mathcal{B})$,

iii.) $L(\mathcal{A}) \cup L(\mathcal{B})$,

iv.) $A^* - L(\mathcal{A})$,

v.) $L(\mathcal{B}) - L(\mathcal{A})$. $\boxed{\text{A.5.1}}$

By virtue of Theorem A.4.4, there exist algorithms analogous to those mentioned in the statement of Theorem A.5.1 whose inputs and outputs are regular expressions rather than FSAs. The same observation applies to the following results:

Theorem A.5.2 (Hopcroft & Ullman 1979, Theorem 3.5). *The class of regular languages is closed under forming homomorphic and inverse homomorphic images. That is, for every homomorphism $\phi : A^* \to B^*$, where A and B are alphabets, and* *for all regular languages $L \subseteq A^*$ and $M \subseteq B^*$, the languages $L\phi$ and $M\phi^{-1}$ are regular. Moreover, there are algorithms that take finite state automata recognizing L and M and the homomorphism ϕ (specified by the image of each letter of A), and return finite state automata recognizing $L\phi$ and $M\phi^{-1}$.* $\boxed{\text{A.5.2}}$

Theorem A.5.3 (Hopcroft & Ullman 1979, Section 3.3). *Let $\mathcal{A}$ be a finite state automaton over an alphabet A. There is an algorithm that tests whether the language recognized by $\mathcal{A}$ is empty.* $\boxed{\text{A.5.3}}$

Theorem A.5.4. *Let $\mathcal{A}$ be a finite state automaton over a finite alphabet A and let u be a word over A. Then there is an algorithm that tests whether u is accepted by $\mathcal{A}$.* $\boxed{\text{A.5.4}}$

Theorem A.5.5 (Hopcroft & Ullman 1979, Theorem 3.8). *Let $\mathcal{A}$ and $\mathcal{B}$ be finite state automata. Then there is an algorithm that tests whether $L(\mathcal{A}) \subseteq L(\mathcal{B})$ and thus an algorithm that tests whether $L(\mathcal{A}) = L(\mathcal{B})$.* $\boxed{\text{A.5.5}}$

For any alphabet A and natural number n, the *padded alphabet* $A(n, \$)$ is the set $\{(a_1, \ldots, a_n) : a_i \in A \cup \{\$\}\} - \{(\$, \ldots, \$)\}$. Define the mapping $\delta_A : (A^+)^n \to A(n, \$)^+$ as follows: to obtain the image of $(w_1, \ldots, w_n) \in (A^+)^n$, pad each word w_i to length $m = \max\{|w_j| : j = 1, \ldots, n\}$ by appending symbols \$. If the resulting tuple is $(a_1 \cdots a_m, b_1 \cdots b_m, \ldots)$, the desired image is

$$(a_1, b_1, \ldots)(a_2, b_2, \ldots) \cdots \in A(n, \$)^+.$$

[The mapping of Section 2.2 is a specialized version of the one just defined. Actually, there is no need for the same alphabet to be used in each component, but this restricted definition suffices for the purposes of this thesis.]

Formally, a *predicate* is a boolean value function $\sigma : ((A^+)^n)\delta_A \to \{\mathrm{T}, \mathrm{F}\}$. Every predicate σ has a corresponding language $\mathrm{T}\sigma^{-1}$ over $A(n, \$)$. A predicate is *regular* if its corresponding language is regular. Theorem A.5.1 implies that the class of regular predicates is closed under the operations $\wedge$, $\vee$, $\neg$, since:

$$\begin{aligned} \mathrm{T}(\sigma \wedge \tau)^{-1} &= (\mathrm{T}\sigma^{-1}) \cap (\mathrm{T}\tau^{-1}), \\ \mathrm{T}(\sigma \vee \tau)^{-1} &= (\mathrm{T}\sigma^{-1}) \cup (\mathrm{T}\tau^{-1}), \\ \mathrm{T}(\neg\sigma)^{-1} &= ((A^+)^n)\delta_A - (\mathrm{T}\sigma^{-1}). \end{aligned}$$

It is also closed under $\implies$ since $\alpha \implies \beta$ is defined as $(\neg\alpha) \vee \beta$.

Let σ be a regular predicate with $\mathrm{T}\sigma^{-1} \subseteq ((A^+)^n)\delta_A$. The predicate $\exists\sigma$ is defined by

$$\begin{aligned} \mathrm{T}(\exists\sigma)^{-1} = \big\{ &(w_1, \ldots, w_{n-1}) \in (A^+)^{n-1} : \\ &(w_1, \ldots, w_{n-1}, w_n)\delta_A \in \mathrm{T}\sigma^{-1} \text{ for some } w_n \big\}\delta_A \subseteq ((A^+)^{n-1})\delta_A. \end{aligned}$$

The predicate $\exists\sigma$ is regular. Since $\forall\sigma = \neg(\exists(\neg\sigma))$, the predicate $\forall\sigma$ is also regular.

Theorem A.5.6 (Epstein et al. 1992, Corollary 1.4.7). *The class of regular predicates is closed under* $\wedge$, $\vee$, $\neg$, $\implies$, $\forall$, *and* $\exists$. $\boxed{\text{A.5.6}}$

The 'syntactic congruence' is used briefly in Section 9.3:

Definition A.5.7 (Howie 1991, Section 3.1). Let L be a language over an alphabet A. The relation

$$\rho_L = \{(x, y) : (\forall u, v \in A^*)(uxv \in L \iff uyv \in L)\}$$

on A^* is called the *syntactic congruence* of L.

As its name suggests, the syntactic congruence is indeed a congruence on the monoid A^*. Actually, it is the maximum congruence on A^* that respects L (Howie 1991, Theorem 3.1.3).

Theorem A.5.8 (Howie 1991, Theorem 3.1.4). *A language L over a finite alphabet A is regular if and only if the number of ρ_L-classes is finite.* $\boxed{\text{A.5.8}}$

A.6. RATIONAL RELATIONS AND ASYNCHRONOUS AUTOMATA

AS ONE MIGHT SUSPECT, a rational relation between A^* and B^*, where A and B are both alphabets, is a relation that forms a rational subset of $A^* \times B^*$. [See Section A.3 for the definition of a rational subset.] As one might hope, these relations possess elegant properties that make them pleasant to work with. They can be recognized by special types of automata called 'asynchronous [two-tape] automata' (Epstein et al. 1992, Shapiro 1992) or '2-automata' (Pelletier & Sakarovitch 1999).

Much of the theory of rational relations is almost 'folklore': whilst the area has been active for some decades (the earliest work seems to be due to Rabin & Scott (1959)), it has not been sufficiently popular to warrant inclusion in standard texts. Perhaps as a consequence of this, there are many slight variations on the definitions. In particular, this section follows Rabin & Scott (1959) in requiring 'endmarkers'. As Pelletier & Sakarovitch (1999, Section 2) observe, the presence or absence of endmarkers makes no difference to the class of relations recognized by asynchronous automata. However, the class of relations recognized by deterministic asynchronous automata *with* endmarkers strictly contains the class recognized by deterministic asynchronous automata *without* endmarkers (see Pelletier & Sakarovitch 1999, Example 2.2). Furthermore, both of these classes are properly contained in the class of relations recognized by arbitrary asynchronous automata, which coincides with the class of rational relations by Theorem A.6.2.

Definition A.6.1. Let A and B be finite alphabets. An *asynchronous [two-tape] automaton [with endmarkers]* over A and B is a sextuple $\mathcal{A} = (Q, A, B, \delta, q_0, Y)$. As with finite state automata, Q is a set of states, q_0 is the initial state, and Y the set of accept states. The transitions in δ are labelled by elements of

$$((A \cup \{\$\}) \times \{\varepsilon_B\}) \cup (\{\varepsilon_A\} \times (B \cup \{\$\})) \cup \{(\varepsilon_A, \varepsilon_B)\},$$

where \$ is a new symbol in neither A nor B. The label of a trail in $\mathcal{A}$ is the product in $A^* \times B^*$ of the labels on its edges.

An element (u, v) is accepted by $\mathcal{A}$ if there is a trail in $\mathcal{A}$ from q_0 to a state of Y whose label is $(u\$, v\$)$. The relation — the subset of $A^* \times B^*$ — recognized by $\mathcal{A}$ is the set of all such (u, v).

The interpretation of a label from $(A \cup \{\$\}) \times \{\varepsilon_B\}$ is that the automaton reads a symbol of the alphabet $A \cup \{\$\}$ from the left-hand input tape; a label from $\{\varepsilon_A\} \times (B \cup \{\$\})$ is a symbol of $B \cup \{\$\}$ read from the right-hand input tape. The symbol \$ serves as an *endmarker*: when the automaton has read \$ from the left tape, it can read no more symbols from that tape. Similarly, when it has read \$ from the right tape, it reads no more symbols from that tape.

Theorem A.6.2. *A subset of $A^* \times B^*$ is rational if and only if it is recognized by an asynchronous two-tape automaton whose edges are labelled by elements of some generating set for $A^* \times B^*$.* $\boxed{\text{A.6.2}}$

Theorem A.6.2 is a special case of a result of Elgot & Mezei (1965). In general, one can recognize a rational subset of an arbitrary monoid M using a finite state automaton whose edges are labelled by elements of any generating set of M. However, for the purposes of this thesis, Definition A.6.1 and Theorem A.6.2 suffice.

Definition A.6.3. An asynchronous [two-tape] automaton $\mathcal{A} = (Q, A, B, \delta, q_0, Y)$ is *deterministic* if it satisfies the following two additional conditions:

i.) There is a partition of Q as $Q_A \cup Q_B$, where all edges whose origin lies in Q_A are labelled by elements of $(A \cup \{\$\}) \times \{\varepsilon_B\}$ and all edges whose origin lies in Q_B are labelled by elements of $\{\varepsilon_A\} \times (B \cup \{\$\})$. (Observe that no edges are labelled by $(\varepsilon_A, \varepsilon_B)$.)

ii.) For each state, no two edges from that state have the same label.

The final definition in this section is rather technical, and is only useful in working with asynchronous automatic structures (see Subsection 2.3.2).

Definition A.6.4. An asynchronous [two-tape] automaton $\mathcal{A}$ is called *boundedly asynchronous* if there exists a constant $k \in \mathbb{N}$ such that $\mathcal{A}$ cannot read more than k letters consecutively from one tape.

Theorem A.6.5 (Hoffmann et al. 2002*a*, Proposition 2.1(3)). *Let $R_1 \subseteq A^* \times B^*$ and $R_2 \subseteq B^* \times C^*$ be rational relations. Then their composition*

$$R_1 \circ R_2 = \{(u, v) \in A^* \times C^* : (\exists w \in B^*)((u, w) \in R_1 \wedge (w, v) \in R_2)\}$$

is a rational relation between A^* and C^*. $\boxed{\text{A.6.5}}$

A.7. PUSHDOWN AUTOMATA

A PUSHDOWN AUTOMATON is, loosely speaking, a finite state automaton augmented by a stack. A stack is a list of potentially arbitrary length, which can store an unlimited amount of information. However, only the 'top' of this list is accessible. One imagines a stack to be just that: a pile of items. One may inspect the item sitting on the top of the pile; one may remove that top item; or one may add further items. This restriction on how one can access the list means that the last item to be added to the stack must be the first to be removed. [Thus stacks are sometimes called LIFO ('last-in-first-out') lists. Knuth (1997, Section 2.2.1) treats of stacks in the context of more general list data structures.]

The extra storage provided by the stack allows pushdown automata to recognize languages that cannot be recognized by finite state automata.

Definition A.7.1. A *pushdown automaton* (PDA) $\mathcal{P}$ over a finite alphabet A is a sextuple $(Q, A, B, \delta, q_0, b_0, Y)$, where Q is a finite set of *states*; B is a finite alphabet, called the *stack alphabet*; δ is a function from $Q \times (A \cup \{\varepsilon\}) \times B$ to finite subsets of $Q \times B^*$, known as the *transition function*; $q_0 \in Q$ is called the *initial state*; $b_0 \in B$ is the *initial stack symbol*; and $Y \subseteq Q$ is the set of *accept states*.

The state of the PDA $\mathcal{P}$ at any given point may be described by a triple (q, w, s), where $q \in Q$ is the internal state; $w \in A^*$ is the input yet to be read; and $s \in B^*$ is the current contents of the stack. Such a triple is called an *instantaneous description*. The PDA can move from (q, aw, bs) to (q', w, ts) (where $q, q' \in Q$, $w \in A^*$, $a \in A \cup \{\varepsilon\}$, and s, $t \in B^*$) if $(q, a, b)\delta$ contains (q', t). The interpretation of such a move is that the automaton has read a whilst in state q with b at the top of the stack, and has moved to state q', popped b from the stack and pushed t onto the top of the stack.

A word $w \in A^*$ is *accepted* by the automaton $\mathcal{P}$ if there is a sequence of instantaneous descriptions $d_1, \ldots, d_m$, with $d_1 = (q_0, w, b_0)$ and $d_m = (q, \varepsilon, s)$, where $q \in Y$, $s \in B^*$ is arbitrary, and the automaton can move from d_i to d_{i+1} for each $i = 1, \ldots, m$.

The language $L(\mathcal{P})$ recognized by the pushdown automaton $\mathcal{P}$ is the set of words $w \in A^*$ recognized by $\mathcal{P}$.

The PDA $\mathcal{P}$ is *deterministic* if only one move is possible given a particular input letter and symbol at the top of the stack. More formally, for $q \in Q$ and $b \in B$, if the set $(q, \varepsilon, b)\delta$ is non-empty, then $(q, a, b)\delta$ is empty for all $a \in A$; and, for $q \in Q$, $a \in A \cup \{\varepsilon\}$, and $b \in B$, the set $(q, a, b)\delta$ contains at most one element of $Q \times B^*$.

A.8. CONTEXT-FREE GRAMMARS

Definition A.8.1. Let A be a finite alphabet. A *context-free grammar* (CFG) Γ over A is a quadruple (N, A, P, O), where N is a finite alphabet of *non-terminals*; P is a finite set of *productions*, defined below; and $O \in N$ is called the *start symbol*. The set of productions P is a finite subset of $N \times (N \cup A)^*$. A pair $(n, r) \in P$ is usually denoted $n \to r$. The symbolism $n \to r_1 \mid \ldots \mid r_k$ abbreviates the k separate productions $n \to r_i$.

Define a relation $\Rightarrow$ on $(N \cup A)^*$ by

$$unv \Rightarrow urv \text{ for all } u, v \in (N \cup A)^* \text{ and } (n \to r) \in P.$$

Let $\overset{*}{\Rightarrow}$ be the reflexive and transitive closure of $\Rightarrow$.

The language defined by the context-free grammar Γ is

$$L(\Gamma) = \{w : O \overset{*}{\Rightarrow} w \text{ and } w \in A^*\}.$$

Any language that can be defined by means of a context-free grammar is called a context-free language (CFL).

It is possible for two different context-free grammars to define the same context-free language. In general, it is undecidable whether two context-free grammars define the same language (Hopcroft & Ullman 1979, Theorem 8.12).

Theorem A.8.2 (Hopcroft & Ullman 1979, Section 5.3). *The class of context-free languages coincides with the class of languages recognized by pushdown automata. Moreover, there is an algorithm that takes a pushdown automaton $\mathcal{P}$ and yields a context-free grammar defining $L(\mathcal{P})$; and an algorithm that takes a context-free grammar Γ and yields a pushdown automaton recognizing $L(\Gamma)$.* $\boxed{\text{A.8.2}}$

Theorem A.8.3 (Hopcroft & Ullman 1979, Theorem 6.6). *There is an algorithm that tests whether $L(\Gamma)$ is empty, where Γ is a context-free grammar.* $\boxed{\text{A.8.3}}$

Theorem A.8.4 (Hopcroft & Ullman 1979, pp. 139–141). *There is an algorithm that tests whether $w \in L(\Gamma)$, where Γ is a context-free grammar over an alphabet A and w is a word in A^*.* $\boxed{\text{A.8.4}}$

Theorem A.8.5 (Hopcroft & Ullman 1979, Theorem 6.5). *The class of context-free languages is closed under intersection with the class of regular languages. That is, if L be a context-free language and R a regular language, then $L \cap R$ is a context-free language. Furthermore, given a pushdown automaton recognizing L and a finite state automaton recognizing R, it is possible to effectively construct a pushdown automaton recognizing $L \cap R$.* $\boxed{\text{A.8.5}}$

Theorem A.8.6 (Hopcroft & Ullman 1979, Theorem 6.3). *The class of context-free languages is closed under forming inverse homomorphic images. That is, for every homomorphism $\phi : A^* \to B^*$, where A and B are alphabets, and for every context-free language $L \subseteq B^*$, the language $L\phi^{-1}$ is context-free. Moreover, there is an algorithm that takes a pushdown automaton $\mathcal{P}$ over B and a homomorphism $\phi : A^* \to B^*$ (specified by the images of each $a \in A$), and returns a pushdown automaton recognizing $(L(\mathcal{P}))\phi^{-1}$.* $\boxed{\text{A.8.6}}$

A context-free language is *deterministic* if it can be recognized by a deterministic pushdown automaton. The class of general context-free languages is not closed under taking complements (Hopcroft & Ullman 1979, Corollary to Theorem 6.4). The situation differs for deterministic context-free languages:

Theorem A.8.7 (Hopcroft & Ullman 1979, Theorem 10.1). *The class of deterministic context-free languages is closed under taking complements.* $\boxed{\text{A.8.7}}$

A.8.1. Derivation trees

Informally, a derivation tree is a directed tree that describes how a word in the language defined by a context-free grammar Γ can be obtained from the start symbol of Γ by replacing left-hand sides of productions by right-hand sides. For further information on derivation trees, see Section 4.3 of Hopcroft & Ullman (1979).

Definition A.8.8. Let $\Gamma = (N, A, P, O)$ be a context-free grammar. A Γ-*derivation tree* (or simply a *derivation tree*) is a directed rooted tree T with the following properties:

i.) The vertices of T are labelled by elements of $N \cup A \cup \{\varepsilon\}$.

ii.) The internal vertices of the tree are labelled by elements of N. In particular, the root of the tree is labelled by the start symbol O.

iii.) The leaves of the tree are labelled by elements of $A \cup \{\varepsilon\}$.

iv.) If a vertex is labelled by $n \in N$, then either the child nodes of that vertex are labelled from left to right by the letters of r, where $n \to r$ is a production in P; or there must be a single child node of that vertex labelled by ε, where $n \to \varepsilon$ is a production. Any leaf node labelled by ε must be the only child of its parent node.

The *yield* of the derivation tree T is the word over A found by reading the leaves of T from left to right. Observe that the yield may be the empty word.

Theorem A.8.9 (Hopcroft & Ullman 1979, Theorem 4.1). *Let Γ be a context-free grammar. A word is in the language $L(\Gamma)$ if and only if there is a Γ-derivation tree whose yield is that word.* $\boxed{\text{A.8.9}}$

Example A.8.10. Let $\Gamma = (N, A, P, O)$ be a context-free grammar, where $N = \{O, M_1, M_2\}$, $A = \{a, b, c, d\}$, and P consists of the productions:

$$\begin{aligned} &O \to M_1cM_1cM_2, && \\ &M_1 \to aM_1b, && M_1 \to d, \\ &M_2 \to bM_2a, && M_2 \to \varepsilon. \end{aligned}$$

The word *adbcdcbbaa* is in the language $L(\Gamma)$, as is shown by the derivation tree in Figure A.1. Reading the leaves from left to right — as indicated by the dotted arrows — gives the yield.

There may be more than one Γ-derivation tree for some word w, in which case Γ is said to be *ambiguous*. Although it is possible that w may have only one Δ-derivation tree, where Δ is another context-free grammar defining the same language as Γ, there exist context-free languages such that any grammar defining one of those languages is ambiguous (Hopcroft & Ullman 1979, Theorem 4.7). Such languages are called *inherently ambiguous*. Both the ambiguity of a given context-free grammar and the inherent ambiguity of a particular context-free language are undecidable (Hopcroft & Ullman 1979, Theorems 8.9 and 8.16).

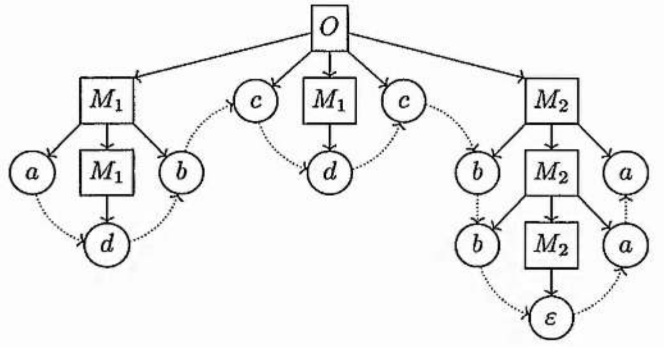

Figure A.1. An example of a derivation tree.

A.9. CONTEXT-SENSITIVE LANGUAGES

THE FINAL CLASS of languages needed is that of the context-sensitive languages. A full definition is included for completeness, although these languages are only mentioned in passing in Section 4.3. [Harrison (1978) treats of context-sensitive languages more fully than Hopcroft & Ullman (1979).]

Definition A.9.1. Let A be a finite alphabet. A *context-sensitive grammar* (CSG) Γ over A is a quadruple (N, A, P, O), where N is a finite alphabet of *non-terminals*; P is a finite set of *productions*, defined below; and $O \in N$ is called the *start symbol*. The set of productions P is a finite set of pairs of the form $O \to \varepsilon$ or $c_1nc_2 \to c_1rc_2$, where $n \in N$, $c_1, c_2 \in (N \cup A)^*$, and $r \in (N \cup A)^+$. [Observe that r cannot be empty.]

The definitions of the relations $\Rightarrow$ and $\stackrel{*}{\Rightarrow}$ are retained from Definition A.8.1. The language defined by the context-sensitive grammar Γ is

$$L(\Gamma) = \{w : O \stackrel{*}{\Rightarrow} w \text{ and } w \in A^*\}.$$

Any language that can be defined by means of a context-sensitive grammar is called a context-sensitive language (CSL).

Observe that the only substantive difference between Definition A.9.1 and Definition A.8.1 is in the productions. A CSG production $c_1nc_2 \to c_1rc_2$ functions like the CFG production $n \to r$, but is restricted to the 'context' c_1–c_2.

Proposition A.9.2. *The class of context-sensitive languages contains the class of context-free languages.* $\boxed{\text{A.9.2}}$

Proposition A.9.3. *The class of context-sensitive languages is closed under complementation.* $\boxed{\text{A.9.3}}$

[Whether the assertion of Proposition A.9.3 holds true was a long-standing open question in formal linguistics. The result was proven independently by Immerman (1988) and Szelepcsényi (1988).]

BIBLIOGRAPHY

And as for the Citation of so many Authors, 'tis the easiest Thing in Nature. Find out one of those Books with an alphabetical Index, and without any farther Ceremony, remove it verbatim *into your own...*

— Miguel de Cervantes,
Don Quixote de la Mancha (1605–15), vol. i, preface
(trans. P. A. Motteux)

ADJAN, S. I. (1966*a*)
'Определяющие соотношения и алгоритмические проблемы для групп и полугрупп' ('Defining relations and algorithmic problems for groups and semigroups'), *Trudy Matematicheskogo Instituta imeni V. A. Steklova* **85**. [In Russian. See Adjan (1966*b*) for a translation.] (Cited on p. 18.)

ADJAN, S. I. (1966*b*)
'Defining relations and algorithmic problems for groups and semigroups', *Proceedings of the Steklov Institute of Mathematics* **85**. [Translated from the Russian by M. Greendlinger.] (Cited on pp. 10, 18, 159.)

ADYAN, S. I. (1960*a*)
'On the embeddability of semigroups in groups', *Doklady Akademii Nauk SSSR* **1**, 819–821. [In Russian. See Adyan (1960*b*) for a translation.] (Cited on p. 10.)

ADYAN, S. I. (1960*b*)
'On the embeddability of semigroups in groups', *Soviet Mathematics Doklady* **1**, 819–821. [Translated from the Russian.] (Cited on p. 159.)

ALBERT, J., CULIK, II, K. & KARHUMÄKI, J. (1982)
'Test sets for context free languages and algebraic systems of equations over a free monoid', *Information and Control* **52**(2), 172–186. (Cited on p. 71.)

ALBERT, M. H. & LAWRENCE, J. (1985)
'A proof of Ehrenfeucht's conjecture', *Theoretical Computer Science* **41**(1), 121–123. (Cited on p. 71.)

ANĪSĪMOV, A. V. (1971)
'The group languages', *Otdelenie Matematiki, Mekhaniki i Kibernetiki Akademii Nauk Ukrainskoĭ SSR. Kibernetika* 1971 (4), 18–24. [In Russian.] (Cited on p. 62.)

ANĪSĪMOV, A. V. (1972)
'Certain algorithmic questions for groups and context-free languages', *Otdelenie Matematiki, Mekhaniki i Kibernetiki Akademii Nauk Ukrainskoĭ SSR. Kibernetika* 1972 (2), 4–11. [In Russian.] (Cited on p. 62.)

BANDYOPADHYAY, G. (1963)
'A simple proof of the decipherability criterion of Sardinas and Patterson', *Information and Control* **6**, 331–336. (Cited on p. 63.)

BAUMSLAG, G. (1962)
'A remark on generalized free products', *Proceedings of the American Mathematical Society* **13**, 53–54. (Cited on pp. 95, 107.)

BAUMSLAG, G. (1974)
'Some problems on one-relator groups', in *Proceedings of the Second International Conference on the Theory of Groups (Australian National University, Canberra, 1973)*, Vol. 372 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 75–81. (Cited on pp. 24, 131.)

BAUMSLAG, G. & ROSEBLADE, J. E. (1984)
'Subgroups of direct products of free groups', *Journal of the London Mathematical Society. Second Series* **30**(1), 44–52. (Cited on p. 25.)

BAUMSLAG, G. & SOLITAR, D. (1962)
'Some two-generator one-relator non-Hopfian groups', *Bulletin of the American Mathematical Society* **68**, 199–201. (Cited on p. 123.)

BAUMSLAG, G., GERSTEN, S. M., SHAPIRO, M. & SHORT, H. (1991)
'Automatic groups and amalgams', *Journal of Pure and Applied Algebra* **76**(3), 229–316. (Cited on p. 34.)

BLUM, E. K. (1965*a*)
'Free subsemigroups of a free semigroup', *Michigan Mathematical Journal* **12**, 179–182. (Cited on p. 63.)

BLUM, E. K. (1965*b*)
'A note on free subsemigroups with two generators', *Bulletin of the American Mathematical Society* **71**, 678–679. (Cited on p. 63.)

BOOK, R. V. & OTTO, F. (1993)
String-Rewriting Systems, Texts and Monographs in Computer Science, Springer-Verlag, New York. (Cited on pp. 4, 5.)

Brooksbank, P., Qin, H., Robertson, E. & Seress, Á. (2004)
'On Dowling geometries of infinite groups', *Journal of Combinatorial Theory. Series A* **108**(1), 155–158. (Cited on p. xiv.)

Budkina, L. G. & Markov, A. A. (1973*a*)
'Об F-полугруппах с тремя образующими' ('F-semigroups with three generators'), *Akademiya Nauk SSSR. Matematicheskie Zametki* **14**, 267–277. [In Russian. See Budkina & Markov (1973*b*) for a translation.] (Cited on p. 79.)

Budkina, L. G. & Markov, A. A. (1973*b*)
'F-semigroups with three generators', *Mathematical Notes* **14**, 267–277. [Translated from the Russian.] (Cited on p. 161.)

Cain, A. J. (2005*a*)
'Automatism of subsemigroups of Baumslag–Solitar semigroups'. In preparation. (Cited on pp. xiii, 123.)

Cain, A. J. (2005*b*)
'A finitely generated group-embeddable non-automatic semigroup whose universal group is automatic'. Submitted. (Cited on pp. xiii, 27, 87, 88, 96.)

Cain, A. J. (2005*c*)
'Subsemigroups of direct products of coherent groups'. In preparation. (Cited on pp. xiii, 103.)

Cain, A. J., Robertson, E. F. & Ruškuc, N. (2005*a*)
'Subsemigroups of groups: presentations, Malcev presentations and automatic structures'. Submitted. (Cited on pp. xiii, xiv, 25, 27, 28, 43, 72, 83, 90, 91, 92, 97.)

Cain, A. J., Robertson, E. F. & Ruškuc, N. (2005*b*)
'Subsemigroups of virtually free groups: finite Malcev presentations and testing for freeness', *Mathematical Proceedings of the Cambridge Philosophical Society*. To appear. (Cited on pp. xiii, 24, 25, 64, 66.)

Campbell, C. M., Robertson, E. F., Ruškuc, N. & Thomas, R. M. (1995)
'Reidemeister–Schreier type rewriting for semigroups', *Semigroup Forum* **51**(1), 47–62. (Cited on p. 133.)

Campbell, C. M., Robertson, E. F., Ruškuc, N. & Thomas, R. M. (2000)
'Direct products of automatic semigroups', *Australian Mathematical Society Journal. Series A — Pure Mathematics and Statistics* **69**(1), 19–24. (Cited on p. 40.)

Campbell, C. M., Robertson, E. F., Ruškuc, N. & Thomas, R. M. (2001)
'Automatic semigroups', *Theoretical Computer Science* **250**(1–2), 365–391. (Cited on pp. xii, 27, 29, 32, 37, 38, 40, 41, 42, 71, 85.)

CAMPBELL, C. M., ROBERTSON, E. F., RUŠKUC, N. & THOMAS, R. M. (2002)
'Automatic completely simple semigroups', *Acta Mathematica Hungarica* **95**(3), 201–215. (Cited on pp. 27, 32, 43.)

CANNON, J. W. (1984)
'The combinatorial structure of cocompact discrete hyperbolic groups', *Geometriae Dedicata* **16**(2), 123–148. (Cited on p. 27.)

CLIFFORD, A. H. & PRESTON, G. B. (1961)
The Algebraic Theory of Semigroups (Vol. I), number 7 in 'Mathematical Surveys', American Mathematical Society, Providence, R.I. (Cited on pp. 1, 9, 10, 81, 84.)

CLIFFORD, A. H. & PRESTON, G. B. (1967)
The Algebraic Theory of Semigroups (Vol. II), number 7 in 'Mathematical Surveys', American Mathematical Society, Providence, R.I. (Cited on pp. 1, 7, 9, 17, 18.)

COHN, P. M. (1962)
'On subsemigroups of free semigroups', *Proceedings of the American Mathematical Society* **13**, 347–351. (Cited on p. 63.)

CULIK, II, K. & SALOMAA, A. (1978)
'On the decidability of homomorphism equivalence for languages', *Journal of Computer and System Sciences* **17**(2), 163–175. (Cited on p. 71.)

CUTTING, A. & SOLOMON, A. (2001)
'Remarks concerning finitely generated semigroups having regular sets of unique normal forms', *Journal of the Australian Mathematical Society* **70**(3), 293–309. (Cited on p. 27.)

DE LUCA, A. (1976)
'A note on variable length codes', *Information and Control* **32**(3), 263–271. (Cited on p. 63.)

DESCALÇO, L. (2002)
Automatic Semigroups: Constructions and Subsemigroups, Ph.D. Thesis, University of St Andrews. (Cited on pp. 31, 96.)

DOMBI, E. R. (2004)
Automatic S-acts and inverse semigroup presentations, Ph.D. Thesis, University of St Andrews. (Cited on pp. 27, 31.)

DUBREIL, P. (1943)
'Sur les problèmes d'immersion et la théorie des modules' ('On problems of embedding and the theory of modules'), *Comptes Rendus Mathématique, Académie des Sciences, Paris* **216**, 625–627. (Cited on p. 10.)

DUBREIL, P. (1954)
Algèbre, Tome I — Equivalences, opérations, groupes, annaux, corps (*Algebra, Vol. I — Equivalences, operations, groups, rings, fields*), Second edition, Gauthier-Villars, Paris. (Cited on p. 10.)

DUNCAN, A. J., ROBERTSON, E. F. & RUŠKUC, N. (1999)
'Automatic monoids and change of generators', *Mathematical Proceedings of the Cambridge Philosophical Society* **127**(3), 403–409. (Cited on pp. 27, 38.)

DUNWOODY, M. J. (1985)
'The accessibility of finitely presented groups', *Inventiones Mathematicae* **81**(3), 449–457. (Cited on p. 62.)

DUNWOODY, M. J. (1993)
'An inaccessible group', in *Geometric group theory, Vol. 1* (Sussex, 1991), Vol. 181 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, pp. 75–78. (Cited on p. 62.)

ELGOT, C. C. & MEZEI, J. E. (1965)
'On relations defined by generalized finite automata', *IBM Journal of Research and Development* **9**, 47–68. (Cited on p. 154.)

EPSTEIN, D. B. A., CANNON, J. W., HOLT, D. F., LEVY, S. V. F., PATERSON, M. S. & THURSTON, W. P. (1992)
Word Processing in Groups, Jones & Bartlett, Boston, Mass. (Cited on pp. 2, 27, 28, 29, 30, 32, 33, 34, 35, 38, 40, 41, 42, 43, 50, 61, 85, 87, 90, 91, 96, 102, 124, 125, 148, 151, 153.)

ERDŐS, P., FÜREDI, Z., HAJNAL, A., KOMJÁTH, P., RÖDL, V. & SERESS, Á. (1986)
'Coloring graphs with locally few colors', *Discrete Mathematics* **59**(1–2), 21–34. (Cited on p. xiv.)

FEIGHN, M. & HANDEL, M. (1999)
'Mapping tori of free group automorphisms are coherent', *Annals of Mathematics. Second Series* **149**(3), 1061–1077. (Cited on p. 24.)

FRALEIGH, J. B. (1998)
A First Course in Abstract Algebra, Addison–Wesley Publishing Co., Reading, Mass. (Cited on p. 84.)

FREYD, P. (1968)
'Redei's finiteness theorem for commutative semigroups', *Procedings of the American Mathematical Society* **19**, 1003. (Cited on p. 85.)

GRAY, R. & RUŠKUC, N. (2005)
'Generators and relations for subsemigroups via boundaries in Cayley graphs'. Submitted. (Cited on p. 133.)

GRILLET, P.-A. (1993)
'A short proof of Rédei's theorem', *Semigroup Forum* **46**(1), 126–127. (Cited on pp. 85, 86.)

GRUNEWALD, F. J. (1978)
'On some groups which cannot be finitely presented', *Journal of the London Mathematical Society. Second Series* **17**(3), 427–436. (Cited on pp. 25, 103.)

GUBA, V. S. (1986)
'Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems', *Akademiya Nauk SSSR. Matematicheskie Zametki* **40**(3), 321–324, 428. [In Russian.] (Cited on p. 71.)

GUBA, V. S. (1994*a*)
'Conditions for the embeddability of semigroups into groups', *Mathematical Notes* **56**(1–2), 763–769. [Translated from the Russian.] (Cited on p. 164.)

GUBA, V. S. (1994*b*)
'Об условиях вложимости полугрупп в группы' ('Conditions for the embeddability of semigroups into groups'), *Rossiĭskaya Akademiya Nauk. Matematicheskie Zametki* **56**(2), 3–14, 158. [In Russian. See Guba (1994*a*) for a translation.] (Cited on pp. 10, 11.)

HALL, JR., M. (1949)
'Subgroups of finite index in free groups', *Canadian Journal of Math.* **1**, 187–190. (Cited on p. 107.)

HARRISON, M. A. (1978)
Introduction to Formal Language Theory, Addison–Wesley Publishing Co., Reading, Mass. (Cited on pp. 147, 158.)

HERBST, T. & THOMAS, R. M. (1993)
'Group presentations, formal languages and characterizations of one-counter groups', *Theoretical Computer Science* **112**(2), 187–213. (Cited on p. 62.)

HOFFMANN, M. (2001)
Automatic Semigroups, Ph.D. Thesis, University of Leicester. (Cited on pp. 32, 90, 124.)

HOFFMANN, M. & THOMAS, R. M. (2002)
'Automaticity and commutative semigroups', *Glasgow Mathematical Journal* **44**(1), 167–176. (Cited on p. 85.)

HOFFMANN, M. & THOMAS, R. M. (2003)
'Notions of automaticity in semigroups', *Semigroup Forum* **66**(3), 337–367. (Cited on pp. 31, 32.)

HOFFMANN, M., KUSKE, D., OTTO, F. & THOMAS, R. M. (2002a)
'Some relatives of automatic and hyperbolic groups', in G. M. S. Gomes, J.-É. Pin & P. V. Silva, editors, *Semigroups, Algorithms, Automata and Languages* (Coimbra, 2001), World Scientific Publishing, River Edge, N.J., pp. 379–406. (Cited on pp. xii, 28, 30, 37, 38, 41, 154.)

HOFFMANN, M., THOMAS, R. M. & RUŠKUC, N. (2002b)
'Automatic semigroups with subsemigroups of finite Rees index', *International Journal of Algebra and Computation* **12**(3), 463–476. (Cited on p. 38.)

HOLT, D. F., REES, S., RÖVER, C. E. & THOMAS, R. M. (2004)
'Groups with context-free co-word problem'. Submitted. Preprint: Techical Report 2004/23, Department of Computer Science, University of Leicester. (Cited on p. 71.)

HOPCROFT, J. E. & ULLMAN, J. D. (1979)
Introduction to Automata Theory, Languages, and Computation, Addison–Wesley Publishing Co., Reading, Mass. (Cited on pp. 54, 59, 63, 147, 150, 151, 152, 156, 157, 158.)

HOPCROFT, J. E., ULLMAN, J. D. & MOTWANI, R. (2001)
Introduction to Automata Theory, Languages, and Computation, Second edition, Addison–Wesley Publishing Co., Reading, Mass. (Cited on p. 147.)

HOWIE, J. M. (1991)
Automata and Languages, Oxford Science Publications, Clarendon Press, Oxford University Press, New York. (Cited on pp. 147, 153.)

HOWIE, J. M. (1995)
Fundamentals of Semigroup Theory, Vol. 12 of *London Mathematical Society Monographs* (New Series), Clarendon Press, Oxford University Press, New York. (Cited on pp. 1, 47, 92.)

HUDSON, J. F. P. (1996)
'Regular rewrite systems and automatic structures', in J. Almeida, G. M. S. Gomes & P. V. Silva, editors, *Semigroups, Automata and Languages* (Porto, 1994), World Scientific Publishing, River Edge, N.J., pp. 145–152. (Cited on p. 27.)

IMMERMAN, N. (1988)
'Nondeterministic space is closed under complementation', *SIAM Journal on Computing* **17**(5), 935–938. (Cited on p. 158.)

JACKSON, D. A. (2002)
'Decision and separability problems for Baumslag–Solitar semigroups', *International Journal of Algebra and Computation* **12**(1–2), 33–49. (Cited on p. 124.)

JOHNSON, D. L. (1997)
Presentations of Groups, Vol. 15 of *London Mathematical Society Student Texts*, Second edition, Cambridge University Press, Cambridge. (Cited on p. 83.)

JURA, A. (1978)
'Determining ideals of a given finite index in a finitely presented semigroup', *Demonstratio Mathematica* **11**(3), 813–827. (Cited on p. 132.)

KAMBITES, M. E. (2003)
Combinatorial Aspects of Partial Algebras, Ph.D. Thesis, University of York. (Cited on p. 90.)

KARRASS, A. & SOLITAR, D. (1970)
'The subgroups of a free product of two groups with an amalgamated subgroup', *Transactions of the American Mathematical Society* **150**, 227–255. (Cited on p. 131.)

KASHINTSEV, E. V. (1992)
'Small cancellation conditions and embeddability of semigroups in groups', *International Journal of Algebra and Computation* **2**(4), 433–441. (Cited on pp. 10, 11.)

KASHINTSEV, E. V. (2001*a*)
'О некоторых условиях вложимости полугрупп в гпуппы' ('On some conditions for the embeddability of semigroups in groups'), *Rossiĭskaya Akademiya Nauk. Matematicheskie Zametki* **70**(5), 705–717. [In Russian. See Kashintsev (2001*b*) for a translation.] (Cited on p. 11.)

KASHINTSEV, E. V. (2001*b*)
'On some conditions for the embeddability of semigroups in groups', *Mathematical Notes* **70**(5–6), 640–650. [Translated from the Russian.] (Cited on p. 166.)

KNUTH, D. E. (1997)
Fundamental Algorithms, Vol. 1 of *The Art of Computer Programming*, Addison–Wesley Publishing Co., Reading, Mass. (Cited on p. 155.)

KNUTH, D. E. (2002)
'Robert W. Floyd, In Memoriam'. Based on a speech to the Stanford Computer Forum, 20th March 2002. [URL: `http://www-cs-faculty.stanford.edu/~knuth/papers/floyd.ps.gz`] (Cited on p. 63.)

KROPHOLLER, P. H. (1990)
'Baumslag–Solitar groups and some other groups of cohomological dimension two', *Commentarii Mathematici Helvetici* **65**(4), 547–558. (Cited on pp. 24, 131.)

LALLEMENT, G. (1979)
Semigroups and Combinatorial Applications, John Wiley & Sons, New York, Chichester, Brisbane. (Cited on p. 64.)

LAMBEK, J. (1951)
'The immersibility of a semigroup into a group', *Canadian Journal of Mathematics* **3**, 34–43. (Cited on p. 9.)

LENNOX, J. C. & WIEGOLD, J. (1974)
'Some remarks on coherent soluble groups', *Bulletin of the Australian Mathematical Society* **10**, 277–279. (Cited on p. 24.)

LEVENŠTEĬN, V. I. (1961*a*)
'Certain properties of code systems', *Doklady Akademii Nauk SSSR* **140**, 1274–1277. [In Russian. See Levenšteĭn (1961*b*) for a translation.] (Cited on p. 63.)

LEVENŠTEĬN, V. I. (1961*b*)
'Certain properties of code systems', *Soviet Physics Doklady* **6**, 858–860. [Translated from the Russian.] (Cited on p. 167.)

LEWIN, J. & LEWIN, T. (1969)
'Semigroup laws in varieties of solvable groups', *Mathematical Proceedings of the Cambridge Philosophical Society* **65**, 1–9. (Cited on p. 80.)

LYNDON, R. C. & SCHUPP, P. E. (1977)
Combinatorial Group Theory, Vol. 89 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer-Verlag, Berlin. (Cited on pp. 3, 8, 11, 25, 42, 61, 63, 64, 92, 95, 108, 124, 132, 138.)

MAGNUS, W. (1932)
'Das Identitätsproblem für Gruppen mit einen definierenden Relation', *Mathematische Annalen* **106**, 295–307. [In German.] (Cited on p. 123.)

MALCEV, A. I. (1937)
'On the immersion of an algebraic ring into a field', *Mathematische Annalen* **113**, 686–691. (Cited on p. 9.)

MALCEV, A. I. (1939)
'О включении ассоииативных систем в группы' ('On the immersion of associative systems in groups'), *Matematicheskiĭ Sbornik* **6**(48), 331–336. [In Russian.] (Cited on pp. 9, 13, 18.)

MALCEV, A. I. (1940)
'О включении ассоииативных систем в группы' ('On the immersion of associative systems in groups II'), *Matematicheskiĭ Sbornik* **8**(50), 251–264. [In Russian.] (Cited on p. 9.)

MALCEV, A. I. (1953)
'Нилпотентные полугруппы' ('Nilpotent semigroups'), *Ivanov. Gos. Ped. Inst. Učen. Zap. Fiz.-Mat. Nauki* **4**, 107–111. [In Russian.] (Cited on pp. 81, 84.)

MARKOV, A. A. (1971/72)
'On finitely generated subsemigroups of a free semigroup', *Semigroup Forum* **3**(3), 251–258. (Cited on p. 79.)

MCCAMMOND, J. P. & WISE, D. T. (2005)
'Coherence, local quasiconvexity, and the perimeter of 2-complexes', *Geometric and Functional Analysis* **15**(4). To appear. (Cited on pp. 24, 131.)

MILLER, III, C. F. (2002)
'Subgroups of direct products with a free group', *The Quarterly Journal of Mathematics* **53**(4), 503–506. (Cited on pp. 103, 107.)

MULLER, D. E. & SCHUPP, P. E. (1983)
'Groups, the theory of ends, and context-free languages', *Journal of Computer and System Sciences* **26**(3), 295–310. (Cited on pp. 62, 63, 64.)

NEUMANN, B. H. (1964)
'"Subsemigroups of nilpotent groups": An acknowledgement', *Proceedings of the Royal Society. Series A* **281**, 436. (Cited on p. 81.)

NEUMANN, B. H. & TAYLOR, T. (1963)
'Subsemigroups of nilpotent groups', *Proceedings of the Royal Society. Series A* **274**, 1–4. (Cited on pp. 81, 84.)

OLSHANSKII, A. Y. & STOROZHEV, A. (1996)
'A group variety defined by a semigroup law', *Australian Mathematical Society Journal. Series A — Pure Mathematics and Statistics* **60**(2), 255–259. (Cited on p. 85.)

ORE, Ø. (1931)
'Linear equations in non-commutative fields', *Annals of Mathematics* **32**, 463–477. (Cited on pp. 9, 10.)

OTTO, F. & RUŠKUC, N. (2000)
[Personal communication/Unpublished] (Cited on p. 47.)

OTTO, F. & SATTLER-KLEIN, A. (1997)
'Some remarks on finitely presented monoids with automatic structure', Mathematische Schriften Kassel 9/97, Universität-GH-Kassel. (Cited on p. 38.)

PELLETIER, M. & SAKAROVITCH, J. (1999)
'On the representation of finite deterministic 2-tape automata', *Theoretical Computer Science* **225**(1–2), 1–63. (Cited on p. 153.)

RABIN, M. O. & SCOTT, D. S. (1959)
'Finite automata and their decision problems', *IBM Journal of Research and Development* **3**, 114–125. (Cited on p. 153.)

RÉDEI, L. (1963)
Theorie der Endlich Erzeugbaren Kommutativen Halbgruppen (*Theory of Finitely Generated Commutative Semigroups*), Vol. 41 of *Hamburger Mathematische Einzelschriften*, Physica-Verlag, Würzburg. [In German. See Rédei (1965) for a translation.] (Cited on p. 85.)

RÉDEI, L. (1965)
The Theory of Finitely Generated Commutative Semigroups, Pergamon Press, Oxford. [Translated from the German. Edited by N. Reilly.] (Cited on p. 169.)

REES, D. (1948)
'On the group of a set of partial transformations', *Journal of the London Mathematical Society* **22**, 281–284. (Cited on p. 10.)

REMMERS, J. H. (1980)
'On the geometry of semigroup presentations', *Advances in Mathematics* **36**(3), 283–296. (Cited on p. 10.)

RILEY, J. A. (1967)
'The Sardinas–Patterson and Levenshtein theorems', *Information and Control* **10**, 120–136. (Cited on p. 63.)

RIPS, E. (1982)
'Subgroups of small cancellation groups', *Bulletin of the London Mathematical Society* **14**(1), 45–47. (Cited on p. 24.)

ROBINSON, D. J. S. (1996)
A Course in the Theory of Groups, Vol. 80 of *Graduate Texts in Mathematics*, Second edition, Springer-Verlag, New York. (Cited on pp. 11, 12, 83.)

ROSALES, J. C. & GARCÍA-SÁNCHEZ, P. A. (1999)
Finitely Generated Commutative Monoids, Nova Science Publishers Inc., Commack, N.Y. (Cited on p. 86.)

ROSENBLATT, J. M. (1974)
'Invariant measures and growth conditions', *Transaction of the American Mathematical Society* **193**, 33–53. (Cited on p. 107.)

RUŠKUC, N. (1995)
Semigroup Presentations, Ph.D. Thesis, University of St Andrews. (Cited on p. 3.)

RUŠKUC, N. (1998)
'On large subsemigroups and finiteness conditions of semigroups', *Proceedings of the London Mathematical Society. Third Series* **76**(2), 383–405. (Cited on pp. 133, 141.)

RUŠKUC, N. & THOMAS, R. M. (1998)
'Syntactic and Rees indices of subsemigroups', *Journal of Algebra* **205**(2), 435–450. (Cited on pp. 132, 135, 136.)

SAKAROVITCH, J. (1987)
'Easy multiplications I. The realm of Kleene's theorem', *Information and Computation* **74**(3), 173–197. (Cited on p. 28.)

SARDINAS, A. A. & PATTERSON, G. W. (1953)
'A necessary and sufficient condition for the unique decomposition of coded messages', in *Convention Records of the Institute of Radio Engineers, 1953 National Convention* (Part 8: Information Theory), pp. 104–108. (Cited on p. 63.)

SCOTT, G. P. (1973)
'Finitely generated 3-manifold groups are finitely presented', *Journal of the London Mathematical Society. Second Series* **6**, 437–440. (Cited on p. 24.)

SERRE, J.-P. (1974)
'Problem section', in J. Cossey, editor, *Proceedings of the Second International Conference on the Theory of Groups (Australian National University, Canberra, 1973)*, Vol. 372 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 733–740. (Cited on p. 24.)

ŠEVRIN, L. N. (1960*a*)
'On subsemigroups of free semigroups', *Doklady Akademii Nauk SSSR* **133**, 537–539. [In Russian. See Ševrin (1960*b*) for a tranlation.] (Cited on p. 64.)

ŠEVRIN, L. N. (1960*b*)
'On subsemigroups of free semigroups', *Soviet Mathematics Doklady* **1**, 892–894. [Translated from the Russian by K. A. Hirsch.] (Cited on p. 170.)

SHAPIRO, M. (1992)
'Deterministic and nondeterministic asynchronous automatic structures', *International Journal of Algebra and Computation* **2**(3), 297–305. (Cited on pp. 30, 33, 34, 36, 153.)

SILVA, P. V. & STEINBERG, B. (2004)
'A geometric characterization of automatic monoids', *The Quarterly Journal of Mathematics* **55**(3), 333–356. (Cited on p. 52.)

SIMS, C. C. (1994)
Computation with Finitely Presented Groups, Vol. 48 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge. (Cited on p. 106.)

SPEHNER, J.-C. (1974/75)
'Quelques constructions et algorithmes relatifs aux sous-monoïdes d'un monoïde

libre' ('Some constructions and relative algorithms for submonoids of a free monoid'), *Semigroup Forum* **9**(4), 334–353. [In French.] (Cited on pp. 21, 64.)

SPEHNER, J.-C. (1977)
'Présentations et présentations simplifiables d'un monoïde simplifiable' ('Presentations and cancellative presentations of a cancellative monoid'), *Semigroup Forum* **14**(4), 295–329. [In French.] (Cited on pp. xii, 13, 18, 21, 24, 25.)

SPEHNER, J.-C. (1981)
'Les présentations des sous-monoïdes de rang 3 d'un monoïde libre' ('Presentations of submonoids of rank 3 of a free monoid'), in *Semigroups* (Oberwolfach, 1978), Vol. 855 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 116–155. [In French.] (Cited on p. 79.)

SPEHNER, J.-C. (1989)
'Every finitely generated submonoid of a free monoid has a finite Malcev's presentation', *Journal of Pure and Applied Algebra* **58**(3), 279–287. (Cited on pp. 13, 18, 24, 25, 66, 71, 79.)

STALLINGS, J. R. (1987)
'A graph-theoretic lemma and group-embeddings', in *Combinatorial Group Theory and Topology* (Alta, Utah, 1984), Vol. 111 of *Annals of Mathematics Studies*, Princeton University Press, Princeton, N.J., pp. 145–155. (Cited on p. 10.)

SZELEPCSÉNYI, R. (1988)
'The method of forced enumeration for nondeterministic automata', *Acta Informatica* **26**(3), 279–284. (Cited on p. 158.)

INDEX

[I] realized I had never used the index of a book fit to read. Who would insult his Decline and Fall, *by consulting it just upon a specific point?*
— T. E. Lawrence,
Seven Pillars of Wisdom (1926)

WHEN A REFERENCE in this index is italicized, the relevant page or range of pages contains the original definition or statement of the concept or result indexed. The ordering of entries is strictly lexicographic, ignoring punctuation and spacing. Symbols outside the Latin alphabet are collected at the start of the index, even if an 'auxiliary' Latin symbol is used: thus $\overline{w}$ is included in this set, since it is the notation $\overline{}$ that is being defined. Brief definitions are given for notation.

This thesis was typeset using LaTeX 2_ε running on TeX 3.14159, using a customized document class file. The $\mathcal{AMS}$-LaTeX package amssymb was used, together with modifed versions of amsmath and amsthm. The body of the thesis is set in Knuth's Computer Modern; some words are set in fonts by the IvriTeX group (Hebrew) and by Yannis Haralambous (Greek). The index was compiled using a custom-written Perl script. The bibliography and references were created using extensions of the BibTeX style agsm and the package harvard. Figures and commutative diagrams were drawn using METAPOST.

The crashing of the sea subsided altogether.

— Giuseppe Tomasi di Lampedusa,
The Leopard (1958), ch. vii
(trans. A. Colquhoun)