Algorithmic Problems for Subsemigroups of Infinite Groups

Ruiwen Dong

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Abstract

This thesis is concerned with several algorithmic problems for subsemigroups of infinite groups. The main objective is to construct algorithms that decide various properties of finitely generated subsemigroups of a given infinite group G (for example, a matrix group). Such problems might not be decidable in general. In fact, they gave rise to some of the earliest undecidability results in algorithmic theory. Nevertheless, when the group G admits additional structures, many algorithmic problems become decidable for its subsemigroups. In this thesis, we study the decidability and the complexity of these algorithmic problems in the cases where G is nilpotent, metabelian, or represented as a low-dimensional matrix group.

Two of the main problems we consider are the *Identity Problem* and the *Group Problem*. Given a finite subset $\mathcal{G}$ of a group G, the Identity Problem asks whether the semigroup $\langle \mathcal{G} \rangle$ generated by $\mathcal{G}$ contains the neutral element, and the Group Problem asks whether this semigroup is a group. We show that both problems are decidable in finitely generated nilpotent groups of class at most ten, and in PTIME if the input is given as unitriangular matrices. We also show the decidability of these problems in finitely generated metabelian groups, as well as their NP-completeness in the special affine group $\mathsf{SA}(2, \mathbb{Z})$.

Apart from the Identity Problem and the Group Problem, we also consider *Semigroup Intersection* and *Orbit Intersection*. Given two finite subsets $\mathcal{G}$ and $\mathcal{H}$ of the group G, Semigroup Intersection asks whether the semigroups $\langle \mathcal{G} \rangle, \langle \mathcal{H} \rangle$ generated by $\mathcal{G}$ and $\mathcal{H}$ have empty intersection, while Orbit Intersection asks whether the sets $S \cdot \langle \mathcal{G} \rangle$ and $T \cdot \langle \mathcal{H} \rangle$ intersect for given elements $S, T \in G$. We show that Semigroup Intersection is decidable in class two nilpotent groups (PTIME if the input is given as unitriangular matrices), and that Orbit Intersection is decidable in the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$.

We apply a large variety of mathematical tools in the study of these problems, ranging from Lie algebra, algebraic geometry and number theory, to combinatorics, graph theory and convex geometry. By adding these techniques into the toolbox, we are able to significantly advance the current state of art in the algorithmic theory of semigroups.

Acknowledgements

I would like to express my gratitude to Christoph Haase and James Worrell for guiding me through this thesis and introducing me to the academic world. Thanks to Joël Ouaknine and Georg Zetzsche for inviting me to visit the Max Planck Institute for Software Systems in Saarbrücken and Kaiserslautern. Markus Schweighofer and David Sawall introduced me to the fascinating area of real algebra. We had inspiring discussions both online and in Konstanz, which motivated some of the work in Chapter 4.

Finally, I would like to thank my parents for their constant support, as well as my friends in Oxford and Paris for the amazing time I spent there.

This thesis was partly supported by the UKRI Frontier Research Grant EP/X033813/1.

Contents

Chapter 1

Introduction

1.1 Algorithmic problems in groups and semigroups

It has been known since the work of Church and Turing from the 1930s that certain decision problems in mathematical logic do not admit algorithmic solutions. For a long period of time, all examples of undecidable problems were derived directly from mathematical logic or the theory of computing, and the notion of undecidability seemed intangible for most mathematicians. It was not until the late 1940s that the Soviet mathematician Andrey Markov produced a concrete undecidable problem using linear algebra. In his seminal work *"On certain insoluble problems concerning matrices"* published in 1948, Markov studied the following decision problem. Its input is a finite set of square matrices $\mathcal{G} = \{A_1, \ldots, A_n\}$ and a matrix T, and the problem is whether or not there exist an integer $p \geq 1$ and a sequence $A_{i_1}, \cdots, A_{i_p}$ of matrices in $\mathcal{G}$ such that $T = A_{i_1} A_{i_2} \cdots A_{i_p}$. Markov showed this problem to be undecidable for integer matrices of dimension at least six, thus marking the first undecidability result obtained outside of mathematical logic and the theory of computing.

Markov's work falls into the area of computational group theory, which is one of the oldest and most well-developed parts of computational algebra. The "official" start of computational group theory dates back to 1911, when Max Dehn formulated three basic problems that would become its foundation. Given a finite presentation of a group G, it is asked whether there are algorithms that solve the *Word Problem* (whether an element is the neutral element), the *Conjugacy Problem* (whether two elements are conjugate in G), and the *Isomorphism Problem* (whether G is isomorphic to another finitely presented group). It was not until the 1950s that the three problem were shown to be undecidable in general groups [1, 82].

Using the language of computational group theory, Markov's problem can be reformulated as deciding *Semigroup Membership* for a matrix (semi)group. Given a finite subset $\mathcal{G}$ of a group

G, denote by $\langle \mathcal{G} \rangle$ the subsemigroup generated by $\mathcal{G}$. Then the Semigroup Membership problem can be formulated as follows.

(i) *(Semigroup Membership)* given a finite set $\mathcal{G}$ and an element T in G, decide whether $T \in \langle \mathcal{G} \rangle$.

Markov's undecidability result as well as the subsequent undecidability results for Dehn's problems generated a surge of research interest in computational group theory. In the 1960s, Mikhailova [74] introduced the *group* version of Semigroup Membership. Given a finite subset $\mathcal{G}$ of a group G, denote by $\langle \mathcal{G} \rangle_{grp}$ the subgroup generated by $\mathcal{G}$.

(ii) *(Group Membership)* given a finite set $\mathcal{G}$ and an element T in G, decide whether $T \in \langle \mathcal{G} \rangle_{grp}$.

Mikhailova [74] showed undecidability of Group Membership when G is the group $\mathsf{SL}(4, \mathbb{Z})$ of 4×4 integer matrices with determinant one. One may note that undecidability of Group Membership subsumes that of Semigroup Membership by including the inverses of the elements in $\mathcal{G}$.

There has been a steady growth in research intensity for Group and Semigroup Membership problems as they establish important connections between algebra and logic. These problems now play an essential role in analysing system dynamics and program termination, and have numerous applications in automata theory, complexity theory, and interactive proof systems [13, 21, 33, 49]. It is worth noting that membership problems are in fact decidable for many classes of groups, such as abelian groups and low dimensional matrix groups [4, 29]. For example, in the matrix group $\mathsf{SL}(2, \mathbb{Z})$, Semigroup Membership is decidable by a classic result of Choffrut [29], and Group Membership is decidable in polynomial time (PTIME) by a recent result of Lohrey [66]. Our interest in computational group theory is two-fold. From an application point of view, we are interested in developing practical algorithms for specific classes of groups. From a theory point of view, we aim to close the gap between decidability and undecidability.

For most classes of groups, Group Membership is much more tractable than Semigroup Membership. For example, Group Membership is decidable in the class of *polycyclic groups* by a classic result of Kopytov [61]; whereas Semigroup Membership is undecidable even in the subclass of *nilpotent groups* [86]. This gap motivated the introduction of two intermediate problems by Choffrut and Karhumäki [29] in 2005:

(iii) *(Identity Problem)* given a finite set $\mathcal{G}$, decide whether $\langle \mathcal{G} \rangle$ contains a neutral element.

(iv) *(Group Problem)* given a finite set $\mathcal{G}$, decide whether $\langle \mathcal{G} \rangle = \langle \mathcal{G} \rangle_{grp}$.

In other words, the Identity Problem asks whether the semigroup $\langle \mathcal{G} \rangle$ is a monoid, and the Group Problem asks whether the semigroup $\langle \mathcal{G} \rangle$ is a group.

The Group Problem is crucial in determining structural properties of a semigroup. For example, given a decision procedure for the Group Problem, one can compute a generating set for the *group of units* of a finitely generated semigroup $\langle \mathcal{G} \rangle$ (see Section 2.2). We also point out that there are significantly more available algorithms for groups than there are for semigroups. Therefore performing preliminary checks using the Group Problem can help decide Semigroup Membership in many special cases. Using the Group Problem, one can also decide the lesser known *Inverse Problem*: given a finite set $\mathcal{G}$ and an element $a \in \mathcal{G}$, decide whether $a^{-1} \in \langle \mathcal{G} \rangle$. As for the Identity Problem, its solution is usually the most essential special case on the way to building an algorithm for Semigroup Membership. The Identity Problem and the Group Problem motivated the development of numerous tools in the study of semigroups, ranging from automata theory to Lie algebra. Both problems are shown to be undecidable in $\mathsf{SL}(4, \mathbb{Z})$ by Bell and Potapov [16] using an embedding of the *Identity Corresponding Problem*. Whereas for $\mathsf{SL}(2, \mathbb{Z})$, they are shown to be NP-complete using techniques in compressed words [14].

Heuristically, the complexity of these intermediate problems tends to lie between Group Membership and Semigroup Membership. For example, in abelian matrix groups, Babai et al. famously reduced algorithmic problems to computation on lattices [4], thus Group Membership reduces to linear algebra over $\mathbb{Z}$, and is decidable in PTIME; Semigroup Membership is equivalent to integer programming, and is hence NP-complete; the Identity Problem and the Group Problem reduce to solving *homogeneous* linear Diophantine equations, placing them in PTIME as well. Unfortunately, decidability of these intermediate problems remain open for larger classes of groups, even in cases where decidability of Group and Semigroup Membership problems already have definitive answers. Notable examples include nilpotent groups, polycyclic groups, metabelian groups and low-dimensional matrix groups.

Beyond the Identity Problem and the Group Problem, we are also interested in another classic problem raised by Markov:

(v) *(Semigroup Intersection)* given two finite sets $\mathcal{G}$ and $\mathcal{H}$, decide whether $\langle \mathcal{G} \rangle \cap \langle \mathcal{H} \rangle = \emptyset$.

In the seminal paper where Markov demonstrated undecidabilty of Semigroup Membership, he also showed undecidability of Semigroup Intersection for integer matrix groups of dimension at least four [72]. Markov's idea is to encode the famous *Post Correspondence Problem*, which can be reformuated as Semigroup Intersection in a direct product of two free monoids.

The final problem we are interested in is a simultaneous generalization of Semigroup Membership and Semigroup Intersection:

(vi) *(Orbit Intersection)* given two finite sets $\mathcal{G}$ and $\mathcal{H}$ as well as two elements S, T in G, decide whether $T \cdot \langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$.

Orbit Intersection was first studied by Babai et al. in the context of abelian matrix groups [4]. By taking T to be the neutral element and $\mathcal{H}$ to be the set containing only the neutral element, Orbit Intersection subsumes Semigroup Membership. By taking both T and S to be the neutral element, Orbit Intersection subsumes Semigroup Intersection.

1.2 Contributions of this thesis

Throughout this thesis we develop various tools to study algorithmic problems for subsemigroups of infinite groups. These tools range from areas in pure mathematics such as Lie algebra, algebraic geometry and number theory, to areas with a more computer science flavour, such as linear programming, graph theory and complexity theory. By adding these deep mathematical tools to the arsenal for algorithmic problems, we are able to tackle several challenging problems. Notably, our main results include:

(1) The Identity Problem and the Group Problem are decidable in nilpotent groups of class at most ten (Corollary 3.1.2). Moreover, if the group is given as an embedding in the group $\mathsf{UT}(n, \mathbb{Q})$ of unitriangular matrices, then both problems are decidable in PTIME (Theorem 3.1.1).

(2) Semigroup Intersection is decidable in nilpotent groups of class two (Corollary 3.1.4). Moreover, if the group is given as an embedding in $\mathsf{UT}(n, \mathbb{Q})$, then it is decidable in PTIME (Theorem 3.1.3).

(3) Orbit Intersection is decidable in the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$ (Theorem 3.1.5).

(4) The Identity Problem and the Group Problem are decidable in finitely generated metabelian groups (Theorem 4.1.2).

(5) The Identity Problem and the Group Problem are decidable and NP-complete in the special affine group $\mathsf{SA}(2, \mathbb{Z})$ (Theorem 5.1.1).

For the results in nilpotent groups, our main tool will be the *Baker-Campbell-Hausdorff formula*. We will reduce semigroup problems to convex geometry on Lie algebra, and use assistance from computer algebra software to prove several technical theorems. For the result in metabelian groups, we employ graph theory techniques to establish a connection between semigroups and polynomial semirings. This connection allows us to solve algorithmic problems for semigroups using algebraic geometry, whose theory is much more mature than the theory of semigroups. For the result in the special affine group, we adopt two different viewpoints on the group $\mathsf{SA}(2, \mathbb{Z})$: one as a transformation group of the lattice $\mathbb{Z}^2$, the other as an extension of the virtually free group $\mathsf{SL}(2, \mathbb{Z})$. This allows us to combine geometric arguments and algebraic arguments to obtain decidability and a complexity upper bound.

A common feature of all our work is the rich interaction between different areas of mathematics and computer science. This demonstrates the power of the interdisciplinary tools we introduce, and paints a promising picture for future progress in the algorithmic theory of semigroups. The following table summarizes some of our results in the context of the current state of art.

Grp. type	Semigrp Mem.	Group Mem.	Identity & Grp. Prb.	Semigrp Intersection
Abelian	NP-complete [4]	PTIME [4]	PTIME [4]	PTIME [4]
Nilpotent	undec. [86]	decidable [61]	?/PTIME for class 10	?/PTIME for class 2
Metabelian	undec. [86]	decidable [87]	decidable	undec.*
$\mathsf{SL}(2,\mathbb{Z})$	NP-complete†	PTIME [66]	NP-complete [14]	NP-complete†
$\mathsf{SA}(2,\mathbb{Z})$	?	decidable [32]	NP-complete	?
$\mathsf{SL}(3,\mathbb{Z})$	?	?	?	?
$\mathsf{SL}(4,\mathbb{Z})$	undec. [74]	undec. [74]	undec. [16]	undec. [16]

Table 1.1: Purple entries = our results. ? = open problem.

*see Section 4.1. † Semigroup Membership and Semigroup Intersection are special cases of the *Rational Subset Membership* problem, which was recently shown to be decidable in NP [15] in $\mathsf{SL}(2,\mathbb{Z})$. Their NP-hardness is subsumed by the Identity Problem.

1.3 Notes for the reader

This thesis is organized as follows. Chapter 2 introduces the preliminaries on group and semigroup theory. We discuss reductions between different semigroup algorithmic problems and their connection to language theory. Chapter 3, 4 and 5 focus on algorithmic problems in specific group classes: nilpotent groups, metabelian groups and the special affine group $\mathsf{SA}(2,\mathbb{Z})$. Chapter 6 discusses possible extensions of our work and poses several open problems that may be interesting for future research.

In order to appeal to a broader audience, we will not assume familiarity with group and semigroup theory beyond the basic notions (normal subgroup, quotient group, generators of a group or semigroup). However, since we will always try to represent abstract groups in a way that is conventional in computational group theory, we will need to employ certain established results from this area. Our strategy is always to first embed an abstract group in a concrete matrix group, and then solve algorithmic problems over the concrete matrix group.

For example, nilpotent groups (Chapter 3) are usually represented by *finite presentations*. However, in Section 3.2 we show that they can be effectively embedded in the unitriangular

matrix group $\mathsf{UT}(n, \mathbb{Q})$ up to a quotient by the torsion subgroup, and we proceed to solve algorithmic problems under this matrix representation for the rest of Chapter 3. Metabelian groups (Chapter 4) are usually represented by *finite metabelian presentations* (or finite presentations in the class of metabelian groups). In Section 4.2 we show that algorithmic problems in metabelian groups reduce to problems in 2×2 matrix groups over finitely presented modules. We then proceed using this matrix representation for the rest of Chapter 4.

The embedding results in Section 3.2 and 4.2 are generally corollaries of deep theorems from group theory. Therefore, readers may choose to consider them as black-boxes, and focus solely on the parts dealing with concrete matrix groups.

Chapter 2

Group and semigroup theory

2.1 Group theory

In this subsection we introduce the necessary concepts in group theory. Inspired by the notation in matrix groups, we denote by I the neutral element of a group.

By an *alphabet*, we mean a set $\mathcal{A}$. Most alphabets considered in this thesis will be finite alphabets. Elements of an alphabet are called *letters*. A *word* over an alphabet $\mathcal{A}$ is a finite string of letters, possibly empty. In particular, the empty string is called the *empty word*. Given any alphabet $\mathcal{A}$, we denote by $\mathcal{A}^*$ the set of words over $\mathcal{A}$:

$$\mathcal{A}^* \coloneqq \{a_1 a_2 \cdots a_m \mid m \geq 0, a_1, \ldots a_m \in \mathcal{A}\}\,.$$

For two words $v, w \in \mathcal{A}^*$, we write vw for the concatenation of v and w, it is again a word over $\mathcal{A}$.

2.1.1 Free groups and finite presentation of groups

Given a (possibly infinite) alphabet Σ, define the corresponding group alphabet $\Sigma^{\pm} \coloneqq \Sigma \cup \{a^{-1} \mid a \in \Sigma\}$, where a^{-1} is a new letter for each $a \in \Sigma$. There is a natural involution $(\cdot)^{-1}$ over $(\Sigma^{\pm})^*$ defined by $(a^{-1})^{-1} = a$ and $(a_1 a_2 \cdots a_m)^{-1} = a_m^{-1} \cdots a_2^{-1} a_1^{-1}$. A word over the alphabet $\Sigma^{\pm}$ is called *reduced* if it does not contain consecutive letters aa^{-1} or $a^{-1}a$. For a word w, define $\mathrm{red}(w)$ to be the reduced word obtained by iteratively replacing consecutive letters aa^{-1} and $a^{-1}a$ with the empty string. The *free group* $F(\Sigma)$ over Σ is then defined as the set of reduced words over the alphabet $\Sigma^{\pm}$, where multiplication is given by $v \cdot w = \mathrm{red}(vw)$, and inversion is given by the involution $(\cdot)^{-1}$. A group is called *free* if it is a free group over some alphabet. The free group $F(\Sigma)$ is abelian if and only if the cardinality of Σ is zero or one: in this case, the

group $F(\Sigma)$ is trivial (when $\Sigma = \emptyset$) or isomorphic to the infinite cyclic group $\mathbb{Z}$ (when $\Sigma = \{a\}$).

In computational group theory, it is customary to represent groups using *finite presentations*:

Definition 2.1.1 (Finite presentation of a group)**.** Let $\Sigma = \{g_1, \ldots, g_n\}$ be a finite alphabet and $r_1, \ldots, r_m$ be elements of the free group $F(\Sigma)$ represented as words over the alphabet $\Sigma^{\pm}$. We say that $\langle g_1, \ldots, g_n \mid r_1, \ldots, r_m \rangle$ is a *finite presentation* of G, if

$$G \cong F(\Sigma) / \operatorname{ncl}_{F(\Sigma)}(r_1, \ldots, r_m).$$

Here, $\operatorname{ncl}_{F(\Sigma)}(r_1, \ldots, r_m)$ denotes the *normal closure*[1] of $\{r_1, \ldots, r_m\}$; that is, the smallest normal subgroup of $F(\Sigma)$ containing $\{r_1, \ldots, r_m\}$. In this case, we often write $G = \langle g_1, \ldots, g_n \mid r_1, \ldots, r_m \rangle$.

Intuitively, G is the group generated by the elements $g_1, \ldots, g_n$, subject to the relations $r_1 = I, r_2 = I, \ldots, r_m = I$. Here are a few examples.

Example 2.1.2. (1) The abelian group $\mathbb{Z}^2$ has the finite presentation

$$\mathbb{Z}^2 = \langle x, y \mid xyx^{-1}y^{-1} \rangle.$$

In particular, $\mathbb{Z}^2$ is the group generated by the two elements x, y. The relation $xyx^{-1}y^{-1} = I$ yields $xy = yx$, meaning x and y commute.

(2) The symmetry group S_3, consisting of permutations of the set $\{1, 2, 3\}$, has the finite presentation

$$S_3 = \langle a, b \mid a^3, b^2, aba^{-2}b \rangle.$$

Here, a represents the cyclic permutation $1 \mapsto 2 \mapsto 3 \mapsto 1$, and b represents the cyclic permutation $1 \mapsto 2 \mapsto 1$. See [51, Section 5.3].

(3) The group $\mathsf{SL}(2, \mathbb{Z})$ of 2×2 integer matrices of determinant one admits the finite presentation

$$\mathsf{SL}(2, \mathbb{Z}) = \langle S, T \mid S^4, (ST)^3 S^{-2} \rangle.$$

Here, the element S represents the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and T represents the matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, see [90, p.81]. Indeed, this presentation is the key to solving many algorithmic problems in $\mathsf{SL}(2, \mathbb{Z})$ [14, 66].

(4) The *integer Heisenberg group* $\mathrm{H}_3(\mathbb{Z})$, consisting of 3×3 upper-triangular integer matrices

[1] In general, $\operatorname{ncl}_{F(\Sigma)}(r_1, \ldots, r_m)$ is *not* the same as the subgroup $\langle r_1, \ldots, r_m \rangle_{grp}$ generated by $r_1, \ldots, r_m$.

with ones on the diagonal, has the finite presentation

$$\mathrm{H}_3(\mathbb{Z}) = \langle x, y, z \mid xyx^{-1}y^{-1}z^{-1}, xzx^{-1}z^{-1}, yzy^{-1}z^{-1} \rangle.$$

Here, the element x represents the matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, the element y represents the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, and z represents the matrix $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. See [34, Example 7.32(7)].

We point out that the finite presentation of a group may not be unique, and a finitely generated group need not admit a finite presentation [6, Theorem 1].

2.1.2 Nilpotent groups

Given a group G and its subgroup H, define the commutator $[G, H]$ to be the group generated by the elements in $\{ghg^{-1}h^{-1} \mid g \in G, h \in H\}$. That is,

$$[G, H] \coloneqq \left\langle \{ghg^{-1}h^{-1} \mid g \in G, h \in H\} \right\rangle_{grp}$$

Definition 2.1.3 (Nilpotent groups)**.** The *lower central series* of a group G is the inductively defined descending sequence of subgroups

$$G = G_1 \geq G_2 \geq G_3 \geq \cdots,$$

in which $G_k = [G, G_{k-1}]$. A group G is called *nilpotent* if its lower central series terminates with $G_{d+1} = \{I\}$ for some d. In this case, the smallest such d is called the *nilpotency class* of G.

In particular, abelian groups are nilpotent of class one, since the commutator $[G, G]$ of an abelian group G is trivial. It is well known that subgroups of nilpotent groups are nilpotent [8], and every finitely generated nilpotent group admits a finite presentation [34, Proposition 13.84]. One of the most important examples of nilpotent groups is the group of unitriangular matrices:

Definition 2.1.4 (Unitriangular matrix groups)**.** Denote by $\mathsf{UT}(n, \mathbb{Q})$ the group of $n \times n$ upper triangular rational matrices with ones on the diagonal:

$$\mathsf{UT}(n, \mathbb{Q}) \coloneqq \left\{ \begin{pmatrix} 1 & * & \cdots & * & * \\ 0 & 1 & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & * \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \text{ where } * \text{ are entries in } \mathbb{Q} \right\}.$$

Then $\mathsf{UT}(n, \mathbb{Q})$ forms a nilpotent group under matrix multiplication.

In particular, the group $\mathsf{UT}(n, \mathbb{Q})$ is of nilpotency class $n - 1$ [34, Examples 13.36]. Note that for $n \geq 2$, the group $\mathsf{UT}(n, \mathbb{Q})$ is *not* finitely generated. Nevertheless, the group $\mathsf{UT}(n, \mathbb{Q})$ plays an important role in the study of finitely generated nilpotent groups by the following fact.

Fact 2.1.5 ([8, Theorem 2.1], [41, Algorithm E])**.** Every finitely generated nilpotent group G is isomorphic to a subgroup of a direct product $\mathsf{UT}(n, \mathbb{Q}) \times F$, where $n \in \mathbb{N}$ and F is a finite group.

The dimension n in Fact 2.1.5 is not directly related to the nilpotency class of G. For example, the direct product $\mathsf{UT}(3, \mathbb{Q}) \times \mathsf{UT}(3, \mathbb{Q})$ naturally embeds into the group $\mathsf{UT}(6, \mathbb{Q})$, but it is only of nilpotency class two.

2.1.3 Metabelian and solvable groups

Definition 2.1.6 (Metabelian groups)**.** A group G is called *metabelian* if the commutator group $[G, G]$ is abelian.

Equivalently, a group G is metabelian if and only if there is an abelian normal subgroup A such that the quotient group G/A is abelian. Here are a few common examples of metabelian groups.

Example 2.1.7. The following groups are metabelian [18, 34]:

(1) all finite groups of order at most 23,

(2) all nilpotent groups of class at most three,

(3) the infinite dihedral group

$$D_\infty = \langle t, r \mid r^2, rtrt \rangle;$$

this is the symmetry group of the line $\mathbb{Z}$, where r represents the reflection and t represents a translation by one;

(4) the group $\mathsf{T}(2, \mathbb{K})$ of 2×2 invertible upper-triangular matrices over any field $\mathbb{K}$:

$$\mathsf{T}(2, \mathbb{K}) \coloneqq \left\{ \begin{pmatrix} x & z \\ 0 & y \end{pmatrix} \;\middle|\; x, y \in \mathbb{K} \setminus \{0\}, z \in \mathbb{K} \right\},$$

(5) all subgroups, quotients and direct products of metabelian groups.

One of the most important examples of finitely generated metabelian groups is the *wreath product* $\mathbb{Z} \wr \mathbb{Z}^n$. This group has various equivalent definitions, we state here an intuitive representation as a matrix group. Denote by $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$ the ring of Laurent polynomials over n

variables with integer coefficients. That is, elements of $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$ are polynomials of the form

$$\sum_{a_1,\ldots,a_n \in \mathbb{Z}} p_{a_1,\ldots,a_n} X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n},$$

where only finitely many coefficients $p_{a_1,\ldots,a_n} \in \mathbb{Z}$ are non-zero.

Definition 2.1.8 (The wreath product $\mathbb{Z} \wr \mathbb{Z}^n$)**.** The wreath product $\mathbb{Z} \wr \mathbb{Z}^n$ is defined as the following matrix group over $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$:

$$\mathbb{Z} \wr \mathbb{Z}^n \coloneqq \left\{ \begin{pmatrix} X_1^{a_1} \cdots X_n^{a_n} & y \\ 0 & 1 \end{pmatrix} \;\middle|\; a_1, \ldots, a_n \in \mathbb{Z}, y \in \mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}] \right\}.$$

The wreath product $\mathbb{Z} \wr \mathbb{Z}^n$ is important among metabelian groups because every finitely generated metabelian group is isomorphic to a quotient G/N, where $N \trianglelefteq G$ are two subgroups of $\mathbb{Z} \wr \mathbb{Z}^n$ for some $n \in \mathbb{N}$ (see Chapter 4). We point out that unlike nilpotent groups, a finitely generated metabelian group does not necessarily admit a finite presentation (in fact, $\mathbb{Z} \wr \mathbb{Z}$ does not admit a finite presentation [6]). Therefore, the convention in representing a metabelian group is to use a *finite metabelian presentation* (see Section 4.2).

Metabelian groups are included in the larger class of groups known as *solvable groups*:

Definition 2.1.9 (Solvable groups)**.** The *derived series* of a group G is the inductively defined descending sequence of subgroups

$$G = G^{(0)} \geq G^{(1)} \geq G^{(2)} \geq \cdots,$$

in which $G^{(k)} = [G^{(k-1)}, G^{(k-1)}]$. A group G is called *solvable* if its derived series terminates with $G^{(d)} = \{I\}$ for some d. In this case, the smallest such d is called the *derived length* of G.

In particular, metabelian groups are exactly those solvable groups of derived length at most two.

2.1.4 Special linear groups and special affine groups

Definition 2.1.10 (Special linear groups)**.** Let $n \in \mathbb{N}$. The *special linear group* $\mathsf{SL}(n, \mathbb{Z})$ is the group of $n \times n$ integer matrices with determinant one:

$$\mathsf{SL}(n, \mathbb{Z}) \coloneqq \{A \in \mathbb{Z}^{n \times n} \mid \det(A) = 1\}.$$

Elements of $\mathsf{SL}(n, \mathbb{Z})$ can be seen as linear transformations $\boldsymbol{x} \mapsto A\boldsymbol{x}$ of the lattice $\mathbb{Z}^n$ that preserves orientation. The affine version of $\mathsf{SL}(n, \mathbb{Z})$, called the *special affine group*, is defined

as follows.

Definition 2.1.11 (Special affine groups)**.** Let $n \in \mathbb{N}$. The *special affine group* $\mathsf{SA}(n, \mathbb{Z})$ is the group of block diagonal matrices of the following form:

$$\mathsf{SA}(n, \mathbb{Z}) \coloneqq \left\{ \begin{pmatrix} A & \boldsymbol{a} \\ 0 & 1 \end{pmatrix} \;\middle|\; A \in \mathsf{SL}(n, \mathbb{Z}), \boldsymbol{a} \in \mathbb{Z}^n \right\}.$$

In particular, $\mathsf{SA}(n, \mathbb{Z})$ contains $\mathsf{SL}(n, \mathbb{Z})$ as a subgroup, and is itself contained as a subgroup in $\mathsf{SL}(n+1, \mathbb{Z})$. Elements of $\mathsf{SA}(n, \mathbb{Z})$ can be seen as *affine* transformations $\boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{a}$ of the lattice $\mathbb{Z}^n$ that preserves orientation.

The group $\mathsf{SA}(n, \mathbb{Z})$ can also be described as the *semidirect product* $\mathbb{Z}^n \rtimes \mathsf{SL}(n, \mathbb{Z})$. In general, given two groups G, H and an action[2] φ of G on H, one can define the *semidirect product* $H \rtimes_\varphi G$:

Definition 2.1.12 (Semidirect product)**.** The semidirect product $H \rtimes_\varphi G$ is the group consisting of pairs (h, g) where $h \in H, g \in G$, where the group law is defined by

$$(h_1, g_1) \cdot (h_2, g_2) = (h_1 \cdot \varphi(g_1)(h_2), g_1 \cdot g_2).$$

When the action φ is clear from the context, we write $H \rtimes G$ instead of $H \rtimes_\varphi G$. Since the group $\mathsf{SL}(n, \mathbb{Z})$ naturally acts on $\mathbb{Z}^n$ by matrix multiplication, one can define the semidirect product $\mathbb{Z}^n \rtimes \mathsf{SL}(n, \mathbb{Z})$ using this action. This corresponds exactly to the special affine group $\mathsf{SA}(n, \mathbb{Z})$.

2.2 Semigroup algorithmic problems

2.2.1 Words and semigroups

In all semigroup algorithmic problems considered in this thesis, we work in some given group G. The group G may be given in one of the two following forms.

(1) G may be explicitly given as a matrix group (for example, $\mathsf{UT}(n, \mathbb{Q}), \mathbb{Z} \wr \mathbb{Z}^n$ or $\mathsf{SL}(n, \mathbb{Z})$). In this case, the elements of G are represented as matrices with binary encoded entries.

(2) G may be given as an abstract group by its generators and its defining relations. For example, a nilpotent group is given by a finite presentation, and a metabelian group is

[2] An *action* φ of the group G on H is defined as follows. For every $g \in G$, there is a group automorphism $\varphi(g) \colon H \to H$, such that:
(i) $\varphi(I)$ is the identity map,
(ii) $\varphi(gg') = \varphi(g) \circ \varphi(g')$ for all $g, g' \in G$.

given by a finite metabelian presentation. In this case, the elements of G are represented as words over its generators.

Most of this thesis deals directly with matrix groups. When G is given as an abstract group, we will first embed it in a concrete matrix group, and then solve algorithmic problems over this matrix group.

Fix a group G. Let $\mathcal{G} = \{g_1, \ldots, g_K\} \subseteq G$ be a finite set of elements, considered as an alphabet. For an arbitrary word $w = g_{i_1} g_{i_2} \cdots g_{i_m} \in \mathcal{G}^*$, by multiplying consecutively the elements appearing in w, we can evaluate w as an element $\pi(w)$ in G. As a convention, the evaluation of the empty word is the neutral element I of G. We say that the word w *represents* the element $\pi(w) \in G$. The semigroup $\langle \mathcal{G} \rangle$ generated by $\mathcal{G}$ is hence the set of elements in G that are represented by *non-empty* words in $\mathcal{G}^*$. The Identity Problem can be reformulated as deciding whether the neutral element of G can be represented by a non-empty word over $\mathcal{G}$. A similar reformulation of the Group Problem can also be obtained as follows. A word w over the alphabet $\mathcal{G}$ is called *full-image* if every letter in $\mathcal{G}$ has at least one occurrence in w.

Lemma 2.2.1. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if the neutral element of G is represented by a full-image word over $\mathcal{G}$.*

Proof. Let $w \in \mathcal{G}^*$ be a full-image word with $\pi(w) = I$. Then for every $g_i \in \mathcal{G}$, the word w can be written as $w = v g_i v'$, so $g_i^{-1} = \pi(v')\pi(v) \in \langle \mathcal{G} \rangle$. Therefore, the semigroup $\langle \mathcal{G} \rangle$ contains all the inverse g_i^{-1}, and is thus a group.

If $\langle \mathcal{G} \rangle$ is a group, then for all $g_i \in \mathcal{G}$, the inverse g_i^{-1} can be written as $\pi(w_i)$ for some word $w_i \in \mathcal{G}^*$. Then $w \coloneqq g_1 w_1 g_2 w_2 \cdots g_a w_a$ is a full-image word with $\pi(w) = \pi(g_1 w_1) \cdots \pi(g_a w_a) = I$. □

2.2.2 Hierarchy of semigroup algorithmic problems

In this subsection we state some classic reductions between different semigroup algorithmic problems. As mentioned in the introduction, here is a list of decision problems that are of interest to us. For all these problems, we work in some fixed group G.

(i) *(Semigroup Membership)* given a finite set $\mathcal{G}$ and an element T in G, decide whether $T \in \langle \mathcal{G} \rangle$.

(ii) *(Group Membership)* given a finite set $\mathcal{G}$ and an element T in G, decide whether $T \in \langle \mathcal{G} \rangle_{grp}$.

(iii) *(Identity Problem)* given a finite set $\mathcal{G}$, decide whether $I \in \langle \mathcal{G} \rangle$.

(iv) *(Group Problem)* given a finite set $\mathcal{G}$, decide whether $\langle \mathcal{G} \rangle = \langle \mathcal{G} \rangle_{grp}$.

(v) *(Semigroup Intersection)* given two finite sets $\mathcal{G}$ and $\mathcal{H}$, decide whether $\langle \mathcal{G} \rangle \cap \langle \mathcal{H} \rangle = \emptyset$.

(vi) *(Orbit Intersection)* given two finite sets $\mathcal{G}$ and $\mathcal{H}$ as well as two elements S, T in G, decide whether $T \cdot \langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$.

First we generalize the Group Problem to the problem of computing *invertible subsets*, whose decidability is equivalent to the Group Problem.

Definition 2.2.2. Let G be a group. Given a finite set of elements $\mathcal{G} = \{g_1, \ldots, g_K\} \subseteq G$, its *invertible subset* $\mathcal{G}_{inv}$ is defined as the set of element in $\mathcal{G}$ who inverse lies in $\langle \mathcal{G} \rangle$:

$$\mathcal{G}_{inv} \coloneqq \left\{ g_i \in \mathcal{G} \;\middle|\; g_i^{-1} \in \langle \mathcal{G} \rangle \right\}.$$

For a semigroup S, the *group of units* of S is the set of elements who have an inverse in S. This set forms a group when it is not empty [48]. The algorithmic problem of computing the group of units of a semigroup is interesting in its own right, and is important for understanding the structure of the semigroup. The following lemma shows that the computation of invertible subsets subsumes the Identity Problem and the Group Problem, as well as gives a generating set for the group of units.

Lemma 2.2.3. *Given a finite set of elements $\mathcal{G} = \{g_1, \ldots, g_K\}$ in a group G.*

(i) The Identity Problem for $\mathcal{G}$ has a positive answer if and only if $\mathcal{G}_{inv}$ is non-empty.

(ii) The group of units of $\langle \mathcal{G} \rangle$ is generated as a semigroup by $\mathcal{G}_{inv}$. In particular, the group of units of $\langle \mathcal{G} \rangle$ is empty if and only if $\mathcal{G}_{inv}$ is empty.

(iii) The Group Problem for $\mathcal{G}$ has a positive answer if and only if $\mathcal{G}_{inv} = \mathcal{G}$.

Proof. (i) If the Identity Problem has a positive answer, let $w \in \mathcal{G}^*$ be a *non-empty* word such that $\pi(w) = I$. Write $w = g_i w'$, (w' could be the empty word), then $g_i^{-1} = \pi(w')$. If $g_i = I$ then obviously $g_i^{-1} = g_i \in \langle \mathcal{G} \rangle$. If $g_i \neq I$ then $\pi(w') \neq I$ so w' is not the empty word and $\pi(w') \in \langle \mathcal{G} \rangle$. Therefore $g_i^{-1} \in \langle \mathcal{G} \rangle$. Hence, $\mathcal{G}_{inv}$ contains g_i and is not empty. Conversely, if $g_i \in \mathcal{G}_{inv}$ for some i, then either $g_i = I$ in which case $I = g_i \in \langle \mathcal{G} \rangle$, or $g_i^{-1} = \pi(w')$ for some non-empty word w', so $I = \pi(g_i w') \in \langle \mathcal{G} \rangle$.

(ii) Since every element in $\mathcal{G}_{inv}$ is invertible in $\langle \mathcal{G} \rangle$, every element in the semigroup $\langle \mathcal{G}_{inv} \rangle$ is also invertible in $\langle \mathcal{G} \rangle$. Hence, it suffices to show that no element in $\langle \mathcal{G} \rangle \setminus \langle \mathcal{G}_{inv} \rangle$ is invertible. Suppose on the contrary that there exists a word $w \in \mathcal{G}^*$ such that $\pi(w) \in \langle \mathcal{G} \rangle \setminus \langle \mathcal{G}_{inv} \rangle$ and $\pi(w)^{-1} \in \langle \mathcal{G} \rangle$. Since $\pi(w) \notin \langle \mathcal{G}_{inv} \rangle$, w must contain a letter $g_i \in \mathcal{G} \setminus \mathcal{G}_{inv}$. But because $\pi(w)^{-1} \in \langle \mathcal{G} \rangle$, there exists a word $v \in \mathcal{G}^*$ such that $\pi(v) = \pi(w)^{-1}$. Thus, $\pi(wv) = I$. But then wv is a word containing the letter g_i. Writing $wv = w_1 g_i w_2$, we have $g_i^{-1} = \pi(w_2)\pi(w_1) \in \langle \mathcal{G} \rangle$, a contradiction to $g_i \notin \mathcal{G}_{inv}$.

(iii) Since a semigroup is a group if and only if it is its own group of units, (iii) is a direct consequence of (ii). □

The following theorem shows that decidability of the Group Problem subsumes the computability of the invertible subset and the decidability of the Identity Problem.

Lemma 2.2.4. *Let G be a group. Suppose the Group Problem is decidable in G. Then one can compute the invertible subset $\mathcal{G}_{inv}$ of any finite set $\mathcal{G} \subseteq G$. In particular, the Identity Problem is also decidable in G.*

Furthermore, if the complexity of the Group Problem in G is in NP, then the Identity Problem is also in NP.

Proof. Let $\mathcal{G} = \{g_1, \ldots, g_K\}$. To compute the invertible subset $\mathcal{G}_{inv}$, it suffices to decide for each $g_i, i = 1, \ldots, K$ whether its inverse is in $\langle \mathcal{G} \rangle$. Without loss of generality suppose we want to decide whether $g_1^{-1} \in \langle \mathcal{G} \rangle$. We claim that $g_1^{-1} \in \langle \mathcal{G} \rangle$ if and only if there is some subset $\mathcal{H}$ of $\mathcal{G}$, such that $\langle \mathcal{H} \cup \{g_1\} \rangle = \langle \mathcal{H} \cup \{g_1\} \rangle_{grp}$. Indeed, if $g_1^{-1} \in \langle \mathcal{G} \rangle$, suppose g_1^{-1} is represented by the word $w \in \mathcal{G}^*$. Let $\mathcal{H}$ be the set of letters appearing in w, then $g_1 w$ is a full-image word in the alphabet $\mathcal{H} \cup \{g_1\}$ representing the neutral element, and hence $\langle \mathcal{H} \cup \{g_1\} \rangle = \langle \mathcal{H} \cup \{g_1\} \rangle_{grp}$ by Lemma 2.2.1. For the opposite implication, if $\langle \mathcal{H} \cup \{g_1\} \rangle = \langle \mathcal{H} \cup \{g_1\} \rangle_{grp}$, then $g_1^{-1} \in \langle \mathcal{H} \cup \{g_1\} \rangle_{grp} = \langle \mathcal{H} \cup \{g_1\} \rangle \subseteq \langle \mathcal{G} \rangle$.

Therefore, to decide whether $g_1^{-1} \in \langle \mathcal{G} \rangle$, it suffices to check the Group Problem for every subset $\mathcal{H} \cup \{g_1\}$ of $\mathcal{G}$ containing g_1. In particular, the Identity Problem is also decidable by Lemma 2.2.3(i).

Note that the Identity Problem has a positive answer if and only if there exists a non-empty subset $\mathcal{H} \subseteq \mathcal{G}$ such that $\langle \mathcal{H} \rangle$ is a group. On one hand, if the semigroup $\langle \mathcal{G} \rangle$ contains the neutral element I, then by Lemma 2.2.3(i) the set $\mathcal{G}_{inv} \subseteq \mathcal{G}$ is non-empty and $\langle \mathcal{G}_{inv} \rangle$ is a group. On the other hand, if there exists a non-empty subset $\mathcal{H} \subseteq \mathcal{G}$ such that $\langle \mathcal{H} \rangle$ is a group, then $I \in \langle \mathcal{H} \rangle \subseteq \langle \mathcal{G} \rangle$. Therefore, if the Group Problem in G is decidable in NP, then by guessing the non-empty subset $\mathcal{H} \subseteq \mathcal{G}$ and checking the Group Problem for $\mathcal{H}$, we can decide the Identity Problem in NP. □

Lemmas 2.2.3 and 2.2.4 show that decidability of the Group Problem is equivalent to the computability of the invertible subset; and they both subsume decidability of the Identity Problem.

The following lemma shows that decidability of Semigroup Intersection or Semigroup Membership subsumes computability of the invertible subset.

Lemma 2.2.5. *Let G be a group.*

(i) If Semigroup Intersection is decidable in G, then one can compute the invertible subset $\mathcal{G}_{inv}$ of any finite set $\mathcal{G} \subseteq G$.

(ii) If Semigroup Membership is decidable in G, then one can compute the invertible subset $\mathcal{G}_{inv}$ of any finite set $\mathcal{G} \subseteq G$.

Proof. (i) Suppose Semigroup Intersection is decidable in G. Let $\mathcal{G} = \{g_1, \ldots, g_K\}$ and $i \in \{1, \ldots, K\}$. We show how to decide whether $g_i^{-1} \in \langle \mathcal{G} \rangle$, this will allow us to compute $\mathcal{G}_{inv}$. We claim that $g_i^{-1} \in \langle \mathcal{G} \rangle$ if and only if $\langle g_i^{-1} \rangle \cap \langle \mathcal{G} \rangle \neq \emptyset$, which is decidable using an algorithm for Semigroup Intersection.

Indeed, if $g_i^{-1} \in \langle \mathcal{G} \rangle$ then $\langle g_i^{-1} \rangle \cap \langle \mathcal{G} \rangle \neq \emptyset$. For the opposite implication, suppose $\langle g_i^{-1} \rangle \cap \langle \mathcal{G} \rangle \neq \emptyset$. So $g_i^{-m} \in \langle \mathcal{G} \rangle$ for some $m \geq 1$. If $m = 1$ then $g_i^{-1} \in \langle \mathcal{G} \rangle$, otherwise $g_i^{-1} = g_i^{m-1} \cdot g_i^{-m} \in \langle \mathcal{G} \rangle$ since both g_i^{m-1} and g_i^{-m} are in $\langle \mathcal{G} \rangle$.

(ii) Suppose Semigroup Membership is decidable in G. Given $\mathcal{G} = \{g_1, \ldots, g_K\}$ and $i \in \{1, \ldots, K\}$, we can decide whether $g_i^{-1} \in \langle \mathcal{G} \rangle$ using the algorithm for Semigroup Membership, this allows us to compute $\mathcal{G}_{inv}$. □

As mentioned in Section 1.1, Orbit Intersection subsumes both Semigroup Intersection and Semigroup Membership; and Semigroup Membership subsumes Group Membership. Reductions between the different algorithmic problems is summarized in the following diagram 2.1.

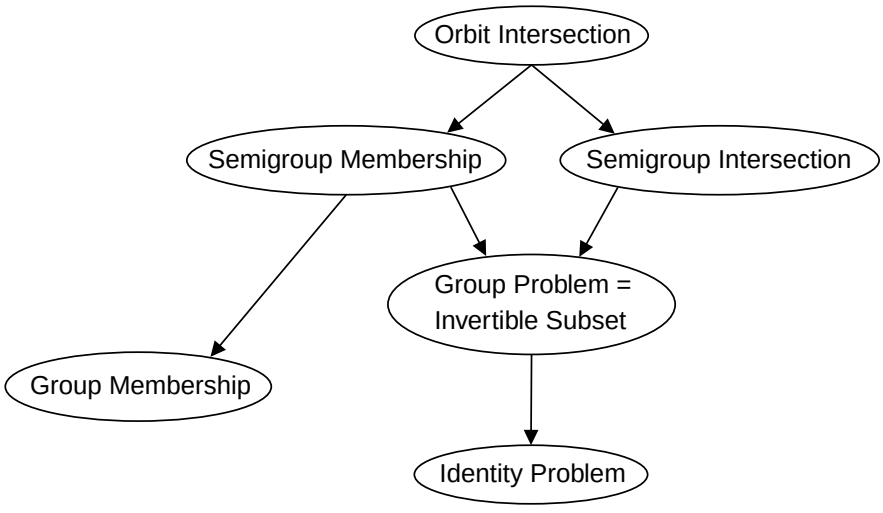

Figure 2.1: Reductions between the different algorithmic problems.

Chapter 3

Nilpotent groups

3.1 Introduction and main results

In this chapter we study algorithmic problems in nilpotent groups of bounded nilpotency class (recall Definition 2.1.3). Nilpotent groups are usually considered as an immediate generalization of abelian groups. Among the classical Max Dehn problems in nilpotent groups, the Word Problem has long been known to be decidable [71], but it was only established comparatively recently that it is decidable in linear time [46]. Decidability of the Conjugacy Problem is due to Blackburn [19], and decidability of the Isomorphism Problem is due to Grunewald and Segal [41]. Grunewald and Segal's solution to the Isomorphism Problem relies on the construction of a full faithful functor from the category of finite presentations of torsion-free nilpotent groups to the category of subgroups of $\mathsf{UT}(n, \mathbb{Z})$. In other words, they constructed a "canonical" embedding of finitely generated nilpotent groups into $\mathsf{UT}(n, \mathbb{Z})$.

In the 1950s, Mal'cev [71] established the decidability of Group Membership in nilpotent groups by showing the property of *subgroup separability*. For an analysis of the complexity of Group Membership in nilpotent groups, see [77]. In 2022, Roman'kov [86] constructed a nilpotent group of class two with undecidable Semigroup Membership. Decidability of the Identity Problem, the Group Problem as well as Semigroup Intersection remains an intricate open problem.

We point out that by the generalized Tits alternative [83], a linear semigroup[1] either contains a finite-index nilpotent subsemigroup[2] or it contains a non-commutative free subsemigroup. Semigroup Intersection in a direct product of two non-commutative free semigroups is undecidable due to an embedding of the Post Correspondence Problem (see Section 4.1). Therefore,

[1] A linear semigroup is a subsemigroup of the group $\mathsf{GL}(n, \mathbb{K})$ of $n \times n$ invertible matrices over a field $\mathbb{K}$.

[2] A semigroup is called nilpotent if the group it generates is nilpotent.

nilpotent groups are the largest "natural" class of groups (i.e. a class closed under direct products) where decidability of Semigroup Intersection remains possible.

Most existing results for semigroup algorithmic problems in nilpotent groups are restricted to the *Heisenberg groups*. The unitriangular matrix group $\mathsf{UT}(3, \mathbb{Q})$ (recall Definition 2.1.4) is commonly called the Heisenberg group over rational numbers and is denoted as $\mathrm{H}_3(\mathbb{Q})$. More generally, the n-dimensional Heisenberg group $\mathrm{H}_n(\mathbb{K})$ over a field or commutative ring $\mathbb{K}$ is defined as

$$\mathrm{H}_n(\mathbb{K}) \coloneqq \left\{ \begin{pmatrix} 1 & \boldsymbol{a}^\top & c \\ 0 & I_{n-2} & \boldsymbol{b} \\ 0 & 0 & 1 \end{pmatrix}, \text{ where } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{K}^{n-2},\, c \in \mathbb{K} \right\},$$

where we use the notation I_{n-2} for the identity matrix of dimension $n-2$. The Heisenberg groups $\mathrm{H}_n(\mathbb{K})$ play an important role in many branches of mathematics, physics and computer science. They first arose in the description of one-dimensional quantum mechanical systems [79, 97], and have now become an important mathematical object connecting domains like representation theory, theta functions, Fourier analysis and quantum algorithms [47, 50, 57, 62, 99]. Heisenberg groups are the simplest non-commutative Lie groups and are of nilpotency class two.

In [58], Ko, Niskanen and Potapov showed PTIME decidability of the Identity Problem in $\mathrm{H}_3(\mathbb{Q})$. Later, using the special structure of the first term in the *Baker-Campbell-Hausdorff formula*, Colcombet, Ouaknine, Semukhin and Worrell proved the decidability of Semigroup Membership in $\mathrm{H}_n(\mathbb{Q})$ for all n by encoding it into a Parikh automaton [30]. Recently, Roman'kov [86] showed undecidability of Semigroup Membership in the direct product $\mathrm{H}_3(\mathbb{Q})^k$ for sufficiently large k. His main idea is an embedding of the Hilbert's tenth problem [73].

In this chapter we extend some of these results to nilpotent groups of higher classes, as well as include new algorithmic problems. By default, a finitely generated nilpotent group is represented using a finite presentation (Definition 2.1.1). Every finitely generated nilpotent group admits a finite presentation [34, Proposition 13.84], making it the natural way to represent nilpotent groups.

Main results

Since every finitely generated nilpotent group admits a quotient that is isomorphic to a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ [8, 53], many decision problems in nilpotent groups reduce to decision problems in $\mathsf{UT}(n, \mathbb{Q})$. In Section 3.2 we will formalize this fact, and the rest of this chapter will only deal with algorithmic problems in the groups $\mathsf{UT}(n, \mathbb{Q})$.

Our first result concerns computing Invertible Subsets (hence also the Identity Problem and the Group Problem) in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most ten.

Theorem 3.1.1. *Let $n \geq 2$ and G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most ten. Given any finite set $\mathcal{G} \subseteq G$, the invertible subset of $\mathcal{G}$ is computable in polynomial time.*

To be precise, elements of G are represented using matrices with *binary encoded* entries. Theorem 3.1.1 implies that the Identity Problem and the Group Problem are decidable in PTIME in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most ten (see Lemma 2.2.3(iii)).

Theorem 3.1.1 can be extended to arbitrary finitely generated nilpotent groups of class ten. However, the complexity will depend on specific group embeddings, which we do not analyse.

Corollary 3.1.2. *Let G be a finitely generated nilpotent group of class at most ten, given by a finite presentation. Then the Group Problem (hence also the Identity Problem) is decidable in G.*

Our second result covers Semigroup Intersection in class two nilpotent subgroups of $\mathsf{UT}(n, \mathbb{Q})$.

Theorem 3.1.3. *Let $n \geq 2$ and let G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most two. Given finite subsets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of G, it is decidable in polynomial time whether $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle = \emptyset$.*

The decidability result for arbitrary nilpotent groups of class two follows as a corollary of Theorem 3.1.3.

Corollary 3.1.4. *Let G be a nilpotent group of class at most two, given by a finite presentation. Given finite subsets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of G, it is decidable whether $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle = \emptyset$.*

Our last result concerns Orbit Intersection in the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$.

Theorem 3.1.5. *Given elements $T, S \in \mathrm{H}_3(\mathbb{Q})$ and two finite subsets $\mathcal{G}, \mathcal{H}$ of $\mathrm{H}_3(\mathbb{Q})$, it is decidable whether $T \cdot \langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$.*

Organization of the chapter

The organization of this chapter is as follows. In Section 3.2, we exhibit a series of embedding theorems and reduce the Group Problem and Semigroup Intersection in arbitrary nilpotent groups to subgroups of $\mathsf{UT}(n, \mathbb{Q})$. This will allow us to deduce Corollary 3.1.2 and 3.1.4 from Theorem 3.1.1 and 3.1.3. We then focus on algorithmic problems in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ in subsequent sections. In Section 3.3, we introduce the necessary tools in linear programming and Lie algebra. Notably, we will state the *Baker-Campbell-Hausdorff formula*. In Section 3.4 we exhibit an algorithm (Algorithm 3.1) that computes invertible subsets in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most ten. Algorithm 3.1 is correct subject to a highly non-trivial

structural theorem of unitriangular matrix semigroups (Theorem 3.4.2). In Section 3.5 we prove this structural theorem. Our proof relies on several technical lemmas proven using assistance from computer algebra software. Their code is provided as additional material in Section 3.10. In Section 3.6 we state a conjecture (Conjecture 3.6.1) under which this structural theorem remains correct for higher nilpotency class. We also give a procedure to verify this conjecture in case it is true[3].

Sections 3.7-3.9 deal with Semigroup Intersection and Orbit Intersection in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ of nilpotency class two. In particular, in Section 3.7 we prove a proposition (Proposition 3.7.2) on the combinatorics of *length-two subwords*. Then, using this proposition, we prove PTIME decidability of Semigroup Intersection in subgroups of $\mathsf{UT}(n, \mathbb{Q})$ of nilpotency class two (Section 3.8), as well as decidability of Orbit Intersection in the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$ (Section 3.9).

3.2 Representing a nilpotent group

The purpose of this section is to prove the following proposition.

Proposition 3.2.1. *Suppose we are given a finite presentation of a nilpotent group G of class at most d. One can compute an integer $n \in \mathbb{N}$ as well as an effective homomorphism $\phi\colon G \to \mathsf{UT}(n, \mathbb{Q})$, such that $\phi(G)$ is nilpotent of class at most d, and*

(1) For any finite set $\mathcal{G} \subseteq G$, the semigroup $\langle \mathcal{G} \rangle$ is a group if and only if the semigroup $\langle \phi(\mathcal{G}) \rangle$ is a group.

(2) For any finite subsets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of G, we have $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle \neq \emptyset$ if and only if $\langle \phi(\mathcal{G}_1) \rangle \cap \cdots \cap \langle \phi(\mathcal{G}_M) \rangle \neq \emptyset$.

Proposition 3.2.1 reduces the Group Problem and Semigroup Intersection in finitely presented nilpotent groups to respective problems in subgroups of $\mathsf{UT}(n, \mathbb{Q})$. Here is some intuition for constructing ϕ: let G be a nilpotent group, we denote by T its subgroup consisting of all torsion elements. The quotient G/T is torsion-free and can be embedded into some $\mathsf{UT}(n, \mathbb{Q})$. Then we can take ϕ as the composition $G \to G/T \hookrightarrow \mathsf{UT}(n, \mathbb{Q})$.

In order to prove Proposition 3.2.1 rigorously, we first recall some well-known properties of finitely generated nilpotent groups.

Lemma 3.2.2 ([34, Lemma 13.56, Theorem 13.57])**.** *(1) Every subgroup of a finitely generated nilpotent group is finitely generated nilpotent.*

[3]Likely due to the lack of computational power, our verification of Conjecture 3.6.1 stops at nilpotency class ten.

(2) If G is finitely generated nilpotent of class d and N is a normal subgroup of G, then the quotient G/N is finitely generated nilpotent of class at most d.

(3) The direct product of a finite family of finitely generated nilpotent groups is again finitely generated nilpotent.

Let G be a nilpotent group with the neutral element I. An element $t \in G$ is called *torsion* if $t^n = I$ for some $n \geq 1$. The group G is called *torsion free* if the only torsion element of G is I.

Lemma 3.2.3 ([34, Theorem 13.64])**.** *Let G be a nilpotent group. The set of all torsion elements of G forms a normal subgroup of G, called the* torsion subgroup *of G.*

Lemma 3.2.4 ([34, Proposition 13.65])**.** *The torsion subgroup of a finitely generated nilpotent group is finite.*

We denote by T the torsion subgroup of G. The quotient group G/T is torsion-free, since every torsion element of G is contained in T. Let $g \mapsto \overline{g} \coloneqq gT$ denote the canonical projection $G \to G/T$. For a set $\mathcal{G} \subseteq G$, denote

$$\overline{\mathcal{G}} \coloneqq \{\overline{g_i} \mid g_i \in \mathcal{G}\}.$$

Lemma 3.2.5. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if $\langle \overline{\mathcal{G}} \rangle$ is a group.*

Proof. Suppose the semigroup $\langle \mathcal{G} \rangle$ is a group, then by Lemma 2.2.1 there exists a full-image word $w = g_{i_1} g_{i_2} \cdots g_{i_m}$ over the alphabet $\mathcal{G}$ that represents the neutral element of G. Then the word $\overline{w} \coloneqq \overline{g_{i_1}} \cdots \overline{g_{i_m}}$ is full-image over the alphabet $\overline{\mathcal{G}}$ and represents the neutral element of G/T. So $\langle \overline{\mathcal{G}} \rangle$ is a group by Lemma 2.2.1.

Suppose the semigroup $\langle \overline{\mathcal{G}} \rangle$ is a group. By Lemma 2.2.1, there exists a full-image word $w = \overline{g_{i_1}}\,\overline{g_{i_2}} \cdots \overline{g_{i_m}}$ over the alphabet $\overline{\mathcal{G}}$ that represents the neutral element of G/T. Then the word $\widetilde{w} \coloneqq g_{i_1} g_{i_2} \cdots g_{i_m}$ is full-image over the alphabet $\mathcal{G}$ and represents some element t in T. Suppose $t^n = I$, then the word

$$\widetilde{w}^n \coloneqq \underbrace{(g_{i_1} g_{i_2} \cdots g_{i_m}) \cdots (g_{i_1} g_{i_2} \cdots g_{i_m})}_{n \text{ times}}$$

is full-image over the alphabet $\mathcal{G}$ and represents $I \in G$. So $\langle \mathcal{G} \rangle$ is a group by Lemma 2.2.1. □

Lemma 3.2.6. *The intersection $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle$ is non-empty if and only if the intersection $\langle \overline{\mathcal{G}_1} \rangle \cap \cdots \cap \langle \overline{\mathcal{G}_M} \rangle$ is non-empty.*

Proof. Suppose the intersection $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle$ is non-empty and contains an element g. Then $\langle \overline{\mathcal{G}_1} \rangle \cap \cdots \cap \langle \overline{\mathcal{G}_M} \rangle$ is non-empty since it contains $\overline{g}$.

Suppose $\langle\overline{\mathcal{G}_1}\rangle\cap\cdots\cap\langle\overline{\mathcal{G}_M}\rangle$ is non-empty. Let $g\in G$ such that $\overline{g}$ is an element in $\langle\overline{\mathcal{G}_1}\rangle\cap\cdots\cap\langle\overline{\mathcal{G}_M}\rangle$, we now show that for a well chosen $N\in\mathbb{N}$, we have $g^N\in\langle\mathcal{G}_1\rangle\cap\cdots\cap\langle\mathcal{G}_M\rangle$.

Let $t_1,\ldots,t_M\in T$ be such that $gt_1\in\langle\mathcal{G}_1\rangle,\ldots gt_M\in\langle\mathcal{G}_M\rangle$. Take any $i\in\{1,\ldots,M\}$, we show that there exists $n_i\geq 1$ such that $(gt_i)^{n_i}=g^{n_i}$. Indeed, consider the sequence $t_i, gt_ig^{-1}, g^2t_ig^{-2},\ldots$. Since T is a normal subgroup (Lemma 3.2.3), every element in the sequence is in T. Since T is finite (Lemma 3.2.4), there exist two elements $g^pt_ig^{-p}, g^qt_ig^{-q}$ where $p<q$ and $g^pt_ig^{-p}=g^qt_ig^{-q}$. Taking $k\coloneqq q-p$ we have $g^kt_ig^{-k}=t_i$ and

$$g^jt_ig^{-j}=g^{j+k}t_ig^{-(j+k)}\quad\text{for all } j\in\mathbb{Z}. \tag{3.1}$$

Consider the element

$$h\coloneqq\prod_{j=1}^{k}g^jt_ig^{-j}\in T.$$

There exists $n\geq 1$ such that $h^n=I$. Then

$$(gt_i)^{kn}g^{-kn}=\left(gt_ig^{-1}\right)\left(g^2t_ig^{-2}\right)\cdots\left(g^{kn}t_ig^{-kn}\right)=\prod_{j=1}^{kn}g^jt_ig^{-j}\overset{(3.1)}{=}\left(\prod_{j=1}^{k}g^jt_ig^{-j}\right)^n=h^n=I.$$

Therefore $n_i\coloneqq kn$ satisfies $(gt_i)^{n_i}=g^{n_i}$.

Now take $N\coloneqq n_1n_2\cdots n_M$, then $(gt_i)^N=g^N$ for $i=1,\ldots,M$. Hence $g^N=(gt_i)^N\in\langle\mathcal{G}_i\rangle$ for all $i\in\{1,\ldots,M\}$, yielding $g^N\in\langle\mathcal{G}_1\rangle\cap\cdots\cap\langle\mathcal{G}_M\rangle$. Thus, $\langle\mathcal{G}_1\rangle\cap\cdots\cap\langle\mathcal{G}_M\rangle$ is non-empty. □

Lemma 3.2.7 ([68, Theorem 8])**.** *Given a finite presentation of a nilpotent group G, one can compute the generators of its torsion subgroup T.*

Note that given a finite presentation of G and the generators of $T\trianglelefteq G$, a finite presentation of G/T can be obtained. In fact, if $G=\langle g_1,\ldots,g_n\mid r_1,\ldots,r_m\rangle$ and the generators of T are $t_1,\ldots,t_p$, then $G/T=\langle g_1,\ldots,g_n\mid r_1,\ldots,r_m,t_1,\ldots,t_p\rangle$ (see, for example, [51, Chapter 4, Proposition 2]).

Lemma 3.2.8 ([41, Algorithm E])**.** *Given a finite presentation of a torsion-free nilpotent group A, one can compute $n\in\mathbb{N}$ and an embedding $\theta: A\hookrightarrow\mathsf{UT}(n,\mathbb{Q})$.*

We can now construct the homomorphism ϕ in Proposition 3.2.1. Since G/T is torsion free, Lemma 3.2.8 gives an effective embedding $\theta\colon G/T\hookrightarrow\mathsf{UT}(n,\mathbb{Q})$ for some n. We compose θ with the canonical projection $G\to G/T, g\mapsto\overline{g}$, and obtain the homomorphism

$$\phi\colon G\to\mathsf{UT}(n,\mathbb{Q}),\quad g\mapsto\theta(\overline{g}). \tag{3.2}$$

This homomorphism is effective by Lemma 3.2.7 and 3.2.8. We can now prove the main proposition of this subsection:

Proposition 3.2.1. *Suppose we are given a finite presentation of a nilpotent group G of class at most d. One can compute an integer $n \in \mathbb{N}$ as well as an effective homomorphism $\phi\colon G \to \mathsf{UT}(n, \mathbb{Q})$, such that $\phi(G)$ is nilpotent of class at most d, and*

(1) For any finite set $\mathcal{G} \subseteq G$, the semigroup $\langle \mathcal{G} \rangle$ is a group if and only if the semigroup $\langle \phi(\mathcal{G}) \rangle$ is a group.

(2) For any finite subsets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of G, we have $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle \neq \emptyset$ if and only if $\langle \phi(\mathcal{G}_1) \rangle \cap \cdots \cap \langle \phi(\mathcal{G}_M) \rangle \neq \emptyset$.

Proof. Let $G = \langle g_1, \ldots, g_n \mid r_1, \ldots, r_m \rangle$ be the given finite presentation of G. Let T be the torsion subgroup of G. By Lemma 3.2.7, the generators of T can be effectively computed. Denote them by $t_1, \ldots, t_p$, then $G/T = \langle g_1, \ldots, g_n \mid r_1, \ldots, r_m, t_1, \ldots, t_p \rangle$. The set $\overline{\mathcal{G}}$ under the canonical projection $G \to G/T$ is given by the same words representing the elements of $\mathcal{G}$.

Take ϕ to be the homomorphism defined in (3.2). By Lemma 3.2.2 (2), the image $\phi(G) \cong G/T$ is of nilpotency class at most d. Since the embedding θ from Lemma 3.2.8 is effective, ϕ is also effective. Since θ is an isomorphism between G/T and $\phi(G)$, Lemma 3.2.5 and 3.2.6 yield respectively the statements (1) and (2) in the proposition. □

By Proposition 3.2.1(1), deciding the Group Problem in a finitely generated nilpotent group G of class at most ten reduces to deciding the Group Problem in the subgroup $\phi(G)$ of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most ten. Since decidability of the Group Problem is equivalent to computability of the Invertible Subset (Lemma 2.2.4), Corollary 3.1.2 follows immediately from Theorem 3.1.1. Similarly by Proposition 3.2.1(2), Corollary 3.1.4 follows immediately from Theorem 3.1.3. Hence, for the rest of this chapter, we will only focus on proving Theorem 3.1.1, 3.1.3 and 3.1.5, and work in subgroups of $\mathsf{UT}(n, \mathbb{Q})$.

3.3 Linear programming and Lie algebra

3.3.1 Convex geometry and linear programming

Let G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$. Given a finite set of matrices $\mathcal{G} = \{A_1, \ldots, A_K\}$ in G, recall that $\mathcal{G}^*$ denotes the set of words over the alphabet $\mathcal{G}$. We now define some concepts necessary for analysing words with linear algebra.

Definition 3.3.1 (Parikh Image)**.** Given a finite alphabet $\mathcal{G} = \{A_1, \ldots, A_K\}$, the *Parikh Image* of a word $w = B_1 \cdots B_m$ in $\mathcal{G}^*$ is the vector $\mathrm{PI}^{\mathcal{G}}(w) \coloneqq \big(\mathrm{PI}_1^{\mathcal{G}}(w), \ldots, \mathrm{PI}_K^{\mathcal{G}}(w)\big) \in \mathbb{Z}_{\geq 0}^K$, where

$\mathrm{PI}_i^{\mathcal{G}}(w)$ is the number of times A_i appears in w. That is, $\mathrm{PI}_i^{\mathcal{G}}(w) \coloneqq \mathrm{card}(\{j \mid B_j = A_i\})$. When the alphabet $\mathcal{G}$ is clear from the context, we write $\mathrm{PI}(w), \mathrm{PI}_i(w)$ instead of $\mathrm{PI}^{\mathcal{G}}(w), \mathrm{PI}_i^{\mathcal{G}}(w)$.

Definition 3.3.2 (Cones)**.** Let V be a $\mathbb{Q}$-linear space. A subset $\mathcal{C} \subseteq V$ is called a $\mathbb{Q}_{\geq 0}$-*cone* if $a \in \mathcal{C} \Longrightarrow a\mathbb{Q}_{\geq 0} \subseteq \mathcal{C}$, and $a, b \in \mathcal{C} \Longrightarrow a + b \in \mathcal{C}$. Given a set of vectors $\mathcal{S} \subseteq V$, denote by $\langle \mathcal{S} \rangle_{\mathbb{Q}_{\geq 0}}$ the $\mathbb{Q}_{\geq 0}$-*cone generated by* $\mathcal{S}$, that is, the smallest $\mathbb{Q}_{\geq 0}$-cone of V containing $\mathcal{S}$. Similarly, denote by $\langle \mathcal{S} \rangle_{\mathbb{Q}}$ the $\mathbb{Q}$-linear space generated by $\mathcal{S}$. These definitions can be naturally extended to $\mathbb{R}_{\geq 0}$-cones and $\mathbb{R}$-linear spaces.

Definition 3.3.3 (Support)**.** A subset $\Lambda \subseteq \mathbb{Z}_{\geq 0}^K$ is called a $\mathbb{Z}_{\geq 0}$-*cone* if $a, b \in \Lambda \Longrightarrow a + b \in \Lambda$, and $\mathbf{0} \in \Lambda$. The *support* of a vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}_{\geq 0}^K$ is defined as the set of indices where the entry of $\boldsymbol{\ell}$ is non-zero:

$$\mathrm{supp}(\boldsymbol{\ell}) \coloneqq \{i \in \{1, \ldots, K\} \mid \ell_i > 0\}.$$

The *support* of a $\mathbb{Z}_{\geq 0}$-cone Λ is defined as the union of supports of all vectors in Λ:

$$\mathrm{supp}(\Lambda) \coloneqq \bigcup_{\boldsymbol{\ell} \in \Lambda} \mathrm{supp}(\boldsymbol{\ell}) = \{i \mid \exists (\ell_1, \ldots, \ell_K) \in \Lambda, \ell_i > 0\}.$$

Let V be a $\mathbb{Q}$-linear subspace of $\mathbb{Q}^K$, represented as the solution set of linear homogeneous equations. Then $\mathbb{Z}_{\geq 0}^K \cap V$ is a $\mathbb{Z}_{\geq 0}$-cone. In this chapter, we will need to compute the support of $\mathbb{Z}_{\geq 0}$-cones of the form $\Lambda = \mathbb{Z}_{\geq 0}^K \cap V$.

Lemma 3.3.4. *Given V represented as the solution set of linear homogeneous equations, one can compute the support of $\Lambda = \mathbb{Z}_{\geq 0}^K \cap V$ in polynomial time.*

Proof. For $i = 1, \ldots, K$, we can check whether $i \in \mathrm{supp}(\Lambda)$ in the following way. By definition, $i \in \mathrm{supp}(\Lambda)$ if and only if the system

$$(\ell_1, \ldots, \ell_K) \in V,\ \ell_1 \geq 0,\ \ldots,\ \ell_i > 0,\ \ldots,\ \ell_K \geq 0, \tag{3.3}$$

has an *integer* solution $(\ell_1, \ldots, \ell_K) \in \mathbb{Z}^K$. By the homogeneity of the system (3.3), it has an *integer* solution if and only if it has a *rational* solution. The existence of a rational solution to system (3.3) can be decided by linear programming in polynomial time. Therefore, the support of Λ can be computed in polynomial time by checking whether $i \in \mathrm{supp}(\Lambda)$ for each $i = 1, \ldots, K$. □

3.3.2 The Baker-Campbell-Hausdorff formula

Definition 3.3.5 (Lie algebra $\mathfrak{u}(n)$)**.** The *Lie algebra* $\mathfrak{u}(n)$ is defined as the $\mathbb{Q}$-linear space of $n \times n$ upper triangular rational matrices with *zeros* on the diagonal:

$$\mathfrak{u}(n) \coloneqq \left\{ \begin{pmatrix} 0 & * & \cdots & * & * \\ 0 & 0 & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & * \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \text{ where } * \text{ are entries in } \mathbb{Q} \right\}.$$

It is a linear space of dimension $\frac{1}{2}n(n-1)$. There exist maps

$$\log : \mathsf{UT}(n, \mathbb{Q}) \to \mathfrak{u}(n), \quad A \mapsto \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} (A - I)^k,$$

and

$$\exp : \mathfrak{u}(n) \to \mathsf{UT}(n, \mathbb{Q}), \quad X \mapsto \sum_{k=0}^{n} \frac{1}{k!} X^k,$$

which are inverse of one another. In particular, $\log I = 0$ and $\exp(0) = I$.

The Lie algebra $\mathfrak{u}(n)$ is equipped with a *Lie bracket* $[\cdot, \cdot] : \mathfrak{u}(n) \times \mathfrak{u}(n) \to \mathfrak{u}(n)$ given by $[X, Y] = XY - YX$. The Lie bracket has the following three properties.

(i) Bilinearity:

$$[aX + bY, Z] = a[X, Z] + b[Y, Z], \quad [Z, aX + bY] = a[Z, X] + b[Z, Y]$$

for all scalars $a, b \in \mathbb{Q}$ and all elements $X, Y, Z \in \mathfrak{u}(n)$.

(ii) Anticommutativity:

$$[X, Y] = -[Y, X]$$

for all elements $X, Y \in \mathfrak{u}(n)$.

(iii) The *Jacobi identity*:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

for all elements $X, Y, Z \in \mathfrak{u}(n)$.

Notation 3.3.6 (Logarithm of sets and long Lie brackets)**.** Given a set of matrices $\mathcal{G} \subseteq \mathsf{UT}(n, \mathbb{Q})$, define the set

$$\log \mathcal{G} \coloneqq \{\log A \mid A \in \mathcal{G}\}$$

of logarithms of matrices in $\mathcal{G}$. It is a subset of $\mathfrak{u}(n)$. If G is a subgroup of $\mathsf{UT}(n, \mathbb{Q})$, then by

slight abuse of notation, $\log G$ is similarly defined by considering G as a set.

Given a set of elements $\mathcal{H} \subseteq \mathfrak{u}(n)$ and an integer $k \geq 2$, define the set

$$[\mathcal{H}]_k \coloneqq \left\{ [\dots[[X_1, X_2], X_3], \dots, X_k] \;\middle|\; X_1, X_2, \dots, X_k \in \mathcal{H} \right\}.$$

That is, $[\mathcal{H}]_k$ is the set of all "left bracketing" of length k of elements in $\mathcal{H}$.

It is a standard result [42, p.576] that, using the bilinearity, anticommutativity and the Jacobi identity, any k-iteration of Lie brackets of elements in $\mathcal{H}$ can be written as a linear combination of elements in $[\mathcal{H}]_k$. For example, for $k = 4$, one can write

$$\begin{aligned}[[X_1, X_2], [X_3, X_4]] &= -[[X_2, [X_3, X_4]], X_1] - [[[X_3, X_4], X_1], X_2] \quad \text{(Jacobi identity)} \\ &= [[[X_3, X_4], X_2], X_1] - [[[X_3, X_4], X_1], X_2] \quad \text{(Anticommutativity)}.\end{aligned}$$

The following lemma relates the nilpotency class of G to the integer d such that the set $[\log G]_{d+1}$ vanishes. This result has deep roots in what is called the Mal'cev correspondence between unipotent Lie groups and nilpotent Lie algebras.

Lemma 3.3.7. *Let G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$. If the nilpotency class of G is d, then $[\log G]_{d+1} = \{0\}$.*

Proof. For an element $g \in G$ and a rational number $q \in \mathbb{Q}$, define $g^q \coloneqq \exp(q \log g)$. A group $G \leq \mathsf{UT}(n, \mathbb{Q})$ is called $\mathbb{Q}$-*powered* if for every element $g \in G$ and $q \in \mathbb{Q}$, we have $g^q \in G$. A subgroup of $\mathsf{UT}(n, \mathbb{Q})$ is torsion-free, because $A^n = I \iff n \log A = 0 \iff \log A = 0 \iff A = I$. Therefore, by [56, Theorem 9.20(a)], G can be embedded in a $\mathbb{Q}$-powered group $\widehat{G}$ of the same nilpotency class d.[4] By [56, Theorem 10.3(d)], $\log \widehat{G}$ is a Lie algebra over $\mathbb{Q}$, and $\log \widehat{G}$ is of nilpotency class d (meaning $[\log \widehat{G}]_{d+1} = \{0\}$). Therefore, $[\log G]_{d+1} \subseteq [\log \widehat{G}]_{d+1} = \{0\}$. □

We now introduce the main tool of this chapter: the *Baker-Campbell-Hausdorff formula*.

Theorem 3.3.8 (Baker-Campbell-Hausdorff formula [5, 26, 45])**.** *Let $n \in \mathbb{N}$ and G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most d. Let $B_1, \dots, B_m$ be elements of G. We have*

$$\log(B_1 B_2 \cdots B_m) = \sum_{i=1}^{m} \log B_i + \sum_{k=2}^{d} H_k(\log B_1, \dots, \log B_m), \tag{3.4}$$

where the terms $H_k(\log B_1, \dots, \log B_m), k = 2, 3, \dots$, can be expressed as $\mathbb{Q}$-linear combinations of elements in $[\{\log B_1, \dots, \log B_m\}]_k$.

[4] The smallest such group $\widehat{G}$ is commonly called the *Mal'cev completion* of G.

In theory, one can compute the expressions H_k effectively using recursion (see, for example [28]). An explicit expression for the term H_k was discovered by Dynkin (see Lemma 3.5.4). However, as k grows, these expressions quickly become highly complicated. For example, here are the explicit expressions of the first two terms.

$$\begin{aligned} H_2(C_1,\ldots,C_m) &= \frac{1}{2}\sum_{i<j}[C_i,C_j], \\ H_3(C_1,\ldots,C_m) &= \sum_{i<j<k}\left(\frac{[C_i,[C_j,C_k]]}{3}+\frac{[[C_i,C_k],C_j]}{6}\right)+\frac{1}{12}\sum_{i<j}\left([C_i,[C_i,C_j]]+[[C_i,C_j],C_j]\right). \end{aligned} \tag{3.5}$$

3.4 Invertible Subset

In this section, we construct the algorithm that proves Theorem 3.1.1. In order to describe our algorithm, we need to introduce the following notation. When $\mathcal{H}$ be a finite set of elements in the Lie algebra $\mathfrak{u}(n)$, denote

$$\mathfrak{L}_{\geq k}(\mathcal{H}) \coloneqq \left\langle \bigcup_{i\geq k}[\mathcal{H}]_i \right\rangle_{\mathbb{Q}}.$$

That is, $\mathfrak{L}_{\geq k}(\mathcal{H})$ is the linear space spanned by the set of all "left bracketings" of length at least k of elements in $\mathcal{H}$. By Lemma 3.3.7, if $G \leq \mathsf{UT}(n,\mathbb{Q})$ has nilpotency class d, then for any $\mathcal{H} \subseteq \log G$, we have $\mathfrak{L}_{\geq k}(\mathcal{H}) = \langle[\mathcal{H}]_k\rangle_{\mathbb{Q}} + \langle[\mathcal{H}]_{k+1}\rangle_{\mathbb{Q}} + \cdots + \langle[\mathcal{H}]_d\rangle_{\mathbb{Q}}$, and $\mathfrak{L}_{\geq d+1}(\mathcal{H}) = \{0\}$. We have thus a descending series of linear spaces $\mathfrak{L}_{\geq 1}(\mathcal{H}) \supseteq \mathfrak{L}_{\geq 2}(\mathcal{H}) \supseteq \cdots \supseteq \mathfrak{L}_{\geq d+1}(\mathcal{H}) = \{0\}$ such that $[\mathfrak{L}_{\geq i}(\mathcal{H}), \mathfrak{L}_{\geq j}(\mathcal{H})] \subseteq \mathfrak{L}_{\geq i+j}(\mathcal{H})$.[5]

Example 3.4.1. Let $G = \mathsf{UT}(4,\mathbb{Q})$, so it is of nilpotency class three. Consider the Lie algebra

$$\log G = \mathfrak{u}(4) = \left\{ \begin{pmatrix} 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 \end{pmatrix}, \text{ where } * \text{ are entries in } \mathbb{Q} \right\}.$$

It is a $\mathbb{Q}$-linear space of dimension six. Let $\mathcal{G} = \{A_1, A_2, A_3\}$, where

$$A_1 = \begin{pmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & -1 & -1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

[5] In mathematical terms, the linear spaces $\mathfrak{L}_{\geq 1}(\mathcal{H}) \supseteq \mathfrak{L}_{\geq 2}(\mathcal{H}) \supseteq \cdots \supseteq \mathfrak{L}_{\geq d+1}(\mathcal{H}) = \{0\}$ give the Lie algebra $\mathfrak{L}_{\geq 1}(\mathcal{H})$ the structure of a *filtered Lie algebra* (for an exact definition, see [60]).

Then letting $\mathcal{H} = \{\log A_1, \log A_2, \log A_3\}$, we have

$$\log A_1 = \begin{pmatrix} 0 & 2 & -3 & \frac{11}{3} \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_2 = \begin{pmatrix} 0 & -1 & -\frac{3}{2} & \frac{3}{2} \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_3 = \begin{pmatrix} 0 & 0 & 3 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.6)$$

We will compute $\mathfrak{L}_{\geq k}(\mathcal{H})$ for $k = 1, 2, 3, \ldots$.

First, we have $\mathfrak{L}_{\geq 4}(\mathcal{H}) = \{0\}$. This can be proved either by showing that all length-four Lie brackets vanish, or by directly applying Lemma 3.3.7 for $d = 3$.

Next, we compute $\mathfrak{L}_{\geq 3}(\mathcal{H})$. This is the subspace generated by the set $\mathfrak{L}_{\geq 4}(\mathcal{H}) = \{0\}$ and the set $[\mathcal{H}]_3$ of length-three brackets:

$$[\mathcal{H}]_3 = \Big\{ [[\log A_1, \log A_2], \log A_1], [[\log A_1, \log A_2], \log A_2], \ldots, [[\log A_3, \log A_2], \log A_2] \Big\}.$$

It is easy to show that all length-three left brackets $[[\log A_i, \log A_j], \log A_k]$ are of the form $\begin{pmatrix} 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, * \in \mathbb{Q}$. For example, $[[\log A_1, \log A_2], \log A_1] = \begin{pmatrix} 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq 0.$

Therefore, we have

$$\mathfrak{L}_{\geq 3}(\mathcal{H}) = \left\{ \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \;\middle|\; a \in \mathbb{Q} \right\}. \quad (3.7)$$

It is a dimension one subspace of $\mathfrak{u}(4)$.

Next, we compute $\mathfrak{L}_{\geq 2}(\mathcal{H})$. This is the subspace generated by $\mathfrak{L}_{\geq 3}(\mathcal{H})$ and the set $[\mathcal{H}]_2$ of length-two Lie brackets:

$$[\mathcal{H}]_2 = \{[\log A_1, \log A_2], [\log A_1, \log A_3], [\log A_2, \log A_3]\}.$$

Direct computation shows

$$[\mathcal{H}]_2 = \left\{ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right\}.$$

Hence,

$$\mathfrak{L}_{\geq 2}(\mathcal{H}) = \left\{ \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \;\middle|\; a, b \in \mathbb{Q} \right\}. \quad (3.8)$$

It is a dimension two subspace of $\mathfrak{u}(4)$.

Finally we compute $\mathfrak{L}_{\geq 1}(\mathcal{H})$. This is the subspace generated by $\mathfrak{L}_{\geq 2}(\mathcal{H})$ and all length-one Lie brackets:

$$[\mathcal{H}]_1 = \mathcal{H} = \{\log A_1, \log A_2, \log A_3\}.$$

Direct computation shows

$$\mathfrak{L}_{\geq 1}(\mathcal{H}) = \left\{ \begin{pmatrix} 0 & c_1 & c_2 & a \\ 0 & 0 & c_1 & b \\ 0 & 0 & 0 & c_3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \;\middle|\; a, b, c_1, c_2, c_3 \in \mathbb{Q} \right\}. \tag{3.9}$$

It is a dimension five subspace of $\mathfrak{u}(4)$. This concludes our example.

Now let $G \leq \mathsf{UT}(n, \mathbb{Q})$ be a group of nilpotency class at most ten, and fix $\mathcal{G} = \{A_1, \ldots, A_K\}$ to be a finite alphabet of elements in G. For any vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}_{\geq 0}^K$, define

$$\mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})} \coloneqq \{A_i \mid A_i \in \mathcal{G}, i \in \mathrm{supp}(\boldsymbol{\ell})\}$$

as the set of matrices in $\mathcal{G}$ whose index appears in $\mathrm{supp}(\boldsymbol{\ell})$.

We now give the key ingredient of an algorithm that computes the invertible subset of $\mathcal{G}$. For a word $w \in \mathcal{G}^*$, we naturally denote by $\log w$ the logarithm of the element represented by w. The key ingredient is the following (highly non-trivial) Theorem 3.4.2, which provides a linear algebra criterion for the existence of a non-empty word $w \in \mathcal{G}^*$ satisfying $\log w = 0$. Note that $\log w = 0$ if and only if w represents the neutral element.

Theorem 3.4.2 (Structural theorem of unitriangular matrix semigroups)**.** *Let $\mathcal{G} = \{A_1, \ldots, A_K\}$ be a finite set of matrices in $\mathsf{UT}(n, \mathbb{Q})$ that satisfies $[\log \mathcal{G}]_{11} = \{0\}$. Given a non-zero vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}_{\geq 0}^K$:*

(i) If there exists a word $w \in \mathcal{G}^*$ with $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$ and $\log w = 0$, then

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}). \tag{3.10}$$

(ii) If $\boldsymbol{\ell}$ satisfies (3.10), then there exists a non-empty word $w \in \mathcal{G}^*$, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$, such that $\log w = 0$.

Part (i) of Theorem 3.4.2 is relatively easy to prove:

Proof of part (i) of Theorem 3.4.2. Let w be a word with $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$. Write $w = B_1 B_2 \cdots B_m$, $B_i \in \mathcal{G}, i = 1, \ldots, m$. Regrouping by letters, we have $\sum_{i=1}^{K} \ell_i \log A_i = \sum_{i=1}^{m} \log B_i$.

If $\log w = 0$, then by the Baker-Campbell-Hausdorff formula (3.4), we have

$$\sum_{i=1}^{m} \log B_i + \sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m) = \log(B_1 B_2 \cdots B_m) = 0.$$

The higher order terms $H_k, k \geq n$ vanish because $[\log \mathcal{G}]_n = \{0\}$ (this is a consequence of $\mathcal{G} \subseteq \mathsf{UT}(n, \mathbb{Q})$). Therefore, $\sum_{i=1}^{K} \ell_i \log A_i = \sum_{i=1}^{m} \log B_i = -\sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m)$.

Since the Parikh Image of the word $B_1 B_2 \cdots B_m$ is $\boldsymbol{\ell}$, the matrices B_i all lie in the set $\mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}$. Therefore, $\log B_i \in \log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}$ for all i. By Theorem 3.3.8, for all $k \geq 2$ we have $-H_k(\log B_1, \ldots, \log B_m) \in \left\langle [\{\log B_i \mid i = 1, \ldots, m\}]_k \right\rangle_{\mathbb{Q}} \subseteq \mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) \subseteq \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. Therefore, we have $\sum_{i=1}^{K} \ell_i \log A_i = -\sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m) \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. □

Part (ii) of Theorem 3.4.2 is highly non-trivial and will be the focus of Section 3.5. We continue Example 3.4.1 to give an intuition of the Condition (3.10) in Theorem 3.4.2.

Example 3.4.1 (continued)**.** Let $\mathcal{G}$ be as in Example 3.4.1. As an example, we show that $\boldsymbol{\ell} = (1, 2, 2)$ satisfies Condition (3.10). In this case, we have $\mathrm{supp}(\boldsymbol{\ell}) = \{1, 2, 3\}$, so (3.8) yields

$$\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) = \left\{ \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \;\middle|\; a, b \in \mathbb{Q} \right\}.$$

Therefore,

$$\sum_{i=1}^{3} \ell_i \log A_i = \log A_1 + 2 \log A_2 + 2 \log A_3 = \begin{pmatrix} 0 & 0 & 0 & \frac{23}{3} \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}),$$

where $\log A_1, \log A_2, \log A_3$ are computed in (3.6). Hence, $\boldsymbol{\ell}$ satisfies Condition (3.10).

Therefore, Theorem 3.4.2 claims that there exists a non-empty word $w \in \mathcal{G}^*$, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot (1, 2, 2)$, such that $\log w = 0$. Indeed, in the beginning of Section 3.5 we will show how to construct this word w. This will serve as an intuition of the proof in the general case. This concludes our example for now.

One can see that the proof of part (i) did not use the condition $[\log \mathcal{G}]_{11} = \{0\}$, hence part (i) would still hold without this condition. However, the condition $[\log \mathcal{G}]_{11} = \{0\}$ will be needed in the proof of part (ii).

Note that finding solutions of Equation (3.10) only relies on linear algebra. Assuming the structural Theorem 3.4.2, we can devise the following Algorithm 3.1 that computes the invertible

subset of any finite set $\mathcal{G} \subseteq G$.

Algorithm 3.1 Computing the invertible subset of $\mathcal{G}$

Input: A finite set of elements $\mathcal{G} = \{A_1, \ldots, A_K\}$ in G.

Output: The invertible subset $\mathcal{G}_{inv}$ of $\mathcal{G}$.

1 **Initialization.** Set $S \coloneqq \{1, \ldots, K\}$.

2 **Main loop.** Repeat the following

(a) Represent the $\mathbb{Q}$-linear subspace of $\mathbb{Q}^K$:

$$V \coloneqq \left\{ (\ell_1, \ldots, \ell_K) \in \mathbb{Q}^K \middle| \sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in S\}) \right\}$$

as the solution set of homogeneous linear equations.

(b) Define $\Lambda \coloneqq \mathbb{Z}^K_{\geq 0} \cap V$ and compute $\mathrm{supp}(\Lambda)$ using Lemma 3.3.4.

(c) If $\mathrm{supp}(\Lambda) = S$, terminate the algorithm and return $\mathcal{G}_{inv} = \{A_i \mid i \in S\}$. Otherwise let $S \coloneqq \mathrm{supp}(\Lambda)$ and continue.

Proof of Theorem 3.1.1 and correctness of Algorithm 3.1 (assuming Theorem 3.4.2). After each iteration of Step 2, the sets $\mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in S\})$ and V do not increase, so the set $S = \mathrm{supp}(\Lambda)$ strictly decreases. Therefore, the algorithm terminates after at most K iterations of Step 2.

Since G has nilpotency class at most ten, by Lemma 3.3.7, its subset $\mathcal{G}$ satisfies $[\log \mathcal{G}]_{11} = \{0\}$. For the correctness of the algorithm, we start by showing that, when the algorithm terminates, every element of $\{A_i \mid i \in S\}$ has an inverse in the semigroup $\langle \mathcal{G} \rangle$. When the algorithm terminates at Step 2(c), we have $\mathrm{supp}(\Lambda) = S$. By the additivity of Λ (that is, $\boldsymbol{a}, \boldsymbol{b} \in \Lambda \implies \boldsymbol{a} + \boldsymbol{b} \in \Lambda$), there exists a vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \Lambda$ such that $\mathrm{supp}(\boldsymbol{\ell}) = \mathrm{supp}(\Lambda) = S$. This vector then satisfies

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in \mathrm{supp}(\boldsymbol{\ell})\})$$

by the definition of V. By Theorem 3.4.2(ii), this shows that there exists a non-empty word w, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$ such that $\log w = 0$. This word w represents I. For any $i \in S$, since $\mathrm{supp}(\boldsymbol{\ell}) = S$, the letter A_i appears in the word w. Write $w = w_1 A_i w_2$, then $w_1 A_i w_2$ represents I. Therefore, A_i^{-1} is represented by the word $w_2 w_1 \in \mathcal{G}^*$, so $A_i^{-1} \in \langle \mathcal{G} \rangle$.

We then show that for every matrix A_i invertible in $\langle \mathcal{G} \rangle$, i is in the set S at the termination of the algorithm. Suppose A_i^{-1} is represented by a non-empty word $w \in \mathcal{G}^*$. Then the word $w' = wA_i$ represents I; that is, $\log w' = 0$. By Theorem 3.4.2(i), the Parikh Image $\boldsymbol{\ell} = \mathrm{PI}^{\mathcal{G}}(w')$ satisfies $\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in \mathrm{supp}(\boldsymbol{\ell})\})$. We show that $\mathrm{supp}(\boldsymbol{\ell}) \subseteq S$ is an invariant of the algorithm.

At initialization, we obviously have $\mathrm{supp}(\boldsymbol{\ell}) \subseteq S$. Before each iteration of Step 2(b), suppose we have $\mathrm{supp}(\boldsymbol{\ell}) \subseteq S$, then

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in \mathrm{supp}(\boldsymbol{\ell})\}) \subseteq \mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in S\}).$$

Hence $\boldsymbol{\ell} \in \Lambda = \mathbb{Z}^K_{\geq 0} \cap V$. Consequently, $\mathrm{supp}(\boldsymbol{\ell}) \subseteq \mathrm{supp}(\Lambda)$ at the beginning of Step 2(c), which shows that $\mathrm{supp}(\boldsymbol{\ell}) \subseteq S$ still holds after the iteration of Step 2. This invariant shows that $i \in \mathrm{supp}(\boldsymbol{\ell}) \subseteq S$ by the end of the algorithm. Combining with the previous implication, we conclude that by the end of the algorithm, S is exactly the set of elements in $\mathcal{G}$ with inverse in the semigroup $\langle \mathcal{G} \rangle$.

For the complexity analysis, recall that the algorithm terminates after at most K iterations of 2. At each iteration of Step 2(b), the support $\mathrm{supp}(\Lambda)$ can be computed in polynomial time by Lemma 3.3.4. The total input size of these linear programming instances is polynomial with respect to the total bit length of the matrix entries in $\mathcal{G}$. Indeed, a $\mathbb{Q}$-basis of $\mathfrak{L}_{\geq 2}(\{\log A_i \mid i \in S\})$ is simply the set $\bigcup_{10 \geq k \geq 2} [\{\log A_i \mid i \in S\}]_k$, whose total bit length is of polynomial size in $\mathcal{G}$. From this, one can express V as the solution set of a system of homogeneous linear equations whose total bit length is polynomial in $\mathcal{G}$ (note that the total bit length of $\log A_i$ is also polynomial in $\mathcal{G}$). Therefore, the overall complexity of Algorithm 3.1 is polynomial with respect to the input $\mathcal{G}$. □

3.5 Structural theorem of unitriangular matrix semigroups

In this section we give the proof of Theorem 3.4.2(ii):

Theorem 3.4.2 (Structural theorem of unitriangular matrix semigroups)**.** *Let $\mathcal{G} = \{A_1, \ldots, A_K\}$ be a finite set of matrices in $\mathsf{UT}(n, \mathbb{Q})$ that satisfies $[\log \mathcal{G}]_{11} = \{0\}$. Given a non-zero vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}^K_{\geq 0}$:*

(i) If there exists a word $w \in \mathcal{G}^*$ with $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$ and $\log w = 0$, then

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}). \tag{3.10}$$

(ii) If $\boldsymbol{\ell}$ satisfies (3.10), then there exists a non-empty word $w \in \mathcal{G}^*$, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$, such that $\log w = 0$.

We first give an intuition of the proof by continuing Example 3.4.1.

Example 3.4.1 (final part)**.** Let $\mathcal{G} = \{A_1, A_2, A_3\}$ be as in Example 3.4.1:

$$A_1 = \begin{pmatrix} 1 & 2 & -1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & -1 & -1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

$$\log A_1 = \begin{pmatrix} 0 & 2 & -3 & \frac{11}{3} \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_2 = \begin{pmatrix} 0 & -1 & -\frac{3}{2} & \frac{3}{2} \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_3 = \begin{pmatrix} 0 & 0 & 3 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let $\boldsymbol{\ell} = (1, 2, 2)$. We have already shown that $\boldsymbol{\ell}$ satisfies Condition (3.10), so Theorem 3.4.2(ii) claims that there exists a non-empty word $w \in \mathcal{G}^*$, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot (1, 2, 2)$, such that $\log w = 0$. We illustrate here how to find this word w.

Step 1. We find elements A_1', A_2', A_3' in $\mathcal{G}^*$, such that $\log A_1', \log A_2', \log A_3'$ generate the subspace

$$\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) = \left\{ \begin{pmatrix} 0 & 0 & 0 & a \\ 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \;\middle|\; a, b \in \mathbb{Q} \right\}$$

as a $\mathbb{Q}_{\geq 0}$-*cone*.

The idea is to take

$$A_1' \coloneqq A_1^t A_2^{2t} A_3^{2t}, \quad A_2' \coloneqq A_2^{2t} A_3^{2t} A_1^t, \quad A_3' \coloneqq A_2^{2t} A_1^t A_3^{2t}, \tag{3.11}$$

for a suitable $t \in \mathbb{N}$. Apply the Baker-Campbell-Hausdorff formula (3.4)

$$\log(B_1 B_2 \cdots B_m) = \sum_{i=1}^{m} \log B_i + \sum_{k=2}^{d} H_k(\log B_1, \ldots, \log B_m)$$

with $m = 3, B_1 \coloneqq A_1^t, B_2 \coloneqq A_2^{2t}, B_3 \coloneqq A_3^{2t}$, we obtain

$$\begin{aligned} \log A_1' &= \log(A_1^t A_2^{2t} A_3^{2t}) \\ &= \log A_1^t + \log A_2^{2t} + \log A_3^{2t} + \sum_{k=2}^{3} H_k(\log A_1^t, \log A_2^{2t}, \log A_3^{2t}) \\ &= t \cdot (\log A_1 + 2 \log A_2 + 2 \log A_3) + \sum_{k=2}^{3} t^k \cdot H_k(\log A_1, 2 \log A_2, 2 \log A_3). \end{aligned} \tag{3.12}$$

The last equality is due to $\log A^t = t \log A$ and because the term H_k is a linear combination of k-iterations of Lie brackets: for example, $[t \log B_1, t \log B_2] = t^2 \cdot [\log B_1, \log B_2]$ and $[[t \log B_1, t \log B_2], t \log B_3] = t^3 \cdot [[\log B_1, \log B_2], \log B_3]$.

The linear term $t\cdot(\log A_1+2\log A_2+2\log A_3)$ in (3.12) falls in the subspace $\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ by Condition (3.10). The non-linear terms $t^k\cdot H_k(\log A_1, 2\log A_2, 2\log A_3), k=2,3$, also fall in the subspace $\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ by Theorem 3.3.8. Therefore, we have $\log A_1' \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. Similarly, $\log A_2'$ and $\log A_3'$ are also in $\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$.

Using the exact expression (3.5) for the terms H_2 and H_3, we obtain

$$\log A_1' = \begin{pmatrix} 0 & 0 & 0 & \frac{4}{3}t^3+\frac{23}{3}t \\ 0 & 0 & 0 & 2t^2-t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_2' = \begin{pmatrix} 0 & 0 & 0 & -\frac{8}{3}t^3+2t^2+\frac{23}{3}t \\ 0 & 0 & 0 & 2t^2-t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\log A_3' = \begin{pmatrix} 0 & 0 & 0 & \frac{4}{3}t^3+\frac{23}{3}t \\ 0 & 0 & 0 & -2t^2-t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We then choose $t=10$. This choice is made so that t is large enough for $\log A_1', \log A_2', \log A_3'$ to exhibit their "asymptotic" behaviour. When $t=10$, we have

$$\log A_1' = \begin{pmatrix} 0 & 0 & 0 & 1410 \\ 0 & 0 & 0 & 190 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_2' = \begin{pmatrix} 0 & 0 & 0 & -2390 \\ 0 & 0 & 0 & 190 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \log A_3' = \begin{pmatrix} 0 & 0 & 0 & 1410 \\ 0 & 0 & 0 & -210 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.13}$$

Then indeed we have

$$\langle \log A_1', \log A_2', \log A_3' \rangle_{\mathbb{Q}_{\geq 0}} = \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$$

by direct computation. Furthermore, the Parikh Images are

$$\mathrm{PI}^{\mathcal{G}}(A_1') = \mathrm{PI}^{\mathcal{G}}(A_2') = \mathrm{PI}^{\mathcal{G}}(A_3') = (10, 20, 20).$$

Step 2. Consider the new alphabet $\mathcal{G}' \coloneqq \{A_1', A_2', A_3'\}$. We now find a non-empty word $A'' \in (\mathcal{G}')^*$, such that $\log A'' \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) = \{0\}$.

By direct computation from (3.13), we obtain

$$117\cdot \log A_1' + 282\cdot \log A_2' + 361\cdot \log A_3' = 0. \tag{3.14}$$

Let

$$A'' \coloneqq \left(A_1'\right)^{117} \cdot \left(A_2'\right)^{282} \cdot \left(A_3'\right)^{361}.$$

By the Baker-Campbell-Hausdorff formula (3.4), we have

$$\log A'' = 117 \cdot \log A'_1 + 282 \cdot \log A'_2 + 361 \cdot \log A'_3 = 0.$$

This is because all the terms $H_k, k \geq 2$ in (3.4) are in

$$\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) \subseteq \mathfrak{L}_{\geq 4}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) = \{0\}. \tag{3.15}$$

Furthermore, the Parikh Image is

$$\mathrm{PI}^{\mathcal{G}}(A'') = 117 \cdot \mathrm{PI}^{\mathcal{G}}(A'_1) + 282 \cdot \mathrm{PI}^{\mathcal{G}}(A'_2) + 361 \cdot \mathrm{PI}^{\mathcal{G}}(A'_3) = 7600 \cdot (1, 2, 2).$$

We have thus found the word $w = A''$ satisfying $\log w = 0$, with Parikh Image $7600 \cdot (1, 2, 2)$. This concludes the example.

The following subsections aim to formalize the idea exhibited in this example and provide a rigorous proof of Theorem 3.4.2(ii). The main difficulties of formalizing a proof in the general case are the following.

(i) In Equation (3.11) we took a specific choice of A'_1, A'_2, A'_3. In the general case, we will use a similar idea of taking A'_i to be words of the form $A^t_{i_1} A^t_{i_2} \cdots A^t_{i_m}$. However, the permutations $(i_1, i_2, \ldots, i_m)$ need to be chosen carefully. We need to show that there exist enough permutations so that the constructed elements $\log A'_1, \log A'_2, \ldots,$ generate the linear space $\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. We achieve this by proving a deep combinatorial property of the terms H_k (Proposition 3.5.1).

(ii) Furthermore, $\log A'_1, \log A'_2, \ldots,$ need to generate $\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ as a *cone*. The coefficients $(117, 282, 361)$ obtained in Equation (3.14) happen to be positive, but this is *a priori* not always true. We need to show that 0 can always be written as a *positive* combination of $\log A'_i$. This is proved by finding identities over the terms H_k using computer assistance (Proposition 3.5.2).

(iii) The exponent t in Equation (3.11) needs to be chosen carefully. In fact, we may need to take several different t. Such t are chosen using techniques from convex geometry (Proposition 3.5.3).

(iv) In the above example the nilpotency class of G is three. This is the reason why in Step 2, Equation (3.15) holds, and the matrices A'_1, A'_2, A'_3 commute with each other. In the general case, we deal with groups of nilpotency class up to ten. Then, Equation (3.15) no longer holds. Hence, we need to repeat the above process for more steps. In general, when G has nilpotency class up to $2^d - 1$, we need to repeat the process for d steps. This corresponds to the derived

length (see Definition 2.1.9) of the group G.

3.5.1 Overview of three technical propositions

The formal proof of Theorem 3.4.2(ii) relies on the three following technical propositions, which address respectively the difficulties (i)-(iii) stated above.

For $k \in \mathbb{Z}_{>0}$, denote by S_k the permutation group of the set $\{1, \dots, k\}$. Theorem 3.3.8 states that the term $H_k(\log B_1, \dots, \log B_m)$ is written as a linear combination of k-iterated Lie brackets $[\dots[[\log B_{i_1}, \log B_{i_2}], \log B_{i_3}], \dots, \log B_{i_k}]$. Our first technical proposition shows that a converse of it is true: that for any $k \geq 2$, the k-iterated Lie bracket $[\dots[[\log B_1, \log B_2], \log B_3], \dots, \log B_k]$ can be written as a linear combination of expressions in H_k.

Proposition 3.5.1. *For every $k \geq 2$, there exists a function $\mu_k \colon \mathrm{S}_k \to \mathbb{Z}$, such that for any sequence of elements $C_1, \dots, C_m, m \geq k$, in the Lie algebra $\mathfrak{u}(n)$ we have*

$$[\dots[[C_1, C_2], C_3], \dots, C_k] = \sum_{\sigma \in \mathrm{S}_k} \mu_k(\sigma) H_k(C_{\sigma(1)}, \dots, C_{\sigma(k)}, C_{k+1}, \dots, C_m). \tag{3.16}$$

Our second technical proposition shows that for $k \leq 10$, one can find a linear combination with *positive* coefficients of the terms H_k that lies in $\mathfrak{L}_{\geq k+1} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\cdot))$. (Note that *a priori* H_k lies in $\mathfrak{L}_{\geq k}(\cdot)$.)

Proposition 3.5.2. *Let $k \leq 10$ and let $\mathcal{H} \subset \mathsf{UT}(n, \mathbb{Q})$ be a finite set of matrices for some $n \geq 2$. Then there exist a non-negative integer r, positive rational numbers $\alpha_1, \dots, \alpha_r$, as well as, for $s = 1, \dots, r$, words $\boldsymbol{j}_s = j_{s,1} j_{s,2} \cdots j_{s,m_s}$ in the alphabet $\mathcal{I} = \{1, 2, \dots, k+1\}$, such that $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) \in \{(1, \dots, 1), (2, \dots, 2)\}$ and*

$$\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \dots, \log B_{\sigma(k+1)}) + \sum_{s=1}^{r} \alpha_s \sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(j_{s,1})}, \dots, \log B_{\sigma(j_{s,m_s})}) \\ \in \mathfrak{L}_{\geq k+1}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})) \tag{3.17}$$

for all matrices $B_1, \dots, B_{k+1}$ in $\mathsf{UT}(n, \mathbb{Q})$ satisfying $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$.

Our third technical proposition concerns convex geometry and will be responsible for finding a suitable t from difficulty (ii).

Proposition 3.5.3. *Let V be a finite dimensional $\mathbb{Q}$-linear space. Let d be a positive integer, $\mathcal{I}$ be a finite index set, and $\boldsymbol{a}_{1i}, \dots, \boldsymbol{a}_{di}, i \in \mathcal{I}$ be vectors in V.*

For any $t \in \mathbb{Z}_{>0}$, define the vectors

$$P_i(t) \coloneqq t\boldsymbol{a}_{1i} + \cdots + t^d \boldsymbol{a}_{di}, \quad \text{for } i \in \mathcal{I}.$$

Suppose the following two conditions hold:

(i) The $\mathbb{Q}_{\geq 0}$-cone $\mathcal{C}_d \coloneqq \langle \boldsymbol{a}_{di} \mid i \in \mathcal{I} \rangle_{\mathbb{Q}_{\geq 0}}$ is a linear space.

(ii) For $k = d-1, d-2, \ldots, 1$, the inductively defined $\mathbb{Q}_{\geq 0}$-cones $\mathcal{C}_k \coloneqq \langle \boldsymbol{a}_{ki} \mid i \in \mathcal{I} \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1}$ are linear spaces.

Then the $\mathbb{Q}_{\geq 0}$-cone $\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0} \rangle_{\mathbb{Q}_{\geq 0}}$ is equal to $\mathcal{C}_1$.

Proposition 3.5.2 is the only one among the three technical propositions that is limited by the nilpotency class. This constitutes the main obstacle to generalizing Theorem 3.4.2 to higher nilpotency classes.

3.5.2 Proof of Proposition 3.5.1

For a permutation $\sigma \in \mathrm{S}_k$, define $d(\sigma)$ to be the number of *descents* in σ, that is, the number of $i \in \{1, \ldots, k-1\}$ such that $\sigma(i) > \sigma(i+1)$. In order to prove Proposition 3.5.1, we need an explicit expression for the terms H_k. This expression is provided by Dynkin[6]:

Lemma 3.5.4 (Dynkin formula [35], [65, Proposition 3.4 and Proposition 4.2])**.** *We have*

$$H_k(C_1, \ldots, C_m) = \sum_{i_1 + \cdots + i_m = k} \frac{1}{i_1! \ldots i_m!} \varphi_k(\underbrace{C_1, \ldots, C_1}_{i_1}, \underbrace{C_2, \ldots, C_2}_{i_2}, \ldots, \underbrace{C_m, \ldots, C_m}_{i_m}), \tag{3.18}$$

where the indices $i_1, \ldots, i_m$ are non-negative integers, and

$$\varphi_k(X_1, \ldots, X_k) = \sum_{\sigma \in \mathrm{S}_k} \frac{(-1)^{d(\sigma)}}{k^2 \binom{k-1}{d(\sigma)}} [\ldots [[X_{\sigma(1)}, X_{\sigma(2)}], X_{\sigma(3)}], \ldots, X_{\sigma(k)}]. \tag{3.19}$$

Define recursively the following maps $\mu_k : \mathrm{S}_k \to \mathbb{Z}, k = 2, 3, \ldots$. For $k = 2$, let $\mu_2(\mathrm{id}) = 1, \mu_2((12)) = -1$, where id is the constant permutation and (12) is the permutation that swaps 1 and 2. For $k \geq 3$, denote by $(j_1 j_2 \cdots j_m)$ the cyclic permutation that sends j_i to j_{i+1}, $i = 1, \ldots, m-1$, and sends j_m to j_1. Suppose μ_{k-1} is already defined, we then define

$$\mu_k(\sigma) \coloneqq \begin{cases} \mu_{k-1}(\sigma) & k = \sigma(k) \\ -\mu_{k-1}(\sigma \circ (12 \cdots k)) & k = \sigma(1) \\ 0 & k = \sigma(i), i = 2, \ldots, k-1. \end{cases} \tag{3.20}$$

[6]Dynkin originally only proved the bivariate case of Lemma 3.5.4. It was later been generalized to the multivariate case without much difficulty.

In the first two cases, the permutation σ or $\sigma \circ (12\cdots k)$ fixes k, so they can be considered as elements in S_{k-1}, hence $\mu_{k-1}(\sigma)$ is well defined. For example, $\mu_3(\sigma) = 1$ when $\sigma = \mathrm{id}$ or (13); $\mu_3(\sigma) = -1$ when $\sigma = (12)$ or (132); and $\mu_3(\sigma) = 0$ otherwise. We will show that, for this μ_k, the Equation (3.16) in Proposition 3.5.1 is satisfied:

Proposition 3.5.1. *For every $k \geq 2$, there exists a function $\mu_k \colon \mathrm{S}_k \to \mathbb{Z}$, such that for any sequence of elements $C_1, \ldots, C_m, m \geq k$, in the Lie algebra $\mathfrak{u}(n)$ we have*

$$[\ldots[[C_1, C_2], C_3], \ldots, C_k] = \sum_{\sigma \in \mathrm{S}_k} \mu_k(\sigma) H_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}, C_{k+1}, \ldots, C_m). \tag{3.16}$$

Proof. Take μ_k to be the function defined recursively in (3.20). For every $j \geq 2$, there is a natural embedding $f_j : \mathrm{S}_j \hookrightarrow \mathrm{S}_{j+1}$, defined by $f_j(\sigma)(i) = \sigma(i), i = 1, \ldots, j$, $f_j(\sigma)(j+1) = j+1$. It is easy to verify that under this natural embedding, μ_j and μ_{j+1} are identified, that is, $\mu_j = \mu_{j+1} \circ f_j$. Therefore, we can denote by μ the map $\cup_{k\geq 2} \mathrm{S}_k \to \mathbb{Z}$ as $\mu(\sigma) = \mu_k(\sigma)$, where $\sigma \in \mathrm{S}_k$. We prove Equation (3.16) in three steps.

(1) First, we simplify the right hand side of Equation (3.16) by showing

$$\sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) H_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}, C_{k+1}, \ldots, C_m) = \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) \varphi_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}), \tag{3.21}$$

where φ_k is defined in Lemma 3.5.4.

Thanks to Lemma 3.5.4, $H_k(C_1, \ldots, C_m)$ can be written as

$$H_k(C_1, \ldots, C_m) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq m} \varphi_k(C_{j_1}, \ldots, C_{j_k}) + \sum_{l=2}^{k-1} \sum_{1 \leq j_1 < j_2 < \cdots < j_l \leq m} H_{k,l}(C_{j_1}, \ldots, C_{j_l}), \tag{3.22}$$

where $H_{k,l}(C_{j_1}, \ldots, C_{j_l})$ is some linear combination of elements in $[\{C_{j_1}, \ldots, C_{j_l}\}]_k$. By abuse of notation, for $\sigma \in \mathrm{S}_k$ and $x > k$, we define $\sigma(x) = \sigma^{-1}(x) = x$. For any $l = 2, \ldots, k-1$, we have

$$\sum_{\sigma \in \mathrm{S}_k} \sum_{1 \leq j_1 < j_2 < \cdots < j_l \leq m} \mu(\sigma) H_{k,l}(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_l)}) = \sum_{\substack{t_1, t_2, \ldots, t_l \in \{1, \ldots, m\} \\ \text{pairwise distinct}}} H_{k,l}(C_{t_1}, \ldots, C_{t_l}) \sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma). \tag{3.23}$$

We claim that, for any pairwise distinct $t_1, t_2, \ldots, t_l \in \{1, \ldots, m\}$, $l < k$, we have

$$\sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma) = 0. \tag{3.24}$$

We show (3.24) by induction on k. When $k = 2$, by the definition of μ, (3.24) holds. Suppose (3.24) holds for $k-1$. Denote by c the cyclic permutation $(12 \cdots k)$, then by the recursive definition of μ,

$$\sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma) = \sum_{\substack{\sigma \in \mathrm{S}_{k-1} \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma) - \sum_{\substack{\sigma \in \mathrm{S}_{k-1} \\ c\circ\sigma^{-1}(t_1) < \cdots < c\circ\sigma^{-1}(t_l)}} \mu(\sigma). \tag{3.25}$$

Without loss of generality, suppose the sum on the left hand side is not empty. That is, there exists at least one permutation $\sigma \in \mathrm{S}_k$ such that $\sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)$. Since $\sigma^{-1} \in \mathrm{S}_k$ does not permute any t_j with $t_j > k$, the elements of $\{t_1, \ldots, t_l\}$ which are larger than k must appear after the elements which are smaller or equal to k, and must appear in increasing order. In other words, there exists some $s \geq 1$, such that $t_i \leq k$ for all $i < s$, and $k < t_s < \cdots < t_l$. (s could be $l+1$, in which case $t_i \leq k$ for all $i = 1, \ldots, l$.) Since $\sigma \in \mathrm{S}_k$ does not change the value of $t_s, \cdots, t_l$, one can discard them without changing the sum. Hence, we suppose without loss of generality $t_1, \ldots, t_l \in \{1, \ldots, k\}$.

(a) **If $t_i = k$ for some $i = 2, \ldots, l-1$.** Then no permutation $\sigma \in \mathrm{S}_{k-1}$ can satisfy $\sigma^{-1}(t_1) < \sigma^{-1}(t_i) = k < \sigma^{-1}(t_l)$ or $c \circ \sigma^{-1}(t_1) < c \circ \sigma^{-1}(t_i) = 1 < c \circ \sigma^{-1}(t_l)$. Hence, both sums on the right hand side of Equation (3.25) are empty. The claim (3.24) follows.

(b) **If $t_1 = k$.** Then no permutation $\sigma \in \mathrm{S}_{k-1}$ can satisfy $\sigma^{-1}(t_1) < \sigma^{-1}(t_i) = k < \sigma^{-1}(t_l)$, so the first sum on the right hand side of Equation (3.25) is empty. As for the second sum, because $c \circ \sigma^{-1}(t_1) = c(k) = 1$, we have $c \circ \sigma^{-1}(t_1) < \cdots < c \circ \sigma^{-1}(t_l)$ if and only if $\sigma^{-1}(t_2) < \cdots < \sigma^{-1}(t_l)$. Hence, using the induction hypothesis on $t_2, \ldots, t_l \in \{1, \ldots, k\}$ yields

$$\sum_{\substack{\sigma \in \mathrm{S}_{k-1} \\ c\circ\sigma^{-1}(t_1) < \cdots < c\circ\sigma^{-1}(t_l)}} \mu(\sigma) = \sum_{\substack{\sigma \in \mathrm{S}_{k-1} \\ \sigma^{-1}(t_2) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma) = 0.$$

Therefore both sums on the right hand side of Equation (3.25) equal zero. The claim (3.24) follows.

(c) **If $t_l = k$.** Similar to the previous case, the second sum on the right hand side

of Equation (3.25) is empty. As for the first sum, because $\sigma^{-1}(t_l) = k$, we have $\sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)$ if and only if $\sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_{l-1})$. Hence, using the induction hypothesis on $t_1, \ldots, t_{l-1} \in \{1, \ldots, k\}$ shows the sum is zero. The claim (3.24) follows.

(d) **If $t_i \neq k$ for all $i = 1, \ldots, l$.** Then $\sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)$ if and only if $c \circ \sigma^{-1}(t_1) < \cdots < c \circ \sigma^{-1}(t_l)$. Hence, the two sums on the right hand side of Equation (3.25) are the same. The claim (3.24) follows.

Using the claim (3.24) on Equation (3.23) yields

$$\sum_{\sigma \in \mathrm{S}_k} \sum_{1 \leq j_1 < j_2 < \cdots < j_l \leq m} \mu(\sigma) H_{k,l}(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_l)}) = 0, \tag{3.26}$$

and this combined with Equation (3.22) yields

$$\begin{aligned}
&\sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) H_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}, C_{k+1}, \ldots, C_m) \\
&= \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) H_k(C_{\sigma(1)}, \ldots, C_{\sigma(m)}) \quad (\text{define } \sigma(s) = s \text{ for } \sigma \in \mathrm{S}_k \text{ and } s > k) \\
&= \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq m} \varphi_k(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_k)}) \quad (\text{by (3.22) and (3.26)}) \\
&= \sum_{\substack{t_1, t_2, \ldots, t_k \in \{1, \ldots, m\} \\ \text{pairwise distinct}}} \varphi_k(C_{t_1}, \ldots, C_{t_k}) \sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_k)}} \mu(\sigma) \\
&= \sum_{l=0}^{k} \sum_{\substack{t_1, t_2, \ldots, t_l \in \{1, \ldots, k\} \\ \text{pairwise distinct}, \\ k < t_{l+1} < \cdots < t_k \leq m}} \varphi_k(C_{t_1}, \ldots, C_{t_k}) \sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma)
\end{aligned} \tag{3.27}$$

Because Equation (3.24) holds for $l < k$, that is, the sum $\sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_l)}} \mu(\sigma)$ vanishes whenever $l < k$, the above expression (3.27) is equal to

$$\sum_{\substack{t_1, t_2, \ldots, t_k \in \{1, \ldots, k\} \\ \text{pairwise distinct}}} \varphi_k(C_{t_1}, \ldots, C_{t_k}) \sum_{\substack{\sigma \in \mathrm{S}_k \\ \sigma^{-1}(t_1) < \cdots < \sigma^{-1}(t_k)}} \mu(\sigma) = \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) \varphi_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}).$$

We have hence shown Equation (3.21):

$$\sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) H_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}, C_{k+1}, \ldots, C_m) = \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) \varphi_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}).$$

(2) The second step is to show

$$\sum_{\sigma\in \mathrm{S}_k} \mu(\sigma)\varphi_k(C_{\sigma(1)},\dots,C_{\sigma(k)}) = \sum_{T\in \mathrm{S}_k} \frac{\mu(T)}{k}[\dots[[C_{T(1)},C_{T(2)}],C_{T(3)}],\dots,C_{T(k)}]. \tag{3.28}$$

Using the exact expression for φ_k in Lemma 3.5.4, we have

$$\begin{aligned}
&\sum_{\sigma\in \mathrm{S}_k} \mu(\sigma)\varphi_k(C_{\sigma(1)},\dots,C_{\sigma(k)}) \\
&= \sum_{\sigma\in \mathrm{S}_k}\sum_{\tau\in \mathrm{S}_k} \frac{(-1)^{d(\tau)}\mu(\sigma)}{k^2\binom{k-1}{d(\tau)}}[\dots[[C_{\sigma\circ\tau(1)},C_{\sigma\circ\tau(2)}],C_{\sigma\circ\tau(3)}],\dots,C_{\sigma\circ\tau(k)}] \\
&= \sum_{T\in \mathrm{S}_k} [\dots[[C_{T(1)},C_{T(2)}],C_{T(3)}],\dots,C_{T(k)}] \sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}}
\end{aligned} \tag{3.29}$$

We will compute the value of $\sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}}$ depending on the permutation T. We show by induction on k that

$$\sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}} = \frac{\mu(T)}{k}. \tag{3.30}$$

When $k=2$, by direct computation, $\sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}}$ is equal to $\frac{1}{2}$ if $T=\mathrm{id}$ and to $-\frac{1}{2}$ if $T=(12)$. This matches the values of $\frac{\mu(T)}{k}$. If $k\geq 3$, suppose (3.30) is proven for $k-1$. Again denote by c the cyclic permutation $(12\cdots k)$, by the recursive definition of μ we have

$$\sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}} = \sum_{\sigma\in \mathrm{S}_{k-1}} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}} - \sum_{\sigma\in \mathrm{S}_{k-1}} \frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(c\circ\sigma^{-1}\circ T)}}. \tag{3.31}$$

(a) **If $T(i)=k$ for some $i=2,\dots,k-1$.** We claim that $d(\sigma^{-1}\circ T)=d(c\circ\sigma^{-1}\circ T)$ for all $\sigma\in \mathrm{S}_{k-1}$. In fact, for $\sigma\in \mathrm{S}_{k-1}$, we have $\sigma^{-1}\circ T(i)=k$ and $c\circ\sigma^{-1}\circ T(i)=1$. Therefore $\sigma^{-1}\circ T(i) > \sigma^{-1}\circ T(i+1)$, $\sigma^{-1}\circ T(i) > \sigma^{-1}\circ T(i-1)$, whereas $c\circ\sigma^{-1}\circ T(i) < c\circ\sigma^{-1}\circ T(i+1)$, $c\circ\sigma^{-1}\circ T(i) < c\circ\sigma^{-1}\circ T(i-1)$. And for $j\neq i-1,i$, we have $\sigma^{-1}\circ T(j) > \sigma^{-1}\circ T(j+1)$ if and only if $c\circ\sigma^{-1}\circ T(j) > c\circ\sigma^{-1}\circ T(j+1)$. This shows $d(\sigma^{-1}\circ T)=d(c\circ\sigma^{-1}\circ T)$. Hence, the two sums on the right hand side of (3.31) are equal, and $\sum_{\sigma\in \mathrm{S}_k} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{k^2\binom{k-1}{d(\sigma^{-1}\circ T)}} = 0 = \frac{\mu(T)}{k}$.

(b) **If $T(1)=k$.** Similar to the above discussion, we can show that $d(\sigma^{-1}\circ T) =$

$d(c \circ \sigma^{-1} \circ T) + 1$. Hence the right hand side of (3.31) is equal to

$$\begin{aligned}
&- \sum_{\sigma \in \mathrm{S}_{k-1}} \left(\frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{k^2 \binom{k-1}{d(c\circ\sigma^{-1}\circ T)+1}} + \frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{k^2 \binom{k-1}{d(c\circ\sigma^{-1}\circ T)}} \right) \\
&= - \sum_{\sigma \in \mathrm{S}_{k-1}} \frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{k(k-1)\binom{k-2}{d(c\circ\sigma^{-1}\circ T)}} \\
&= \frac{-(k-1)}{k} \sum_{\sigma \in \mathrm{S}_{k-1}} \frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{(k-1)^2 \binom{k-2}{d(c\circ\sigma^{-1}\circ T)}}
\end{aligned}$$

We claim that $d(c \circ \sigma^{-1} \circ T) = d(\sigma^{-1} \circ T \circ c)$. This is because $\sigma^{-1} \circ T \circ c(k-1) < \sigma^{-1} \circ T \circ c(k) = k$, $1 = c \circ \sigma^{-1} \circ T(1) < c \circ \sigma^{-1} \circ T(2)$, and $c \circ \sigma^{-1} \circ T(i+1) > c \circ \sigma^{-1} \circ T(i)$ if and only if $\sigma^{-1} \circ T \circ c(i) > \sigma^{-1} \circ T \circ c(i-1)$, for $i = 2, 3, \ldots, k-1$. Hence,

$$\begin{aligned}
&\frac{-(k-1)}{k} \sum_{\sigma \in \mathrm{S}_{k-1}} \frac{(-1)^{d(c\circ\sigma^{-1}\circ T)}\mu(\sigma)}{(k-1)^2 \binom{k-2}{d(c\circ\sigma^{-1}\circ T)}} \\
&= \frac{-(k-1)}{k} \sum_{\sigma \in \mathrm{S}_{k-1}} \frac{(-1)^{d(\sigma^{-1}\circ T\circ c)}\mu(\sigma)}{(k-1)^2 \binom{k-2}{d(\sigma^{-1}\circ T\circ c)}} \\
&= \frac{-(k-1)}{k} \frac{\mu(T \circ c)}{k-1} && \text{(by induction hypothesis)} \\
&= \frac{(k-1)}{k} \frac{\mu(T)}{k-1} && \text{(by definition of } \mu) \\
&= \frac{\mu(T)}{k}.
\end{aligned}$$

(c) **If $T(k) = k$.** Similar to the above discussion, we can show that $d(c \circ \sigma^{-1} \circ T) = d(\sigma^{-1} \circ T) + 1$. And hence the right hand side of (3.31) is equal to

$$\frac{(k-1)}{k} \sum_{\sigma \in \mathrm{S}_{k-1}} \frac{(-1)^{d(\sigma^{-1}\circ T)}\mu(\sigma)}{(k-1)^2 \binom{k-2}{d(\sigma^{-1}\circ T)}} = \frac{(k-1)}{k} \frac{\mu(T)}{k-1} = \frac{\mu(T)}{k}$$

by the induction hypothesis, where T can be considered as an element in S_{k-1} since it stabilizes k.

We have thus shown the claim (3.30). Putting this into Equation (3.29) shows Equation (3.28):

$$\sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) \varphi_k(C_{\sigma(1)}, \ldots, C_{\sigma(k)}) = \sum_{T \in \mathrm{S}_k} \frac{\mu(T)}{k} [\ldots[[C_{T(1)}, C_{T(2)}], C_{T(3)}], \ldots, C_{T(k)}].$$

(3) The third and last step is to show[7]

$$\sum_{T\in \mathrm{S}_k} \mu(T)[\dots[[C_{T(1)}, C_{T(2)}], C_{T(3)}], \dots, C_{T(k)}] = k[\dots[[C_1, C_2], C_3], \dots, C_k]. \tag{3.32}$$

First, using induction on k, we will show that

$$\sum_{T\in \mathrm{S}_k} \mu(T)[\dots[[C_{k+1}, C_{T(1)}], C_{T(2)}], \dots, C_{T(k)}] = -[\dots[[C_1, C_2], C_3], \dots, C_{k+1}]. \tag{3.33}$$

The case where $k = 2$ follows directly from the Jacobi identity. Suppose Equation (3.33) holds for $k-1$, then

$$\begin{aligned}\sum_{T\in \mathrm{S}_k} \mu(T)&[\dots[[C_{k+1}, C_{T(1)}], C_{T(2)}], \dots, C_{T(k)}] \\ &= \sum_{T\in \mathrm{S}_{k-1}} \mu(T)[[\dots[[C_{k+1}, C_{T(1)}], C_{T(2)}], \dots, C_{T(k-1)}], C_k] \\ &\qquad - \sum_{T\in \mathrm{S}_{k-1}} \mu(T)[\dots[[C_{k+1}, C_k], C_{T(1)}], \dots, C_{T(k-1)}]. \end{aligned}\tag{3.34}$$

By the induction hypothesis, the first sum on the right hand side is equal to

$$-[[[\dots[[C_1, C_2], C_3], \dots, C_{k-1}], C_{k+1}], C_k],$$

and the second sum on the right hand side is equal to

$$-[[\dots[[C_1, C_2], C_3], \dots, C_{k-1}], [C_{k+1}, C_k]].$$

Using the Jacobi identity and the anticommutativity of Lie brackets, we have

$$\begin{aligned}-[[[\dots[[C_1, C_2], C_3], \dots, C_{k-1}], C_{k+1}], C_k] + [[\dots[[C_1, C_2], C_3], \dots, C_{k-1}], [C_{k+1}, C_k]] \\ = -[[[\dots[[C_1, C_2], C_3], \dots, C_{k-1}], C_k], C_{k+1}].\end{aligned}$$

[7] A direct way of proving Equation (3.32) is to use the Dynkin-Specht-Wever theorem [35], which states that if a non-commutative polynomial $f \in \mathbb{Q}\langle C_1, \dots, C_k\rangle$ is *Lie*, then one can replace all monomials $C_{i_1}C_{i_2}\cdots C_{i_k}$ by $[\dots[C_{i_1}, C_{i_2}], \dots, C_{i_k}]/k$ without changing its value. Writing the right hand side of (3.32) as an element in $\mathbb{Q}\langle C_1, \dots, C_k\rangle$ gives $k\sum_{\sigma\in \mathrm{S}_k} \mu(\sigma)C_{\sigma(1)}C_{\sigma(2)}\cdots C_{\sigma(k)}$ (we can check this using the definition of μ), which is equal to the left hand side by replacing the monomials $C_{\sigma(1)}C_{\sigma(2)}\cdots C_{\sigma(k)}$ by the Lie brackets $[\dots[[C_{\sigma(1)}, C_{\sigma(2)}], C_{\sigma(3)}], \dots, C_{\sigma(k)}]/k$. Nevertheless, here we will give a self-contained proof without using the Dynkin-Specht-Wever theorem.

Hence, Equation (3.34) yields

$$\sum_{T\in \mathrm{S}_k} \mu(T)[\dots[[C_{k+1}, C_{T(1)}], C_{T(2)}], \dots, C_{T(k)}] = -[[\dots[[C_1, C_2], C_3], \dots, C_k], C_{k+1}],$$

concluding the proof by induction for Equation (3.33).

Next, we will again use induction on k to prove Equation (3.32):

$$\sum_{T\in \mathrm{S}_k} \mu(T)[\dots[[C_{T(1)}, C_{T(2)}], C_{T(3)}], \dots, C_{T(k)}] = k[\dots[[C_1, C_2], C_3], \dots, C_k].$$

The case of $k = 2$ results from direct computation. Suppose (3.32) hold for $k-1$, then

$$\begin{aligned}
&\sum_{T\in \mathrm{S}_k} \mu(T)[\dots[[C_{T(1)}, C_{T(2)}], C_{T(3)}], \dots, C_{T(k)}] \\
&= \sum_{T\in \mathrm{S}_{k-1}} \mu(T)[[\dots[[C_{T(1)}, C_{T(2)}], C_{T(3)}], \dots, C_{T(k-1)}], C_k] \\
&\quad - \sum_{T\in \mathrm{S}_{k-1}} \mu(T)[\dots[[C_k, C_{T(1)}], C_{T(2)}], \dots, C_{T(k-1)}] \\
&= (k-1)[\dots[[C_1, C_2], C_3], \dots, C_k] \\
&\quad - \sum_{T\in \mathrm{S}_{k-1}} \mu(T)[\dots[[C_k, C_{T(1)}], C_{T(2)}], \dots, C_{T(k-1)}] \qquad \text{(by induction hypothesis)} \\
&= k[\dots[[C_1, C_2], C_3], \dots, C_k] \qquad \text{(by Equation (3.33) for } k-1).
\end{aligned}$$

We have thus shown Equation (3.32).

Combining the Equations (3.21), (3.28) and (3.32) obtained in the three steps gives us

$$\sum_{\sigma\in \mathrm{S}_k} \mu(\sigma) H_k(C_{\sigma(1)}, \dots, C_{\sigma(k)}, C_{k+1}, \dots, C_m) = [\dots[[C_1, C_2], C_3], \dots, C_k].$$

□

3.5.3 Proof of Proposition 3.5.2

In this subsection we prove Proposition 3.5.2. Again, the key is understanding the structure of the expressions for H_k. For even k, the following lemma shows that the expression $H_k(C_1, \dots, C_m)$ is "antisymmetric", and immediately yields Proposition 3.5.2.

Lemma 3.5.5. *When k is even, we have*

$$H_k(C_1, \dots, C_m) = -H_k(C_m, \dots, C_1).$$

Proof. Define a new variable t. Replacing B_i by $\exp(tC_i)$ in the Baker-Campbell-Hausdorff formula (3.4), we have

$$\log(\exp(tC_1)\cdots\exp(tC_m)) = t\sum_{i=1}^{m} C_i + t^k \sum_{k=2}^{d-1} H_k(C_1,\ldots,C_m). \tag{3.35}$$

Now, replace B_i by $\exp(-tC_{m+1-i})$, $i = 1,\ldots,m$, in the Baker-Campbell-Hausdorff Formula (3.4), we obtain

$$\log(\exp(-tC_m)\cdots\exp(-tC_1)) = -t\sum_{i=1}^{m} C_i + (-t)^k \sum_{k=2}^{d-1} H_k(C_m,\ldots,C_1). \tag{3.36}$$

Since $\log(\exp(tC_1)\cdots\exp(tC_m)) = -\log(\exp(-tC_m)\cdots\exp(-tC_1))$, comparing the coefficients of t^k in (3.35) and (3.36) yields

$$H_k(C_1,\ldots,C_m) = -H_k(C_m,\ldots,C_1)$$

for even k. □

Next, we need the following lemmas regarding the odd terms H_3, H_5, H_7 and H_9. These correspond to Proposition 3.5.2 for $k = 3, 5, 7, 9$.

Lemma 3.5.6. *Let $\mathcal{H} \subset \mathsf{UT}(n,\mathbb{Q})$ be a finite set of matrices. Given matrices $B_1,\ldots,B_m$ in $\mathsf{UT}(n,\mathbb{Q})$ such that $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$, $i = 1,\ldots,m$, and $\sum_{i=1}^{m} \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, then*

$$\sum_{\sigma\in \mathrm{S}_m} H_3(\log B_{\sigma(1)},\ldots,\log B_{\sigma(m)}) \in \mathfrak{L}_{\geq 4}(\log \mathcal{H}).$$

Proof. Denote $C_i \coloneqq \log B_i, i = 1,\ldots,m$, we will show the following identity

$$\sum_{\sigma\in \mathrm{S}_m} H_3(C_{\sigma(1)},\ldots,C_{\sigma(m)}) = \frac{m!}{12}\sum_{i=1}^{m}\left[C_i,\left[C_i,\sum_{j=1}^{m} C_j\right]\right]. \tag{3.37}$$

Write

$$H_3(C_{\sigma(1)},\ldots,C_{\sigma(m)}) = \sum_{i<j<k} H_{3,3}(C_{\sigma(i)},C_{\sigma(j)},C_{\sigma(k)}) + \sum_{i<j} H_{3,2}(C_{\sigma(i)},C_{\sigma(j)}),$$

where

$$H_{3,3}(X,Y,Z) = \frac{1}{3}[X,[Y,Z]] + \frac{1}{6}[[X,Z],Y],$$

$$H_{3,2}(X,Y) = \frac{1}{12}([X,[X,Y]] + [[X,Y],Y]).$$

Using the Jacobi identity, we have

$$\begin{aligned}
&H_{3,3}(C_i, C_j, C_k) + H_{3,3}(C_j, C_k, C_i) + H_{3,3}(C_k, C_i, C_j) \\
&= \frac{1}{3}\left([C_i, [C_j, C_k]] + [C_j, [C_k, C_i]] + [C_k, [C_i, C_j]]\right) \\
&\quad + \frac{1}{6}\left([[C_i, C_k], C_j] + [[C_j, C_i], C_k] + [[C_k, C_j], C_i]\right) \\
&= 0
\end{aligned}$$

for any i, j, k. Similarly,

$$H_{3,3}(C_k, C_j, C_i) + H_{3,3}(C_j, C_i, C_k) + H_{3,3}(C_i, C_k, C_j) = 0.$$

Hence,

$$\begin{aligned}
&\sum_{\sigma \in \mathrm{S}_m} \sum_{i<j<k} H_{3,3}(C_{\sigma(i)}, C_{\sigma(j)}, C_{\sigma(k)}) \\
&= \frac{m!}{6} \sum_{i<j<k} \left(H_{3,3}(C_i, C_j, C_k) + H_{3,3}(C_j, C_k, C_i) + H_{3,3}(C_k, C_i, C_j)\right) \\
&\quad + \frac{m!}{6} \sum_{i<j<k} \left(H_{3,3}(C_k, C_j, C_i) + H_{3,3}(C_j, C_i, C_k) + H_{3,3}(C_i, C_k, C_j)\right) \\
&= 0.
\end{aligned}$$

Whereas

$$\begin{aligned}
&\sum_{\sigma \in \mathrm{S}_m} \sum_{i<j} H_{3,2}(C_{\sigma(i)}, C_{\sigma(j)}) \\
&= \frac{m!}{2} \sum_{i \neq j} H_{3,2}(C_i, C_j) \\
&= \frac{m!}{2} \sum_{i \neq j} \left(\frac{1}{12}[C_i, [C_i, C_j]] + \frac{1}{12}[[C_i, C_j], C_j]\right) \\
&= \frac{m!}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\frac{1}{12}[C_i, [C_i, C_j]] + \frac{1}{12}[[C_i, C_j], C_j]\right) \\
&= \frac{m!}{2} \sum_{i=1}^{m} \frac{1}{12} \left[C_i, \left[C_i, \sum_{j=1}^{m} C_j\right]\right] + \frac{m!}{2} \sum_{j=1}^{m} \frac{1}{12} \left[\left[\sum_{i=1}^{m} C_i, C_j\right], C_j\right] \\
&= \frac{m!}{12} \sum_{i=1}^{m} \left[C_i, \left[C_i, \sum_{j=1}^{m} C_j\right]\right].
\end{aligned}$$

Adding up the two above expressions yields Equation (3.37). Since $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ for all i

and $\sum_{i=1}^m \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, Equation (3.37) yields

$$\begin{aligned}
&\sum_{\sigma \in \mathrm{S}_m} H_3(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(m)}) \\
&= \frac{m!}{12} \sum_{i=1}^m \left[\log B_i, \left[\log B_i, \sum_{j=1}^m \log B_j \right] \right] \\
&\in \frac{m!}{12} \sum_{i=1}^m [\log B_i, [\log B_i, \mathfrak{L}_{\geq 2}(\log \mathcal{H})]] \\
&\in \mathfrak{L}_{\geq 4}(\log \mathcal{H})
\end{aligned}$$

□

The following Lemmas 3.5.7, 3.5.9 and 3.5.10 regarding H_5, H_7, H_9 are proven using computer assistance from the software SageMath [95]. In what follows, we give a sketch of their proof. Details of the full proof along with the algorithm used for computer assistance are given in Section 3.10. Links to the code can be found in the respective proofs.

Lemma 3.5.7. *Let $\mathcal{H} \subset \mathsf{UT}(n, \mathbb{Q})$ be a finite set of matrices. There exists a permutation $(j_1, j_2, \ldots, j_{12})$ of the tuple $(1, 1, 2, 2, \ldots, 6, 6)$, such that for any given set of matrices $B_1, \ldots, B_6$ in $\mathsf{UT}(n, \mathbb{Q})$ with $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^6 \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, we have*

$$\begin{aligned}
&\sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(6)}) + \sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(j_1)}, \ldots, \log B_{\sigma(j_{12})}) \\
&\qquad \in \mathfrak{L}_{\geq 6}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})). \quad (3.38)
\end{aligned}$$

Namely, we can take $(j_1, j_2, \ldots, j_{12}) = (1, 2, 3, 4, 4, 5, 5, 6, 6, 1, 2, 3)$.

Sketch of proof of Lemma 3.5.7. For $x, y \in \mathfrak{u}(n)$, denote $x \sim y$ if

$$x - y \in \mathfrak{L}_{\geq 6}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})).$$

The claim (3.38) can be written as

$$\sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(6)}) + \sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(j_1)}, \ldots, \log B_{\sigma(j_{12})}) \sim 0$$

By the Dynkin formula (Lemma 3.5.4), the expressions $\sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(6)})$

and $\sum_{\sigma\in S_6} H_5(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_{12})})$ can be expressed as a sum in the form of

$$\sum_{\boldsymbol{j}=(j_1,\dots,j_5)\in\{1,\dots,6\}^5} \alpha_{\boldsymbol{j}} \sum_{\sigma\in S_6} \varphi_5(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_5)}), \tag{3.39}$$

where $\alpha_{\boldsymbol{j}}$ are rational numbers.

Since $\sum_{i=1}^6 \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, for any tuple $\boldsymbol{j}=(j_1,\dots,j_5)\in\{1,\dots,6\}^5$, the expression $\sum_{\sigma\in S_6}\varphi_5(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_5)})$ is equivalent (under $\sim$) to a rational multiple of $\sum_{i\neq j}[[[[\log B_i,\log B_j],\log B_j],\log B_j],\log B_i],\log B_i]$. (See Section 3.10 for detailed justification.) In particular, using computer assistance, we can compute these rational multiples and show

$$\sum_{\sigma\in S_6} H_5(\log B_{\sigma(1)},\dots,\log B_{\sigma(6)}) \sim \sum_{i\neq j}[[[[\log B_i,\log B_j],\log B_j],\log B_i],\log B_i],$$
$$\sum_{\sigma\in S_6} H_5\big(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_{12})}\big) \sim -\sum_{i\neq j}[[[[\log B_i,\log B_j],\log B_j],\log B_i],\log B_i].$$

This yields

$$\sum_{\sigma\in S_6} H_5(\log B_{\sigma(1)},\dots,\log B_{\sigma(6)}) + \sum_{\sigma\in S_6} H_5(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_{12})}) \sim 0.$$

The code for computer assistance can be found at https://doi.org/10.6084/m9.figshare.20124146.v1. □

Remark 3.5.8. The added expression of $\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log\mathcal{H}))$ on the right hand side of Equation (3.38) is crucial for its correctness. In fact, we can consider Equation (3.38) in the quotient Lie algebra $L \coloneqq \mathfrak{L}_{\geq 1}(\log\mathcal{H})/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log\mathcal{H}))$. The Lie algebra L is *metabelian*, meaning $[[L,L],[L,L]]=0$. (Free) metabelian Lie algebras have significantly fewer dimensions compared to (free) Lie algebras having the same number of generators. Moreover, free metabelian Lie algebras admit a relatively simple basis (sometimes called the *Gröbner-Shirshov basis*) [22], making it computationally viable to find identities such as Equation (3.38). In our computer assisted proofs, we are using a heavily modified version of this basis to compute Equation (3.38) as well as Equations (3.40) and (3.41) in the following lemmas.

Lemma 3.5.9. *Let $\mathcal{H}\subset \mathsf{UT}(n,\mathbb{Q})$ be a finite set of matrices. There exist positive rational numbers α_1,α_2, as well as, for $s=1,2$, permutations $(j_{s,1},j_{s,2},\dots,j_{s,16})$ of the tuple $(1,1,2,2,\dots,8,8)$, such that for any given set of matrices $B_1,\dots,B_8$ in $\mathsf{UT}(n,\mathbb{Q})$ with $\log B_i\in\mathfrak{L}_{\geq 1}(\log\mathcal{H})$ and $\sum_{i=1}^8\log B_i\in\mathfrak{L}_{\geq 2}(\log\mathcal{H})$, we have*

$$\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(1)},\dots,\log B_{\sigma(8)}) + \sum_{s=1}^{2}\alpha_s \sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(j_{s,1})},\dots,\log B_{\sigma(j_{s,16})})$$
$$\in \mathfrak{L}_{\geq 8}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})). \quad (3.40)$$

Namely, we can take $\alpha_1 = \frac{1}{15}, \alpha_2 = \frac{8}{15}$*, and*

$$(j_{1,1}, j_{1,2}, \dots, j_{1,16}) = (1,2,3,4,5,5,6,6,7,7,8,8,1,2,3,4),$$
$$(j_{2,1}, j_{2,2}, \dots, j_{2,16}) = (1,2,3,4,5,4,6,7,1,2,8,3,5,6,7,8).$$

Sketch of proof of Lemma 3.5.9. Similar to Lemma 3.5.7, define the equivalence relation

$$x \sim y \iff x - y \in \mathfrak{L}_{\geq 8}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})).$$

By the Dynkin formula (Lemma 3.5.4), the expressions $\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(1)},\dots,\log B_{\sigma(8)})$ and $\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_{16})})$ can be expressed as a sum in the form of

$$\sum_{\boldsymbol{j}=(j_1,\dots,j_7)\in\{1,\dots,8\}^7} \alpha_{\boldsymbol{j}} \sum_{\sigma\in \mathrm{S}_8} \varphi_7(\log B_{\sigma(j_1)},\dots,\log B_{\sigma(j_7)}),$$

where $\alpha_{\boldsymbol{j}}$ are rational numbers.

Denote $C_i \coloneqq \log B_i, i = 1,\dots,m$. Since $\sum_{i=1}^{8} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, for any tuple $\boldsymbol{j} = (j_1,\dots,j_7) \in \{1,\dots,8\}^7$, the expression $\sum_{\sigma\in \mathrm{S}_8} \varphi_7(C_{\sigma(j_1)},\dots,C_{\sigma(j_7)})$ is equivalent to a linear combination (with rational coefficient) of

$$\sum_{i\neq j} [[[[[[C_i, C_j], C_j], C_i], C_i], C_i], C_i],$$
$$\sum_{i\neq j} [[[[[[C_i, C_j], C_j], C_j], C_i], C_i], C_i],$$

and

$$\sum_{\substack{i,j,k\\ \text{distinct}}} [[[[[[C_i, C_j], C_j], C_k], C_k], C_i], C_i].$$

(See Section 3.10 for detailed justification.) In fact, using computer assistance, we show that

$$\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(1)},\dots,\log B_{\sigma(8)}) \sim \frac{34}{15}\cdot \sum_{i\neq j} [[[[[[C_i, C_j], C_j], C_i], C_i], C_i], C_i]$$
$$- \frac{34}{45}\cdot \sum_{i\neq j} [[[[[[C_i, C_j], C_j], C_j], C_i], C_i], C_i] + \frac{68}{15}\cdot \sum_{\substack{i,j,k\\ \text{distinct}}} [[[[[[C_i, C_j], C_j], C_k], C_k], C_i], C_i],$$

$$\sum_{\sigma\in \mathrm{S}_8} H_7\big(\log B_{\sigma(j_{1,1})},\dots,\log B_{\sigma(j_{1,16})}\big) \sim \frac{34}{15}\cdot\sum_{i\neq j}[[[[[[C_i,C_j],C_j],C_i],C_i],C_i],C_i]$$
$$+\frac{238}{45}\cdot\sum_{i\neq j}[[[[[[C_i,C_j],C_j],C_j],C_i],C_i],C_i]-\frac{68}{5}\cdot\sum_{\substack{i,j,k\\ \text{distinct}}}[[[[[[C_i,C_j],C_j],C_k],C_k],C_i],C_i],$$

and

$$\sum_{\sigma\in \mathrm{S}_8} H_7\big(\log B_{\sigma(j_{2,1})},\dots,\log B_{\sigma(j_{2,16})}\big) \sim -\frac{68}{15}\cdot\sum_{i\neq j}[[[[[[C_i,C_j],C_j],C_i],C_i],C_i],C_i]$$
$$+\frac{34}{45}\cdot\sum_{i\neq j}[[[[[[C_i,C_j],C_j],C_j],C_i],C_i],C_i]-\frac{34}{5}\cdot\sum_{\substack{i,j,k\\ \text{distinct}}}[[[[[[C_i,C_j],C_j],C_k],C_k],C_i],C_i].$$

This yields

$$\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(1)},\dots,\log B_{\sigma(8)})+\sum_{s=1}^{2}\alpha_s\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(j_{s,1})},\dots,\log B_{\sigma(j_{s,16})})\sim 0,$$

where $\alpha_1=\frac{1}{15},\alpha_2=\frac{8}{15}$. The code can be found at https://doi.org/10.6084/m9.figshare.20124113.v1. □

Lemma 3.5.10. *Let $\mathcal{H}\subset \mathsf{UT}(n,\mathbb{Q})$ be a finite set of matrices. There exist positive rational numbers $\alpha_1,\dots,\alpha_6$, as well as, for $s=1,\dots,6$, permutations $(j_{s,1},j_{s,2},\dots,j_{s,20})$ of the tuple $(1,1,2,2,\dots,10,10)$, such that for any given set of matrices $B_1,\dots,B_{10}$ in $\mathsf{UT}(n,\mathbb{Q})$ with $\log B_i\in\mathfrak{L}_{\geq 1}(\log\mathcal{H})$ and $\sum_{i=1}^{10}\log B_i\in\mathfrak{L}_{\geq 2}(\log\mathcal{H})$, we have*

$$\sum_{\sigma\in \mathrm{S}_{10}} H_9(\log B_{\sigma(1)},\dots,\log B_{\sigma(10)})+\sum_{s=1}^{6}\alpha_s\sum_{\sigma\in \mathrm{S}_{10}} H_9(\log B_{\sigma(j_{s,1})},\dots,\log B_{\sigma(j_{s,20})})$$
$$\in \mathfrak{L}_{\geq 10}(\log\mathcal{H})+\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log\mathcal{H})). \quad (3.41)$$

Namely, we can take $\alpha_1=\frac{44566633}{13702661},\alpha_2=\frac{557040}{13702661},\alpha_3=\frac{205175}{3915046},\alpha_4=\frac{1307207}{13702661},\alpha_5=\frac{86275275}{27405322}$, $\alpha_6=\frac{4105194}{1957523}$, and

$$(j_{1,1},j_{1,2},\dots,j_{1,20})=(5,4,7,10,2,8,3,8,1,9,7,6,5,6,2,3,9,10,1,4),$$
$$(j_{2,1},j_{2,2},\dots,j_{2,20})=(8,3,5,7,10,6,8,2,1,10,2,4,9,1,5,9,3,6,7,4),$$
$$(j_{3,1},j_{3,2},\dots,j_{3,20})=(7,10,2,6,4,9,6,4,1,5,3,5,1,9,3,7,10,2,8,8),$$
$$(j_{4,1},j_{4,2},\dots,j_{4,20})=(10,2,2,6,7,1,9,3,9,4,8,7,8,5,5,1,4,10,6,3),$$
$$(j_{5,1},j_{5,2},\dots,j_{5,20})=(3,5,10,1,4,8,6,9,3,2,7,6,1,10,9,7,2,4,5,8),$$

$$(j_{6,1}, j_{6,2}, \ldots, j_{6,20}) = (4, 7, 2, 10, 2, 1, 3, 5, 8, 1, 6, 9, 10, 7, 6, 8, 3, 5, 9, 4).$$

Sketch of proof of Lemma 3.5.10. Similar to Lemma 3.5.7, and Lemma 3.5.9, denote $C_i = \log B_i$ for $i = 1, \ldots, m$. Since $\sum_{i=1}^{10} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$, for any tuple $\boldsymbol{j} = (j_1, \ldots, j_9) \in \{1, \ldots, 10\}^9$, the expression $\sum_{\sigma \in \mathrm{S}_{10}} \varphi_9(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_9)})$ is equivalent to a linear combination (with rational coefficient) of

$$\sum_{i \neq j} [[[[[[[[C_i, C_j], C_j], C_i], C_i], C_i], C_i], C_i], C_i],$$
$$\sum_{i \neq j} [[[[[[[[C_i, C_j], C_j], C_j], C_i], C_i], C_i], C_i], C_i],$$
$$\sum_{i \neq j} [[[[[[[[C_i, C_j], C_j], C_j], C_j], C_i], C_i], C_i], C_i],$$
$$\sum_{\substack{i,j,k \\ \text{distinct}}} [[[[[[[[C_i, C_j], C_j], C_k], C_k], C_i], C_i], C_i], C_i],$$
$$\sum_{\substack{i,j,k \\ \text{distinct}}} [[[[[[[[C_i, C_j], C_j], C_j], C_k], C_k], C_i], C_i], C_i],$$

and

$$\sum_{\substack{i,j,k \\ \text{distinct}}} [[[[[[[[C_i, C_j], C_j], C_k], C_k], C_k], C_i], C_i], C_i].$$

Similar to the previous lemmas, the rest of the proof can be done by computer assistance. The code can be found at https://doi.org/10.6084/m9.figshare.20122979.v1. □

Combining Lemma 3.5.5-3.5.10, we obtain Proposition 3.5.2.

Proposition 3.5.2. *Let $k \leq 10$ and let $\mathcal{H} \subset \mathsf{UT}(n, \mathbb{Q})$ be a finite set of matrices for some $n \geq 2$. Then there exist a non-negative integer r, positive rational numbers $\alpha_1, \ldots, \alpha_r$, as well as, for $s = 1, \ldots, r$, words $\boldsymbol{j}_s = j_{s,1} j_{s,2} \cdots j_{s,m_s}$ in the alphabet $\mathcal{I} = \{1, 2, \ldots, k+1\}$, such that $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) \in \{(1, \ldots, 1), (2, \ldots, 2)\}$ and*

$$\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(k+1)}) + \sum_{s=1}^{r} \alpha_s \sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(j_{s,1})}, \ldots, \log B_{\sigma(j_{s,m_s})})$$
$$\in \mathfrak{L}_{\geq k+1}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})) \quad (3.17)$$

for all matrices $B_1, \ldots, B_{k+1}$ in $\mathsf{UT}(n, \mathbb{Q})$ satisfying $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$.

Proof. For even k, Equation (3.17) is satisfied by Lemma 3.5.5 by taking $r = 0$ and pairing each permutation σ with its *reversal*:

$$\begin{aligned}
&\sum_{\sigma\in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \dots, \log B_{\sigma(k+1)}) \\
&= \frac{1}{2}\left(\sum_{\sigma\in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \dots, \log B_{\sigma(k+1)}) + \sum_{\sigma\in \mathrm{S}_{k+1}} H_k(\log B_{\mathrm{rev}(\sigma)(1)}, \dots, \log B_{\mathrm{rev}(\sigma)(k+1)})\right) \\
&= \frac{1}{2}\sum_{\sigma\in \mathrm{S}_{k+1}} \left(H_k(\log B_{\sigma(1)}, \dots, \log B_{\sigma(k+1)}) + H_k(\log B_{\mathrm{rev}(\sigma)(1)}, \dots, \log B_{\mathrm{rev}(\sigma)(k+1)})\right) \\
&= 0.
\end{aligned}$$

Here $\mathrm{rev}(\sigma) \in \mathrm{S}_{k+1}$ is the reversal of σ, meaning $\mathrm{rev}(\sigma)(i) = \sigma(k+2-i), i = 1, \dots, k+1$. For $k = 3, 5, 7, 9$, Equation (3.17) is satisfied by Lemma 3.5.6, 3.5.7, 3.5.9 and 3.5.10 respectively. □

3.5.4 Proof of Proposition 3.5.3

In this subsection, we give a proof of Proposition 3.5.3.

Proposition 3.5.3. *Let V be a finite dimensional $\mathbb{Q}$-linear space. Let d be a positive integer, $\mathcal{I}$ be a finite index set, and $\boldsymbol{a}_{1i}, \dots, \boldsymbol{a}_{di}, i \in \mathcal{I}$ be vectors in V.*

For any $t \in \mathbb{Z}_{>0}$, define the vectors

$$P_i(t) \coloneqq t\boldsymbol{a}_{1i} + \cdots + t^d \boldsymbol{a}_{di}, \quad \text{for } i \in \mathcal{I}.$$

Suppose the following two conditions hold:

(i) The $\mathbb{Q}_{\geq 0}$-cone $\mathcal{C}_d \coloneqq \langle \boldsymbol{a}_{di} \mid i \in \mathcal{I}\rangle_{\mathbb{Q}_{\geq 0}}$ is a linear space.

(ii) For $k = d-1, d-2, \dots, 1$, the inductively defined $\mathbb{Q}_{\geq 0}$-cones $\mathcal{C}_k \coloneqq \langle \boldsymbol{a}_{ki} \mid i \in \mathcal{I}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1}$ are linear spaces.

Then the $\mathbb{Q}_{\geq 0}$-cone $\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}}$ is equal to $\mathcal{C}_1$.

Proof. For convenience, define $\mathcal{C}_{d+1} \coloneqq \{0\}$. We will prove that, for all $k = 2, \dots, d+1$, the cone $\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_k$ is equal to $\mathcal{C}_1$. Notice that the claim in the proposition is the case where $k = d+1$. We use induction on k.

For $k = 2$, since $\boldsymbol{a}_{ki} \in \mathcal{C}_2$ for $k \geq 2$, we have $P_i(t) + \mathcal{C}_2 = t\boldsymbol{a}_{1i} + \mathcal{C}_2$, so

$$\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_2 = \langle t\boldsymbol{a}_{1i} \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_2 \overset{\text{(ii)}}{=} \mathcal{C}_1.$$

For the induction step, suppose now that the cone $\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_k$ is equal

to $\mathcal{C}_1$, we want to prove that $\langle P_i(t) \mid i \in \mathcal{I}, t \in \mathbb{Z}_{>0} \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1}$ is equal to $\mathcal{C}_1$.

By the induction hypothesis, there exist indices $i_1, \ldots, i_m \in \mathcal{I}$ as well as positive integers $t_1, \ldots, t_m \in \mathbb{Z}_{>0}$, such that

$$\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_k = \mathcal{C}_1.$$

Condition (ii) of the proposition shows that there exist indices $i'_1, \ldots, i'_{m'} \in \mathcal{I}$ such that

$$\langle \boldsymbol{a}_{ki'_j} \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \mathcal{C}_k.$$

Hence

$$\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \langle \boldsymbol{a}_{ki'_j} \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \mathcal{C}_1. \tag{3.42}$$

We show that there exists $t \in \mathbb{Z}_{>0}$ such that

$$\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \langle P_{i'_j}(t) \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \mathcal{C}_1.$$

Suppose the contrary, that for every $t \in \mathbb{Z}_{>0}$,

$$\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \langle P_{i'_j}(t) \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} \subsetneq \mathcal{C}_1.$$

For any $\mathbb{Q}_{\geq 0}$-cone $\mathcal{C}$, define the *normal cone* of $\mathcal{C}$ as the set of vectors $\boldsymbol{v} \in V$ such that $\boldsymbol{v}^\top \boldsymbol{c} \leq 0$ for all $\boldsymbol{c} \in \mathcal{C}$. For every t, take a normalized vector $\boldsymbol{v}_t \in \mathcal{C}_1$ (meaning the norm of $\boldsymbol{v}_t$ is 1) in the normal cone of $\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \langle P_{i'_j}(t) \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1}$. That is,

$$\boldsymbol{v}_t^\top P_{i_j}(t_j) \leq 0 \text{ for all } j, \quad \boldsymbol{v}_t^\top P_{i'_j}(t) \leq 0 \text{ for all } j, \quad \boldsymbol{v}_t \perp \mathcal{C}_{k+1}. \tag{3.43}$$

Such a vector must exist because $\langle P_{i_j}(t_j) \mid j = 1, \ldots, m \rangle_{\mathbb{Q}_{\geq 0}} + \langle P_{i'_j}(t) \mid j = 1, \ldots, m' \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1}$ is a strict sub-cone of the linear space $\mathcal{C}_1$. The $\mathbb{R}$-linear space $V_{\mathbb{R}} = V \otimes_{\mathbb{Q}} \mathbb{R}$ is finite dimensional and hence compact. Embed V into $V_{\mathbb{R}}$ canonically, then the sequence $\{\boldsymbol{v}_t\}_{t \in \mathbb{Z}_{>0}}$ has a limit point in $V_{\mathbb{R}}$. Denote by $\boldsymbol{v}_{lim}$ this limit point. As all the vectors $\boldsymbol{v}_t$ are in $\mathcal{C}_1$, $\boldsymbol{v}_{lim}$ must be in $\mathcal{C}_1 \otimes_{\mathbb{Q}} \mathbb{R}$. Since the inner product of V canonically extends to the inner product of $V_{\mathbb{R}}$, taking the limit of (3.43), we have

$$\boldsymbol{v}_{lim}^\top \cdot P_{i_j}(t_j) \leq 0 \text{ for all } j, \quad \boldsymbol{v}_{lim} \perp \mathcal{C}_{k+1}, \tag{3.44}$$

and

$$\begin{aligned}\boldsymbol{v}_{lim}^{\top}\cdot\boldsymbol{a}_{ki'_j} &= \boldsymbol{v}_{lim}^{\top}\cdot\lim_{t\to\infty}\left(\frac{P_{i'_j}(t)}{t^k} - t\boldsymbol{a}_{k+1,i'_j} - \cdots - t^{d-k}\boldsymbol{a}_{d,i'_j}\right)\\ &= \boldsymbol{v}_{lim}^{\top}\cdot\lim_{t\to\infty}\frac{P_{i'_j}(t)}{t^k} \leq 0, \quad j=1,\ldots,m'. \end{aligned} \tag{3.45}$$

The second equality is due to $\boldsymbol{a}_{k+1,i'_j},\ldots,\boldsymbol{a}_{d,i'_j} \in \mathcal{C}_{k+1} \perp \boldsymbol{v}_{lim}$. Hence, (3.44) and (3.45) show that $\boldsymbol{v}_{lim}^{\top}\cdot\boldsymbol{v}\leq 0$ for all $\boldsymbol{v}$ in the $\mathbb{R}_{\geq 0}$-cone

$$\langle P_{i_j}(t_j) \mid j=1,\ldots,m\rangle_{\mathbb{Q}_{\geq 0}} + \langle \boldsymbol{a}_{ki'_j} \mid j=1,\ldots,m'\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} \overset{\text{Eq. (3.42)}}{=} \mathcal{C}_1.$$

Since $\mathcal{C}_1$ is a linear space, $\boldsymbol{v}_{lim}$ is non-zero (it has norm one) and is in $\mathcal{C}_1\otimes_{\mathbb{Q}}\mathbb{R}$, this yields a contradiction. We have thus shown that there exists $t\in\mathbb{Z}_{>0}$ such that

$$\langle P_{i_j}(t_j) \mid j=1,\ldots,m\rangle_{\mathbb{Q}_{\geq 0}} + \langle P_{i'_j}(t) \mid j=1,\ldots,m'\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \mathcal{C}_1.$$

Since $P_i(t)\in\mathcal{C}_1, i\in\mathcal{I}, t\in\mathbb{Z}_{>0}$, this means

$$\langle P_i(t) \mid i\in\mathcal{I}, t\in\mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \mathcal{C}_1,$$

concluding the induction.

Finally, take $k=d+1$. This yields $\langle P_i(t) \mid i\in\mathcal{I}, t\in\mathbb{Z}_{>0}\rangle_{\mathbb{Q}_{\geq 0}} = \mathcal{C}_1$. □

3.5.5 Full proof of Theorem 3.4.2

In this subsection, with Propositions 3.5.1 - 3.5.3 at our disposal, we will show the proof of Theorem 3.4.2. First, we need the following lemma.

Lemma 3.5.11. *Let $\mathcal{H}$ be a finite subset of the Lie algebra $\mathfrak{u}(n)$. Let W, V be linear subspaces of $\mathfrak{L}_{\geq 1}(\mathcal{H})$ such that $W+\mathfrak{L}_{\geq 2}(V)=V$, then $\mathfrak{L}_{\geq 2}(W)=\mathfrak{L}_{\geq 2}(V)$.*

Proof. Since $W+\mathfrak{L}_{\geq 2}(V)=V$, we have $W\subseteq V$, and thus $\mathfrak{L}_{\geq 2}(W)\subseteq\mathfrak{L}_{\geq 2}(V)$. Therefore, it suffices to prove the opposite inclusion $\mathfrak{L}_{\geq 2}(W)\supseteq\mathfrak{L}_{\geq 2}(V)$.

Note that since W, V are linear spaces, the sets $[V]_k, [W]_k$ are also linear spaces for all $k=1,\ldots,n$. We use induction on k to show that

$$[V]_k \subseteq [W]_k + \mathfrak{L}_{\geq k+1}(V). \tag{3.46}$$

For $k=1$ this immediately results from the equation $W+\mathfrak{L}_{\geq 2}(V)=V$. Suppose Equation (3.46)

hold for $k-1$. Then, take any elements $x \in V, y \in [V]_{k-1}$, by the induction hypothesis and by $W + \mathfrak{L}_{\geq 2}(V) = V$, there exist $x' \in W, y' \in [W]_{k-1}$, such that $x - x' \in \mathfrak{L}_{\geq 2}(V), y - y' \in \mathfrak{L}_{\geq k}(V)$. Then,

$$\begin{aligned}
[y, x] &= [y', x'] + [y - y', x'] + [y, x - x'] \\
&\in [[W]_{k-1}, W] + [\mathfrak{L}_{\geq k}(V), W] + [[V]_{k-1}, \mathfrak{L}_{\geq 2}(V)] \\
&\subseteq [W]_k + [\mathfrak{L}_{\geq k}(V), V] + [\mathfrak{L}_{\geq k-1}(V), \mathfrak{L}_{\geq 2}(V)] \\
&\subseteq [W]_k + \mathfrak{L}_{\geq k+1}(V).
\end{aligned}$$

Taking the linear span for all $x \in V, y \in [V]_{k-1}$ shows $[V]_k \subseteq [W]_k + \mathfrak{L}_{\geq k+1}(V)$, concluding the induction.

Now, for any $l = 2, \ldots, d$, take the sum of Equation (3.46) for $k = l, \ldots, d$, we have

$$\mathfrak{L}_{\geq l}(V) = \sum_{k \geq l} [V]_k \subseteq \sum_{k \geq l} [W]_k + \sum_{k \geq l} \mathfrak{L}_{\geq k+1}(V) = \mathfrak{L}_{\geq l}(W) + \mathfrak{L}_{\geq l+1}(V).$$

Therefore,

$$\begin{aligned}
&\mathfrak{L}_{\geq 2}(V) \\
\subseteq\ &\mathfrak{L}_{\geq 2}(W) + \mathfrak{L}_{\geq 3}(V) \\
\subseteq\ &\mathfrak{L}_{\geq 2}(W) + \mathfrak{L}_{\geq 3}(W) + \mathfrak{L}_{\geq 4}(V) \\
&\vdots \\
\subseteq\ &\mathfrak{L}_{\geq 2}(W) + \mathfrak{L}_{\geq 3}(W) + \cdots + \mathfrak{L}_{\geq n}(W) \\
=\ &\mathfrak{L}_{\geq 2}(W).
\end{aligned}$$

This shows the inclusion $\mathfrak{L}_{\geq 2}(W) \supseteq \mathfrak{L}_{\geq 2}(V)$. □

Let $\mathcal{G} = \{A_1, \ldots, A_K\}$ be a finite alphabet of elements in $\mathsf{UT}(n, \mathbb{Q})$. For any vector $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}^K_{\geq 0}$, define inductively the following $\mathbb{Q}$-cones $\mathcal{R}_k(\boldsymbol{\ell})$ for $k = 11, 10, \ldots, 2$:

$$\mathcal{R}_{11}(\boldsymbol{\ell}) \coloneqq \{0\}, \tag{3.47}$$

$$\mathcal{R}_k(\boldsymbol{\ell}) \coloneqq \mathcal{R}_{k+1}(\boldsymbol{\ell}) + \left\langle H_k(\log B_1, \ldots, \log B_m) \;\middle|\; B_i \in \mathcal{G}^*, \sum_{i=1}^m \mathrm{PI}^{\mathcal{G}}(B_i) \in \{\boldsymbol{\ell}, 2\boldsymbol{\ell}\} \right\rangle_{\mathbb{Q}_{\geq 0}}. \tag{3.48}$$

That is, $\mathcal{R}_k(\boldsymbol{\ell})$ is the $\mathbb{Q}_{\geq 0}$-cone generated by the elements $H_j(\log B_1, \ldots, \log B_m), j \geq k$, where $B_1, \ldots, B_m$ are *words* in $\mathcal{G}^*$, and the Parikh Images of B_i sum up to $\boldsymbol{\ell}$ or $2\boldsymbol{\ell}$. Recall the definition

of

$$\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})} \coloneqq \{\log A_i \mid i \in \mathrm{supp}(\boldsymbol{\ell})\}$$

as the set of logarithm of matrices in $\mathcal{G}$ whose index appears in $\mathrm{supp}(\boldsymbol{\ell})$. Combining Proposition 3.5.1 and 3.5.2, we can show the following proposition that characterizes the cones $\mathcal{R}_k(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ up to the quotient by $\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$.

Proposition 3.5.12. *Let $\mathcal{G} = \{A_1, \ldots, A_K\}$ be a finite set of matrices in $\mathsf{UT}(n, \mathbb{Q})$ that satisfies $[\log \mathcal{G}]_{11} = \{0\}$. Let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}_{\geq 0}^K$ be a non-zero vector that satisfies $\sum_{i=1}^K \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ as well as $\ell_i \geq 10$ for all $i \in \mathrm{supp}(\boldsymbol{\ell})$. Consider the quotient linear space $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$.*

For any set $\mathcal{C} \subseteq \mathfrak{u}(n)$, denote by $\overline{\mathcal{C}}$ the subset of $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$ consisting of the equivalence classes $c + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})), c \in \mathcal{C}$. Then for all $k \leq 11$, the cone $\overline{\mathcal{R}_k(\boldsymbol{\ell})}$ is equal to the linear space $\overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}$.

Proof. We show that the claim is true for $k = 11, 10, \ldots, 2$, using induction with reverse order on k. For $k = 11$, we have $\overline{\mathcal{R}_{11}(\boldsymbol{\ell})} = \overline{\mathfrak{L}_{\geq 11}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} = \{0\}$ because $[\log \mathcal{G}]_{11} = \{0\}$. Now for some $10 \geq k \geq 2$, suppose $\overline{\mathcal{R}_{k+1}(\boldsymbol{\ell})} = \overline{\mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}$ by induction hypothesis. We will show that $\overline{\mathcal{R}_k(\boldsymbol{\ell})} = \overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}$.

First, we show that for any $i_1, i_2, \ldots, i_k \in \mathrm{supp}(\boldsymbol{\ell})$, we have

$$[\ldots[[\log A_{i_1}, \log A_{i_2}], \log A_{i_3}], \ldots, \log A_{i_k}] \in \mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})).$$

Take a tuple of words $(B'_1, \ldots, B'_{k+1})$ with $B'_1 = A_{i_1}, B'_2 = A_{i_2}, \ldots, B'_k = A_{i_k}, B'_{k+1} \in \mathcal{G}^*$, such that $\sum_{i=1}^{k+1} \mathrm{PI}^{\mathcal{G}}(B'_i) = \boldsymbol{\ell}$. Such a tuple can always be found because $\boldsymbol{\ell}$ satisfies $\ell_i \geq 10 \geq k, i \in \mathrm{supp}(\boldsymbol{\ell})$. For this tuple, the Baker-Campbell-Hausdorff formula gives us

$$\sum_{i=1}^{k+1} \log B'_i \in \sum_{i=1}^{K} \ell_i \log A_i + \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) \subseteq \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}).$$

Hence, for any $\sigma \in \mathrm{S}_k$, Proposition 3.5.2 shows that

$$\begin{aligned}
&- H_k\left(\log B'_{\sigma(1)}, \log B'_{\sigma(2)}, \ldots, \log B'_{\sigma(k)}, \log B'_{k+1}\right) \\
\in &\left\langle H_k(\log B_1, \ldots, \log B_{k+1}) \;\middle|\; B_i \in \mathcal{G}^*, \sum_{i=1}^{k+1} \mathrm{PI}^{\mathcal{G}}(B_i) = \boldsymbol{\ell} \right\rangle_{\mathbb{Q}_{\geq 0}} \\
&+ \left\langle H_k(\log B_1, \ldots, \log B_{2k+2}) \;\middle|\; B_i \in \mathcal{G}^*, \sum_{i=1}^{2k+2} \mathrm{PI}^{\mathcal{G}}(B_i) = 2\boldsymbol{\ell} \right\rangle_{\mathbb{Q}_{\geq 0}} \\
&+ \mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))
\end{aligned}$$

$$\subseteq \mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$$
$$= \mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})). \tag{3.49}$$

The last equality come from $\overline{\mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} = \overline{\mathcal{R}_{k+1}(\boldsymbol{\ell})} \subseteq \overline{\mathcal{R}_k(\boldsymbol{\ell})}$ by the induction hypothesis.

Hence, by Proposition 3.5.1,

$$\begin{aligned}
&[\dots[[\log A_{i_1}, \log A_{i_2}], \log A_{i_3}], \dots, \log A_{i_k}] \\
&= [\dots[[\log B'_1, \log B'_2], \log B'_3], \dots, \log B'_k] \\
&= \sum_{\sigma \in \mathrm{S}_k} \mu(\sigma) H_k(\log B'_{\sigma(1)}, \log B'_{\sigma(2)}, \dots, \log B'_{\sigma(k)}, \log B'_{k+1}) \\
&\in \mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})).
\end{aligned}$$

The last inclusion comes from the fact that both $H_k(\log B'_{\sigma(1)}, \dots, \log B'_{\sigma(k)}, \log B'_{k+1})$ and $-H_k(\log B'_{\sigma(1)}, \dots, \log B'_{\sigma(k)}, \log B'_{k+1})$ are in the cone $\mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$ (by Equation (3.49)). Therefore for every $\sigma \in \mathrm{S}_k$, regardless of which sign $\mu(\sigma)$ takes, the summand $\mu(\sigma) H_k(\log B'_{\sigma(1)}, \dots, \log B'_{\sigma(k)}, \log B'_{k+1})$ is in the cone $\mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$.

Therefore, $[\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}]_k \subseteq \mathcal{R}_k(\boldsymbol{\ell}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$; that is, $\overline{[\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}]_k} \subseteq \overline{\mathcal{R}_k(\boldsymbol{\ell})}$. And since $\overline{\mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} \subseteq \overline{\mathcal{R}_{k+1}(\boldsymbol{\ell})} \subseteq \overline{\mathcal{R}_k(\boldsymbol{\ell})}$ by the induction hypothesis, we have

$$\overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} = \overline{\langle [\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}]_k \rangle_{\mathbb{Q}}} + \overline{\mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} \subseteq \overline{\mathcal{R}_k(\boldsymbol{\ell})}. \tag{3.50}$$

Next, take any tuple $(B_1, \dots, B_m) \in (\mathcal{G}^*)^m$, $\sum_{i=1}^m \mathrm{PI}^{\mathcal{G}}(B_i) = \boldsymbol{\ell}$ or $2\boldsymbol{\ell}$. Note that $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$, $i = 1, \dots, m$, by the Baker-Campbell-Hausdorff formula. Hence, the expression $H_k(\log B_1, \dots, \log B_m)$ can be written as a linear combination of elements in $\left[\mathfrak{L}_{\geq 1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})\right]_k$. That is,

$$\begin{aligned}
\mathcal{R}_k(\boldsymbol{\ell}) &\subseteq \left\langle \left[\mathfrak{L}_{\geq 1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})\right]_k \right\rangle_{\mathbb{Q}} + \mathcal{R}_{k+1}(\boldsymbol{\ell}) \\
&\subseteq \mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathcal{R}_{k+1}(\boldsymbol{\ell}) = \mathfrak{L}_{\geq k+1}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}).
\end{aligned} \tag{3.51}$$

Combining (3.50) and (3.51) we have the desired equality. This concludes the induction and thus the whole proof. □

We now prove Theorem 3.4.2. Although part (i) has already been proven when the theorem is first stated, we will restate it for the sake of completeness.

Theorem 3.4.2 (Structural theorem of unitriangular matrix semigroups)**.** *Let* $\mathcal{G} = \{A_1, \dots, A_K\}$

be a finite set of matrices in $\mathsf{UT}(n,\mathbb{Q})$ *that satisfies* $[\log \mathcal{G}]_{11} = \{0\}$. *Given a non-zero vector* $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}^K_{\geq 0}$*:*

(i) If there exists a word $w \in \mathcal{G}^*$ *with* $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$ *and* $\log w = 0$*, then*

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}). \tag{3.10}$$

(ii) If $\boldsymbol{\ell}$ *satisfies* (3.10)*, then there exists a non-empty word* $w \in \mathcal{G}^*$*, with* $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$*, such that* $\log w = 0$.

Proof. (i) Let w be a word with $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$. Write $w = B_1 B_2 \cdots B_m$ $B_i \in \mathcal{G}, i = 1, \ldots, m$. Regrouping by letters, we have $\sum_{i=1}^{K} \ell_i \log A_i = \sum_{i=1}^{m} \log B_i$.

If $\log w = 0$, then by the Baker-Campbell-Hausdorff formula, we have

$$\sum_{i=1}^{m} \log B_i + \sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m) = \log(B_1 B_2 \cdots B_m) = 0.$$

The higher order terms $H_k, k \geq n$ vanish because $[\log \mathcal{G}]_n = \{0\}$ (a consequence of $\mathcal{G} \subseteq \mathsf{UT}(n,\mathbb{Q})$). Therefore, $\sum_{i=1}^{K} \ell_i \log A_i = -\sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m)$.

Since the Parikh Image of the word $B_1 \cdots B_m$ is $\boldsymbol{\ell}$, the matrices B_i all lie in the subset $\{A_i \mid i \in \mathrm{supp}(\boldsymbol{\ell})\}$ of $\mathcal{G}$. Therefore, $\log B_i \in \log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}$ for all i. By Theorem 3.3.8, for all $k \geq 2$ we have $-H_k(\log B_1, \ldots, \log B_m) \in \left\langle [\{\log B_i \mid i = 1, \ldots, m\}]_k \right\rangle_{\mathbb{Q}} \subseteq \mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) \subseteq \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. Therefore, we have $\sum_{i=1}^{K} \ell_i \log A_i = -\sum_{k=2}^{n-1} H_k(\log B_1, \ldots, \log B_m) \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$.

(ii) Suppose condition (3.10) hold for the vector $\boldsymbol{\ell}$. Resonating Example 3.4.1, our proof for (ii) proceeds in four steps. Now we give an overview of each step. As the first step, we want to construct some matrices $A'_1, \ldots, A'_{K'} \in \langle \mathcal{G} \rangle$, such that

$$\langle \log A'_i \mid i = 1, \ldots, K' \rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) = \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})). \tag{3.52}$$

The candidates for the matrices $A'_1, \ldots, A'_{K'}$ are of the form $B_1^t \cdots B_m^t$, where $m \geq 1$, $t \in \mathbb{Z}_{>0}$, $B_i \in \mathcal{G}^*, i = 1, \ldots, m$ and $\sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}}(B_i) = \boldsymbol{\ell}$ or $2\boldsymbol{\ell}$. The general strategy is to invoke Proposition 3.5.3 while using Proposition 3.5.12 to guarantee that the conditions (i) and (ii) of Proposition 3.5.3 are satisfied.

As the second step, we work in the new alphabet $\mathcal{G}' = \{A'_1, \ldots, A'_{K'}\}$ of matrices found in the previous step. We want to fabricate some matrices $A''_1, \ldots, A''_{K''} \in \langle \mathcal{G}' \rangle$, such that

$$\langle \log A''_i \mid i = 1, \ldots, K'' \rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})))$$

$$= \mathfrak{L}_{\geq 2}\left(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})\right) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))). \quad (3.53)$$

The candidates for the matrices $A''_1, \ldots, A''_{K''}$ are of the form $B_1^t \cdots B_m^t$, where $B_i \in (\mathcal{G}')^*, i = 1, \ldots, m$. The idea is to again invoke Proposition 3.5.3 and to use Proposition 3.5.12 for the new alphabet $\mathcal{G}'$ and a suitable vector $\boldsymbol{\ell}'$.

As the third step, we work in the new alphabet $\mathcal{G}'' = \{A''_1, \ldots, A''_{K''}\}$ of matrices found in the previous step. We want to fabricate some matrices $A'''_1, \ldots, A'''_{K'''} \in \langle \mathcal{G}'' \rangle$, such that

$$\langle \log A'''_i \mid i = 1, \ldots, K'' \rangle_{\mathbb{Q}_{\geq 0}} = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))). \quad (3.54)$$

(Note that $\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})))) = \{0\}$.) The candidates for $A'''_1, \ldots, A'''_{K'''}$ are of the form $B_1^t \cdots B_m^t$, where $B_i \in (\mathcal{G}'')^*, i = 1, \ldots, m$. The idea is to again invoke Proposition 3.5.3 and to use Proposition 3.5.12 for the new alphabet $\mathcal{G}''$ and a suitable vector $\boldsymbol{\ell}''$.

As the fourth and last step, we work in the new alphabet $\mathcal{G}''' = \{A'''_1, \ldots, A'''_{K'''}\}$ of matrices found in the previous step. We then observe that the matrices $A'''_1, \ldots, A'''_{K'''}$ commute with each other, because $\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})))) = \{0\}$. Hence, it is very easy to search for the desired non-empty word $w \in (\mathcal{G}''')^*$ with $\log w = 0$.

We now give the detailed account of each step.

(1) **Find matrices $A'_1, \ldots, A'_{K'} \in \langle \mathcal{G} \rangle$ satisfying condition (3.52).** Since the right hand side of Equation (3.10) is a linear space, we can replace $\boldsymbol{\ell}$ by $10\boldsymbol{\ell}$, and thus suppose $\boldsymbol{\ell}$ satisfy $\ell_i \geq 10, i \in \mathrm{supp}(\boldsymbol{\ell})$. Since $\mathbb{Z}_{>0} \cdot 10\boldsymbol{\ell} \subseteq \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$, the resulting word w will still satisfy $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$.

Since $\boldsymbol{\ell}$ satisfies $\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$, we are able to use Proposition 3.5.12 for the vector $\boldsymbol{\ell}$. Our aim is to apply Proposition 3.5.3 in the quotient space

$$V \coloneqq \mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})),$$

for the index set

$$\mathcal{I} \coloneqq \left\{ (B_1, \ldots, B_m) \;\middle|\; m \geq 1, B_i \in \mathcal{G}^*, \sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}}(B_i) \in \{\boldsymbol{\ell}, 2\boldsymbol{\ell}\} \right\},$$

that is, the set of tuples of words whose concatenation has Parikh Image $\boldsymbol{\ell}$ or $2\boldsymbol{\ell}$. For any element $\boldsymbol{x} \in \mathfrak{u}(n)$, denote by $\overline{\boldsymbol{x}} \coloneqq \boldsymbol{x} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ its equivalence class in

$$V = \mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})).$$

For any tuple $b = (B_1, \dots, B_m) \in \mathcal{I}$, consider the vectors in V:

$$\begin{aligned}
\boldsymbol{a}_{1b} &\coloneqq \overline{\sum_{i=1}^{m} \log B_i}, \\
\boldsymbol{a}_{kb} &\coloneqq \overline{H_k(\log B_1, \dots, \log B_m)}, \quad k = 2, \dots, 10,
\end{aligned}$$

and

$$P_b(t) \coloneqq \overline{\log(B_1^t \cdots B_m^t)} = t\boldsymbol{a}_{1b} + \sum_{k=2}^{10} t^k \boldsymbol{a}_{kb},$$

coming from the Baker-Campbell-Hausdorff formula for $B_1^t, \dots, B_m^t$. We now apply Proposition 3.5.3 to these vectors: we need to verify that the cones $\mathcal{C}_k, k = 10, \dots, 1$ as defined in Proposition 3.5.3 are indeed linear spaces. Proposition 3.5.12 shows that

$$\mathcal{C}_{10} = \langle \boldsymbol{a}_{10b} \mid b \in \mathcal{I} \rangle_{\mathbb{Q}_{\geq 0}} = \overline{\mathcal{R}_{10}(\boldsymbol{\ell})}$$

and

$$\mathcal{C}_k = \langle \boldsymbol{a}_{kb} \mid b \in \mathcal{I} \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_{k+1} = \overline{\mathcal{R}_k(\boldsymbol{\ell})}, \quad k = 9, \dots, 2,$$

are linear subspaces of $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$. Furthermore, by the condition $\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$, we have

$$\boldsymbol{a}_{1b} \in \left\{ \overline{\sum_{i=1}^{K} \ell_i \log A_i}, 2 \cdot \overline{\sum_{i=1}^{K} \ell_i \log A_i} \right\} \subseteq \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})} = \overline{\mathcal{R}_2(\boldsymbol{\ell})}$$

for all $b \in \mathcal{I}$. Hence,

$$\mathcal{C}_1 = \langle \boldsymbol{a}_{1b} \mid b \in \mathcal{I} \rangle_{\mathbb{Q}_{\geq 0}} + \mathcal{C}_2 = \overline{\mathcal{R}_2(\boldsymbol{\ell})} = \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}$$

is also a linear space. The conditions (i) and (ii) in Proposition 3.5.3 are thus satisfied. We can thus apply Proposition 3.5.3, which yields

$$\langle P_b(t) \mid b \in \mathcal{I}, t \in \mathbb{Z}_{\geq 0} \rangle_{\mathbb{Q}_{\geq 0}} = \mathcal{C}_1 = \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}.$$

In other words,

$$\left\langle \overline{\log(B_1^t \cdots B_m^t)} \;\middle|\; t \in \mathbb{Z}_{\geq 0}, m \geq 1, B_i \in \mathcal{G}^*, \sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}}(B_i) \in \{\boldsymbol{\ell}, 2\boldsymbol{\ell}\} \right\rangle_{\mathbb{Q}_{\geq 0}} = \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}.$$

Since $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$ is of finite dimension, this shows that there exist $K' > 0$

tuples of words $(B_{11},\dots,B_{1m})$, …, $(B_{K'1},\dots,B_{K'm})$ with $\sum_{i=1}^{m}\mathrm{PI}^{\mathcal{G}}(B_{ji})=\boldsymbol{\ell}$ or $2\boldsymbol{\ell}$ for all $j\in\{1,\dots,K'\}$, as well as positive integers $t_1,\dots,t_{K'}\in\mathbb{Z}_{>0}$, such that

$$\langle \log(B_{i1}^{t_i}\cdots B_{im}^{t_i}) \mid i=1,\dots,K'\rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) \\ = \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})).$$

Hence, the matrices $A'_i=B_{i1}^{t_i}\cdots B_{im}^{t_i}, i=1,\dots,K'$ satisfy the Equation (3.52). Define a new alphabet $\mathcal{G}'=\{A'_1,\dots,A'_{K'}\}\subseteq G$.

(2) **Find matrices $A''_1,\dots,A''_{K''}\in\langle\mathcal{G}'\rangle$ satisfying condition** (3.53)**.** Since the right hand side of Equation (3.52) is a linear space, we have $-\log A'_j\in\langle \log A'_i \mid i=1,\dots,K'\rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$ for $j=1,\dots,K'$. Hence, there exists a non-zero vector $\boldsymbol{\ell}'=(\ell'_1,\dots,\ell'_{K'})$ in $\mathbb{Z}_{\geq 0}^{K'}$, satisfying $\mathrm{supp}(\boldsymbol{\ell}')=\{1,\dots,K'\}$, $\ell'_i\geq 10$ for all $i\in\{1,\dots,K'\}$, and

$$\sum_{i=1}^{K'}\ell'_i\log A'_i\in\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})). \tag{3.55}$$

Define $\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')} \coloneqq \{\log A'_i \mid i\in\mathrm{supp}(\boldsymbol{\ell}')\} = \log\mathcal{G}'$, because $\mathrm{supp}(\boldsymbol{\ell}')=\{1,\dots,K'\}$. First, we claim that

$$\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) = \mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}). \tag{3.56}$$

Indeed, Equation (3.52) shows that

$$\left\langle \log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}\right\rangle_{\mathbb{Q}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) \\ = \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) = \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}).$$

Applying Lemma 3.5.11 with $W=\left\langle \log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}\right\rangle_{\mathbb{Q}}, V=\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$ to the above equation yields the equality (3.56). Consequently, we have

$$\sum_{i=1}^{K'}\ell'_i\log A'_i\in\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})$$

by (3.55). Apply Proposition 3.5.12 for the alphabet $\mathcal{G}'$ and the vector $\boldsymbol{\ell}'$, then we have that, in the quotient space

$$\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})),$$

the equations $\overline{\mathcal{R}_k(\boldsymbol{\ell}')} = \overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})}, k = 10, \ldots, 2$, hold. Then, applying Proposition 3.5.3 in the quotient linear space

$$V \coloneqq \mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}))$$

as in the previous step, we have

$$\left\langle \overline{\log(B_1^t \cdots B_m^t)} \;\middle|\; t \in \mathbb{Z}_{\geq 0}, m \geq 1, B_i \in \left(\mathcal{G}'\right)^*, \sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}'}(B_i) \in \{\boldsymbol{\ell}', 2\boldsymbol{\ell}'\} \right\rangle_{\mathbb{Q}_{\geq 0}} = \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})}.$$

Hence, there exist $K'' > 0$ tuples of words $(B'_{11}, \ldots, B'_{1m})$, …, $(B'_{K''1}, \ldots, B'_{K''m})$ in $(\mathcal{G}')^*$ with $\sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}'}(B'_{ji}) = \boldsymbol{\ell}'$ or $2\boldsymbol{\ell}'$ for all j, as well as positive integers $t'_1, \ldots, t'_{K''} \in \mathbb{Z}_{>0}$, such that

$$\langle \log({B'}_{i1}^{t'_i} \cdots {B'}_{im}^{t'_i}) \mid i = 1, \ldots, K'' \rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})) = \mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})). \quad (3.57)$$

Substituting with $\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$, Equation (3.57) can be rewritten as

$$\langle \log(B_{i1}^{t'_i} \cdots B_{im}^{t'_i}) \mid i = 1, \ldots, K'' \rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))).$$

Hence, the matrices $A''_i = {B'}_{i1}^{t'_i} \cdots {B'}_{im}^{t'_i}, i = 1, \ldots, K''$ satisfy the Equation (3.53). Define the new alphabet $\mathcal{G}'' = \{A''_1, \ldots, A''_{K''}\}$.

(3) **Find matrices $A'''_1, \ldots, A'''_{K'''} \in \langle \mathcal{G}'' \rangle$ satisfying condition (3.54).** Similar to the previous step, one can find a vector $\boldsymbol{\ell}'' = (\ell''_1, \ldots, \ell''_{K''}) \in \mathbb{Z}_{\geq 0}^{K''}$, satisfying $\mathrm{supp}(\boldsymbol{\ell}'') = \{1, \ldots, K''\}$, $\ell''_i \geq 10, i = 1, \ldots, K''$, and

$$\sum_{i=1}^{K''} \ell''_i \log A''_i \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})).$$

Define $\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')} \coloneqq \{\log A''_i \mid i \in \mathrm{supp}(\boldsymbol{\ell}'')\} = \log \mathcal{G}''$. As in the previous step, we have

$$\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')})).$$

Combining it with $\mathfrak{L}_{\geq 2}(\log \mathcal{G}'_{\mathrm{supp}(\boldsymbol{\ell}')}) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$, we have

$$\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))).$$

Apply Proposition 3.5.12 for the alphabet $\mathcal{G}''$ and the vector $\boldsymbol{\ell}''$, then we have that, in the quotient space $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}))$, the equations $\overline{\mathcal{R}_k(\boldsymbol{\ell}'')} = \overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')})}$, $k = 10, \ldots, 2$, hold.

Then, applying Proposition 3.5.3 in the quotient linear space

$$V \coloneqq \mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}))$$

as in the previous steps, we have

$$\left\langle \overline{\log(B_1^t \cdots B_m^t)} \;\middle|\; t \in \mathbb{Z}_{\geq 0}, m \geq 1, B_i \in \left(\mathcal{G}'\right)^*, \sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}''}(B_i) \in \{\boldsymbol{\ell}'', 2\boldsymbol{\ell}''\} \right\rangle_{\mathbb{Q}_{\geq 0}} = \overline{\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')})}.$$

Hence, there exist $K''' > 0$ tuples of words $(B''_{11}, \ldots, B''_{1m})$, …, $(B''_{K'''1}, \ldots, B''_{K'''m})$ in $(\mathcal{G}'')^*$ with $\sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}''}(B''_{ji}) = \boldsymbol{\ell}''$ or $2\boldsymbol{\ell}''$ for all j, as well as positive integers $t''_1, \ldots, t''_{K'''} \in \mathbb{Z}_{>0}$, such that

$$\begin{aligned}\langle \log({B''}_{i1}^{t''_i} \cdots {B''}_{im}^{t''_i}) \mid i = 1, \ldots, K''' \rangle_{\mathbb{Q}_{\geq 0}} + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')})) \\ = \mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')})). \quad (3.58)\end{aligned}$$

Since $\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')}) = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})))$, we have

$$\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}''_{\mathrm{supp}(\boldsymbol{\ell}'')})) \subseteq \mathfrak{L}_{\geq 16}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) = \{0\}.$$

Thus, Equation (3.58) can be rewritten as

$$\langle \log({B''}_{i1}^{t''_i} \cdots {B''}_{im}^{t''_i}) \mid i = 1, \ldots, K''' \rangle_{\mathbb{Q}_{\geq 0}} = \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))).$$

Hence, the matrices $A'''_i = {B''}_{i1}^{t''_i} \cdots {B''}_{im}^{t''_i}, i = 1, \ldots, K'''$ satisfy the Equation (3.54). Define the new alphabet $\mathcal{G}''' = \{A'''_1, \ldots, A'''_{K'''}\}$.

(4) **Find a word $w \in \langle \mathcal{G}''' \rangle$ with $\log w = 0$.** Since the right hand side of Equation (3.54) is a linear space, we have $-\log A'''_j \in \langle \log A'''_i \mid i = 1, \ldots, K''' \rangle_{\mathbb{Q}_{\geq 0}}$ for $j = 1, \ldots, K'''$. Hence,

there exists a non-zero vector $\boldsymbol{\ell}''' = (\ell'''_1, \ldots, \ell'''_{K'}) \in \mathbb{Z}_{\geq 0}^{K'''}$, satisfying

$$\sum_{i=1}^{K'''} \ell'''_i \log A'''_i = 0.$$

Since $\log \mathcal{G}''' \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))) \subseteq \mathfrak{L}_{\geq 8}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})$, we have

$$\mathfrak{L}_{\geq 2}(\log \mathcal{G}''') \subseteq \mathfrak{L}_{\geq 16}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}) = \{0\}.$$

Hence, by the Baker-Campbell-Hausdorff formula,

$$\log({A'''}_1^{\ell'''_1} \cdots {A'''}_{K'''}^{\ell'''_{K'''}}) = \sum_{i=1}^{K'''} \ell'''_i \log A'''_i = 0,$$

because the terms $H_k, k \geq 2$ are in $\mathfrak{L}_{\geq 2}(\log \mathcal{G}''')$, which vanishes. Therefore, we have found the non-empty word $w = {A'''}_1^{\ell'''_1} \cdots {A'''}_{K'''}^{\ell'''_{K'''}} \in (\mathcal{G}''')^*$ satisfying $\log w = 0$. By replacing A'''_i with their corresponding words ${B''}_{i1}^{t''_i} \cdots {B''}_{im}^{t''_i}$ in $(\mathcal{G}'')^*$, then replacing A''_i with corresponding words in $(\mathcal{G}')^*$, then replacing A'_i with corresponding words in $\mathcal{G}^*$, we see that w considered as a word in $\mathcal{G}^*$ has Parikh Image in $\mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$, because the words $B_{i1}^{t_i} \cdots B_{im}^{t_i}$ corresponding to A'_i all have Parikh Image in $\mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$.

□

3.6 Conjecture for higher nilpotency class

In Sections 3.4 and 3.5, we showed that the invertible subset of any finite set $\mathcal{G} \subseteq G$ is computable in polynomial time, where G is a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ of nilpotency class at most ten. The only obstacle for generalizing this result to higher nilpotency class is to prove Proposition 3.5.2 for $k \geq 11$. If the identities (3.17) exist for $k \geq 11$, then they can be found with the same computer aided procedure as the one used in the proof of Lemma 3.5.7-3.5.9 (see Section 3.10). Following this idea, given $k \geq 11$, we propose the following conjecture, which generalizes Proposition 3.5.2:

Conjecture 3.6.1. *Let $\mathcal{H} \subset \mathsf{UT}(n, \mathbb{Q})$ be any finite set of matrices. There exist an integer $r \geq 0$, positive rational numbers $\alpha_1, \ldots, \alpha_r$, as well as, for $s = 1, \ldots, r$, words $\boldsymbol{j}_s = j_{s,1} j_{s,2} \cdots j_{s,m_s}$ in the alphabet $\mathcal{I} = \{1, 2, \ldots, k+1\}$, such that $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) \in \mathbb{Z}_{>0} \cdot (1, 1, \ldots, 1)$ and*

$$\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(k+1)}) + \sum_{s=1}^{r} \alpha_s \sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(j_{s,1})}, \ldots, \log B_{\sigma(j_{s,m_s})})$$
$$\in \mathfrak{L}_{\geq k+1}(\log \mathcal{H}) + \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})) \quad (3.59)$$

for all matrices $B_1, \dots, B_{k+1}$ *in* $\mathsf{UT}(n, \mathbb{Q})$ *satisfying* $\log B_i \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ *and* $\sum_{i=1}^{k+1} \log B_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$.

For even k, Conjecture 3.6.1 is correct by the antisymmetry of H_k (Lemma 3.5.5). For $k = 3$, it is correct by taking $r = 0$ and using Lemma 3.5.6. For $k = 5, 7, 9$, it is verified by Lemma 3.5.7, 3.5.9, 3.5.10, where the words $\boldsymbol{j}_s, s = 1, \dots, r$, all satisfy $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) = (2, 2, \dots, 2)$.

For odd k larger than 10, using Algorithm 3.3 in Section 3.10, we can search for words $\boldsymbol{j}_s$ that potentially verify Conjecture 3.6.1. Namely, starting with $q = 2$, take all the words $\boldsymbol{j}_s$ satisfying $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) = (p, p, \dots, p), 2 \leq p \leq q$. Under the equivalence relation $\sim$ (defined in the proof of Lemma 3.5.7), we can write each expression

$$h_k(\boldsymbol{j}_s) \coloneqq \sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(j_{s,1})}, \dots, \log B_{\sigma(j_{s,m_s})})$$

as a linear combination of expressions $\widehat{M}(P, c)$ (see Section 3.10) using Algorithm 3.3. Then, writing $-\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(\log B_{\sigma(1)}, \dots, \log B_{\sigma(k+1)})$ also as a linear combination of $\widehat{M}(P, c)$, we can verify whether it is in the $\mathbb{Q}_{\geq 0}$-cone generated by the elements $h_k(\boldsymbol{j}_s)$. If this is the case, then there exist positive rational numbers $\alpha_s, s = 1, 2, \dots$, satisfying Equation (3.59). If this is not the case, we can increase q and repeat the above procedure.

If there exists a relation of the form (3.59), then the above procedure terminates for some q and returns this relation. Otherwise it does not terminate. In practice, it is more computationally viable to not take all the words $\boldsymbol{j}_s$ satisfying $\mathrm{PI}^{\mathcal{I}}(\boldsymbol{j}_s) = (p, p, \dots, p)$, but only a small amount of them chosen randomly.

Due to the restraint on computational power, we have only verified Conjecture 3.6.1 for all $k \leq 10$, this is the reason why the our result stops at nilpotency class ten. However, if we can verify Conjecture 3.6.1 for larger k (it suffices to verify for odd k), then we can extend our result to higher nilpotency class. This is formalized by the following theorem.

Theorem 3.6.2. *Let* G *be a subgroup of* $\mathsf{UT}(n, \mathbb{Q})$ *whose nilpotency class is at most* d. *If Conjecture 3.6.1 holds for all* $k \leq d$, *then Algorithm 3.1 correctly computes the invertible subset of any finite set* $\mathcal{G} \subseteq G$ *in polynomial time.*

Proof. For any $\boldsymbol{\ell} \in \mathbb{Z}_{\geq 0}^K$, similar to Equation (3.47), define recursively the cones

$$\mathcal{R}_{d+1}(\boldsymbol{\ell}) \coloneqq \{0\},$$
$$\mathcal{R}_k(\boldsymbol{\ell}) \coloneqq \mathcal{R}_{k+1}(\boldsymbol{\ell}) + \left\langle H_k(\log B_1, \dots, \log B_m) \;\middle|\; m \geq 1, B_i \in \mathcal{G}^*, \sum_{i=1}^{m} \mathrm{PI}^{\mathcal{G}}(B_i) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell} \right\rangle_{\mathbb{Q}_{\geq 0}},$$
$$k = d, d-1, \dots, 3, 2,$$

and the set

$$\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})} \coloneqq \{\log A_i \mid A_i \in \mathcal{G}, i \in \mathrm{supp}(\boldsymbol{\ell})\}.$$

Suppose $\boldsymbol{\ell}$ satisfies $\ell_i \geq d, i \in \mathrm{supp}(\boldsymbol{\ell})$. Consider the quotient space $\mathfrak{u}(n)/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}))$. Following the pattern in the proof of Proposition 3.5.12, we can show that for all $k \leq d+1$, the cone $\overline{\mathcal{R}_k(\boldsymbol{\ell})}$ is equal to the linear space $\overline{\mathfrak{L}_{\geq k}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})})}$.

Then, using the same arguments as Theorem 3.4.2, we can show the following generalization of Theorem 3.4.2:

(i) If there exists a word $w \in \mathcal{G}^*$ with $\mathrm{PI}^{\mathcal{G}}(w) = \boldsymbol{\ell}$ and $\log w = 0$, then

$$\sum_{i=1}^{K} \ell_i \log A_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{G}_{\mathrm{supp}(\boldsymbol{\ell})}). \tag{3.60}$$

(ii) If $\boldsymbol{\ell}$ satisfies (3.60), then there exists a non-empty word $w \in \mathcal{G}^*$, with $\mathrm{PI}^{\mathcal{G}}(w) \in \mathbb{Z}_{>0} \cdot \boldsymbol{\ell}$, such that $\log w = 0$.

From here, the proof of correctness of Algorithm 3.1 and its complexity analysis is identical to the proof of Theorem 3.1.1, replacing the property $[\log \mathcal{G}]_{11} = \{0\}$ by $[\log \mathcal{G}]_{d+1} = \{0\}$. □

A natural question is whether our result can be extended to arbitrary nilpotency class d. This can either be done by proving Conjecture 3.6.1 for higher k or by finding another way to approach this problem. In particular, similar to Corollary 3.1.2, this would yield the decidability for the Identity Problem and the Group Problem for arbitrary finitely generated nilpotent groups of class at most d.

3.7 Combinatorics for length two subwords

Starting from this section we study Semigroup Intersection and Orbit Intersection in subgroups of $\mathsf{UT}(n, \mathbb{Q})$. In this section we introduce tools from combinatorics which will be crucial in solving these two problems.

Fix a finite alphabet $\mathcal{G} = \{A_1, \ldots, A_K\}$. For $1 \leq i < j \leq K$, let w be a word over the alphabet $\mathcal{G}$, denote by $\delta_{ij}^{\mathcal{G}}(w)$ the number of occurrences of the subword $\cdots A_i \cdots A_j \cdots$ minus the number of occurrences of the subword $\cdots A_j \cdots A_i \cdots$ in w. That is, writing $w = B_1 B_2 \cdots B_s$, we have

$$\delta_{ij}^{\mathcal{G}}(w) \coloneqq \delta_{ij}^{\mathcal{G},+}(w) - \delta_{ij}^{\mathcal{G},-}(w),$$

where

$$\delta_{ij}^{\mathcal{G},+}(w) \coloneqq \{(u,v) \mid 1 \leq u < v \leq s, B_u = A_i, B_v = A_j\},$$
$$\delta_{ij}^{\mathcal{G},-}(w) \coloneqq \{(u,v) \mid 1 \leq u < v \leq s, B_u = A_j, B_v = A_i\}$$

If the alphabet $\mathcal{G}$ is clear from the context, we write $\delta_{ij}(w), \delta_{ij}^{\pm}(w)$ instead of $\delta_{ij}^{\mathcal{G}}(w), \delta_{ij}^{\mathcal{G},\pm}(w)$. Obviously, we have the parity constraint

$$\delta_{ij}(w) \equiv \delta_{ij}^{+}(w) + \delta_{ij}^{-}(w) = \mathrm{PI}_i(w) \cdot \mathrm{PI}_j(w) \mod 2. \tag{3.61}$$

Let G be a class two nilpotent subgroup of $\mathsf{UT}(n,\mathbb{Q})$. Fix a finite alphabet $\mathcal{G} = \{A_1, \ldots, A_K\}$ in G. Given a word $w \coloneqq B_1 B_2 \cdots B_s$ in $\mathcal{G}^*$, specializing the Baker-Campbell-Hausdorff formula (3.4) to nilpotency class two yields

$$\log(B_1 B_2 \cdots B_m) = \sum_{i=1}^{m} \log B_i + \frac{1}{2} \sum_{1 \leq i < j \leq s} [\log B_i, \log B_j]. \tag{3.62}$$

Let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K)$ be the Parikh Image of w. Regrouping the terms in Equation (3.62) yields

$$\begin{aligned}\log w = \sum_{i=1}^{K} \ell_i \log A_i + \frac{1}{2} \sum_{1 \leq i < j \leq K} \Big(\delta_{ij}^{+}(w)[\log A_i, \log A_j] + \delta_{ij}^{-}(w)[\log A_j, \log A_i] \Big) \\ = \sum_{i=1}^{K} \ell_i \log A_i + \frac{1}{2} \sum_{1 \leq i < j \leq K} \delta_{ij}(w)[\log A_i, \log A_j]. \end{aligned} \tag{3.63}$$

The second equality follows from the anticommutativity of the Lie bracket.

Now let us describe the general strategy for solving intersection-type decision problems. Consider a simple example: given two alphabets $\mathcal{G} = \{A_1, \ldots, A_K\}$, $\mathcal{H} = \{B_1, \ldots, B_M\}$ in a class two nilpotent subgroup of $\mathsf{UT}(n,\mathbb{Q})$, we want to decide whether $\langle \mathcal{G} \rangle \cap \langle \mathcal{H} \rangle \neq \emptyset$. This boils down to finding two non-empty words v, w respectively in the alphabets $\mathcal{G}$ and $\mathcal{H}$, such that $\log v = \log w$. Denote by $\boldsymbol{x} = (x_1, \ldots, x_K)$ the Parikh Image of v, and by $\boldsymbol{y} = (y_1, \ldots, y_M)$ the Parikh Image of w, then formula (3.63) yields the equivalence between $\log v = \log w$ and

$$\begin{aligned}\sum_{i=1}^{K} x_i \log A_i + \sum_{i<j} \frac{\delta_{ij}^{\mathcal{G}}(v)}{2} [\log A_i, \log A_j] = \sum_{i=1}^{M} y_i \log B_i + \sum_{i<j} \frac{\delta_{ij}^{\mathcal{H}}(w)}{2} [\log B_i, \log B_j], \\ \boldsymbol{x} \in \mathbb{Z}_{\geq 0}^{K},\ \boldsymbol{y} \in \mathbb{Z}_{\geq 0}^{M}, \quad \mathrm{PI}^{\mathcal{G}}(v) = \boldsymbol{x},\ \mathrm{PI}^{\mathcal{H}}(w) = \boldsymbol{y}. \end{aligned} \tag{3.64}$$

Hence, deciding whether $\langle \mathcal{G} \rangle \cap \langle \mathcal{H} \rangle \neq \emptyset$ boils down to solving Equation (3.64) in the numerical

variables $\boldsymbol{x}, \boldsymbol{y}$ and the word variables v, w over alphabets $\mathcal{G}, \mathcal{H}$.

Consider a "relaxed" version of this problem. That is, we replace $\delta_{ij}^{\mathcal{G}}(v)$ and $\delta_{ij}^{\mathcal{H}}(w)$ by new variables c_{ij}, d_{ij} over integers, without imposing any constraint. This gives the equation

$$\sum_{i=1}^{K} x_i \log A_i + \sum_{1 \le i < j \le K} \frac{c_{ij}}{2} [\log A_i, \log A_j] = \sum_{i=1}^{M} y_i \log B_i + \sum_{1 \le i < j \le M} \frac{d_{ij}}{2} [\log B_i, \log B_j],$$
$$\boldsymbol{x} \in \mathbb{Z}_{\ge 0}^{K},\ \boldsymbol{y} \in \mathbb{Z}_{\ge 0}^{M}, \quad c_{ij}, d_{ij} \in \mathbb{Z} \text{ for all } i, j. \tag{3.65}$$

Obviously, if Equation (3.64) has a solution, then the relaxed version (3.65) will also admit a solution. The converse is not necessarily true. The implicit constraints imposed by the word combinatorial variables $\delta_{ij}^{\mathcal{G}}(v), \delta_{ij}^{\mathcal{H}}(w)$ in Equation (3.64) are highly non-trivial. (For example, one should at least have $|\delta_{ij}^{\mathcal{G}}(v)| \le x_i x_j$ for all i, j). However, these constraints are not reflected by the numerical variables c_{ij} and d_{ij} in Equation (3.65).

The key idea of subsequent sections is the following surprising fact. For the two problems we consider (Semigroup Intersection and Orbit Intersection), it is sufficient to solve the relaxed version of the equation, plus several simple constraints (such as the "modulo 2" constraint in Equation (3.61)). In particular, given a "suitable" solution to the relaxed Equation (3.65), we can always construct a solution to Equation (3.64). *A priori*, the values of $\delta_{ij}^{\mathcal{G}}(v)$ cannot reach all integers like the free variables c_{ij}; nevertheless, when $x_1, \ldots, x_K$ tend towards infinity, the vector $\left(\delta_{ij}^{\mathcal{G}}(v)\right)_{1 \le i < j \le K}$ can in fact reach every value within a ball of radius size $O(|\boldsymbol{x}|^2)$, satisfying modulo 2 constraints. This will suffice to construct a suitable word v, as the quadratic radius will eventually dominate the linear term $\sum_{i=1}^{K} x_i \log A_i$.

This section aims to formalize this idea. The main result of this section will be Proposition 3.7.2. First, we prove a simple case where the alphabet consists of two letters.

Lemma 3.7.1. *Given an alphabet $\mathcal{G} = \{A_i, A_j\}$ and non-negative integers $s_i, s_j \in \mathbb{Z}_{\ge 0}$, then for every $C \in \mathbb{Z}$ satisfying*

$$|C| \le s_i s_j \quad \textit{and} \quad C \equiv s_i s_j \mod 2, \tag{3.66}$$

there exists a permutation w of the word $A_i^{s_i} A_j^{s_j}$ such that $\delta_{ij}(w) = C$.

Proof. For an illustration of the proof, see Figure 3.1. We start with the word $w = A_i^{s_i} A_j^{s_j}$, which satisfies $\delta_{ij}(w) = s_i s_j$. We gradually swap pairs of consecutive letters in w: each time we replace an occurrence of consecutive $A_i A_j$ with $A_j A_i$. An occurrence of $A_i A_j$ can always be found unless we have reached the "final" permutation $A_j^{s_j} A_i^{s_i}$. It is easy to see that each swap reduces the value of $\delta_{ij}(w)$ by 2. Therefore, by swapping consecutive $A_i A_j$ one by one, $\delta_{ij}(w)$

can reach every value between $\delta_{ij}(A_i^{s_i}A_j^{s_j}) = s_is_j$ and $\delta_{ij}(A_j^{s_j}A_i^{s_i}) = -s_is_j$ that has the same parity with s_is_j. This proves the lemma. □

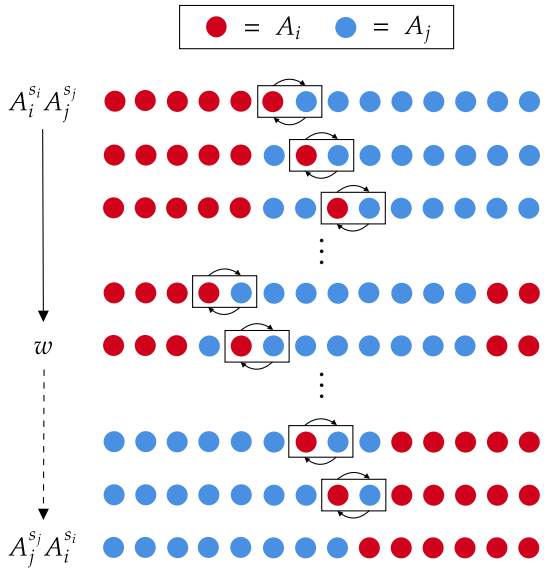

Figure 3.1: Illustration for the proof of Lemma 3.7.1.

We then prove the main result of this section, which generalizes Lemma 3.7.1 to alphabets of more than two letters.

Proposition 3.7.2. *Fix a finite alphabet $\mathcal{G}$ of size $K \geq 2$. Then for any tuples $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_K) \in \mathbb{Z}_{\geq 0}^K$ and $\{C_{ij}\}_{1 \leq i < j \leq K} \in \mathbb{Z}_{\geq 0}^{K(K-1)/2}$ satisfying*

$$|C_{ij}| \leq \frac{\ell_i \ell_j}{4K^2} - 2K(\ell_i + \ell_j) - 4K^2, \quad \text{for all } 1 \leq i < j \leq K, \tag{3.67}$$

and

$$C_{ij} \equiv \ell_i \ell_j \mod 2, \quad \text{for all } 1 \leq i < j \leq K, \tag{3.68}$$

there exists a word w with Parikh Image $\boldsymbol{\ell}$ such that

$$\delta_{ij}(w) = C_{ij}, \quad \text{for all } 1 \leq i < j \leq K. \tag{3.69}$$

Proof. For an illustration of the proof, see Figure 3.2. For all i, write $\ell_i = 2(K-1)s_i + r_i$ with $0 \leq r_i < 2(K-1)$. Consider the word $W_{init} \coloneqq W_{res} \cdot W \cdot W_{rev}$, where

$$\begin{aligned}
W_{res} &\coloneqq A_1^{r_1} A_2^{r_2} \cdots A_K^{r_K}, \\
W &\coloneqq (A_1^{s_1} A_2^{s_2})\,(A_1^{s_1} A_3^{s_3}) \cdots (A_1^{s_1} A_K^{s_K})\,(A_2^{s_2} A_3^{s_3}) \cdots (A_2^{s_2} A_K^{s_K})\,(A_3^{s_3} A_4^{s_4}) \cdots (A_{K-1}^{s_{K-1}} A_K^{s_K}), \\
W_{rev} &\coloneqq (A_K^{s_K} A_{K-1}^{s_{K-1}})\,(A_K^{s_K} A_{K-2}^{s_{K-2}}) \cdots (A_K^{s_K} A_1^{s_1})\,(A_{K-1}^{s_{K-1}} A_{K-2}^{s_{K-2}})\,(A_{K-1}^{s_{K-1}} A_{K-3}^{s_{K-3}}) \cdots (A_2^{s_2} A_1^{s_1}).
\end{aligned}$$

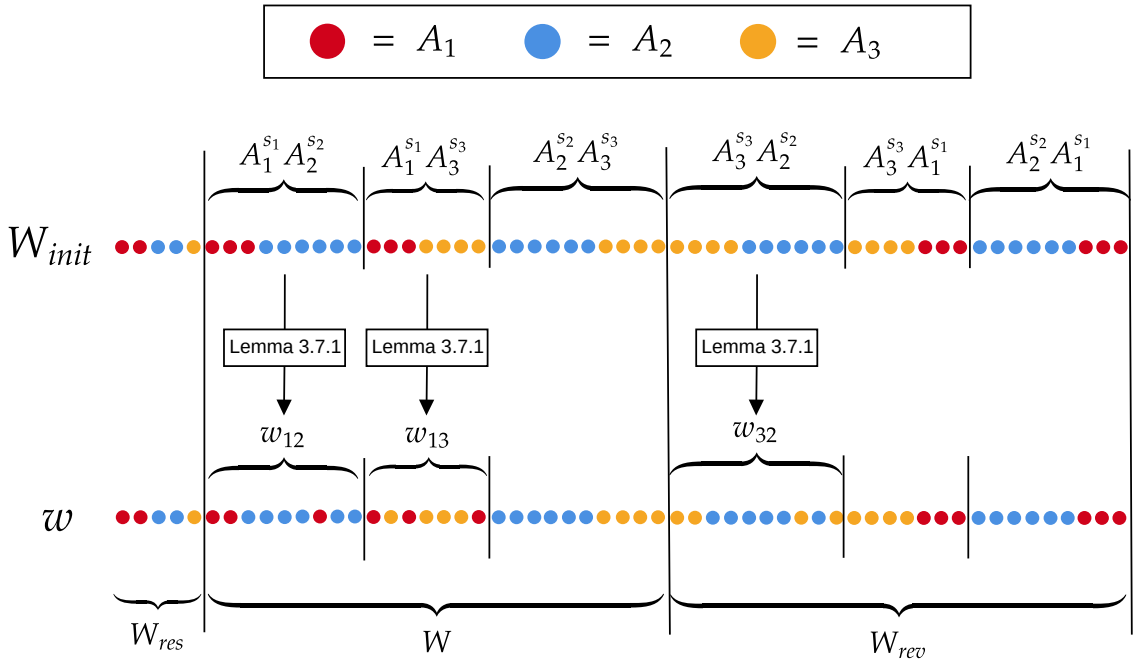

Figure 3.2: Illustration for the proof of Proposition 3.7.2.

In particular, W is the concatenation of all words of the form $A_i^{s_i}A_j^{s_j}$ where $i < j$, and W_{rev} is the reverse of W. It is easy to verify that W_{init} contains ℓ_i occurrences of the letter A_i, so its Parikh Image is exactly $\boldsymbol{\ell}$.

We now compute $\delta_{ij}(W_{init})$ for $i < j$. Since $W \cdot W_{rev}$ is a palindrome, we have $\delta_{ij}(W \cdot W_{rev}) = 0$, so

$$\begin{aligned}\delta_{ij}(W_{init}) &= \delta_{ij}(W_{res}) + \mathrm{PI}_i(W_{res})\,\mathrm{PI}_j(W \cdot W_{rev}) - \mathrm{PI}_j(W_{res})\,\mathrm{PI}_i(W \cdot W_{rev}) \\ &= r_i r_j + r_i \cdot 2(K-1)s_j - r_j \cdot 2(K-1)s_i. \end{aligned} \tag{3.70}$$

In particular, since $0 \le r_i < 2(K-1)$, we have

$$|\delta_{ij}(W_{init})| \le 4(K-1)^2 + 2(K-1)^2(s_j + s_i) < 4K^2 + 2(K-1)(\ell_i + \ell_j) \tag{3.71}$$

By Condition (3.67), we have

$$\begin{aligned}|\delta_{ij}(W_{init}) - C_{ij}| &\le |\delta_{ij}(W_{init})| + |C_{ij}| \\ &< 4K^2 + 2(K-1)(\ell_i + \ell_j) + \frac{\ell_i \ell_j}{4K^2} - 2K(\ell_i + \ell_j) - 4K^2 \\ &= \frac{\ell_i \ell_j}{4K^2} - 2(\ell_i + \ell_j) \\ &< \frac{\ell_i \ell_j}{4(K-1)^2} - \frac{\ell_i + \ell_j}{2(K-1)} + 1\end{aligned}$$

$$= \left(\frac{\ell_i}{2(K-1)} - 1\right)\left(\frac{\ell_j}{2(K-1)} - 1\right)$$
$$< s_i s_j. \tag{3.72}$$

Since Equation (3.70) yields $\delta_{ij}(W_{init}) \equiv r_i r_j \equiv \ell_i \ell_j \mod 2$, Condition (3.68) then gives

$$\delta_{ij}(W_{init}) \equiv C_{ij} \mod 2. \tag{3.73}$$

We now show how to construct the word w. Starting with the word W_{init}, for every pair $i < j$ perform the following:

1. If $\delta_{ij}(W_{init}) > C_{ij}$. By Lemma 3.7.1 there exists a permutation w_{ij} of the word $A_i^{s_i} A_j^{s_j}$ such that $\delta_{ij}(w_{ij}) = s_i s_j + C_{ij} - \delta_{ij}(W_{init})$. Indeed, Equations (3.72) and (3.73) guarantee that the conditions (3.66) in Lemma 3.7.1 are satisfied. We then replace the subword $A_i^{s_i} A_j^{s_j}$ in the W-part of W_{init} with the word w_{ij}. The resulting new word W'_{init} will satisfy

   $$\delta_{ij}(W'_{init}) = \delta_{ij}(W_{init}) - \delta_{ij}(A_i^{s_i} A_j^{s_j}) + \delta_{ij}(w_{ij}) = C_{ij}.$$

   This replacement does not change $\delta_{uv}(W_{init})$ for $(u, v) \neq (i, j)$.

2. If $\delta_{ij}(W_{init}) < C_{ij}$. By Lemma 3.7.1 there exists a permutation w_{ji} of the word $A_j^{s_j} A_i^{s_i}$ such that $\delta_{ij}(w_{ji}) = -s_i s_j + C_{ij} - \delta_{ij}(W_{init})$. Again, Equations (3.72) and (3.73) guarantee that the conditions (3.66) in Lemma 3.7.1 are satisfied. We then replace the subword $A_j^{s_j} A_i^{s_i}$ in the W_{rev}-part of W_{init} with the word w_{ji}. The resulting new word W'_{init} will satisfy

   $$\delta_{ij}(W'_{init}) = \delta_{ij}(W_{init}) - \delta_{ij}(A_j^{s_j} A_i^{s_i}) + \delta_{ij}(w_{ji}) = C_{ij}.$$

   This replacement does not change $\delta_{uv}(W_{init})$ for $(u, v) \neq (i, j)$.

3. If $\delta_{ij}(W_{init}) = C_{ij}$, do not perform any change.

Performing all these replacements on W_{init} for all pairs $i < j$ simultaneously, the resulting word w then satisfies $\delta_{ij}(w) = C_{ij}$ for all $1 \leq i < j \leq K$. □

3.8 Semigroup Intersection

We prove Theorem 3.1.3 in this section. Let G be subgroup of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most two. Let

$$\mathcal{G}_1 = \{A_{11}, A_{12}, \ldots, A_{1K_1}\}, \ldots, \mathcal{G}_M = \{A_{M1}, A_{M2}, \ldots, A_{MK_M}\}$$

be M sets of matrices in G. The following proposition shows that Semigroup Intersection can be reduced to solving homogeneous linear Diophantine equations with extra constraints on supports.

Proposition 3.8.1. *We have $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle \neq \emptyset$ if and only if there exist non-zero vectors $\boldsymbol{\ell}_1 \in \mathbb{Z}_{\geq 0}^{K_1} \setminus \{\mathbf{0}\}, \cdots, \boldsymbol{\ell}_M \in \mathbb{Z}_{\geq 0}^{K_M} \setminus \{\mathbf{0}\}$ as well as rational numbers c_{mij} for $1 \leq m \leq M, i, j \in \operatorname{supp}(\boldsymbol{\ell}_m)$, such that*

$$\begin{aligned}\sum_{j=1}^{K_1} \ell_{1j} \log A_{1j} + \sum_{\substack{i<j \\ i,j \in \operatorname{supp}(\boldsymbol{\ell}_1)}} c_{1ij} [\log A_{1i}, \log A_{1j}] \\ = \sum_{j=1}^{K_2} \ell_{2j} \log A_{2j} + \sum_{\substack{i<j \\ i,j \in \operatorname{supp}(\boldsymbol{\ell}_2)}} c_{2ij} [\log A_{2i}, \log A_{2j}] = \cdots \\ = \sum_{j=1}^{K_M} \ell_{Mj} \log A_{Mj} + \sum_{\substack{i<j \\ i,j \in \operatorname{supp}(\boldsymbol{\ell}_M)}} c_{Mij} [\log A_{Mi}, \log A_{Mj}]. \quad (3.74)\end{aligned}$$

Proof. If $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle \neq \emptyset$, let g be an element in the intersection. There exist non-empty words $w_1, \ldots, w_M$ over the alphabets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ such that $\log g = \log w_1 = \cdots = \log w_M$. By the Baker-Campbell-Hausdorff formula (3.63),

$$\log g = \sum_{j=1}^{K_m} \mathrm{PI}_j(w_m) \log A_{mj} + \sum_{\substack{i<j \\ i,j \in \operatorname{supp}(\boldsymbol{\ell}_m)}} \frac{\delta_{ij}(w_m)}{2} [\log A_{mi}, \log A_{mj}], \quad \text{for } m = 1, \ldots, M.$$

This shows that (3.74) is satisfied by $\boldsymbol{\ell}_m \coloneqq \mathrm{PI}^{\mathcal{G}_m}(w_m)$ and $c_{mij} \coloneqq \delta_{ij}(w_m)/2$ for $1 \leq m \leq M, i, j \in \operatorname{supp}(\boldsymbol{\ell}_m)$.

For the other implication, suppose such non-zero vectors $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_M$ and the rational numbers c_{mij} exist. Then there exists $H \in \mathfrak{u}(n)$ such that

$$\sum_{j=1}^{K_m} \ell_{mj} \log A_{mj} + \sum_{\substack{i<j \\ i,j \in \operatorname{supp}(\boldsymbol{\ell}_m)}} \frac{2c_{mij}}{2} [\log A_{mi}, \log A_{mj}] = H, \quad \text{for } m = 1, \ldots, M. \quad (3.75)$$

Note that if $i, j \in \operatorname{supp}(\boldsymbol{\ell}_m)$ then $\ell_{mi}\ell_{mj} \neq 0$.

By homogeneity, for any $N \in \mathbb{Z}_{>0}$, the vectors $N\boldsymbol{\ell}_1, \ldots, N\boldsymbol{\ell}_M$ and Nc_{mij} also satisfy Condition (3.74). Hence, multiplying all ℓ_{ij}, c_{mij} and H by a common denominator, we can suppose all ℓ_{ij} and c_{mij} to be integers. Denote $K = \max_{1 \leq m \leq M} K_m$, then there exists a large enough

even integer $N \in \mathbb{Z}_{>0}$ such that

$$|N \cdot 2c_{mij}| \leq \frac{N^2 \ell_{mi}\ell_{mj}}{4K^2} - 2NK(\ell_i + \ell_j) - 4K^2 \tag{3.76}$$

for $1 \leq m \leq M, i, j \in \mathrm{supp}(\boldsymbol{\ell}_m)$. This is because $\ell_{mi}\ell_{mj} > 0$, so the right hand side of (3.76) is quadratic in N and dominates the linear term on the left for large enough N. Replace all ℓ_{ij} with $N\ell_{ij}$, all c_{mij} with Nc_{mij}, and H with $N \cdot H$, then the new variables satisfy $2c_{mij} \equiv 0 \equiv \ell_{mi}\ell_{mj}$ mod 2, and

$$|2c_{mij}| \leq \frac{\ell_{mi}\ell_{mj}}{4K^2} - 2K(\ell_i + \ell_j) - 4K^2 \leq \frac{\ell_{mi}\ell_{mj}}{4K_m^2} - 2K_m(\ell_i + \ell_j) - 4K_m^2 \tag{3.77}$$

for all i, j, m. Equation (3.75) is still satisfied after the variable replacements. Therefore, by Proposition 3.7.2, there exist words $w_1, \ldots, w_M$ over the alphabets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ such that $\mathrm{PI}(w_m) = \boldsymbol{\ell}_m$ and $\delta_{ij}(w_m) = 2c_{mij}$ for all $1 \leq m \leq M, i, j \in \mathrm{supp}(\boldsymbol{\ell}_m)$. These words are non-empty since $\boldsymbol{\ell}_m \neq \mathbf{0}$. Plugging into the Baker-Campbell-Hausdorff formula (3.63), we have

$$\log w_m = \sum_{j=1}^{K_m} \ell_{mj} \log A_{mj} + \sum_{\substack{i<j \\ i,j \in \mathrm{supp}(\boldsymbol{\ell}_m)}} \frac{2c_{mij}}{2}[\log A_{mi}, \log A_{mj}] = H \quad \text{for } m = 1, \ldots, M.$$

This shows that $H \in \bigcap_{i=1}^M \log\langle \mathcal{G}_i \rangle$, so $(\exp H) \in \bigcap_{i=1}^M \langle \mathcal{G}_i \rangle \neq \emptyset$. □

Using Proposition 3.8.1, we devise Algorithm 3.2 that decides Semigroup Intersection.

Proposition 3.8.2. *Algorithm 3.2 is correct and terminates in polynomial time.*

Proof. We prove that Algorithm 3.2 outputs **False** if and only if $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle \neq \emptyset$.

After each iteration of Step 2, $\mathrm{card}(S_1) + \cdots + \mathrm{card}(S_M)$ strictly decreases. Therefore, the algorithm terminates after at most $K_1 + \cdots + K_M$ iterations of Step 2.

We now show correctness of the algorithm. We first show that if Algorithm 3.2 returns **False**, then $\bigcap_{i=1}^M \langle \mathcal{G}_i \rangle \neq \emptyset$. Suppose the algorithm terminates with output **False**, the condition in Step 2(d) shows that $\mathrm{supp}(\Lambda) \cap S_m = S_m$ for all $1 \leq m \leq M$. By the additivity of Λ (that is, $\boldsymbol{a}, \boldsymbol{b} \in \Lambda \implies \boldsymbol{a}+\boldsymbol{b} \in \Lambda$), there exists a vector $\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_M) \in \Lambda$ such that $\mathrm{supp}(\boldsymbol{\ell}) = \mathrm{supp}(\Lambda)$. This yields $\mathrm{supp}(\boldsymbol{\ell}_m) = \mathrm{supp}(\Lambda) \cap S_m = S_m$ for all m. Since $\mathrm{supp}(\boldsymbol{\ell}_m) = S_m \neq \emptyset$, we have $\boldsymbol{\ell}_m \neq \mathbf{0}$ for all $1 \leq m \leq M$. By the definition (3.79) of $\pi_{\boldsymbol{\ell}}(W)$, there exist rational numbers $(c_{mij})_{1 \leq m \leq M, i,j \in S_m}$ such that

$$\sum_{j=1}^{K_1} \ell_{1j} \log A_{1j} + \sum_{\substack{i<j \\ i,j \in S_1}} c_{1ij}[\log A_{1i}, \log A_{1j}] = \sum_{j=1}^{K_2} \ell_{2j} \log A_{2j} + \sum_{\substack{i<j \\ i,j \in S_2}} c_{2ij}[\log A_{2i}, \log A_{2j}]$$

Algorithm 3.2 Algorithm for Semigroup Intersection

Input: M finite sets of matrices $\mathcal{G}_1 = \{A_{11}, A_{12}, \ldots, A_{1K_1}\}, \ldots, \mathcal{G}_M = \{A_{M1}, A_{M2}, \ldots, A_{MK_M}\}$ in the group G.

Output: **True** (intersection is empty) or **False** (intersection is not empty).

1 **Initialization.** Set $S_1 \coloneqq \{1, 2, \ldots, K_1\}, \ldots, S_M \coloneqq \{1, 2, \ldots, K_M\}$.

2 **Main loop.** Repeat the following

(a) Represent the $\mathbb{Q}$-linear subspace of $V \coloneqq \mathbb{Q}^{\sum_{m=1}^{M} K_m + \sum_{m=1}^{M} \mathrm{card}(S_m)(\mathrm{card}(S_m)-1)/2}$:

$$W \coloneqq \Bigg\{ \big((\ell_{mj})_{1\le m\le M, 1\le j\le K_m}, (c_{mij})_{1\le m\le M, i,j\in S_m}\big) \in V \;\Bigg|\; \sum_{j=1}^{K_1} \ell_{1j} \log A_{1j} + \sum_{\substack{i<j \\ i,j\in S_1}} c_{1ij}[\log A_{1i}, \log A_{1j}] = \cdots = \sum_{j=1}^{K_M} \ell_{Mj} \log A_{Mj} + \sum_{\substack{i<j \\ i,j\in S_M}} c_{Mij}[\log A_{Mi}, \log A_{Mj}] \Bigg\} \quad (3.78)$$

as the solution set of homogeneous linear equations.

(b) Compute the projection of W onto the coordinates $(\ell_{mj})_{1\le m\le M, 1\le j\le K_m}$:

$$\pi_{\boldsymbol{\ell}}(W) \coloneqq \Bigg\{ (\ell_{mj})_{1\le m\le M, 1\le j\le K_m} \in \mathbb{Q}^{\sum_{m=1}^{M} K_m} \;\Bigg|\; \exists (c_{mij})_{1\le m\le M, i,j\in S_m}, \big((\ell_{mj})_{1\le m\le M, 1\le j\le K_m}, (c_{mij})_{1\le m\le M, i,j\in S_m}\big) \in W \Bigg\} \quad (3.79)$$

represented as the solution set of homogeneous linear equations.

(c) Define $\Lambda \coloneqq \mathbb{Z}_{\ge 0}^{\sum_{m=1}^{M} K_m} \cap \pi_{\boldsymbol{\ell}}(W)$ and compute $\mathrm{supp}(\Lambda)$ using Lemma 3.3.4.

(d) If $\mathrm{supp}(\Lambda) \cap S_m = S_m$ for all $1 \le m \le M$, terminate the loop and go to Step 3. Otherwise, let $S_m \coloneqq \mathrm{supp}(\Lambda) \cap S_m$ for every m, and continue with Step 2.

3 **Output.**

(a) If $S_m = \emptyset$ for any $1 \le m \le M$, return **True**.

(b) Otherwise return **False**.

$$= \cdots = \sum_{j=1}^{K_M} \ell_{Mj} \log A_{Mj} + \sum_{\substack{i<j \\ i,j\in S_M}} c_{Mij}[\log A_{Mi}, \log A_{Mj}]. \quad (3.80)$$

Since $S_m = \mathrm{supp}(\boldsymbol{\ell}_m)$ for all m, Equation (3.80) is identical to Equation (3.74) in Proposition 3.8.1. Therefore Proposition 3.8.1 shows $\bigcap_{i=1}^{M} \langle \mathcal{G}_i \rangle \neq \emptyset$.

Next, we show that if $\bigcap_{i=1}^{M} \langle \mathcal{G}_i \rangle \neq \emptyset$, then Algorithm 3.2 returns **False**. Suppose $\bigcap_{i=1}^{M} \langle \mathcal{G}_i \rangle \neq \emptyset$. By Proposition 3.8.1, there exist $\boldsymbol{\ell}_1 = (\ell_{1j})_{1\le j\le K_1} \in \mathbb{Z}_{\ge 0}^{K_1} \setminus \{\mathbf{0}\}, \ldots, \boldsymbol{\ell}_M = (\ell_{Mj})_{1\le j\le K_M} \in \mathbb{Z}_{\ge 0}^{K_M} \setminus \{\mathbf{0}\}$, and rational numbers $(c_{mij})_{1\le m\le M, i,j\in \mathrm{supp}(\boldsymbol{\ell}_m)}$ that satisfies Equation (3.74) in Propo-

sition 3.8.1. We show that "$\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq S_m$ for all $1 \leq m \leq M$" is an invariant of the algorithm.

At initialization, we obviously have $\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq S_m = \{1, \ldots, K_m\}$. Before each iteration of Step 2(d), suppose we have $\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq S_m$ for all m, then Equation (3.74) shows that

$$(\ell_{mj})_{1 \leq m \leq M, 1 \leq j \leq K_m} \in \pi_{\boldsymbol{\ell}}(W).$$

Consequently, $\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq \mathrm{supp}(\Lambda)$, meaning $\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq S_m$ still holds after Step 2(d).

This invariant shows that $\mathrm{supp}(\boldsymbol{\ell}_m) \subseteq S_m$ for all m by the start of Step 3. Since $\boldsymbol{\ell}_m \in \mathbb{Z}_{\geq 0}^{K_m} \setminus \{\mathbf{0}\}$, the set $\mathrm{supp}(\boldsymbol{\ell}_m)$ is non-empty for every m. We conclude that $S_m \neq \emptyset$ for all m by the start of Step 3. Therefore, Algorithm 3.2 returns **False**.

Finally, we show that Algorithm 3.2 terminates in polynomial time. Recall that the algorithm terminates after at most $K_1 + \cdots + K_M$ iterations of Step 2. At each iteration of Step 2(b), the projection can be computed in polynomial time by eliminating the variables $(c_{mij})_{1 \leq m \leq M, i, j \in S_m}$ from the equations defining W. Then, at each iteration of Step 2(c) the support $\mathrm{supp}(\Lambda)$ is computed by Lemma 3.3.4. The total input size of the linear programming instances is polynomial with respect to the total bit length of the matrix entries in $\mathcal{G}_1, \ldots, \mathcal{G}_M$. Indeed, the total bit length of $\log A_{mi}$ and $[\log A_{mi}, \log A_{mj}]$ is at most of quadratic size in $\mathcal{G}_m$; and the projection performed in Step 2(b) can only alter the total entry bit size at most polynomially. From this, one can express $\pi_{\boldsymbol{\ell}}(W)$ as the solution set of a system of homogeneous linear equations whose total bit length is polynomial in $\mathcal{G}_1, \ldots, \mathcal{G}_M$. Hence Lemma 3.3.4 computes the support of $\Lambda \coloneqq \mathbb{Z}_{\geq 0}^{\sum_{m=1}^{M} K_m} \cap \pi_{\boldsymbol{\ell}}(W)$ in polynomial time. Therefore, each iteration of Step 2 takes polynomial time, and thus the overall complexity of Algorithm 3.2 is polynomial with respect to the input $\mathcal{G}_1, \ldots, \mathcal{G}_M$. □

Theorem 3.1.3. *Let $n \geq 2$ and let G be a subgroup of $\mathsf{UT}(n, \mathbb{Q})$ with nilpotency class at most two. Given finite subsets $\mathcal{G}_1, \ldots, \mathcal{G}_M$ of G, it is decidable in polynomial time whether $\langle \mathcal{G}_1 \rangle \cap \cdots \cap \langle \mathcal{G}_M \rangle = \emptyset$.*

Proof. Theorem 3.1.3 follows from the correctness and polynomial time complexity of Algorithm 3.2 (Proposition 3.8.2). □

3.9 Orbit Intersection

We prove Theorem 3.1.5 in this section. Let $\mathcal{G}$ and $\mathcal{H}$ be finite sets of matrices in the group $\mathrm{H}_3(\mathbb{Q})$, and T, S be matrices in $\mathrm{H}_3(\mathbb{Q})$. Our goal is to decide whether $T \cdot \langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$. Multiplying both $T \cdot \langle \mathcal{G} \rangle$ and $S \cdot \langle \mathcal{H} \rangle$ on the left by T^{-1}, one can without loss of generality

suppose $T = I$. That is, it suffices to consider the problem of deciding whether $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$. Denote by $\varphi\colon \log \mathrm{H}_3(\mathbb{Q}) \to \mathbb{Q}^2$ the projection onto the superdiagonal, and by $\phi\colon \log \mathrm{H}_3(\mathbb{Q}) \to \mathbb{Q}$ the projection onto the upper right entry:

$$\varphi\colon \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \end{pmatrix}; \qquad \phi\colon \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \mapsto c.$$

One easily verifies that for matrices $X, Y \in \mathrm{H}_3(\mathbb{Q})$, we have $[\log X, \log Y] = 0$ if and only if $\varphi(\log X)$ and $\varphi(\log Y)$ are linearly dependent. Indeed, write $\varphi(\log X) = (x_a, x_b)^\top$ and $\varphi(\log Y) = (y_a, y_b)^\top$. Then $\varphi([\log X, \log Y]) = (0,0)^\top$ and $\phi([\log X, \log Y]) = x_a y_b - y_a x_b$. Therefore, we have $[\log X, \log Y] = 0$ if and only if the vectors $\varphi(\log X)$ and $\varphi(\log Y)$ are linearly dependent. Define the cones

$$\mathcal{C}_\mathcal{G} \coloneqq \langle \varphi(\log \mathcal{G}) \rangle_{\mathbb{Q}_{\geq 0}}, \quad \mathcal{C}_\mathcal{H} \coloneqq \langle \varphi(\log \mathcal{H}) \rangle_{\mathbb{Q}_{\geq 0}}.$$

3.9.1 Easy case: The cone $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$ has dimension zero or one

The situation in this case is similar to the one discussed in [30, Section 3, Case I].

Proposition 3.9.1. *Suppose the cone $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$ has dimension zero or one. Deciding whether $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$ can be done by solving finitely many linear Diophantine equations.*

Proof. Let $\mathcal{L} \subseteq \mathbb{Q}^2$ be a linear space of dimension one that contains $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$. We decompose $\mathcal{G}$ and $\mathcal{H}$ into disjoint subsets: $\mathcal{G} = \mathcal{G}_0 \cup \mathcal{G}_+$, $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_+$, where

$$\mathcal{G}_0 \coloneqq \{A_i \in \mathcal{G} \mid \varphi(\log A_i) \in \mathcal{L}\}, \quad \mathcal{G}_+ \coloneqq \mathcal{G} \setminus \mathcal{G}_0;$$
$$\mathcal{H}_0 \coloneqq \{B_i \in \mathcal{H} \mid \varphi(\log B_i) \in \mathcal{L}\}, \quad \mathcal{H}_+ \coloneqq \mathcal{H} \setminus \mathcal{H}_0.$$

The key observation is that all matrices in $\mathcal{G}_0$ and in $\mathcal{H}_0$ commute with each other. Indeed, for all $A_i \in \mathcal{G}_0, B_i \in \mathcal{H}_0$, the vectors $\varphi(\log A_i)$ and $\varphi(\log B_i)$ all fall in $\mathcal{L}$ and are hence linearly dependant. Therefore $[\log A_i, \log A_j] = [\log B_i, \log B_j] = 0$ for all $A_i, A_j \in \mathcal{G}_0, B_i, B_j \in \mathcal{H}_0$.

Suppose $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$, that is, there exist non-empty words v over the alphabet $\mathcal{G}$ and w over the alphabet $\mathcal{H}$ such that $\log v = \log Sw$. We show that the number of occurrences of letters of $\mathcal{G}_+$ in v is effectively bounded; similarly, the number of occurrences of letters of $\mathcal{H}_+$ in w is also effectively bounded.

Let $\boldsymbol{n}$ be a non-zero vector orthogonal to $\mathcal{L}$, then $\boldsymbol{x} \mapsto \boldsymbol{n}^\top \boldsymbol{x}$ is the projection parallel to $\mathcal{L}$. Since $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H} \subseteq \mathcal{L}$, the values $\boldsymbol{n}^\top \varphi(\log A_i), A_i \in \mathcal{G}$ have signs opposite to that of $\boldsymbol{n}^\top \varphi(\log B_j), B_j \in \mathcal{H}$. Without loss of generality, suppose $\boldsymbol{n}^\top \varphi(\log A_i) \geq 0$ for all $A_i \in \mathcal{G}$ and

$\boldsymbol{n}^\top \varphi(\log B_j) \leq 0$ for all $B_j \in \mathcal{H}$. Since $\boldsymbol{n}$ is orthogonal to $\mathcal{L}$, we have furthermore $\boldsymbol{n}^\top \varphi(\log A_i) > 0$ for all $A_i \in \mathcal{G}_+$ and $\boldsymbol{n}^\top \varphi(\log B_j) < 0$ for all $B_j \in \mathcal{H}_+$; as well as $\boldsymbol{n}^\top \varphi(\log X) = 0$ for all $X \in \mathcal{G}_0 \cup \mathcal{H}_0$.

Now, $\log v = \log Sw$ yields $\varphi(\log v) = \varphi(\log S) + \varphi(\log w)$. Projecting onto $\boldsymbol{n}$, this shows

$$\sum_{i, A_i \in \mathcal{G}_+} \mathrm{PI}_i^{\mathcal{G}}(v) \cdot \boldsymbol{n}^\top \varphi(\log A_i) = \boldsymbol{n}^\top \varphi(\log S) + \sum_{i, B_i \in \mathcal{H}_+} \mathrm{PI}_i^{\mathcal{H}}(w) \cdot \boldsymbol{n}^\top \varphi(\log B_i).$$

This yields

$$\mathrm{PI}_i^{\mathcal{G}}(v) \leq \frac{\boldsymbol{n}^\top \varphi(\log S)}{\boldsymbol{n}^\top \varphi(\log A_i)}, \quad \mathrm{PI}_j^{\mathcal{H}}(w) \leq \frac{-\boldsymbol{n}^\top \varphi(\log S)}{\boldsymbol{n}^\top \varphi(\log B_j)}, \tag{3.81}$$

for all $A_i \in \mathcal{G}_+$ and $B_j \in \mathcal{H}_+$. This gives bounds

$$\beta_{\mathcal{G}} \coloneqq \sum_{i, A_i \in \mathcal{G}_+} \frac{\boldsymbol{n}^\top \varphi(\log S)}{\boldsymbol{n}^\top \varphi(\log A_i)} \quad \text{and} \quad \beta_{\mathcal{H}} \coloneqq \sum_{i, B_i \in \mathcal{H}_+} \frac{-\boldsymbol{n}^\top \varphi(\log S)}{\boldsymbol{n}^\top \varphi(\log B_i)},$$

such that if $\log v = \log Sw$, then the number of letters of $\mathcal{G}_+$ in v is bounded by $\beta_{\mathcal{G}}$, and the number of letters of $\mathcal{H}_+$ in w is bounded by $\beta_{\mathcal{H}}$.

Write $v = v_0 C_1 v_1 C_2 \cdots v_{s-1} C_s v_s$, where $C_1, \ldots, C_s$ are matrices in $\mathcal{G}_+$, and $v_0, \ldots, v_s$ are words in the alphabet $\mathcal{G}_0$. Similarly, write $w = w_0 D_1 w_1 D_2 \cdots w_{t-1} D_t w_t$, where $D_1, \ldots, D_t$ are matrices in $\mathcal{H}_+$, and $w_0, \ldots, w_t$ are words in the alphabet $\mathcal{H}_0$. Write $\mathcal{G}_0 = \{A'_1, \ldots, A'_{K'}\}$ and $\mathcal{H}_0 = \{B'_1, \ldots, B'_{M'}\}$. Define $x_{ij} \coloneqq \mathrm{PI}_j^{\mathcal{G}_0}(v_i)$ for $0 \leq i \leq s, 1 \leq j \leq K'$, and $y_{ij} \coloneqq \mathrm{PI}_j^{\mathcal{H}_0}(w_i)$ for $0 \leq i \leq t, 1 \leq j \leq M'$. Then $\log v = \log Sw$ is equivalent to

$$\begin{aligned}
\sum_{i=1}^{s} \log C_i + \sum_{i=0}^{s} \sum_{j=1}^{K'} x_{ij} \log A'_j + \frac{1}{2} \sum_{0 \leq i < k \leq s} \sum_{j=1}^{K'} x_{ij} [\log A'_j, \log C_k] + \frac{1}{2} \sum_{1 \leq k \leq i \leq s} \sum_{j=1}^{K'} x_{ij} [\log C_k, \log A'_j] \\
= \log S + \sum_{i=1}^{t} (\log D_i + \frac{1}{2} [\log S, \log D_i]) + \frac{1}{2} \sum_{i=0}^{t} \sum_{j=1}^{M'} y_{ij} [\log S, \log B'_j] \\
+ \frac{1}{2} \sum_{0 \leq i < k \leq t} \sum_{j=1}^{M'} y_{ij} [\log B'_j, \log D_k] + \frac{1}{2} \sum_{1 \leq k \leq i \leq t} \sum_{j=1}^{M'} y_{ij} [\log D_k, \log B'_j]
\end{aligned} \tag{3.82}$$

All other terms are of the form $[\log A'_i, \log A'_j]$ or $[\log B'_i, \log B'_j]$ and hence vanish by the commutativity of $\mathcal{G}_0$ and $\mathcal{H}_0$. Note that Equation (3.82) is a linear Diophantine equation in the variables x_{ij}, y_{ij}. Therefore, $\log v = \log Sw$ has a solution if and only if there exist matrices $C_1, \ldots, C_s$ in $\mathcal{G}_+$ and matrices $D_1, \ldots, D_t$ in $\mathcal{H}_+$, such that Equation (3.82) has a solution in non-negative integers, with the additional constraint that, if $s = 0$, then $(x_{01}, \ldots, x_{0K'}) \neq \mathbf{0}$; and if $t = 0$, then $(y_{01}, \ldots, y_{0M'}) \neq \mathbf{0}$. This additional constraint comes from the condition that v, w are not empty words. Recall the bounds $s \leq \beta_{\mathcal{G}}$ and $t \leq \beta_{\mathcal{H}}$. Hence, deciding whether

$\log v = \log Sw$ has a solution amounts to solving finitely many linear Diophantine equations of the form (3.82). □

In theory, it is possible to give a bound on the complexity of the procedure described in Proposition 3.9.1. The size of the each bound in Equation (3.81) is exponential in the bit size of the entries $S, \mathcal{G}, \mathcal{H}$. Hence the procedure consists of solving exponentially many linear Diophantine equations.

3.9.2 Hard case: The cone $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$ has dimension two

Suppose now that the cone $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$ spans a linear space of dimension two. We have $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$ if and only if there exist words v in the alphabet $\mathcal{G}$ and w in the alphabet $\mathcal{H}$ such that $\log v = \log Sw$. Let $\boldsymbol{x} = (x_1, \ldots, x_K)$ be the Parikh Image of v, and $\boldsymbol{y} = (y_1, \ldots, y_M)$ be the Parikh Image of w. By the Baker-Campbell-Hausdorff formulas (3.62) and (3.63), the equality $\log v = \log Sw$ is equivalent to

$$\sum_{i=1}^{K} x_i \log A_i + \frac{1}{2} \sum_{1 \leq i < j \leq K} \delta_{ij}^{\mathcal{G}}(v)[\log A_i, \log A_j] = \log S + \sum_{i=1}^{M} y_i (\log B_i + \frac{1}{2}[\log S, \log B_i]) + \frac{1}{2} \sum_{1 \leq i < j \leq M} \delta_{ij}^{\mathcal{H}}(w)[\log B_i, \log B_j] \quad (3.83)$$

The following proposition shows that it suffices to solve a relaxed version of Equation (3.83).

Proposition 3.9.2. *Suppose the cone $\mathcal{C}_\mathcal{G} \cap \mathcal{C}_\mathcal{H}$ has dimension two. We have $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$ if and only if there exist integers $x_i, 1 \leq i \leq K$ and $y_j, 1 \leq j \leq M$ and $c_{ij}, 1 \leq i < j \leq K$ and $d_{ij}, 1 \leq i < j \leq M$, satisfying*

$$\sum_{i=1}^{K} x_i \varphi(\log A_i) = \varphi(\log S) + \sum_{i=1}^{M} y_i \varphi(\log B_i), \quad (3.84)$$

and

$$\sum_{i=1}^{K} x_i \phi(\log A_i) + \frac{1}{2} \sum_{1 \leq i < j \leq K} c_{ij} \phi([\log A_i, \log A_j]) = \phi(\log S) + \sum_{i=1}^{M} y_i \phi(\log B_i + \frac{1}{2}[\log S, \log B_i]) + \frac{1}{2} \sum_{1 \leq i < j \leq M} d_{ij} \phi([\log B_i, \log B_j]), \quad (3.85)$$

and

$$c_{ij} \equiv x_i x_j \mod 2, \quad 1 \leq i < j \leq K; \qquad d_{ij} \equiv y_i y_j \mod 2, \quad 1 \leq i < j \leq M. \quad (3.86)$$

Proof. If $\langle\mathcal{G}\rangle \cap S\cdot\langle\mathcal{H}\rangle \neq \emptyset$, then let v, w be non-empty words over the respectively alphabets $\mathcal{G}$ and $\mathcal{H}$, such that $\log v = \log Sw$. Let $(x_1, \ldots, x_K)$ and $(y_1, \ldots, y_M)$ be the respective Parikh Image of v and w, and let $c_{ij} \coloneqq \delta^{\mathcal{G}}_{ij}(v)$ and $d_{ij} \coloneqq \delta^{\mathcal{H}}_{ij}(w)$ for all i, j. Since Equation (3.83) is satisfied, projecting it under φ and ϕ gives respectively (3.84) and (3.85). The parity condition (3.86) is obviously due to Equation (3.61). Hence we have found the integers x_i, y_j, c_{ij}, d_{ij} satisfying Equations (3.84), (3.85) and (3.86).

For the other implication direction, let x_i, y_j, c_{ij}, d_{ij} be integers that satisfy Equations (3.84), (3.85) and (3.86). Since $\mathcal{C}_{\mathcal{G}}$ and $\mathcal{C}_{\mathcal{H}}$ have dimension two, the commutators $[\log A_i, \log A_j]$ and $[\log B_i, \log B_j]$ are not all zero (since $\varphi(A_i)$ are not all linearly dependant, same for $\varphi(B_i)$). Hence, there exist integers C_{ij}, D_{ij} such that

$$D \coloneqq \sum_{1\leq i<j\leq K} C_{ij}\phi([\log A_i, \log A_j]) + \sum_{1\leq i<j\leq M} D_{ij}\phi([\log B_i, \log B_j])$$

is a strictly positive rational number. Denote by E a common denominator of all the entries of the matrices $\log A_i$, $\log B_i$, $\log S$, $\frac{1}{2}[\log S, \log B_i]$, $\frac{1}{2}[\log A_i, \log A_j]$ and $\frac{1}{2}[\log B_i, \log B_j]$. In particular, DE is a positive integer.

Since the cone $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$ has dimension two, there exist *strictly positive* integers $X_1, \ldots, X_K$ and $Y_1, \ldots, Y_M$, such that

$$\sum_{i=1}^{K} X_i\varphi(\log A_i) = \sum_{i=1}^{M} Y_i\varphi(\log B_i). \tag{3.87}$$

This is because, taking $\boldsymbol{v}$ to be a vector in the *interior* of $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$ (i.e. $\boldsymbol{v}$ admits an open neighbourhood contained in $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$), then $\boldsymbol{v}$ is in the interior of both $\mathcal{C}_{\mathcal{G}}$ and $\mathcal{C}_{\mathcal{H}}$. Hence, there exist strictly positive rational numbers $X'_1, \ldots, X'_K$ and $Y'_1, \ldots, Y'_M$, such that

$$\sum_{i=1}^{K} X'_i\varphi(\log A_i) = \boldsymbol{v} = \sum_{i=1}^{M} Y'_i\varphi(\log B_i).$$

Multiplying $X'_1, \ldots, X'_K$ and $Y'_1, \ldots, Y'_M$ by their common denominator gives positive integers satisfying Equation (3.87).

For any $N \in \mathbb{Z}_{>0}$, the integers x_i, y_i, c_{ij}, d_{ij} can be replaced by the integers

$$\begin{aligned}
x'_i &\coloneqq x_i + 2NDEX_i \\
y'_i &\coloneqq y_i + 2NDEY_i \\
c'_{ij} &\coloneqq c_{ij} - 4NEC_{ij}\left(\sum_{k=1}^{K} X_k\phi(\log A_k) - \sum_{k=1}^{M} Y_k\phi(\log B_k + \frac{1}{2}[\log S, \log B_k])\right)
\end{aligned}$$

$$d'_{ij} \coloneqq d_{ij} + 4NED_{ij}\left(\sum_{k=1}^{K} X_k \phi(\log A_k) - \sum_{k=1}^{M} Y_k \phi(\log B_k + \frac{1}{2}[\log S, \log B_k])\right)$$

for all i, j, while still satisfying Equations (3.84), (3.85) and (3.86). Furthermore, when N is large enough, we have

$$x'_i > 0, y'_j > 0, \quad 1 \le i \le K, 1 \le j \le M, \tag{3.88}$$

$$|c'_{ij}| \le \frac{x'_i x'_j}{4K^2} - 2K(x'_i + x'_j) - 4K^2, \quad 1 \le i < j \le K, \tag{3.89}$$

and

$$|d'_{ij}| \le \frac{y'_i y'_j}{4M^2} - 2M(y'_i + y'_j) - 4M^2, \quad 1 \le i < j \le M. \tag{3.90}$$

This is because the right hand sides of the inequalities (3.89) and (3.90) are quadratic in N, whereas the left hand sides grow linearly in N.

Fix an N such that the inequalities (3.88), (3.89) and (3.90) are satisfied. Then, by Proposition 3.7.2, there exist non-empty words v, w over the alphabets $\mathcal{G}$ and $\mathcal{H}$, such that

$$\begin{aligned} \mathrm{PI}^{\mathcal{G}}(v) &= (x'_1, \ldots, x'_K), \quad \delta^{\mathcal{G}}_{ij}(v) = c'_{ij}, \quad \text{for } 1 \le i < j \le K, \\ \mathrm{PI}^{\mathcal{H}}(w) &= (y'_1, \ldots, y'_K), \quad \delta^{\mathcal{H}}_{ij}(v) = d'_{ij}, \quad \text{for } 1 \le i < j \le M. \end{aligned}$$

(Note that Condition (3.68) is guaranteed by Equation (3.86).) For these words v, w, we have

$$\varphi(\log v) = \sum_{i=1}^{K} x'_i \varphi(\log A_i) = \varphi(\log S) + \sum_{i=1}^{M} y'_i \varphi(\log B_i) = \varphi(\log Sw),$$

as well as

$$\begin{aligned} \phi(\log v) = \sum_{i=1}^{K} x'_i \phi(\log A_i) + \frac{1}{2} \sum_{1 \le i < j \le K} c'_{ij} \phi([\log A_i, \log A_j]) = \\ \phi(\log S) + \sum_{i=1}^{M} y'_i \phi(\log B_i + \frac{1}{2}[\log S, \log B_i]) + \frac{1}{2} \sum_{1 \le i < j \le M} d'_{ij} \phi([\log B_i, \log B_j]) = \phi(\log Sw). \end{aligned}$$

This shows $\log v = \log Sw$, hence $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$. □

Combining the two cases in Subsections 3.9.1 and 3.9.2 solves Orbit Intersection in $\mathrm{H}_3(\mathbb{Q})$.

Theorem 3.1.5. *Given elements $T, S \in \mathrm{H}_3(\mathbb{Q})$ and two finite subsets $\mathcal{G}, \mathcal{H}$ of $\mathrm{H}_3(\mathbb{Q})$, it is decidable whether $T \cdot \langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle = \emptyset$.*

Proof. As mentioned in the beginning of Section 3.9, one can without loss of generality suppose

$T = I$, and decide whether $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$. Given $\mathcal{G}$ and $\mathcal{H}$, one can effectively compute $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$ and its dimension using linear programming [89].

If $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$ has dimension zero or one, then Proposition 3.9.1 shows we can decide whether $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$ by solving a finite number of linear Diophantine equations of the form (3.82).

If $\mathcal{C}_{\mathcal{G}} \cap \mathcal{C}_{\mathcal{H}}$ has dimension two, then Proposition 3.9.2 shows we can decide whether $\langle \mathcal{G} \rangle \cap S \cdot \langle \mathcal{H} \rangle \neq \emptyset$ by solving Equations (3.84), (3.85) and (3.86). Equation (3.86) can be replaced by a boolean combination of conditions of the form "$x_i \equiv 0 \mod 2$", "$x_i \equiv 1 \mod 2$", "$y_i \equiv 0 \mod 2$", "$y_i \equiv 1 \mod 2$", "$c_{ij} \equiv 0 \mod 2$", "$c_{ij} \equiv 1 \mod 2$", "$d_{ij} \equiv 0 \mod 2$", and "$d_{ij} \equiv 1 \mod 2$". Each of these conditions can be expressed as a linear equation over integers, for example "$x_i \equiv 1 \mod 2$" is equivalent to "$x_i = 2x_i' + 1, x_i' \in \mathbb{Z}$". Therefore, solving Equations (3.84), (3.85) and (3.86) is equivalent to solving a boolean combination of linear equations over integers, which is decidable by integer programming. □

3.10 (Additional material) Computer-assisted proofs

In this section we give the detailed account for the proof of Lemma 3.5.7-3.5.10 using computer assistance.

We fix an integer k for the whole section. Let $\mathcal{H}$ be a subset of $\mathsf{UT}(n, \mathbb{Q})$. For $x, y \in \mathfrak{u}(n)$, we write

$$x \overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim} y$$

if $x - y \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))$, and

$$x \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} y$$

if $x - y \in \mathfrak{L}_{\geq k+1}(\log \mathcal{H})$. Obviously, $\overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim}$ and $\overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim}$ are equivalence relations and we denote by $\sim$ the transitive closure of these two relations.

The following lemma shows the effect of the relation $\overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim}$. In fact, the quotient Lie algebra $L \coloneqq \mathfrak{L}_{\geq 1}(\log \mathcal{H})/\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))$ is *metabelian*, meaning $[[L, L], [L, L]] = 0$. This property allows us the permute elements in iterated Lie brackets:

Lemma 3.10.1. *For $C_1, \ldots, C_k \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $i = 3, \ldots, k-1$, we have*

$$[\ldots[[\ldots[C_1, C_2], \ldots, C_i], C_{i+1}], \ldots, C_k] \overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim} [\ldots[[\ldots[C_1, C_2], \ldots, C_{i+1}], C_i], \ldots, C_k].$$

Proof. For $i = 3, \ldots, k-1$, by the Jacobi identity,

$$[\ldots[[\ldots[C_1, C_2], \ldots, C_i], C_{i+1}], \ldots, C_k] - [\ldots[[\ldots[C_1, C_2], \ldots, C_{i+1}], C_i], \ldots, C_k]$$

$$
\begin{aligned}
&= [\dots[[\dots[C_1, C_2], \dots, C_{i-1}], [C_i, C_{i+1}]], \dots, C_k] \\
&\in [\dots[\mathfrak{L}_{\geq 2}(\log \mathcal{H}), \mathfrak{L}_{\geq 2}(\log \mathcal{H})], \dots, C_k]. \\
&\subseteq [\dots[\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})), C_{i+2}], \dots, C_k]. \qquad (3.91)
\end{aligned}
$$

We then show that

$$
X \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})), Y \in \mathfrak{L}_{\geq 1}(\log \mathcal{H}) \implies [X, Y] \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})). \qquad (3.92)
$$

Since X is in $\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))$, it can be written as a linear combination of elements of the form $[\dots[X_1, X_2], \dots, X_s]$ where $s \geq 2$, $X_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H}), i = 1, \dots, s$. Therefore it suffices to show the implication (3.92) for the case $X = [\dots[X_1, X_2], \dots, X_s]$ where $X_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H}), i = 1, \dots, s$. Let

$$
X' \coloneqq [\dots[X_1, X_2], \dots, X_{s-1}] \in \mathfrak{L}_{\geq 2(s-1)}(\log \mathcal{H}) \subseteq \mathfrak{L}_{\geq 2}(\log \mathcal{H}),
$$

so $X = [X', X_s]$ with $X', X_s \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. Then by the Jacobi identity,

$$
[X, Y] = [[X', X_s], Y] = -[[X_s, Y], X'] - [[Y, X'], X_s],
$$

where

$$
\begin{aligned}
[[X_s, Y], X'] \in [[\mathfrak{L}_{\geq 2}(\log \mathcal{H}), \mathfrak{L}_{\geq 1}(\log \mathcal{H})], \mathfrak{L}_{\geq 2}(\log \mathcal{H})] \\
\subseteq [\mathfrak{L}_{\geq 2}(\log \mathcal{H}), \mathfrak{L}_{\geq 2}(\log \mathcal{H})] \subseteq \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))
\end{aligned}
$$

and

$$
\begin{aligned}
[[Y, X'], X_s] \in [[\mathfrak{L}_{\geq 1}(\log \mathcal{H}), \mathfrak{L}_{\geq 2}(\log \mathcal{H})], \mathfrak{L}_{\geq 2}(\log \mathcal{H})] \\
\subseteq [\mathfrak{L}_{\geq 2}(\log \mathcal{H}), \mathfrak{L}_{\geq 2}(\log \mathcal{H})] \subseteq \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})).
\end{aligned}
$$

Therefore $[X, Y] \in \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))$, showing the implication (3.92).

Applying this implication with $Y = C_{i+2}, C_{i+3}, \dots, C_k$ in Equation (3.91) shows

$$
\begin{aligned}
&[\dots[\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})), C_{i+2}], \dots, C_k] \\
\subseteq\ &[\dots[\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})), C_{i+3}], \dots, C_k] \\
&\vdots \\
\subseteq\ &[\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H})), C_k]
\end{aligned}
$$

$$\subseteq \mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))$$

Hence Equation (3.91) yields

$$[\dots[[\dots[C_1, C_2], \dots, C_i], C_{i+1}], \dots, C_k] \overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim} [\dots[[\dots[C_1, C_2], \dots, C_{i+1}], C_i], \dots, C_k].$$

□

Fix an integer k. Define an *integer partition* P (of k) to be a series of numbers $(a_1, \dots, a_s)$ such that $a_1 \geq a_2 \geq \cdots \geq a_s \geq 1$ and $k = a_1 + \cdots + a_s$. Define $\max(P) \coloneqq a_1, \min(P) \coloneqq a_s$ and $\operatorname{set}(P) \coloneqq \{t \mid \exists a_i = t\}$. Define a *set partition* S (of $\{1, \dots, k\}$) to be a set of non-empty disjoint sets $S = \{A_1, \dots, A_s\}$ such that $A_1 \cup \cdots \cup A_s = \{1, \dots, k\}$. For any k-tuple $\boldsymbol{j} = (j_1, \dots, j_k) \in \{1, \dots, k+1\}^k$, define the *associated set partition* of $\boldsymbol{j}$ the set partition consisting of sets of indices of its distinct elements

$$\operatorname{SP}(\boldsymbol{j}) \coloneqq \left\{ A_i \coloneqq \{l \mid j_l = i\} \;\middle|\; i = 1, \dots, k+1, A_i \neq \emptyset \right\}.$$

For example, if $k = 6$, $\boldsymbol{j} = (4, 2, 7, 2, 2, 4)$, then $\operatorname{SP}(\boldsymbol{j}) = \{\{1, 6\}, \{2, 4, 5\}, \{3\}\}$.

Define the *associated integer partition* $\operatorname{IP}(S)$ of a set partition S to be the series of set cardinalities in S in decreasing order. For example, if $k = 6$, $S = \{\{1, 6\}, \{2, 4, 5\}, \{3\}\}$, then $\operatorname{IP}(S) = (3, 2, 1)$. In particular, in this example we have $\max(\operatorname{IP}(S)) = 3, \min(\operatorname{IP}(S)) = 1$ and $\operatorname{set}(P) = \{3, 2, 1\}$.

We now fix elements $C_1, \dots, C_k \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$. For a given tuple $\boldsymbol{j} = (j_1, \dots, j_k) \in \{1, \dots, k+1\}^k$, define the symmetric sums

$$\Phi(\boldsymbol{j}) \coloneqq \frac{1}{(k+1-\operatorname{card}(\operatorname{SP}(\boldsymbol{j})))!} \sum_{\sigma \in \mathrm{S}_{k+1}} \varphi_k(C_{\sigma(j_1)}, C_{\sigma(j_2)}, \dots, C_{\sigma(j_k)}),$$
$$M(\boldsymbol{j}) \coloneqq \frac{1}{(k+1-\operatorname{card}(\operatorname{SP}(\boldsymbol{j})))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\dots[C_{\sigma(j_1)}, C_{\sigma(j_2)}], \dots, C_{\sigma(j_k)}].$$

Here, φ_k is the expression defined in the Dynkin formula (3.19). The relation between $\Phi(\boldsymbol{j})$ and $M(\boldsymbol{j})$ can be computed as follows.

$$\begin{aligned}\Phi(\boldsymbol{j}) =& \frac{1}{(k+1-\operatorname{card}(\operatorname{SP}(\boldsymbol{j})))!} \sum_{\sigma \in \mathrm{S}_{k+1}} \varphi_k(C_{\sigma(j_1)}, C_{\sigma(j_2)}, \dots, C_{\sigma(j_k)}) \\ =& \frac{1}{(k+1-\operatorname{card}(\operatorname{SP}(\boldsymbol{j})))!} \sum_{\sigma \in \mathrm{S}_{k+1}} \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} [\dots[C_{\sigma(j_{\tau(1)})}, C_{\sigma(j_{\tau(2)})}], \dots, C_{\sigma(j_{\tau(k)})}]\end{aligned}$$

$$= \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot \frac{1}{(k+1-\mathrm{card}(\mathrm{SP}(\boldsymbol{j})))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[C_{\sigma(j_{\tau(1)})}, C_{\sigma(j_{\tau(2)})}], \ldots, C_{\sigma(j_{\tau(k)})}]$$

$$= \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot M(\boldsymbol{j}_\tau), \tag{3.93}$$

where $\boldsymbol{j}_\tau \coloneqq (j_{\tau(1)}, j_{\tau(2)}, \ldots, j_{\tau(k)})$.

From the definition of $\Phi(\boldsymbol{j})$ and $M(\boldsymbol{j})$ it follows that that for any $\sigma \in \mathrm{S}_{k+1}$, writing $\sigma(\boldsymbol{j}) \coloneqq (\sigma(j_1), \ldots, \sigma(j_k))$, we have $\Phi(\sigma(\boldsymbol{j})) = \Phi(\boldsymbol{j})$ and $M(\sigma(\boldsymbol{j})) = M(\boldsymbol{j})$. By this symmetry, $\Phi(\boldsymbol{j})$ and $M(\boldsymbol{j})$ only depend on their associated set partition $\mathrm{SP}(\boldsymbol{j})$. Hence for any set partition S, we can define

$$\Phi(S) \coloneqq \Phi(\boldsymbol{j}), \quad M(S) \coloneqq M(\boldsymbol{j}), \quad \text{where } \mathrm{SP}(\boldsymbol{j}) = S.$$

From Equation (3.93) we get

$$\Phi(S) = \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot M(S_\tau), \tag{3.94}$$

where S_τ is the set partition obtained by replacing i by $\tau(i)$ in all sets of S for all $i = 1, \ldots, k$:

$$S_\tau \coloneqq \left\{ \{\tau(j) \mid j \in A\} \;\middle|\; A \in S \right\}.$$

For two set partitions S_1 and S_2, S_2 is called a *coarsening* of S_1 if for every $A \in S_1$, there exists $A' \in S_2$ such that $A \subseteq A'$. For example, $\{\{1,3,4\},\{2,5,6\}\}$ is a coarsening of $\{\{1,3,4\},\{2\},\{5,6\}\}$. In particular, any set partition is a coarsening of itself. Denote by $S_2 \succcurlyeq S_1$ if S_2 is a coarsening of S_1.

The next lemma shows the effect of the relation $\overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim}$ for sums over coarsenings.

Lemma 3.10.2. *Let $\mathcal{H}$ be a subset of $\mathsf{UT}(n, \mathbb{Q})$. Suppose $C_1, \ldots, C_{k+1} \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. If a set partition S satisfies $\min(S) = 1$, then*

$$\sum_{S' \succcurlyeq S} M(S') \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} 0. \tag{3.95}$$

Proof. First let us illustrate the intuition with an example. Let $k = 6$, $S = \{\{1,3,4\},\{2\},\{5,6\}\}$, then there are five coarsenings of S, which are:

$$S, \{\{1,3,4\},\{2,5,6\}\}, \{\{1,3,4,2\},\{5,6\}\}, \{\{1,3,4,5,6\},\{2\}\}, \{\{1,3,4,2,5,6\}\}.$$

Correspondingly,

$$
\begin{aligned}
&M(S)+M(\{\{1,3,4\},\{2,5,6\}\})+M(\{\{1,3,4,2\},\{5,6\}\})+M(\{\{1,3,4,5,6\},\{2\}\})\\
&\quad+M(\{\{1,3,4,2,5,6\}\})\\
=&\frac{1}{4!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(1)},C_{\sigma(2)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(3)}],C_{\sigma(3)}]\\
&\quad+\frac{1}{5!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(1)},C_{\sigma(2)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(2)}],C_{\sigma(2)}]\\
&\quad+\frac{1}{5!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(1)},C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(2)}],C_{\sigma(2)}]\\
&\quad+\frac{1}{5!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(1)},C_{\sigma(2)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}]\\
&\quad+\frac{1}{6!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(1)},C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}],C_{\sigma(1)}]\\
=&\sum_{i,j,k\text{ distinct}}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]+\sum_{i\neq j=k}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]\\
&\quad+\sum_{i=j\neq k}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]+\sum_{i=k\neq j}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]\\
&\quad+\sum_{i=j=k}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]\\
=&\sum_{i=1}^{7}\sum_{j=1}^{7}\sum_{k=1}^{7}[[[[[[C_i,C_j],C_i],C_i],C_k],C_k]\\
=&\sum_{i=1}^{7}\sum_{k=1}^{7}[[[[[[C_i,\sum_{j=1}^{7}C_j],C_i],C_i],C_k],C_k]\\
\in&\sum_{i=1}^{7}\sum_{k=1}^{7}[[[[[[C_i,\mathfrak{L}_{\geq 2}(\log\mathcal{H})],C_i],C_i],C_k],C_k]\\
\subseteq&\mathfrak{L}_{\geq 7}(\log\mathcal{H}).
\end{aligned}
$$

So $\sum_{S'\succcurlyeq S}M(S')\overset{\mathfrak{L}_{\geq k+1}(\log\mathcal{H})}{\sim}0$ for this particular example.

For the general case, write $S=\{A_1,\ldots,A_s\}$ with $\mathrm{card}(A_1)=1$, then

$$
\begin{aligned}
&\sum_{S'\succcurlyeq S}M(S')\\
=&\sum_{S'\succcurlyeq S}\sum_{\substack{\boldsymbol{j}\in\{1,\ldots,k+1\}^k\\ \mathrm{SP}(\boldsymbol{j})=S'}}[\ldots[C_{j_1},C_{j_2}],\ldots,C_{j_k}]\\
=&\sum_{\substack{(j_1,\ldots,j_k)\in\{1,\ldots,k+1\}^k\\ j_i=j_{i'}\text{ if }i,i'\text{ are in the same set of }S}}[\ldots[C_{j_1},C_{j_2}],\ldots,C_{j_k}]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i_1=1}^{k+1} \cdots \sum_{i_s=1}^{k+1} [\ldots[C_{i_{f(1)}}, C_{i_{f(2)}}], \ldots, C_{i_{f(k)}}] \quad \text{where } f(r) \text{ is defined by } r \in A_{f(r)}. \\
&= \sum_{i_2=1}^{k+1} \cdots \sum_{i_s=1}^{k+1} [\ldots[\ldots[C_{i_{f(1)}}, C_{i_{f(2)}}], \ldots, \sum_{i_1=1}^{k+1} C_{i_1}], \ldots, C_{i_{f(k)}}] \\
&\in \sum_{i_2=1}^{k+1} \cdots \sum_{i_s=1}^{k+1} [\ldots[\ldots[C_{i_{f(1)}}, C_{i_{f(2)}}], \ldots, \mathfrak{L}_{\geq 2}(\log \mathcal{H})], \ldots, C_{i_{f(k)}}] \\
&\subseteq \mathfrak{L}_{\geq k+1}(\log \mathcal{H}).
\end{aligned}
$$

Hence $\sum_{S' \succcurlyeq S} M(S') \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} 0$. □

Using Equation (3.94), Lemma 3.10.2 gives the following corollaries.

Corollary 3.10.3. *Let $\mathcal{H}$ be a subset of $\mathsf{UT}(n, \mathbb{Q})$. Suppose $C_1, \ldots, C_{k+1} \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. If a set partition S satisfies $\min(S) = 1$, then*

$$
\sum_{S' \succcurlyeq S} \Phi(S') \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} 0. \tag{3.96}
$$

Proof. For any $\tau \in \mathrm{S}_k$, we have that $S'_\tau \succcurlyeq S_\tau$ if and only if $S' \succcurlyeq S$. Therefore by Equation (3.94),

$$
\begin{aligned}
\sum_{S' \succcurlyeq S} \Phi(S') = \sum_{S' \succcurlyeq S} \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot M(S'_\tau) &= \sum_{S'_\tau \succcurlyeq S_\tau} \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot M(S'_\tau) \\
&= \sum_{\tau \in \mathrm{S}_k} \frac{(-1)^{d(\tau)}}{k^2 \binom{k-1}{d(\tau)}} \cdot \sum_{S'_\tau \succcurlyeq S_\tau} M(S'_\tau) \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} 0.
\end{aligned}
$$

□

Corollary 3.10.4. *Let $\mathcal{H}$ be a subset of $\mathsf{UT}(n, \mathbb{Q})$. Suppose $C_1, \ldots, C_{k+1} \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. For any set partition S, the symmetric sum $\Phi(S)$ is equivalent under $\overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim}$ to a linear combination of $\Phi(S')$ where $\min(\mathrm{IP}(S')) \geq 2$ (that is, every set in the partitions S' has cardinality at least two).*

In other words, there exist integers $\alpha_{S'}$, where S' ranges over all set partitions satisfying $\min(\mathrm{IP}(S')) \geq 2$, such that

$$
\Phi(S) \overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim} \sum_{S', \min(\mathrm{IP}(S')) \geq 2} \alpha_{S'} \Phi(S').
$$

Proof. Corollary 3.10.3 shows that if $\min(\mathrm{IP}(S)) = 1$, then under the equivalence $\overset{\mathfrak{L}_{\geq k+1}(\log \mathcal{H})}{\sim}$, we can replace $\Phi(S)$ by $-\sum_{S' \succcurlyeq S, S' \neq S} \Phi(S')$. Repeat this "coarsening" procedure for all $\Phi(S')$,

$\min(\mathrm{IP}(S')) = 1$, for sufficiently many times, we can rewrite $\Phi(S)$ as a linear combination of expressions $\Phi(S')$ where $\min(\mathrm{IP}(S')) \geq 2$. $\square$

Define a *partition-integer pair* to be a pair (P, c), where P is an integer partition and c is a number in $\mathrm{set}(P)$. For a partition-integer pair (P, c), define the following symmetric sum.

$$\widehat{M}(P, c) \coloneqq M(S),$$

where S is a set partition such that $\mathrm{IP}(S) = P$, and $1 \in A \in S$ with $\mathrm{card}(A) = \max(P)$ and $2 \in A' \in S$ with $\mathrm{card}(A') = c$. For example, a possible definition of $\widehat{M}((3,2,1),1)$ can be

$$\begin{aligned}\widehat{M}((3,2,1),1) &\coloneqq M(\{\{1,3,4\},\{2\},\{5,6\}\}) \\ &= \frac{1}{4!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(2)}, C_{\sigma(7)}], C_{\sigma(2)}], C_{\sigma(2)}], C_{\sigma(4)}], C_{\sigma(4)}] \\ &= \sum_{1\leq i,j,k\leq 7, i,j,k \text{ distinct}} [[[[[[C_i, C_j], C_i], C_i], C_k], C_k].\end{aligned}$$

Note that this definition *a priori* depends on the choice of the set partition S. However, under the equivalence relation $\overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim}$, different choices of S result in the same equivalence class. Indeed, let $\boldsymbol{j}$ be a tuple whose associated set partition is S. By Lemma 3.10.1, any exchange of order among the elements $j_3, j_4, \ldots, j_k$ will not change the equivalence class of $[\ldots[C_{\sigma(j_1)}, C_{\sigma(j_2)}], \ldots, C_{\sigma(j_k)}]$, so it will not change the equivalence class of $M(\boldsymbol{j})$. This means that the equivalent class of $M(S)$ does not change when we permute the numbers $3, 4, \ldots, k$. For example, $M(\{\{1,3,4\},\{2\},\{5,6\}\}) \sim M(\{\{1,3,5\},\{2\},\{4,6\}\})$, because

$$\begin{aligned}M(\{\{1,3,4\},\{2\},\{5,6\}\}) &= \frac{1}{4!}\sum_{\sigma\in \mathrm{S}_7}[[[[[[C_{\sigma(2)}, C_{\sigma(7)}], C_{\sigma(2)}], C_{\sigma(2)}], C_{\sigma(4)}], C_{\sigma(4)}] \\ \overset{\mathfrak{L}_{\geq 2}(\mathfrak{L}_{\geq 2}(\log \mathcal{H}))}{\sim} \frac{1}{4!}\sum_{\sigma\in \mathrm{S}_7}&[[[[[[C_{\sigma(2)}, C_{\sigma(7)}], C_{\sigma(2)}], C_{\sigma(4)}], C_{\sigma(2)}], C_{\sigma(4)}] = M(\{\{1,3,5\},\{2\},\{4,6\}\}).\end{aligned}$$

Hence, the equivalence class of $M(S)$ only depends on the integer partition $\mathrm{IP}(S)$ as well as the cardinality of the sets where 1 and 2 belong. This is uniquely determined by the partition-cardinality pair (P, c).

Lemma 3.10.5. *Let $\mathcal{H}$ be a subset of G. Suppose $C_1, \ldots, C_{k+1} \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ and $\sum_{i=1}^{k+1} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. For any set partition S satisfying $\min(\mathrm{IP}(S)) \geq 2$, the symmetric sum $M(S)$ is equivalent (under $\sim$) to a linear combination of $\widehat{M}(P, c)$, where (P, c) are partition-integer pairs satisfying $c \neq \max(P)$ and $\min(P) \geq 2$.*

In other words, there exist integers $\beta_{(P,c)}$, where (P, c) ranges over all partition-integer pairs

with $c \neq \max(P)$ *and* $\min(P) \geq 2$*, such that*

$$M(S) \sim \sum_{(P,c)} \beta_{(P,c)} \widehat{M}(P,c).$$

Proof. Write $S = \{A_1, \ldots, A_s\}$ with $\mathrm{card}(A_1) = \max(\mathrm{IP}(S))$. By Lemma 3.10.1, the equivalence class of $M(S)$ does not change when we permute the numbers $3, 4, \ldots, k$. We can therefore suppose $3 \in A_1$. Take any tuple $\boldsymbol{j} = (j_1, \ldots, j_k) \in \{1, \ldots, k+1\}^k$ with $\mathrm{SP}(\boldsymbol{j}) = S$. By the Jacobi identity,

$$\begin{aligned}[\ldots[[C_{\sigma(j_1)}, C_{\sigma(j_2)}], C_{\sigma(j_3)}], \ldots, C_{\sigma(j_k)}] = \\ [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_2)}], C_{\sigma(j_1)}], \ldots, C_{\sigma(j_k)}] - [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_1)}], C_{\sigma(j_2)}], \ldots, C_{\sigma(j_k)}]. \quad (3.97)\end{aligned}$$

Summing up for $\sigma \in \mathrm{S}_{k+1}$, the expression $\sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_2)}], C_{\sigma(j_1)}], \ldots, C_{\sigma(j_k)}]$ is equivalent to $(k+1-\mathrm{card}(S))! \cdot \widehat{M}(\mathrm{IP}(S), c)$, with $c = \mathrm{card}(A_i)$ where $j_2 \in A_i$. Similarly, the expression

$$\sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_1)}], C_{\sigma(j_2)}], \ldots, C_{\sigma(j_k)}]$$

is equivalent to $(k+1-\mathrm{card}(S))! \cdot \widehat{M}(\mathrm{IP}(S), c')$, with $c' = \mathrm{card}(A_{i'})$ where $j_1 \in A_{i'}$.

We claim that if $c = \max(\mathrm{IP}(S))$, then $\widehat{M}(\mathrm{IP}(S), c) \sim 0$. This is because, writing

$$\widehat{M}(\mathrm{IP}(S), c) = \frac{1}{(k+1-\mathrm{card}(S))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_2)}], C_{\sigma(j_1)}], \ldots, C_{\sigma(j_k)}],$$

if $j_2 \in A_i$ with $\mathrm{card}(A_i) = \max(\mathrm{IP}(S))$, then swapping 2 and 3 in the set partition $\mathrm{SP}(\boldsymbol{j})$ does not change its associated integer partition. Therefore, we have

$$\begin{aligned}\widehat{M}(\mathrm{IP}(S), \max(S')) = \frac{1}{(k+1-\mathrm{card}(S))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_3)}, C_{\sigma(j_2)}], C_{\sigma(j_1)}], \ldots, C_{\sigma(j_k)}] \sim \\ - \frac{1}{(k+1-\mathrm{card}(S))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_2)}, C_{\sigma(j_3)}], C_{\sigma(j_1)}], \ldots, C_{\sigma(j_k)}] \sim -\widehat{M}(\mathrm{IP}(S'), \max(S')),\end{aligned}$$

so $\widehat{M}(\mathrm{IP}(S), \max(S)) \sim 0$. This proves that if $c = \max(\mathrm{IP}(S))$, then $\widehat{M}(\mathrm{IP}(S), c) \sim 0$.

Summing up Equation (3.97) for $\sigma \in \mathrm{S}_{k+1}$, we conclude that

$$\begin{aligned}M(S) = \frac{1}{(k+1-\mathrm{card}(S))!} \sum_{\sigma \in \mathrm{S}_{k+1}} [\ldots[[C_{\sigma(j_1)}, C_{\sigma(j_2)}], C_{\sigma(j_3)}], \ldots, C_{\sigma(j_k)}] \\ = \widehat{M}(\mathrm{IP}(S), c) - \widehat{M}(\mathrm{IP}(S), c')\end{aligned}$$

is equivalent (under $\sim$) to a linear combination of expressions $\widehat{M}(\mathrm{IP}(S), c)$, where $c \neq \max(S)$. $\square$

For any k, all partition-integer pairs satisfying $c \neq \max(P)$ and $\min(P) \geq 2$ can be effectively listed. For example, when $k = 5$, there is only one pair $((3,2),2)$. When $k = 7$, there are three pairs

$$((5,2),2), ((4,3),3), ((3,2,2),2).$$

When $k = 9$, there are six pairs

$$((7,2),2), ((6,3),3), ((5,4),4), ((5,2,2),2), ((4,3,2),3), ((4,3,2),2).$$

Combining Corollary 3.10.4, Equation (3.94) and Lemma 3.10.5, we obtain the following proposition.

Proposition 3.10.6. *Suppose* $C_1, \ldots, C_{k+1} \in \mathfrak{L}_{\geq 1}(\log \mathcal{H})$ *and* $\sum_{i=1}^{k+1} C_i \in \mathfrak{L}_{\geq 2}(\log \mathcal{H})$. *Let* $m \geq 2$ *and* $\boldsymbol{j} = (j_1, \ldots, j_m) \in \{1, \ldots, k+1\}^m$. *The expression* $\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_m)})$ *is equivalent (under* $\sim$*) to a linear combination of* $\widehat{M}(P, c)$, *where* (P, c) *ranges over all partition-integer pairs with* $c \neq \max(P)$ *and* $\min(P) \geq 2$. *Furthermore, this linear combination can be effectively computed.*

In other words, one can effectively compute rational numbers $\gamma_{(P,c)}$, *such that*

$$\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_m)}) \sim \sum_{(P,c)} \gamma_{(P,c)} \widehat{M}(P, c).$$

Proof. By the Dynkin formula (Lemma 3.5.4), the expression $\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_m)})$ can be rewritten into a linear combination of $\Phi(\mathrm{SP}(\boldsymbol{j}'))$, where $\boldsymbol{j}'$ are subsequences (with possible repetition) of $\boldsymbol{j}$. Then, Corollary 3.10.4 shows that each $\Phi(\mathrm{SP}(\boldsymbol{j}'))$ is equivalent (under $\sim$) to a linear combination of $\Phi(S')$ with $\min(\mathrm{IP}(S')) \geq 2$. Next, Equation (3.94) shows that each $\Phi(S'), \min(\mathrm{IP}(S')) \geq 2$ is equal to a linear combination of $M(S'')$ with $\min(\mathrm{IP}(S'')) \geq 2$. The condition $\min(\mathrm{IP}(S'')) \geq 2$ is due to the fact that for any $\tau \in \mathrm{S}_k$ we have $\mathrm{IP}(S_\tau) = \mathrm{IP}(S)$. Finally, by Lemma 3.10.5, each $M(S''), \min(\mathrm{IP}(S'')) \geq 2$ is equivalent (under $\sim$) to a linear combination of $\widehat{M}(P, c)$ with $c \neq \max(P)$ and $\min(P) \geq 2$.

In summary, any expression $\sum_{\sigma \in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)}, \ldots, C_{\sigma(j_r)})$ is equivalent to a linear combination of $\widehat{M}(P, c)$, where (P, c) ranges over all partition-integer pairs with $c \neq \max(P)$ and $\min(P) \geq 2$. Furthermore, the proof of Corollary 3.10.4, Equation (3.94) and Lemma 3.10.5 give an effective procedure that computes the coefficients of this linear combination. $\square$

The effective procedure of Proposition 3.10.6 is summarized by Algorithm 3.3. Note that for the algorithm we fix the integer k, so all set partitions in the algorithm refer to set partitions of k.

We can now give computer assisted proofs of Lemmas 3.5.7 - 3.5.10 based on Algorithm 3.3.

Proof of Lemma 3.5.7. (The SageMath [95] code can be found at https://doi.org/10.6084/m9.figshare.20124146.v1.) Set $k = 5$. Using Algorithm 3.3 on the tuples $(1, 2, 3, 4, 5, 6)$ and

$$\boldsymbol{j} = (1, 2, 3, 4, 4, 5, 5, 6, 6, 1, 2, 3),$$

we get

$$\sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(6)}) \sim \widehat{M}((3, 2), 2),$$
$$\sum_{\sigma \in \mathrm{S}_6} H_5\big(\log B_{\sigma(j_1)}, \ldots, \log B_{\sigma(j_{12})}\big) \sim -\widehat{M}((3, 2), 2).$$

Therefore,

$$\sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(6)}) + \sum_{\sigma \in \mathrm{S}_6} H_5(\log B_{\sigma(j_1)}, \ldots, \log B_{\sigma(j_{12})}) \sim 0.$$

□

Proof of Lemma 3.5.9. (The SageMath [95] code can be found at https://doi.org/10.6084/m9.figshare.20124113.v1.) Set $k = 7$. Using Algorithm 3.3 on the tuples $(1, 2, \ldots, 8)$ and

$$\boldsymbol{j}_1 = (j_{1,1}, j_{1,2}, \ldots, j_{1,16}) = (1, 2, 3, 4, 5, 5, 6, 6, 7, 7, 8, 8, 1, 2, 3, 4),$$
$$\boldsymbol{j}_2 = (j_{2,1}, j_{2,2}, \ldots, j_{2,16}) = (1, 2, 3, 4, 5, 4, 6, 7, 1, 2, 8, 3, 5, 6, 7, 8).$$

We get

$$\sum_{\sigma \in \mathrm{S}_8} H_7(\log B_{\sigma(1)}, \ldots, \log B_{\sigma(8)}) \sim \frac{34}{15}\widehat{M}((5, 2), 2) - \frac{34}{45}\widehat{M}((4, 3), 3) + \frac{68}{15}\widehat{M}((3, 2, 2), 2),$$

$$\sum_{\sigma \in \mathrm{S}_8} H_7\big(\log B_{\sigma(j_{1,1})}, \ldots, \log B_{\sigma(j_{1,16})}\big) \sim \frac{34}{15}\widehat{M}((5, 2), 2) + \frac{238}{45}\widehat{M}((4, 3), 3) - \frac{68}{5}\widehat{M}((3, 2, 2), 2),$$

$$\sum_{\sigma \in \mathrm{S}_8} H_7\big(\log B_{\sigma(j_{2,1})}, \ldots, \log B_{\sigma(j_{2,16})}\big) \sim -\frac{68}{15}\widehat{M}((5, 2), 2) + \frac{34}{45}\widehat{M}((4, 3), 3) - \frac{34}{5}\widehat{M}((3, 2, 2), 2).$$

Therefore,

$$\sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(1)},\dots,\log B_{\sigma(8)}) + \sum_{s=1}^{2}\alpha_s \sum_{\sigma\in \mathrm{S}_8} H_7(\log B_{\sigma(j_{s,1})},\dots,\log B_{\sigma(j_{s,16})}) \sim 0$$

with $\alpha_1 = \frac{1}{15}, \alpha_2 = \frac{8}{15}$. □

Proof of Lemma 3.5.10. (The SageMath [95] code can be found at https://doi.org/10.6084/m9.figshare.20122979.v1). Set $k = 9$. Using Algorithm 3.3 on the tuples $(1, 2, \dots, 10)$ and

$$\begin{aligned}
(j_{1,1}, j_{1,2}, \dots, j_{1,20}) &= (5, 4, 7, 10, 2, 8, 3, 8, 1, 9, 7, 6, 5, 6, 2, 3, 9, 10, 1, 4),\\
(j_{2,1}, j_{2,2}, \dots, j_{2,20}) &= (8, 3, 5, 7, 10, 6, 8, 2, 1, 10, 2, 4, 9, 1, 5, 9, 3, 6, 7, 4),\\
(j_{3,1}, j_{3,2}, \dots, j_{3,20}) &= (7, 10, 2, 6, 4, 9, 6, 4, 1, 5, 3, 5, 1, 9, 3, 7, 10, 2, 8, 8),\\
(j_{4,1}, j_{4,2}, \dots, j_{4,20}) &= (10, 2, 2, 6, 7, 1, 9, 3, 9, 4, 8, 7, 8, 5, 5, 1, 4, 10, 6, 3),\\
(j_{5,1}, j_{5,2}, \dots, j_{5,20}) &= (3, 5, 10, 1, 4, 8, 6, 9, 3, 2, 7, 6, 1, 10, 9, 7, 2, 4, 5, 8),\\
(j_{6,1}, j_{6,2}, \dots, j_{6,20}) &= (4, 7, 2, 10, 2, 1, 3, 5, 8, 1, 6, 9, 10, 7, 6, 8, 3, 5, 9, 4).
\end{aligned}$$

We get

$$\begin{aligned}
\sum_{\sigma\in \mathrm{S}_{10}} H_9(\log B_{\sigma(1)},\dots,\log B_{\sigma(10)}) &\sim \frac{347}{105}\widehat{M}((7,2),2) + \frac{347}{315}\widehat{M}((6,3),3)\\
&+ \frac{347}{105}\widehat{M}((5,4),4) + \frac{1388}{105}\widehat{M}((5,2,2),2) - \frac{347}{21}\widehat{M}((4,3,2),3) + \frac{347}{21}\widehat{M}((4,3,2),2),
\end{aligned}$$

$$\begin{aligned}
\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{1,1})},\dots,\log B_{\sigma(j_{1,20})}\big) &\sim -\frac{347}{105}\widehat{M}((7,2),2) + \frac{21167}{945}\widehat{M}((6,3),3)\\
&- \frac{4511}{315}\widehat{M}((5,4),4) + 0\cdot\widehat{M}((5,2,2),2) + \frac{3817}{63}\widehat{M}((4,3,2),3) + \frac{1735}{63}\widehat{M}((4,3,2),2),
\end{aligned}$$

$$\begin{aligned}
\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{2,1})},\dots,\log B_{\sigma(j_{2,20})}\big) &\sim \frac{347}{45}\widehat{M}((7,2),2) + \frac{18391}{945}\widehat{M}((6,3),3)\\
&+ \frac{347}{14}\widehat{M}((5,4),4) - \frac{1388}{315}\widehat{M}((5,2,2),2) + \frac{9022}{63}\widehat{M}((4,3,2),3) - \frac{694}{63}\widehat{M}((4,3,2),2),
\end{aligned}$$

$$\begin{aligned}
\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{3,1})},\dots,\log B_{\sigma(j_{3,20})}\big) &\sim \frac{16309}{42}\widehat{M}((7,2),2) + \frac{85709}{630}\widehat{M}((6,3),3)\\
&+ \frac{241859}{1260}\widehat{M}((5,4),4) + \frac{30883}{126}\widehat{M}((5,2,2),2) - \frac{8675}{63}\widehat{M}((4,3,2),3) + \frac{94037}{630}\widehat{M}((4,3,2),2),
\end{aligned}$$

$$\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{4,1})},\dots,\log B_{\sigma(j_{4,20})}\big) \sim \frac{20473}{210}\widehat{M}((7,2),2) - \frac{314729}{1890}\widehat{M}((6,3),3)$$
$$+ \frac{4511}{140}\widehat{M}((5,4),4) + \frac{137759}{630}\widehat{M}((5,2,2),2) - \frac{23249}{315}\widehat{M}((4,3,2),3) + \frac{33659}{210}\widehat{M}((4,3,2),2),$$

$$\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{5,1})},\dots,\log B_{\sigma(j_{5,20})}\big) \sim \frac{347}{210}\widehat{M}((7,2),2) + \frac{35741}{1890}\widehat{M}((6,3),3)$$
$$- \frac{18391}{1260}\widehat{M}((5,4),4) + \frac{1041}{70}\widehat{M}((5,2,2),2) - \frac{347}{63}\widehat{M}((4,3,2),3) + \frac{1735}{126}\widehat{M}((4,3,2),2),$$

$$\sum_{\sigma\in \mathrm{S}_{10}} H_9\big(\log B_{\sigma(j_{6,1})},\dots,\log B_{\sigma(j_{6,20})}\big) \sim -\frac{1388}{105}\widehat{M}((7,2),2) - \frac{56561}{945}\widehat{M}((6,3),3)$$
$$+ \frac{4511}{126}\widehat{M}((5,4),4) - \frac{3123}{70}\widehat{M}((5,2,2),2) - \frac{28454}{315}\widehat{M}((4,3,2),3) - \frac{51703}{630}\widehat{M}((4,3,2),2).$$

Therefore,

$$\sum_{\sigma\in \mathrm{S}_{10}} H_9(\log B_{\sigma(1)},\dots,\log B_{\sigma(10)}) + \sum_{s=1}^{6}\alpha_s \sum_{\sigma\in \mathrm{S}_{10}} H_9(\log B_{\sigma(j_{s,1})},\dots,\log B_{\sigma(j_{s,20})}) \sim 0$$

with $\alpha_1 = \frac{44566633}{13702661}, \alpha_2 = \frac{557040}{13702661}, \alpha_3 = \frac{205175}{3915046}, \alpha_4 = \frac{1307207}{13702661}, \alpha_5 = \frac{86275275}{27405322}, \alpha_6 = \frac{4105194}{1957523}$. □

Algorithm 3.3 Computing $\gamma_{(P,c)}$ where $\sum_{\sigma\in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)},\ldots,C_{\sigma(j_m)}) \sim \sum_{(P,c)} \gamma_{(P,c)}\widehat{M}(P,c)$

Input: an integer k and a tuple $\boldsymbol{j} = (j_1,\ldots,j_m) \in \{1,\ldots,k+1\}^m$.

Output: rational numbers $\gamma_{(P,c)}$, where (P,c) ranges over all partition-integer pairs with $c \neq \max(P)$ and $\min(P) \geq 2$.

1. **Compute rational numbers a_S such that**

$$\sum_{\sigma\in \mathrm{S}_{k+1}} H_k(C_{\sigma(j_1)},\ldots,C_{\sigma(j_m)}) = \sum_{\text{set partition } S} a_S \Phi(S) \tag{3.98}$$

in the following way:

(a) Initialize with $a_S := 0$ for all set partitions S.

(b) For every tuple $(i_1,\ldots,i_m) \in \mathbb{Z}^m_{\geq 0}$ such that $i_1+\cdots+i_m = k$, compute the sequence

$$\iota := (\underbrace{j_1,\ldots,j_1}_{i_1},\underbrace{j_2,\ldots,j_2}_{i_2},\ldots,\underbrace{j_m,\ldots,j_m}_{i_m})$$

and update $a_{\mathrm{SP}(\iota)} := a_{\mathrm{SP}(\iota)} + \frac{(k+1-\mathrm{card}(\mathrm{SP}(\iota)))!}{i_1!\cdots i_m!}$.

2. **Compute rational numbers b_S such that**

$$\sum_{\text{set partition } S} a_S \Phi(S) = \sum_{\substack{\text{set partition } S,\\ \min(\mathrm{IP}(S))\geq 2}} b_S \Phi(S) \tag{3.99}$$

in the following way:

(a) Initialize with $b_S := a_S$ for all set partitions S.

(b) Order all set partitions S into $S_1, S_2, \ldots, S_p$, such that if $S_j \succcurlyeq S_i$ then $j \geq i$.

(c) For $i = 1, 2, \ldots, p$:

If $\min(\mathrm{IP}(S_i)) = 1$, then update $b_{S_i} := 0$ and $b_{S_j} := b_{S_j} - b_{S_i}$ for all $S_j \succcurlyeq S_i$.

3. **Compute rational numbers g_S such that**

$$\sum_{\substack{\text{set partition } S,\\ \min(\mathrm{IP}(S))\geq 2}} b_S \Phi(S) = \sum_{\substack{\text{set partition } S,\\ \min(\mathrm{IP}(S))\geq 2}} g_S M(S) \tag{3.100}$$

in the following way:

(a) Initialize with $g_S := 0$ for all set partitions $S, \min(\mathrm{IP}(S)) \geq 2$.

(b) For every set partition S and every permutation $\sigma \in \mathrm{S}_k$, compute the set partition

$$S_\sigma := \Big\{ \{\sigma(j) \mid j \in A\} \;\Big|\; A \in S \Big\}$$

and update $g_{S_\sigma} := g_{S_\sigma} + b_S \cdot \frac{(-1)^{d(\sigma)}}{k^2 \binom{k-1}{d(\sigma)}}$ (where $d(\cdot)$ denotes the number of descents).

(To be continued in the next page)

Algorithm 3.3 (continued)

4. **Compute all partition-integer pairs** (P, c) **with** $c \neq \max(P)$ **and** $\min(P) \geq 2$**.**
5. **Compute rational numbers** $\gamma_{(P,c)}$ **such that**

$$\sum_{\substack{\text{set partition } S \\ \min(\mathrm{IP}(S))\geq 2}} g_S M(S) = \sum_{\substack{(P,c) \\ c\neq\max(P),\min(P)\geq 2}} \gamma_{(P,c)} \widehat{M}(P,c) \tag{3.101}$$

in the following way:

(a) Initialize with $\gamma_{(P,c)} \coloneqq 0$ for all (P, c), $c \neq \max(P)$ and $\min(P) \geq 2$.

(b) For all set partitions S with $\min(\mathrm{IP}(S)) \geq 2$:

i. If $1 \in A$, $\mathrm{card}(A) = \max(\mathrm{IP}(S))$ and $2 \in B$, $\mathrm{card}(B) \neq \max(\mathrm{IP}(S))$, then update $\gamma_{(\mathrm{IP}(S),\mathrm{card}(B))} \coloneqq \gamma_{(\mathrm{IP}(S),\mathrm{card}(B))} + g_S$.

ii. If $1 \in A$, $\mathrm{card}(A) \neq \max(\mathrm{IP}(S))$ and $2 \in B$, $\mathrm{card}(B) = \max(\mathrm{IP}(S))$, then update $\gamma_{(\mathrm{IP}(S),\mathrm{card}(A))} \coloneqq \gamma_{(\mathrm{IP}(S),\mathrm{card}(A))} - g_S$.

iii. If $1 \in A$, $\mathrm{card}(A) \neq \max(\mathrm{IP}(S))$ and $2 \in B$, $\mathrm{card}(B) \neq \max(\mathrm{IP}(S))$, then update $\gamma_{(\mathrm{IP}(S),\mathrm{card}(A))} \coloneqq \gamma_{(\mathrm{IP}(S),\mathrm{card}(A))} - g_S$, $\gamma_{(\mathrm{IP}(S),\mathrm{card}(B))} \coloneqq \gamma_{(\mathrm{IP}(S),\mathrm{card}(B))} + g_S$.

6. **Return the numbers** $\gamma_{(P,c)}$**.**

Chapter 4

Metabelian groups

4.1 Introduction and main result

In this chapter we study algorithmic problems in finitely generated *metabelian groups* (recall Definition 2.1.6). Developing a complete algorithmic theory for finitely generated metabelian groups has been the focus of extensive research since the 1950s. For recent advancements in this area, we refer readers to surveys [9, 55]. As the convention in computational group theory, a finitely generated metabelian group G is typically represented by a *finite metabelian presentation*[1] (see Section 4.2 for its definition). This differs from the usual *finite presentation* of groups, since not all metabelian groups admit a finite presentation [6]. Every finitely generated metabelian group admits a finite metabelian presentation, making them a natural target for algorithmic methods [9, p.629].

Among the classic Max Dehn problems for finitely generated metabelian groups, decidability of the Word Problem is known since the 1950s following the seminal work of Hall [43]. The Conjugacy Problem is shown to be decidable by Noskov [81]. The Isomorphism Problem remains an outstanding open problem [10]. We may note that in the hierarchy of solvable groups, metabelian groups are on the fringe of decidability. By a celebrated result of Kharlampovich, all three Max Dehn problems are undecidable in solvable groups of derived length three (recall Definition 2.1.9) [54], [55, Theorem 6.17, Section 6.8].

In finitely generated metabelian groups, decidability of Group Membership is a classic result of Romanovskii [87]. Romanovskii's solution is based on a reduction from Group Membership to deciding membership in a finitely presented module over polynomials rings. On the other hand, Semigroup Membership is undecidable for many instances of metabelian group. For example, Lohrey, Steinberg, and Zetzsche [67] showed undecidability of Semigroup Membership in

[1] also called an $\mathscr{A}^2$-presentation in literature.

the wreath product $\mathbb{Z} \wr \mathbb{Z}$ by embedding the halting problem for two-counter machines. More recently, as mentioned in the previous chapter, Roman'kov [86] showed undecidability of Semigroup Membership in $\mathrm{H}_3(\mathbb{Q})^k$ for sufficiently large k by embedding the Hilbert's tenth problem.

While the undecidability of Semigroup Intersection in general metabelian groups has never been explicitly stated in literature, it can be easily deduced from the undecidability of the Post Correspondence Problem. Indeed, the free monoid $\{a, b\}^*$ over two generators can be embedded as a submonoid of a finitely generated metabelian group (such as the wreath product $\mathbb{Z} \wr \mathbb{Z}$ or the free metabelian group over two generators [11, 70]). Let G be this metabelian group, then the direct product $G \times G$ is metabelian and contains as a submonoid the direct product $\{a, b\}^* \times \{a, b\}^*$. Hence, Semigroup Intersection is undecidable in the finitely generated metabelian group $G \times G$ by the following classic result of Emil Post.

Theorem 4.1.1 (Post Correspondence Problem [84])**.** *The following problem is undecidable: given as input a set of elements $S = \{(v_1, w_1), \ldots, (v_K, w_K)\} \subset \{a, b\}^* \times \{a, b\}^*$, decide whether the semigroup $\langle S \rangle$ intersects the set of diagonals $\{(x, x) \mid x \in \{a, b\}^*\}$.*

Note that the set of diagonals is the semigroup generated by the neutral element and the elements $(a, a), (b, b)$. Therefore, the Post Correspondence Problem shows it is undecidable whether the intersection $\langle S \rangle \cap \langle (I, I), (a, a), (b, b) \rangle$ is empty.

Despite the undecidability of Semigroup Membership and Semigroup Intersection in finitely generated metabelian groups, the decidability status of the Identity Problem and the Group Problem remained open. The main result of this chapter gives a positive answer to this open problem.

Main result

Theorem 4.1.2. *The Group Problem (hence also the Identity Problem) is decidable in all finitely generated metabelian groups.*

Here, the metabelian group is represented by a finite metabelian presentation.

It has been noticed since the work of Hall that metabelian groups have natural connections with polynomials rings. Indeed, this connection is the key to deciding many *group* algorithmic problems in metabelian groups. However, a parallel theory for *semigroups* was yet to be developed. In this chapter, we establish a connection between sub-*semigroups* of metabelian groups and polynomial *semirings*, utilizing graph theory as an intermediate step. The connection will be the key in solving the Group Problem, and consequently, the Identity Problem.

Organization of the chapter

The organization of this chapter is as follows. In Section 4.2, we exhibit an embedding (Proposition 4.2.2) and reduce the Group Problem in metabelian groups to the Group Problem in semidirect products of the form $\mathcal{Y} \rtimes \mathbb{Z}^n$. We focus on algorithmic problems in $\mathcal{Y} \rtimes \mathbb{Z}^n$ in subsequent sections. In Section 4.3, we introduce *$\mathcal{G}$-graphs* as means to describe words over a finite alphabet $\mathcal{G} \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$. Section 4.4 outlines a solution for the Group Problem in $\mathcal{Y} \rtimes \mathbb{Z}^n$, relying on three key theorems: Theorem 4.4.3, which relates connectivity of a graph to its "face-accessibility"; Theorem 4.4.8, which establishes a local-global principle for solving linear equations over polynomial semirings; and Theorem 4.4.10, which shows decidability of the "local" conditions in the previous theorem. Section 4.5 presents the proof of Theorem 4.4.3 using a combination of convex geometry and graph theory. Section 4.6 proves Theorem 4.4.8 by generalizing a result of Einsiedler, Mouat and Tuncel [36, Theorem 1.3] using algebraic geometry techniques. Finally, Section 4.7 provides the proof of Theorem 4.4.10, employing a combination of the first order theory of reals, Gröbner basis over modules and number theory.

4.2 Representing a metabelian group

Let R be a commutative ring (such as $\mathbb{Z}$ or $\mathbb{R}$) or semiring (such as $\mathbb{N}$ or $\mathbb{R}_{\geq 0}$). Denote by $R[X_1^{\pm}, \ldots, X_n^{\pm}]$ the Laurent polynomial ring or semiring over n variables with coefficients in R: this is the set of polynomials of the form

$$\sum_{a_1,\ldots,a_n \in \mathbb{Z}} c_{a_1,\ldots,a_n} X_1^{a_1} \cdots X_n^{a_n},$$

where $c_{a_1,\ldots,a_n} \in R$ and $(a_1, \ldots, a_n)^\top$ ranges over a finite subset of $\mathbb{Z}^n$. Unless otherwise specified, all polynomials considered in this chapter are Laurent polynomials. When n is fixed, we denote

$$R[\overline{X}^{\pm}] \coloneqq R[X_1^{\pm}, \ldots, X_n^{\pm}], \quad R[\overline{X}^{\pm}]^* \coloneqq R[\overline{X}^{\pm}] \setminus \{0\}.$$

For $a = (a_1, \ldots, a_n)^\top \in \mathbb{Z}^n$, denote by $\overline{X}^a$ the monomial $X_1^{a_1} X_2^{a_2} \cdots X_n^{a_n}$. Given $f \in R[\overline{X}^{\pm}]^*$ and a vector $v \in (\mathbb{R}^n)^* \coloneqq \mathbb{R}^n \setminus \{0\}$, define the *weighted degree*

$$\deg_v(f) \coloneqq \max\{v^\top a \mid a \in \mathbb{Z}^n, c_{a_1,\ldots,a_n} \neq 0\}, \quad \text{where } f = \sum c_{a_1,\ldots,a_n} \overline{X}^a \neq 0.$$

Additionally, define $\deg_v(0) = -\infty$ for all $v \in (\mathbb{R}^n)^*$.

Let R be a commutative ring. An *$R[\overline{X}^{\pm}]$-module* is an abelian group $(M, +)$ along with an

operation $\cdot : R[\overline{X}^{\pm}] \times M \to M$ satisfying $f \cdot (a+b) = f \cdot a + f \cdot b$, $(f+g) \cdot a = f \cdot a + g \cdot a$, $fg \cdot a = f \cdot (g \cdot a)$ and $1 \cdot a = a$. For example, for any $d \in \mathbb{N}$, $R[\overline{X}^{\pm}]^d$ is an $R[\overline{X}^{\pm}]$-module by $f \cdot (g_1, \ldots, g_d) = (fg_1, \ldots, fg_d)$. Throughout this chapter, we use the bold symbol $\boldsymbol{f}$ to denote a vector $(f_1, \ldots, f_d) \in R[\overline{X}^{\pm}]^d$.

Given $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in R[\overline{X}^{\pm}]^d$, we say they *generate* the $R[\overline{X}^{\pm}]$-module $\sum_{i=1}^m R[\overline{X}^{\pm}] \cdot \boldsymbol{g}_i \coloneqq \{\sum_{i=1}^m p_i \cdot \boldsymbol{g}_i \mid p_1, \ldots, p_m \in R[\overline{X}^{\pm}]\}$. A module is called *finitely generated* if it can be generated by a finite number of elements. Given two finitely generated submodules N, M of $R[\overline{X}^{\pm}]^d$ such that $N \subseteq M$, we can define the quotient $M/N \coloneqq \{\overline{m} \mid m \in M\}$ where $\overline{m_1} = \overline{m_2}$ iff $m_1 - m_2 \in N$. This quotient is also an $R[\overline{X}^{\pm}]$-module. We say that an $R[\overline{X}^{\pm}]$-module $\mathcal{Y}$ is *finitely presented* if it can be written as a quotient M/N for two finitely generated submodules $N \subseteq M$ of $R[\overline{X}^{\pm}]^d$ for some $d \geq 1$. We call a *finite presentation* of $\mathcal{Y}$ the respective generators of such M, N.

Metabelian groups are usually represented by a *finite metabelian presentation*. We now give its formal definition. Understanding the technical details in the definition is not essential, since we will only be using the more intuitive representation given by Equations (4.1), (4.2) and Proposition 4.2.2 throughout this chapter.

Let F_s be the free group over $s \geq 2$ generators. The quotient

$$M_s \coloneqq F_s / [[F_s, F_s], [F_s, F_s]]$$

is metabelian and is called the *free metabelian group* over s generators. Let $\{x_1, \ldots, x_s\}$ be the generators of F_s, then their equivalence classes $\{\overline{x}_1, \ldots, \overline{x}_s\}$ are the generators of M_s. An element of M_s is represented as a word over $\{\overline{x}_1, \ldots, \overline{x}_s\}$.

Definition 4.2.1 (Finite metabelian presentation)**.** A *finite metabelian presentation* of a group G is a free metabelian group $M_s, s \geq 2$, along with a finite set of elements $r_1, \ldots, r_m \in M_s$, such that

$$G = M_s / \operatorname{ncl}_{M_s}(r_1, \ldots, r_m).$$

Here, $\operatorname{ncl}_{M_s}(r_1, \ldots, r_m)$ denotes the *normal closure* of $\{r_1, \ldots, r_m\}$, that is, the smallest normal subgroup of M_s containing $\{r_1, \ldots, r_m\}$.

By [43, Corollary 1] or [9, p.629], every finitely generated metabelian group admits a finite metabelian presentation. Finite metabelian presentations are not to be confused with the usual finite presentation of groups (Definition 2.1.1). Every *finite presentation* of a metabelian group G naturally induces a *finite metabelian presentation* of G. Indeed, if $\langle x_1, \ldots, x_s \mid r_1, \ldots, r_m \rangle$ is a finite presentation of a metabelian group G, then $G \cong F_s / \operatorname{ncl}_{F_s}(r_1, \ldots, r_m)$, where F_s denotes the free group generated by $x_1, \ldots, x_s$. Since G is metabelian, we have $[[F_s, F_s], [F_s, F_s]] \leq$

$\mathrm{ncl}_{F_s}(r_1, \ldots, r_m)$. Therefore, letting $M_s = F_s/[[F_s, F_s], [F_s, F_s]]$, we obtain a finite metabelian presentation

$$G \cong M_s / \mathrm{ncl}_{M_s}(r_1, \ldots, r_m)$$

of the group G. Nevertheless, while every finitely generated metabelian group admits a *finite metabelian presentation*, some of them (such as the wreath product $\mathbb{Z} \wr \mathbb{Z}$) do not admit a *finite presentation* [6, Theorem 1].

Given a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module $\mathcal{Y}$, define the following semidirect product:

$$\mathcal{Y} \rtimes \mathbb{Z}^n := \{(y, a) \mid y \in \mathcal{Y}, a \in \mathbb{Z}^n\}; \tag{4.1}$$

this is a group where multiplication and inversion are defined by

$$(y, a) \cdot (y', a') = (y + \overline{X}^a \cdot y', a + a'), \quad (y, a)^{-1} = (-\overline{X}^{-a} \cdot y, -a). \tag{4.2}$$

The neutral element of $\mathcal{Y} \rtimes \mathbb{Z}^n$ is $(0, 0)$. Intuitively, the element (y, a) can be seen as a 2×2 matrix $\begin{pmatrix} \overline{X}^a & y \\ 0 & 1 \end{pmatrix}$, where the group law is represented by matrix multiplication.[2] There is a canonical projection π of the group $\mathcal{Y} \rtimes \mathbb{Z}^n$ onto $\mathbb{Z}^n$:

$$\pi \colon \mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n,$$
$$(y, a) \mapsto a.$$

By Lemma 2.2.4, solving the Group Problem in metabelian groups will also provide a solution for the Identity Problem. The following proposition shows that in order to solve the Group Problem in metabelian groups, it suffices to solve it in groups of the form $\mathcal{Y} \rtimes \mathbb{Z}^n$.

Proposition 4.2.2. *Suppose we are given a finite metabelian presentation of a group G as well as a finite set $\mathcal{G} \subseteq G$. One can effectively construct a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module $\mathcal{Y}$ for some $n \in \mathbb{N}$, as well as a subset $\widetilde{\mathcal{G}}$ of the group $\mathcal{Y} \rtimes \mathbb{Z}^n$, such that $\langle \mathcal{G} \rangle$ is a group if and only if $\langle \widetilde{\mathcal{G}} \rangle$ is a group. Furthermore, the constructed set $\widetilde{\mathcal{G}}$ satisfies $\pi(\langle \widetilde{\mathcal{G}} \rangle_{grp}) = \mathbb{Z}^n$ under the canonical projection $\pi \colon \mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$.*

The rest of this section aims to prove Proposition 4.2.2. Readers interested only in the decision procedure for the Group Problem in $\mathcal{Y} \rtimes \mathbb{Z}^n$ may skip the rest of this section and accept Proposition 4.2.2 as a black-box.

[2] When $n = 0$, the polynomial ring $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$ becomes $\mathbb{Z}$, and the group $\mathcal{Y} \rtimes \mathbb{Z}^n$ degenerates into the $\mathbb{Z}$-module $\mathcal{Y}$ itself, which is an abelian group.

Let T be an arbitrary group. The group ring $\mathbb{Z}[T]$ is defined as

$$\mathbb{Z}[T] \coloneqq \left\{ \sum_{t\in T} z_t t \;\middle|\; z_t \in \mathbb{Z}, \text{ and } z_t = 0 \text{ except for a finite number of } t \in T \right\}.$$

Here, $\sum_{t\in T} z_t t$ denotes a formal sum with finite support. The ring structure on $\mathbb{Z}[T]$ is given by $\sum_{t\in T} y_t t + \sum_{t\in T} z_t t = \sum_{t\in T}(y_t + z_t)t$ and $\sum_{t\in T} y_t t \cdot \sum_{t\in T} z_t t = \sum_{t\in T}\sum_{uv=t} y_u z_v t$.

We start by stating a lemma of Baumslag, Cannonito, and Robinson:

Lemma 4.2.3 (Reformulation of [9, Theorem 3.3], [64, Theorem 9.5.3])**.** *There is an algorithm which, when a finite metabelian presentation of G is given, together with a finite subset $\mathcal{G} \subseteq G$, finds a finite metabelian presentation of the subgroup $\langle \mathcal{G} \rangle_{grp}$. Furthermore, under this finite metabelian presentation of $\langle \mathcal{G} \rangle_{grp}$, the set $\mathcal{G}$ itself is explicitly given as the generators.*

Proof. The first part of the statement is simply [64, Theorem 9.5.3]. We now retrace its proof to show that the set $\mathcal{G}$ is explicitly given as the generators under this finite metabelian presentation.

Denote $H = \langle \mathcal{G} \rangle_{grp}$, $G' = [G, G]$, and $\bar{H} = HG'/G' \cong H/H \cap G'$. The abelian group $H \cap G'$ is a $\mathbb{Z}[\bar{H}]$-module by action of conjugation. That is, for $hG' \in \bar{H}$, the action of hG' on $H \cap G'$ is defined as $(hG') \cdot x \coloneqq h^{-1}xh$. This action is well-defined (not depending on the choice of h) since G' is normal and abelian.

Following the proof of [64, Theorem 9.5.3], we have:

(i) Write $\mathcal{G} = \{g_1, \ldots, g_K\}$, then the elements $g_1G', \ldots, g_KG'$ generate the abelian group $\bar{H}$. We can compute a finite set of elements

$$y_i \coloneqq (g_1G')^{n_{i1}} \cdots (g_KG')^{n_{iK}}, i = 1, \ldots, S, \tag{4.3}$$

such that $\bar{H}$ is the abelian group generated by $\{g_1G', \ldots, g_KG'\}$ modulo the relations $\{y_1, \ldots, y_S\}$.

(ii) The $\mathbb{Z}[\bar{H}]$-module $H \cap G'$ is generated by the set

$$\mathcal{C} \coloneqq \{[g_i, g_j] \mid g_i, g_j \in \mathcal{G}\} \cup \{g_1^{n_{i1}} \cdots g_K^{n_{iK}} \mid i = 1, \ldots, S\}.$$

A finite presentation of the $\mathbb{Z}[\bar{H}]$-module $H \cap G'$ under the generators $\mathcal{C}$ can be effectively computed (see [64, Theorem 9.4.6]). That is, for each $c \in \mathcal{C}$ let e_c be a new variable, we can compute a finite set of elements $f_1, \ldots, f_T$ in the free $\mathbb{Z}[\bar{H}]$-module $\sum_{c\in\mathcal{C}} \mathbb{Z}[\bar{H}] \cdot e_c$, such that $H \cap G'$ is isomorphic to the quotient of $\sum_{c\in\mathcal{C}} \mathbb{Z}[\bar{H}] \cdot e_c$ by the submodule generated by $f_1, \ldots, f_T$. This isomorphism is given by $c \mapsto e_c$.

The proof of [64, Theorem 9.5.3] then states that these data naturally give us a finite metabelian presentation of H. We now explicitly give this finite metabelian presentation. Let $M(\mathcal{G})$ be the free metabelian group generated by the set $\mathcal{G}$, and $r_1, \ldots, r_T$ be the elements in $M(\mathcal{G})$ obtained as follows. For each $f_i, i = 1, \ldots, T$, defined above, substitute e_c with c for all $c \in \mathcal{C}$, and write the action of $\mathbb{Z}[\bar{H}]$ using conjugation by elements of $\mathcal{G} \cup \mathcal{G}^{-1}$. We then obtain the elements $r_i, i = 1, \ldots, T$, in $M(\mathcal{G})$. For example, if $f_1 = (g_1 G') \cdot e_c$, then r_1 is obtained as $g_1^{-1} c g_1$. Then

$$H \cong M(\mathcal{G}) / \operatorname{ncl}_{M(\mathcal{G})}(r_1, \ldots, r_T)$$

is a finite metabelian presentation of H, where the set $\mathcal{G}$ is explicitly given as the generators.

Indeed, the elements $r_1, \ldots, r_T$ are obviously relations in H. On the other hand, let w be a word over $\mathcal{G} \cup \mathcal{G}^{-1}$ that is a relation in H. We show that w is normally generated by the relations $r_1, \ldots, r_T$ (and the double-commutators).

Note that wH' represents the neutral element in the abelianization H/H', this means that wH' can be written as $(r_1 H')^{p_1} \cdots (r_T H')^{p_T}$. This is because $r_1, \ldots, r_T$ generate $H \cap G'$ as a $\mathbb{Z}[\bar{H}]$-module, but conjugacy by $\mathcal{G}$ does not change the class of element in $H \cap G'/H'$, so $r_1 H', \ldots, r_T H'$ generate $H \cap G'/H'$ as an abelian group. In particular, by pushing the elements of H' to the right, w can be rewritten in the form

$$r_1^{p_1} \cdots r_T^{p_T} w_1 \cdots w_q, \tag{4.4}$$

where each $w_1, \ldots, w_q$ is a conjugate of some commutator $[g_i, g_j], g_i, g_j \in \mathcal{G}$, by some word over $\mathcal{G} \cup \mathcal{G}^{-1}$. Hence, it suffices to show that $w_1 \cdots w_q$ is normally generated by the relations $r_1, \ldots, r_T$. Note that $w_1, \ldots, w_q$ commute with each other by the double-commutator relations. Since $w, r_1, \ldots, r_T$ all represent the neutral element in H, the word $w_1 \cdots w_q$ also represents the neutral element in H. Therefore $w_1 \cdots w_q$ must be in the $\mathbb{Z}[\bar{H}]$-module generated by $r_1, \ldots, r_T$, so it can be written as a product of conjugates of $r_1, \ldots, r_T$. We conclude that w is normally generated by the relations $r_1, \ldots, r_T$. □

Since our goal is to decide the Group Problem (whether $\langle \mathcal{G} \rangle = \langle \mathcal{G} \rangle_{grp}$), we can without loss of generality suppose $G = \langle \mathcal{G} \rangle_{grp}$ by Lemma 4.2.3.

We now recall the definition of the *wreath product*. Given two groups A, T, their (restricted) wreath product $A \wr T$ is defined as a semidirect product $A^T \rtimes T$ (recall Definition 2.1.12). Here, A^T is the direct sum of A over the index set T and is called the *base group*. That is, A^T is the set of sequences $(a_s)_{s \in T}, a_s \in A$ where a_s is the neutral element for all but finitely many $s \in T$. It is a group by pointwise multiplication. There is a natural action φ of T on A^T by

$\varphi(t)\colon (a_s)_{s\in T} \mapsto (a_{t^{-1}s})_{s\in T}, t \in T$. Hence, the wreath product $A \wr T = A^T \rtimes T$ is the set of pairs $((a_s)_{s\in T}, t)$ with $(a_s)_{s\in T} \in A^T, t \in T$, where multiplication is defined by

$$((a_s)_{s\in T}, t) \cdot \left((a'_s)_{s\in T}, t'\right) = \left((a_s a'_{t^{-1}s})_{s\in T}, tt'\right).$$

The wreath product $A \wr T$ canonically contains as a subgroup $T \cong \{(I_{A^T}, t) \mid t \in T\}$, where I_{A^T} is the neutral element of A^T, as well as $A^T \cong \{((a_s)_{s\in T}, I_T) \mid (a_s)_{s\in T} \in A^T\}$, where I_T is the neutral element of T.

An important special case of the wreath product is when $A = \mathbb{Z}^n$ and T is abelian. In this case, the base group A^T is isomorphic to the direct power $(\mathbb{Z}[T])^n$. The wreath product $A \wr T$ then becomes the semidirect product $(\mathbb{Z}[T])^n \rtimes T$ consisting of the pairs (y, t), where $y \in (\mathbb{Z}[T])^n, t \in T$, with multiplication given by $(y, t) \cdot (y', t') = (y + t \cdot y', tt')$.

Furthermore, if $A = \mathbb{Z}^n$ and $T = \mathbb{Z}^d$, then the wreath product $A \wr T$ is simply the semidirect product $\left(\mathbb{Z}[X_1^{\pm}, \ldots, X_d^{\pm}]\right)^n \rtimes \mathbb{Z}^d$ defined in Equation (4.1). The following classic result of Magnus gives an explicit embedding of the quotient of a free group into a wreath product.

Lemma 4.2.4 (Magnus Embedding Theorem [7, Lemma 2], [69])**.** *Let F be a free group over the alphabet $X = \{x_i \mid i \in I\}$, and let R be a normal subgroup of F. Let the mapping*

$$x_i R \mapsto t_i, \quad i \in I,$$

define an isomorphism from F/R to a group T generated by $t_i, i \in I$. Furthermore, let A be a free abelian group, freely generated by the elements $a_i, i \in I$. Then the mapping

$$x_i [R, R] \mapsto a_i t_i, \quad i \in I,$$

defines an injection of $F/[R, R]$ into the wreath product $W = A \wr T$.

Recall that $\mathcal{Y} \rtimes \mathbb{Z}^n$ canonically contains the subgroup $\mathbb{Z}^n \cong \{(0, a) \mid a \in \mathbb{Z}^n\}$. The next lemma shows that a finitely generated metabelian group can be effectively embedded in a quotient $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$, where H is a subgroup of $\mathbb{Z}^n \cong \{(0, a) \mid a \in \mathbb{Z}^n\}$.

Lemma 4.2.5 (Corollary of [7, Lemma 3])**.** *Let G be a finitely generated metabelian group. Then G is isomorphic to a subgroup $\widetilde{G}$ of a quotient $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$, where*

(i) $n \in \mathbb{N}$ and $\mathcal{Y}$ is a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module.

(ii) H is a subgroup of $\mathbb{Z}^n \leq \mathcal{Y} \rtimes \mathbb{Z}^n$, and elements of H commute with all elements in $\mathcal{Y} \rtimes \mathbb{Z}^n$.

(iii) the image $\pi(\widetilde{G})$ under the projection $\pi : (\mathcal{Y} \rtimes \mathbb{Z}^n)/H \to \mathbb{Z}^n/H$ is equal to $\mathbb{Z}^n/H$.

Furthermore, given a finite metabelian presentation of G, the integer n, the finite presentation of $\mathcal{Y}$, the generators of H and the isomorphism $G \xrightarrow{\sim} \widetilde{G}$ can all be effectively computed.

Proof. This lemma is a simple extension of [7, Lemma 3]. We give a recount of its proof to show effectiveness and the conditions (ii) and (iii).

Let $\{g_1, \ldots, g_n\}$ be the generators of G, let F be the free group over the an alphabet $\{x_1, \ldots, x_n\}$, such that $M_n = F/[[F,F],[F,F]]$ and $G = M_n/\operatorname{ncl}_{M_n}(\widetilde{r_1}, \ldots, \widetilde{r_m})$ is the given finite metabelian presentation of G. Here, $\widetilde{r_i} = r_i[[F,F],[F,F]], i = 1, \ldots, m$ where r_i is given as an element of the free group F. Let ϕ be the epimorphism from F to G defined by

$$\phi\colon x_i \mapsto g_i, \quad (i = 1, \ldots, n).$$

Let K be the kernel of ϕ and let R be the inverse image of $[G,G]$ under ϕ. Since $\phi([F,F]) = [G,G]$ we have $R = \phi^{-1}([G,G]) = K[F,F]$. Then $R/K \cong [G,G]$ and hence is abelian. Therefore $[R,R] \leq K$, which means that

$$[R,R] \leq K \leq R.$$

Note that K is the normal closure of $r_1, \ldots, r_m$ and $[[F,F],[F,F]]$ in the group F.

Now let A be a free abelian group generated by the elements $a_1, \ldots, a_n$ and consider the wreath product $A \wr T$, where $T = F/R$. The structure of the abelian group T can be effectively computed by $T = F/R = F/\left(K[F,F]\right) \cong \left(F/[F,F]\right)/\left(K[F,F]/[F,F]\right)$. In other words, writing $r_j = x_{i_{j_1}}^{e_{j_1}} \cdots x_{i_{j_{\ell_j}}}^{e_{j_{\ell_j}}}, j = 1, \ldots, m$, and writing $t_i = x_i R, i = 1, \ldots, n$, then T is the quotient F_{ab}/H, where $F_{ab} = F/[F,F]$ is the free abelian group generated by $t_1, \ldots, t_n$, and $H = K[F,F]/[F,F]$ is the subgroup of F_{ab} generated by $t_{i_{j_1}}^{e_{j_1}} \cdots t_{i_{j_{\ell_i}}}^{e_{j_{\ell_i}}}, j = 1, \ldots, m$.

Let ψ be the homomorphism of F into $A \wr T$ defined by

$$\psi(x_i) = a_i t_i, \quad (i = 1, \ldots, n).$$

By Lemma 4.2.4 the kernel of ψ is $[R,R]$. Hence ψ induces an isomorphism $\psi_* : x_i[R,R] \mapsto a_i t_i$, of $F/[R,R]$ into $A \wr T$.

We put $N = \psi_*(K/[R,R])$. Now $N \leq \psi_*(F/[R,R])$. Therefore N is normalized by the elements $a_i t_i, i = 1, \ldots, n$. But it follows from the definition of ψ_* that N is contained in the base group $B = A^T$ of $A \wr T$. Since B is abelian, N is normalized by B and hence by the elements t_i, and therefore by all of $A \wr T$. In other words N is a normal subgroup of $A \wr T$, and is the normal closure of $\psi_*(r_1[R,R]), \ldots, \psi_*(r_m[R,R])$. This is because $N = \psi_*(K/[R,R])$ and K is

the normal closure of $\{r_1, \ldots, r_m\} \cup [[F,F],[F,F]]$ in F. Note that we can effectively write

$$\psi_*(r_j[R,R]) = \left(a_{i_{j_1}} t_{i_{j_1}}\right)^{e_{j_1}} \cdots \left(a_{i_{j_{\ell_j}}} t_{i_{j_{\ell_j}}}\right)^{e_{j_{\ell_j}}}, j = 1, \ldots, m. \tag{4.5}$$

Now G is isomorphic to the subgroup $\widetilde{G} \coloneqq \psi_*(F/[R,R])/N$ of $(A \wr T)/N$ by the map $g_i \mapsto a_i t_i N$.

Note that $(A \wr T)/N = \left(A^T/N\right) \rtimes T$ where N is normal, so N is the $\mathbb{Z}[T]$-module generated by $\psi_*(r_1[R,R]), \ldots, \psi_*(r_m[R,R])$. Furthermore, since $\widetilde{G}$ contains the elements $a_i t_i N, i = 1, \ldots, n$, its projection onto T contains the elements $t_i, i = 1, \ldots, n$ and hence is T itself.

We write $T = F_{ab}/H = \mathbb{Z}^n/H$ since F_{ab} is the free abelian group over n generators. The canonical projection $\mathbb{Z}^n \to T$ induces a ring homomorphism $\mathbb{Z}[\mathbb{Z}^n] \to \mathbb{Z}[T]$, so the $\mathbb{Z}[T]$-module A^T/N is naturally also a $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}] = \mathbb{Z}[\mathbb{Z}^n]$-module, where elements of the form $\overline{X}^h, h \in H$ act trivially. Taking $\mathcal{Y} \coloneqq A^T/N$, we define the semidirect product $\mathcal{Y} \rtimes \mathbb{Z}^n$ by considering $\mathcal{Y} = A^T/N$ as a $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module. We show that elements of H commute with every element in $\mathcal{Y} \rtimes \mathbb{Z}^n$. This is rather straightforward: since elements of the form $\overline{X}^h, h \in H$ act trivially on $\mathcal{Y} = A^T/N$, we have $(0,h)(y,a)(0,h^{-1}) = (\overline{X}^h \cdot y, a) = (y,a)$. Therefore, elements of H commute with every element in $\mathcal{Y} \rtimes \mathbb{Z}^n$, proving (ii). *A fortiori*, this shows that H is a normal subgroup of $\mathcal{Y} \rtimes \mathbb{Z}^n$.

Finally, we have

$$\widetilde{G} = \left(A^T/N\right) \rtimes T = \mathcal{Y} \rtimes (\mathbb{Z}^n/H) = (\mathcal{Y} \rtimes \mathbb{Z}^n)/H,$$

and (iii) follows directly from the fact that the projection $\widetilde{G} = \left(A^T/N\right) \rtimes T \to T = \mathbb{Z}^n/H$ has full image. We now show that $\mathcal{Y} = A^T/N$ can be effectively written as a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module. First, we write A^T as a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module in the following way. Let $h_1, \ldots, h_m$ the generators of H as a subgroup of $\mathbb{Z}^n$. Since $A \cong \mathbb{Z}^n$, we have $A^T = (\mathbb{Z}[T])^n$, and $\mathbb{Z}[T]$ is the quotient of $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}] = \mathbb{Z}[\mathbb{Z}^n]$ by the ideal generated by elements $\overline{X}^{h_i} - 1, i = 1, \ldots, m$. Hence we obtain a finite presentation of the $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module $\mathbb{Z}[T]$, and from it a finite presentation of the module $A^T = (\mathbb{Z}[T])^n$. The generators of $N \subseteq A^T$ as a $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module is the same as its generators as a $\mathbb{Z}[T]$-module, which are given by the elements in (4.5). Therefore, we obtain a finite presentation of the $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module $\mathcal{Y} = A^T/N$, and (i) follows. □

By Lemma 4.2.5 we can hence suppose G is given as a subgroup of $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$ and the generator set $\mathcal{G}$ is given as a subset of $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$. Write $\mathcal{G} = \{g_1 H, \ldots, g_k H\}$ where $g_1, \ldots, g_k$ are elements of $\mathcal{Y} \rtimes \mathbb{Z}^n$, and let $h_1, \ldots, h_M$ be the generators of $H \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$ as a *semi*group.

Lemma 4.2.6. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if the semigroup generated by the set $\{g_1, \dots, g_k, h_1, \dots, h_M\} \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$ is a group.*

Proof. Suppose $\langle \mathcal{G} \rangle$ is a group. Then by Lemma 2.2.1 there exists a full-image word

$$w = g_{i_1} H \cdot g_{i_2} H \cdot \cdots g_{i_p} H$$

over the alphabet $\mathcal{G}$ representing the neutral element in $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$. Since elements of H commute with every element of $\mathcal{Y} \rtimes \mathbb{Z}^n$, this means $g_{i_1} g_{i_2} \cdots g_{i_p} \in H$. Therefore there exists a word v over the alphabet $\mathcal{H} \coloneqq \{h_1, \dots, h_M\}$ such that $g_{i_1} g_{i_2} \cdots g_{i_p} \cdot v$ represents the neutral element in $\mathcal{Y} \rtimes \mathbb{Z}^n$. Since $\mathcal{H}$ generates the group H as a semigroup, there exists a full-image word v' over the alphabet $\mathcal{H}$ representing the neutral element of $H \leq \mathcal{Y} \rtimes \mathbb{Z}^n$. Then, the word $g_{i_1} g_{i_2} \cdots g_{i_p} \cdot v \cdot v'$ is full-image over the alphabet $\{g_1, \dots, g_k, h_1, \dots, h_M\}$ and it represents the neutral element in $\mathcal{Y} \rtimes \mathbb{Z}^n$. By Lemma 2.2.1, the semigroup generated by $\{g_1, \dots, g_k, h_1, \dots, h_M\}$ is a group.

For the other implication, suppose now the semigroup generated by $\{g_1, \dots, g_k, h_1, \dots, h_M\}$ is a group, so there exists a full-image word $\widetilde{w}$ over the alphabet $\{g_1, \dots, g_k, h_1, \dots, h_M\}$ representing the neutral element (Lemma 2.2.1). Since the elements $h_1, \dots, h_M$ commute with all other elements, we can move them to the rightmost side of $\widetilde{w}$ and suppose

$$\widetilde{w} = g_{i_1} g_{i_2} \cdots g_{i_p} h_{j_1} h_{j_2} \cdots h_{j_q}.$$

Then the word $g_{i_1} H \cdot g_{i_2} H \cdot \cdots g_{i_p} H$ is full-image over the alphabet $\mathcal{G}$ and represents the neutral element of $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$. By Lemma 2.2.1, the semigroup $\langle \mathcal{G} \rangle$ is a group. □

Proposition 4.2.2 follows from Lemmas 4.2.3, 4.2.5 and 4.2.6:

Proposition 4.2.2. *Suppose we are given a finite metabelian presentation of a group G as well as a finite set $\mathcal{G} \subseteq G$. One can effectively construct a finitely presented $\mathbb{Z}[X_1^{\pm}, \dots, X_n^{\pm}]$-module $\mathcal{Y}$ for some $n \in \mathbb{N}$, as well as a subset $\widetilde{\mathcal{G}}$ of the group $\mathcal{Y} \rtimes \mathbb{Z}^n$, such that $\langle \mathcal{G} \rangle$ is a group if and only if $\langle \widetilde{\mathcal{G}} \rangle$ is a group. Furthermore, the constructed set $\widetilde{\mathcal{G}}$ satisfies $\pi(\langle \widetilde{\mathcal{G}} \rangle_{grp}) = \mathbb{Z}^n$ under the canonical projection $\pi \colon \mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$.*

Proof. By Lemma 4.2.3, we can compute a finite metabelian presentation for the group $\langle \mathcal{G} \rangle_{grp}$. Since the generators $\mathcal{G}$ are explicitly given under this presentation, we can without loss of generality suppose $G = \langle \mathcal{G} \rangle_{grp}$. By Lemma 4.2.5, G can be effectively embedded as a subgroup of a quotient $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$, where H is a subgroup of $\mathbb{Z}^n \leq \mathcal{Y} \rtimes \mathbb{Z}^n$, and elements of H commute with all elements of $\mathcal{Y} \rtimes \mathbb{Z}^n$. We can hence suppose $G = \widetilde{G}$ is a subgroup of $(\mathcal{Y} \rtimes \mathbb{Z}^n)/H$ and

the generator set $\mathcal{G}$ is given as $\{g_1H, \ldots, g_kH\}$ where $g_1, \ldots, g_k \in \mathcal{Y} \rtimes \mathbb{Z}^n$. Let $h_1, \ldots, h_M$ be the generators of $H \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$ as a semigroup.

Let $\widetilde{\mathcal{G}} \coloneqq \{g_1, \ldots, g_k, h_1, \ldots, h_M\} \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$. By Lemma 4.2.5, the image of $\langle \mathcal{G} \rangle_{grp} = \widetilde{G}$ under the projection $\pi : (\mathcal{Y} \rtimes \mathbb{Z}^n)/H \to \mathbb{Z}^n/H$ is equal to $\mathbb{Z}^n/H$. Since $h_1, \ldots, h_M$ generate H as a semigroup, the group generated by $\widetilde{\mathcal{G}}$ admits image $(\mathbb{Z}^n/H) + H = \mathbb{Z}^n$ under the canonical projection $\mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$. Finally, by Lemma 4.2.6, the semigroup $\langle \mathcal{G} \rangle$ is a group if and only if the semigroup generated by $\widetilde{\mathcal{G}}$ is a group. This proves the proposition. □

4.3 $\mathcal{G}$-graphs

Thanks to Proposition 4.2.2, for the rest of this chapter we will focus on solving the Group Problem in $\mathcal{Y} \rtimes \mathbb{Z}^n$. We now fix the set of elements

$$\mathcal{G} \coloneqq \{(y_1, a_1), \ldots, (y_K, a_K)\} \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n.$$

In this section, we define the notion of $\mathcal{G}$-graphs.

Definition 4.3.1 ($\mathcal{G}$-graphs)**.** Suppose a finite set $\mathcal{G} \coloneqq \{(y_1, a_1), \ldots, (y_K, a_K)\} \subseteq \mathcal{Y} \rtimes \mathbb{Z}^n$ is given. A *$\mathcal{G}$-graph* is a directed multigraph Γ, whose set of vertices is a finite subset of $\mathbb{Z}^n$, each adjacent to at least one edge. The edges of Γ are each labeled with an index in $\{1, \ldots, K\}$. Furthermore, if an edge from vertex v to vertex w has label i, then $v = w + a_i$.

For a graph Γ, we denote by $V(\Gamma)$ its set of vertices and by $E(\Gamma)$ its set of edges. For a (directed) edge e, we denote by $s(e)$ its starting vertex and by $d(e)$ its destination vertex. A *loop* is an edge that starts and ends at the same vertex. In particular, if $a_i = (0, \ldots, 0)$, then edges of label i in a $\mathcal{G}$-graph are loops.

A *circuit* of a graph is a path that starts and ends at the same vertex. An *Euler path* of a graph G is a path that uses each edge exactly once. An *Euler circuit* is an Euler path that starts and ends at the same vertex. We call a graph *Eulerian* if it contains an Euler circuit. A directed graph is called *symmetric* if for each vertex, its out-degree equals its in-degree. It is a well known fact that a directed graph is Eulerian if and only if it is symmetric and connected. Note that in a symmetric directed graph, strong connectivity (reachable from s to t for every pair (s, t) of vertices) and weak connectivity (connected as an undirected graph) are equivalent, so connectivity can refer to any one of the two notions.

Given $z \in \mathbb{Z}^n$ and a $\mathcal{G}$-graph Γ, its *translation* $\Gamma + z$ is a graph obtained by moving everything in Γ by the vector z. See Figure 4.1 for an illustration.

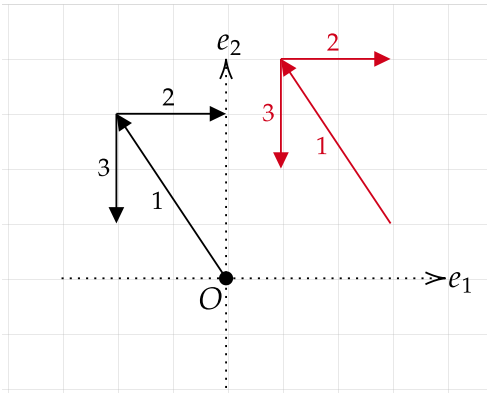

Figure 4.1: A $\mathcal{G}$-graph Γ (in black) and its translation $\Gamma + (3,1)$ (in red).

Definition 4.3.2 (Element represented by a $\mathcal{G}$-graph)**.** For an edge $e \in E(\Gamma)$, denote by $\ell(e)$ the label of e. We say Γ *represents* the following element of $\mathcal{Y} \rtimes \mathbb{Z}^n$:

$$\left(\sum_{e \in E(\Gamma)} \overline{X}^{s(e)} \cdot y_{\ell(e)}, \sum_{e \in E(\Gamma)} a_{\ell(e)} \right). \tag{4.6}$$

For a word w over the alphabet $\mathcal{G}$, we associate to it a unique $\mathcal{G}$-graph $\Gamma(w)$, defined as follows. Write $w = (y_{i_1}, a_{i_1})(y_{i_2}, a_{i_2}) \cdots (y_{i_p}, a_{i_p})$. For each $j = 1, \ldots, p$, we add an edge starting at the vertex $a_{i_1} + \cdots + a_{i_{j-1}}$, ending at the vertex $a_{i_1} + \cdots + a_{i_j}$, with the label i_j. (If $j = 1$ then the edge starts at 0 and ends at a_{i_1}.) The graph $\Gamma(w)$ is then obtained by taking the connected component of the vertex 0. See Figure 4.2 for an illustration.

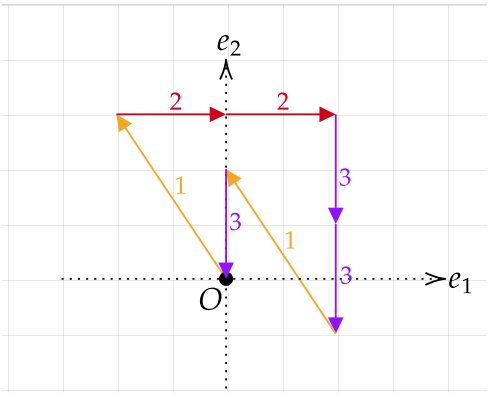

Figure 4.2: The graph $\Gamma(w)$ where $a_1 = (-2, 3)$, $a_2 = (2, 0)$, $a_3 = (0, -2)$ and $w = (y_1, a_1)(y_2, a_2)(y_2, a_2)(y_3, a_3)(y_3, a_3)(y_1, a_1)(y_3, a_3)$.

Fact 4.3.3. For a word w over the alphabet $\mathcal{G}$, the element of $\mathcal{Y} \rtimes \mathbb{Z}^n$ represented by its associated graph $\Gamma(w)$ is equal to the element of $\mathcal{Y} \rtimes \mathbb{Z}^n$ represented by the word w.

By reading the letters in w one by one and tracing the corresponding edges of $\Gamma(w)$, we obtain an Euler path of $\Gamma(w)$. Furthermore, if the word w represents the neutral element (or

any element of the form $(y, 0)$), then this Euler path is an Euler circuit. Conversely, given an Eulerian $\mathcal{G}$-graph Γ containing the vertex 0, we can follow an Euler circuit starting from 0 and read a word w such that $\Gamma(w) = \Gamma$.

A $\mathcal{G}$-graph Γ is called *full-image* if it contains an edge with label i for every $i \in \{1, \dots, K\}$. Note that for a word w over the alphabet $\mathcal{G}$, its associated graph $\Gamma(w)$ is full-image if and only if the word w is full-image. Combining Lemma 2.2.1 with the above correspondence between words and Eulerian graphs, we immediately obtain the following lemma.

Lemma 4.3.4. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if there exists a full-image Eulerian $\mathcal{G}$-graph that represents the neutral element.*

Proof. If the semigroup $\langle \mathcal{G} \rangle$ is a group, then by Lemma 2.2.1 there exists a full-image word w representing the neutral element. Then its associated $\mathcal{G}$-graph $\Gamma(w)$ is full-image, Eulerian and represents the neutral element.

If Γ is a full-image Eulerian $\mathcal{G}$-graph representing the neutral element, then let $z \in V(\Gamma)$ be any vertex of Γ. Consider the translation $\Gamma - z$: it represents the element $(\overline{X}^{-z} \cdot 0, 0) = (0, 0)$ and contains an Eulerian circuit starting from 0. We read from this Eulerian circuit a word w, then w represents the neutral element. Furthermore, w is full-image because $\Gamma - z$ is full-image. Therefore by Lemma 2.2.1 the semigroup $\langle \mathcal{G} \rangle$ is a group. □

4.4 The Group Problem

Recall that the Group Problem in metabelian groups reduces to the Group Problem in $\mathcal{Y} \rtimes \mathbb{Z}^n$ (Proposition 4.2.2). Therefore, our main decidability result on the Group Problem in metabelian groups (Theorem 4.1.2) boils down to proving the following technical theorem.

Theorem 4.4.1. *Let $\mathcal{Y}$ be a $\mathbb{Z}[X_1^{\pm}, \dots, X_n^{\pm}]$-module with a given finite presentation. Suppose we are given a finite subset $\mathcal{G}$ of the semidirect product $\mathcal{Y} \rtimes \mathbb{Z}^n$, such that the subgroup $\langle \mathcal{G} \rangle_{grp}$ of $\mathcal{Y} \rtimes \mathbb{Z}^n$ admits the image $\mathbb{Z}^n$ under the canonical projection $\mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$. Then it is decidable whether the semigroup $\langle \mathcal{G} \rangle$ is a group.*

In this section we summarize the proof of Theorem 4.4.1. In Subsection 4.4.1 we introduce the notion of "face-accessibility" of a $\mathcal{G}$-graph to replace the property of being Eulerian. In Subsection 4.4.2 we introduce "position polynomials" to reduce problems on $\mathcal{G}$-graphs to solving linear equations over polynomial semirings. In Subsection 4.4.3 we state a local-global principle for modules over polynomial semirings. In Subsection 4.4.4 we state our decidability result. Proofs of the stated theorems will be given in the later Sections 4.5,4.6 and 4.7.

4.4.1 From semigroups to face-accessible graphs

By Lemma 4.3.4, deciding the Group Problem boils down to deciding existence of an Eulerian graph with given properties. There is one downside of this reduction: being Eulerian is hard to characterize as a property of the graph. In fact, being Eulerian is equivalent to being symmetric and connected; while symmetry is easy to describe *locally* (i.e. at each vertex), connectivity of a graph is a *global* property and is hence hard to describe. The key idea of this subsection is to introduce a *local* property called "face-accessibility" to replace connectivity.

For a detailed reference on convex polytopes, see [3]. Let C be a bounded closed convex polytope in $\mathbb{R}^n$. That is, C is the convex hull of a finite number of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ in $\mathbb{R}^n$:

$$C = \{r_1\boldsymbol{x}_1 + \cdots + r_m\boldsymbol{x}_m \mid r_1, \ldots, r_m \in \mathbb{R}_{\geq 0},\ r_1 + \cdots + r_m = 1\}.$$

The *dimension* of a polytope is the dimension of the smallest linear space containing it. All polytopes considered in this chapter will be bounded closed convex polytopes.

A *face* of a polytope is any intersection of the polytope with a halfspace such that none of the interior points of the polytope lies on the boundary of the halfspace. A *strict face* of a polytope C is a face that is not the empty set or C itself. For example, if C is of dimension two, then the strict faces of C are its edges and its vertices. (While the non-strict faces of C also include the empty set and C itself.)

Let Γ be a $\mathcal{G}$-graph. Denote by C the convex hull of $V(\Gamma)$. A strict face F of C is called *accessible* if there exists an edge of Γ starting in F and ending in $C \setminus F$. That is, F is accessible if there is an edge $e \in E(\Gamma)$ starting from a vertex $s(e) \in F \cap V(\Gamma)$ and ending at a vertex $d(e) \in (C \setminus F) \cap V(\Gamma)$. See Figure 4.3 for an example. Note that Γ may contain loops as shown in Figure 4.3, even though they do not affect the accessibility of a face.

The graph Γ is called *face-accessible* if every strict face of C is accessible.

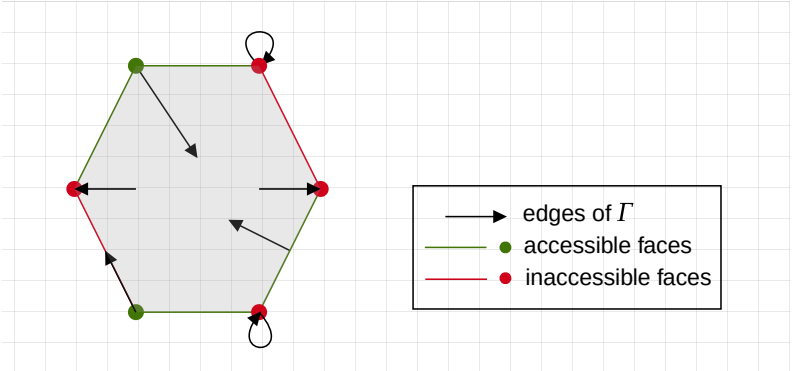

Figure 4.3: Accessible and inaccessible faces.

Observation 4.4.2. *An Eulerian graph is face-accessible.*

Indeed, if Γ is an Eulerian graph and F is a strict face of the convex hull of $V(\Gamma)$, then there must be a path of Γ starting in F and ending outside F. This path must contain an edge that starts in F and ends outside F, so F is accessible.

On the contrary, a symmetric face-accessible graph need not be Eulerian, as it might not be connected. Moreover, a symmetric graph need not be face-accessible. See Figures 4.4 and 4.5.

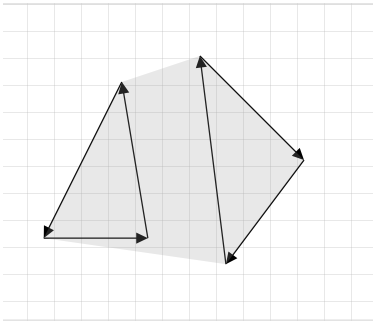

Figure 4.4: A symmetric and face-accessible graph that is not Eulerian (due to being disconnected).

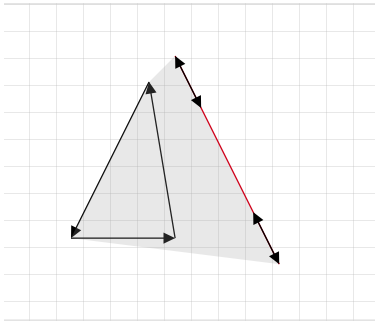

Figure 4.5: A symmetric graph that is not face-accessible (the red face is not accessible).

A graph Γ with vertices in $\mathbb{Z}^n$ is called $\mathbb{Z}^n$-*generating* if the set of vectors

$$\{d(e) - s(e) \mid e \in E(\Gamma)\}$$

generates $\mathbb{Z}^n$ as a semigroup. If Γ is symmetric, then Γ can be decomposed into a finite number of circuits. Then for any edge e from $d(e)$ to $s(e)$, Γ contains a path from $s(e)$ to $d(e)$. The sum of the edge vectors in this path ammounts to $s(e) - d(s)$, which is the inverse of $d(e) - s(e)$. Therefore, if Γ is symmetric, then the semigroup generated by $\{d(e)-s(e) \mid e \in E(\Gamma)\}$ is a group. Hence, a full-image symmetric $\mathcal{G}$-graph is $\mathbb{Z}^n$-generating if and only if $\{a_1, \ldots, a_K\}$ generates $\mathbb{Z}^n$ as a group. When $\mathcal{G}$ satisfies the assumption in Theorem 4.4.1, any full-image symmetric $\mathcal{G}$-graph is $\mathbb{Z}^n$-generating.

The main theorem of this subsection is the following, it shows that face-accessibility can characterize connectivity up to taking a finite union of translations.

Theorem 4.4.3. *Let Γ be a $\mathcal{G}$-graph that is symmetric, face-accessible and $\mathbb{Z}^n$-generating. Then there exist $z_1, \ldots, z_m \in \mathbb{Z}^n$, such that the union of translations $\widehat{\Gamma} \coloneqq \bigcup_{i=1}^{m} (\Gamma + z_i)$ is an Eulerian graph.*

See Figures 4.6 and 4.7 for an illustration of Theorem 4.4.3. The proof of Theorem 4.4.3 uses a combination of convex geometry and graph theory, and will be given in Section 4.5. Note

that the face-accessibility condition in Theorem 4.4.3 is necessary. The example in Figures 4.8 and 4.9 shows that Theorem 4.4.3 does not hold without face-accessibility.

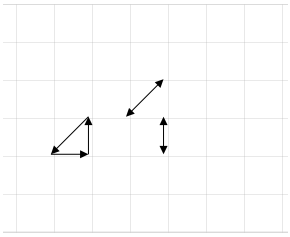

Figure 4.6: A symmetric face-accessible and $\mathbb{Z}^n$-generating graph Γ from Theorem 4.4.3.

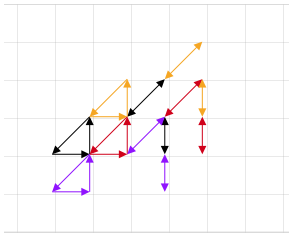

Figure 4.7: An Eulerian graph $\widehat{\Gamma}$ constructed in Theorem 4.4.3, consisting of four translations of Γ, each noted with a different colour.

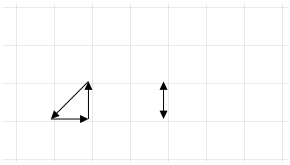

Figure 4.8: A Γ that is not face-accessible.

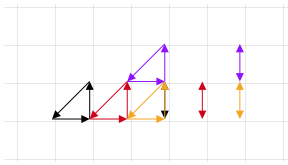

Figure 4.9: A union of translations of Γ is disconnected due to isolation of the right-most component.

If Γ represents the neutral element, then the union $\widehat{\Gamma} \coloneqq \bigcup_{i=1}^{m} (\Gamma + z_i)$ represents the element $(\sum_{i=1}^{m} \overline{X}^{z_i} \cdot 0, \sum_{i=1}^{m} 0) = (0, 0)$. Therefore, from Lemma 4.3.4 and Theorem 4.4.3 we immediately obtain the following.

Proposition 4.4.4. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if there exists a full-image symmetric face-accessible $\mathcal{G}$-graph that represents the neutral element.*

Proof. By Lemma 4.3.4 and Observation 4.4.2, it suffices to prove the "if" direction. Suppose Γ is a full-image symmetric face-accessible $\mathcal{G}$-graph representing the neutral element. By Theorem 4.4.3, there exist $z_1, \ldots, z_m \in \mathbb{Z}^n$, such that $\widehat{\Gamma} \coloneqq \bigcup_{i=1}^{m} (\Gamma + z_i)$ is an Eulerian graph. Since $\widehat{\Gamma}$ contains at least one translation of Γ, it is full-image. The graph $\widehat{\Gamma}$ represents the neutral element $(\sum_{i=1}^{m} \overline{X}^{z_i} \cdot 0, \sum_{i=1}^{m} 0) = (0, 0)$. Therefore, by Lemma 4.3.4, the semigroup $\langle \mathcal{G} \rangle$ is a group. □

4.4.2 From face-accessible graphs to positive polynomials

In this subsection we introduce "position polynomials" to describe a $\mathcal{G}$-graph, and reduce the graph theory problem in Proposition 4.4.4 to a computational problem over polynomial semirings. Recall that for an edge e in a $\mathcal{G}$-graph, $\ell(e)$ denotes the label of e, and $s(e) \in \mathbb{Z}^n$ denotes the starting vertex of e. Given a $\mathcal{G}$-graph Γ, define its tuple of *position polynomials* $\boldsymbol{f} = (f_1, \ldots, f_K)$

in the following way:

$$f_i \coloneqq \sum_{e \in E(\Gamma), \ell(e)=i} \overline{X}^{s(e)}, \quad i = 1, \ldots, K. \tag{4.7}$$

That is, f_i is the sum of monomials $\overline{X}^s$, where s ranges over the starting vertex of all label i edges in Γ. These polynomials have only non-negative coefficients, and are hence in $\mathbb{N}[\overline{X}^{\pm}]$. See Figure 4.10 for an example. Note that Γ is a multigraph, so the same edge can appear more than once, giving the polynomials f_i a coefficient larger than one.

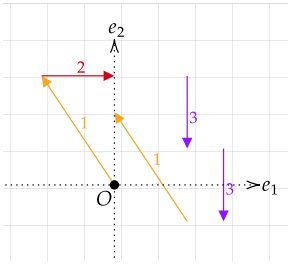

Figure 4.10: In this example, $f_1 = 1 + X_1^2 X_2^{-1}$, $f_2 = X_1^{-2} X_2^3$, $f_3 = X_1^2 X_2^3 + X_1^3 X_2$.

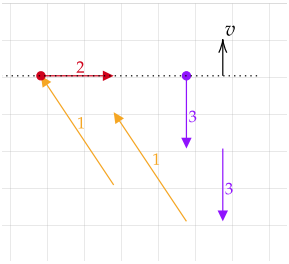

Figure 4.11: Let $v = (0,1)^\top$, we have $M_v(\{1,2,3\}, \boldsymbol{f}) = \{2,3\}$, $O_v = \{1,3\}$.

Conversely, given any tuple of polynomials $\boldsymbol{f} = (f_1, \ldots, f_K) \in \mathbb{N}[\overline{X}^{\pm}]^K$, one can construct a $\mathcal{G}$-graph Γ such that $\boldsymbol{f}$ is exactly its tuple of position polynomials. Indeed, for each monomial $c\overline{X}^b$ of $f_i, i = 1, \ldots, K$, we draw c edges of label i starting at vertex b. The resulting $\mathcal{G}$-graph will have position polynomials $(f_1, \ldots, f_K)$. Note that it is crucial for all f_i to be elements of $\mathbb{N}[\overline{X}^{\pm}]$ instead of $\mathbb{Z}[\overline{X}^{\pm}]$, so that $c \geq 0$ for all monomials $c\overline{X}^b$ of $f_i, i = 1, \ldots, K$.

For a vector $v \in (\mathbb{R}^n)^*$ and a set $I \subseteq \{1, \ldots, K\}$, define

$$M_v(I, \boldsymbol{f}) \coloneqq \left\{ i \in I \;\middle|\; \deg_v(f_i) = \max_{i' \in I} \{\deg_v(f_{i'})\} \right\}.$$

This is the set of indices $i \in I$ such that $\deg_v(f_i)$ is maximal among $i \in I$. Define

$$O_v \coloneqq \{i \in \{1, \ldots, K\} \mid a_i \not\perp v\}.$$

This is the set of indices $i \in \{1, \ldots, K\}$ such that a_i is not orthogonal to v. See Figure 4.11 for an example.

The following proposition shows that we can completely characterize the graph theoretic properties from Proposition 4.4.4 using position polynomials and the sets M_v, O_v. The key point is how we characterize face-accessibility. As a comparison, no good characterization of *connectivity* can be obtained from position polynomials.

Proposition 4.4.5. *Let Γ be a $\mathcal{G}$-graph with position polynomials $f_i \in \mathbb{N}[\overline{X}^{\pm}], i = 1, \ldots, K$.*

(i) Γ *is full-image if and only if* $f_i \neq 0$ *for all* $i = 1, \ldots, K$.

(ii) Γ *is symmetric if and only if* $\sum_{i=1}^{K} f_i \cdot (\overline{X}^{a_i} - 1) = 0$.

(iii) Γ *is face-accessible if and only if*

$$O_v \cap M_v(\{1, \ldots, K\}, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^* . \tag{4.8}$$

(iv) Suppose Γ *is symmetric, then it represents the neutral element if and only if* $\sum_{i=1}^{K} f_i \cdot y_i = 0$.

Proof. (i) Γ is full-image if and only if each label appears at least once, meaning $f_i \neq 0$ for all i.

(ii) We have

$$\sum_{i=1}^{K} f_i \cdot (\overline{X}^{a_i} - 1) = \sum_{i=1}^{K} \sum_{e \in E(\Gamma), \ell(e)=i} \overline{X}^{s(e)} (\overline{X}^{a_i} - 1) = \sum_{i=1}^{K} \sum_{e \in E(\Gamma), \ell(e)=i} (\overline{X}^{s(e)+a_i} - \overline{X}^{s(e)})$$
$$= \sum_{e \in E(\Gamma)} (\overline{X}^{d(e)} - \overline{X}^{s(e)}) = \sum_{e \in E(\Gamma)} \overline{X}^{d(e)} - \sum_{e \in E(\Gamma)} \overline{X}^{s(e)}.$$

This is equal to zero if and only if the in-degree equals the out-degree at every vertex.

(iii) Let C be the convex hull of $V(\Gamma)$. Suppose Γ is face-accessible. Take an arbitrary vector $v \in (\mathbb{R}^n)^*$. The set F of all points x in C such that $v^\top x$ is maximal forms a strict face of C. Since F is accessible, there is an edge $e \in E(\Gamma)$ such that $s(e) \in F$ and $d(e) \in C \setminus F$. By the definition of F, the value of $v^\top \cdot s(e)$ is maximal among all edges in $E(\Gamma)$. Since the monomial $\overline{X}^{s(e)}$ appears in $f_{\ell(e)}$, this means $\ell(e) \in M_v(\{1, \ldots, K\}, \boldsymbol{f})$. Since $d(e) \in C \setminus F$, we have $v^\top \cdot d(e) < v^\top \cdot s(e)$. Therefore $a_{\ell(e)} = d(e) - s(e)$ is not orthogonal to v, so $\ell(e) \in O_v$. Hence, we have $\ell(e) \in O_v \cap M_v(\{1, \ldots, K\}, \boldsymbol{f})$. This yields Property (4.8).

For the other implication, suppose Property (4.8) hold. Take any arbitrary strict face F of C. There is a vector $v \in (\mathbb{R}^n)^*$ such that F consists of all points $x \in C$ where $v^\top x$ is maximal. Let ℓ be any element in $O_v \cap M_v(\{1, \ldots, K\}, \boldsymbol{f})$. Then f_ℓ contains a monomial $\overline{X}^{s(e)}$ corresponding to some edge $e \in E(\Gamma)$, such that $\deg_v(\overline{X}^{s(e)})$ is maximal among all monomials of $f_1, \ldots, f_K$. In particular, $v^\top \cdot s(e)$ is maximal among all edges in $E(\Gamma)$, so $s(e) \in F$. The condition $\ell \in O_v$ shows that $a_\ell \not\perp v$, so $v^\top \cdot d(e) \neq v^\top \cdot s(e)$. Hence, the edge e starts in F and ends in $C \setminus F$. This makes the face F accessible. We conclude that Γ is face-accessible.

(iv) Suppose Γ is symmetric, then $\sum_{e \in E(\Gamma)} a_{\ell(e)} = 0$. By Equation (4.6), Γ represents the element

$$\left(\sum_{e \in E(\Gamma)} \overline{X}^{s(e)} \cdot y_{\ell(e)}, \sum_{e \in E(\Gamma)} a_{\ell(e)} \right) = \left(\sum_{i=1}^{K} \sum_{e \in E(\Gamma), \ell(e)=i} \overline{X}^{s(e)} \cdot y_i, 0 \right) = \left(\sum_{i=1}^{K} f_i \cdot y_i, 0 \right),$$

which is the neutral element if and only if $\sum_{i=1}^{K} f_i \cdot y_i = 0$. □

Let $\mathcal{M}_{\mathbb{Z}}$ be the $\mathbb{Z}[\overline{X}^{\pm}]$-module consisting of all $\boldsymbol{f} \in \mathbb{Z}[\overline{X}^{\pm}]^K$ satisfying $\sum_{i=1}^{K} f_i \cdot (\overline{X}^{a_i} - 1) = 0$ and $\sum_{i=1}^{K} f_i \cdot y_i = 0$:

$$\mathcal{M}_{\mathbb{Z}} \coloneqq \left\{ \boldsymbol{f} \in \mathbb{Z}[\overline{X}^{\pm}]^K \;\middle|\; \sum_{i=1}^{K} f_i \cdot (\overline{X}^{a_i} - 1) = 0 \text{ and } \sum_{i=1}^{K} f_i \cdot y_i = 0 \right\}. \tag{4.9}$$

Then Proposition 4.4.5 shows the following: there exists a full-image symmetric face-accessible $\mathcal{G}$-graph that represents the neutral element if and only if $\mathcal{M}_{\mathbb{Z}}$ contains an element $\boldsymbol{f} \in \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.8). Using linear algebra over $\mathbb{Z}[\overline{X}^{\pm}]$, generators of $\mathcal{M}_{\mathbb{Z}}$ can be effectively computed:

Lemma 4.4.6. *A finite set of generators $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathbb{Z}[\overline{X}^{\pm}]^K$ of the $\mathbb{Z}[\overline{X}^{\pm}]$-module $\mathcal{M}_{\mathbb{Z}}$ can be computed from $\{(y_1, a_1), \ldots, (y_K, a_K)\}$.*

Proof. Recall that $\mathcal{Y}$ is given as a quotient M/N where N and M are $\mathbb{Z}[\overline{X}^{\pm}]$-submodules of $\mathbb{Z}[\overline{X}^{\pm}]^d$ respectively generated by $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{L'}$ and $\boldsymbol{n}_1, \ldots, \boldsymbol{n}_L \in \mathbb{Z}[\overline{X}^{\pm}]^d$. For $i = 1, \ldots, K$, the element y_i is given as $y_i = \widetilde{\boldsymbol{y}}_i + N$ where $\widetilde{\boldsymbol{y}}_i \in M \subseteq \mathbb{Z}[\overline{X}^{\pm}]^d$.

The equation $\sum_{i=1}^{K} f_i \cdot y_i = 0$ can be written as $\sum_{i=1}^{K} f_i \cdot \widetilde{\boldsymbol{y}}_i \in N$, which is equivalent to

$$\sum_{i=1}^{K} f_i \cdot \widetilde{\boldsymbol{y}}_i = \sum_{j=1}^{L} h_j \cdot \boldsymbol{n}_j \text{ for some } h_1, \ldots, h_L \in \mathbb{Z}[\overline{X}^{\pm}]. \tag{4.10}$$

Let $\widetilde{\mathcal{M}}$ be the set of solutions $(f_1, \ldots, f_K, h_1, \ldots, h_L) \in \mathbb{Z}[\overline{X}^{\pm}]^{K+L}$ of the following system of homogeneous linear equations:

$$\begin{aligned} \sum_{i=1}^{K} f_i \cdot \widetilde{\boldsymbol{y}}_i - \sum_{j=1}^{L} h_j \cdot \boldsymbol{n}_j &= 0 \\ \sum_{i=1}^{K} f_i \cdot (\overline{X}^{a_i} - 1) &= 0. \end{aligned}$$

The set $\widetilde{\mathcal{M}}$ is also known as a syzygy module. It is a classic result from linear algebra over Noetherian rings that a finite set of generators $(\boldsymbol{f}_1, \boldsymbol{h}_1), \ldots, (\boldsymbol{f}_s, \boldsymbol{h}_s)$ for $\widetilde{\mathcal{M}}$ can be effectively computed (see [17, 88] or [37, Theorem 15.10]). Let

$$\begin{aligned} \pi \colon \mathbb{Z}[\overline{X}^{\pm}]^{K+L} &\to \mathbb{Z}[\overline{X}^{\pm}]^K, \\ (f_1, \ldots, f_K, h_1, \ldots, h_L) &\mapsto (f_1, \ldots, f_K) \end{aligned}$$

be the projection onto the first K coordinates. Then $\mathcal{M}_{\mathbb{Z}} = \pi(\mathcal{M})$, and a finite set of generators for $\mathcal{M}_{\mathbb{Z}}$ is simply $\{\pi((\boldsymbol{f}_1, \boldsymbol{h}_1)), \ldots, \pi((\boldsymbol{f}_s, \boldsymbol{h}_s))\} = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_s\}$. □

4.4.3 Local-global principle for positive polynomials

We now start constructing an algorithm that decides whether $\mathcal{M}_{\mathbb{Z}}$ contains an element $\boldsymbol{f} \in \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.8). This problem is highly non-trivial due to the polynomials having coefficients in $\mathbb{N}$ instead of $\mathbb{Z}$. In fact, solving systems of non-homogeneous linear equations over the semiring $\mathbb{N}[\overline{X}^{\pm}]$ is known to be *undecidable* [78]. Our key to obtaining a decidability result is to exploit the *homogeneity* of our linear equations.

The first step is to generalize Property (4.8). Given two sets $I, J \subseteq \{1, \ldots, K\}$, our new goal is to decide whether $\mathcal{M}_{\mathbb{Z}}$ contains an element $\boldsymbol{f} \in \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying the following condition.

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset, \quad \text{for every } v \in (\mathbb{R}^n)^* . \tag{4.11}$$

Note that Property (4.8) can be considered as a special case of (4.11) by taking $I = \{1, \ldots, K\}$, $J = \emptyset$. Considering the sets I and J as variables instead of fixing them as $\{1, \ldots, K\}$ and $\emptyset$ will be crucial to our subsequent results (Theorems 4.4.8 and 4.4.10). Intuitively, edges with labels in J can be considered to be "going out into an $(n+1)$-th dimension"; and edges with labels outside of I can be considered to "exist in an $(n+1)$-th dimension".

The second step is to pass from polynomial semirings over $\mathbb{Z}$ and $\mathbb{N}$ to polynomial semirings over $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ in order to facilitate subsequent usage of analytic methods. Recall from Lemma 4.4.6 that a finite set of generators $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathbb{Z}[\overline{X}^{\pm}]^K$ of the $\mathbb{Z}[\overline{X}^{\pm}]$-module $\mathcal{M}_{\mathbb{Z}}$ can be computed. Let $\mathcal{M}$ be the $\mathbb{R}[\overline{X}^{\pm}]$-submodule of $\mathbb{R}[\overline{X}^{\pm}]^K$ generated by $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$, that is,

$$\mathcal{M} \coloneqq \left\{ h_1 \cdot \boldsymbol{g}_1 + \cdots + h_m \cdot \boldsymbol{g}_m \;\middle|\; h_1, \ldots, h_m \in \mathbb{R}[\overline{X}^{\pm}] \right\}.$$

Lemma 4.4.7. *Fix two sets $I, J \subseteq \{1, \ldots, K\}$. There exists an element $\widetilde{\boldsymbol{f}} \in \mathcal{M}_{\mathbb{Z}} \cap \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property* (4.11)*, if and only if there exists an element $\boldsymbol{f} \in \mathcal{M} \cap \left(\mathbb{R}_{\geq 0}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property* (4.11)*.*

Proof. An element $\widetilde{\boldsymbol{f}} \in \mathcal{M}_{\mathbb{Z}} \cap \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.11) is obviously an element in $\mathcal{M} \cap \left(\mathbb{R}_{\geq 0}[\overline{X}^{\pm}]^*\right)^K$. Therefore it suffices to prove the "if" implication.

Suppose we have an element $\boldsymbol{f} \in \mathcal{M} \cap \left(\mathbb{R}_{\geq 0}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.11), we show that there is an element $\widetilde{\boldsymbol{f}} \in \mathcal{M}_{\mathbb{Z}} \cap \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.11).

Write $\boldsymbol{f} = (f_1, \ldots, f_K)$ where for $i = 1, \ldots, K$,

$$f_i = \sum_{b \in B_i} c_{i,b} \overline{X}^b.$$

Here, the *support* B_i is a non-empty finite subset of $\mathbb{Z}^n$, and $c_{i,b} \in \mathbb{R}_{>0}$ for all $b \in B_i$. Since Property (4.11) depends only on the supports $B_1, \ldots, B_K$, it suffices to show that there exists $\widetilde{\boldsymbol{f}} = (\widetilde{f}_1, \ldots, \widetilde{f}_K) \in \mathcal{M}_{\mathbb{Z}} \cap \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ where

$$\widetilde{f}_i = \sum_{b \in B_i} \widetilde{c}_{i,b} \overline{X}^b,$$

and $\widetilde{c}_{i,b} \in \mathbb{Z}_{>0}$ for all $b \in B_i$.

Since $\boldsymbol{f} \in \mathcal{M}$, we have $\boldsymbol{f} = \sum_{j=1}^m h_j \cdot \boldsymbol{g}_j$ for some $h_1, \ldots, h_m \in \mathbb{R}[\overline{X}^{\pm}]$. For each $j \in \{1, \ldots, m\}$, write $h_j = \sum_{b \in H_j} h_{j,b} \overline{X}^b$, where H_j is a finite subset of $\mathbb{Z}^n$. Then the equation $\boldsymbol{f} = \sum_{j=1}^m h_j \cdot \boldsymbol{g}_j$ can be rewritten as a finite system of linear equations over $\mathbb{R}$, where the left hand sides are 0 or the variables $c_{i,b}, b \in B_i, i = 1, \ldots, K$, and the right hand sides are $\mathbb{Z}$-linear combinations of the variables $h_{j,b}, j \in \{1, \ldots, m\}, b \in H_j$ (because the coefficients of $\boldsymbol{g}_j$ are integers for all j).

Note that this system of linear equations is homogeneous over the variables $c_{i,b}, h_{j,b}$, and the coefficients are all in $\mathbb{Z}$. Therefore, it has a solution $h_{j,b} \in \mathbb{R}, j \in \{1, \ldots, m\}, b \in H_j$ and $c_{i,b} \in \mathbb{R}_{>0}, b \in B_i, i = 1, \ldots, K$, if and only if it has a solution with $h_{j,b} \in \mathbb{Q}, c_{i,b} \in \mathbb{Q}_{>0}$ for all i, j, b. By multiplying all $h_{j,b}, c_{i,b}$ with their common denominator, we obtain a solution $\widetilde{h}_{j,b} \in \mathbb{Z}, \widetilde{c}_{i,b} \in \mathbb{Z}_{>0}$ for all i, j, b. Then, $\widetilde{f}_i \coloneqq \sum_{b \in B_i} \widetilde{c}_{i,b} \overline{X}^b, i = 1, \ldots, K$ and $\widetilde{h}_j = \sum_{b \in H_j} \widetilde{h}_{j,b} \overline{X}^b, j = 1, \ldots, m$, satisfy $\widetilde{\boldsymbol{f}} = \sum_{j=1}^m \widetilde{h}_j \cdot \boldsymbol{g}_j$. Hence, $\widetilde{\boldsymbol{f}} = (\widetilde{f}_1, \ldots, \widetilde{f}_K) \in \mathcal{M}_{\mathbb{Z}} \cap \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$. The element $\widetilde{\boldsymbol{f}}$ satisfies Property (4.11) since the condition depends only on the supports $B_1, \ldots, B_K$. □

Denote

$$\mathbb{A} \coloneqq \mathbb{R}[\overline{X}^{\pm}], \quad \mathbb{A}^+ \coloneqq \mathbb{R}_{\geq 0}[\overline{X}^{\pm}]^*.$$

Given $f \in \mathbb{A}$ and $v \in (\mathbb{R}^n)^*$, the *initial polynomial* of f at direction v is defined as the sum of all monomials in f having the maximal degree $\deg_v(\cdot)$:

$$\mathrm{in}_v(f) \coloneqq \sum\nolimits_{\deg_v(\overline{X}^b) = \deg_v(f)} c_b \overline{X}^b, \quad \text{where } f = \sum c_b \overline{X}^b.$$

For $\boldsymbol{f} = (f_1, \ldots, f_K) \in \mathbb{A}^K$, we naturally denote $\mathrm{in}_v(\boldsymbol{f}) \coloneqq (\mathrm{in}_v(f_1), \ldots, \mathrm{in}_v(f_K)) \in \mathbb{A}^K$. The key result of this subsection is the following local-global principle, which generalizes a deep result of Einsiedler, Mouat and Tuncel [36, Theorem 1.3].

Theorem 4.4.8. *Let $\mathcal{M}$ be a $\mathbb{A}$-submodule of $\mathbb{A}^K$ and I, J be two subsets of $\{1, \ldots, K\}$. There exists $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfying*

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^*, \tag{4.12}$$

if and only if the two following conditions are satisfied:

1. **(LocR):** *For every $r \in \mathbb{R}^n_{>0}$, there exists $\boldsymbol{f}_r \in \mathcal{M}$ such that $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$.*
2. **(LocInf):** *For every $v \in (\mathbb{R}^n)^*$, there exists $\boldsymbol{f}_v \in \mathcal{M}$, such that*
   (a) $\mathrm{in}_v(\boldsymbol{f}_v) \in (\mathbb{A}^+)^K$.
   (b) *Denote $I' \coloneqq M_v(I, \boldsymbol{f}_v), J' \coloneqq O_v \cup J$. We have*

$$\big(O_w \cup J'\big) \cap M_w(I', \mathrm{in}_v(\boldsymbol{f}_v)) \neq \emptyset \quad \text{for every } w \in (\mathbb{R}^n)^*. \tag{4.13}$$

The full proof of Theorem 4.4.8 is highly non-trivial and is given in Section 4.6. As a comparison, the original result of Einsiedler, Mouat and Tuncel does not include Property (4.12) or (4.13):

Theorem 4.4.9 (Einsiedler, Mouat, Tuncel [36, Theorem 1.3])**.** *Let $\mathcal{M}$ be a $\mathbb{A}$-submodule of $\mathbb{A}^K$. Then $\mathcal{M} \cap (\mathbb{A}^+)^K \neq \emptyset$ if and only if the two following conditions are satisfied:*

1. *For every $r \in \mathbb{R}^n_{>0}$, there exists $\boldsymbol{f}_r \in \mathcal{M}$ such that $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$.*
2. *For every $v \in (\mathbb{R}^n)^*$, there exists $\boldsymbol{f}_v \in \mathcal{M}$, such that $\mathrm{in}_v(\boldsymbol{f}_v) \in (\mathbb{A}^+)^K$.*

Intuitively, Theorem 4.4.9 states that the module $\mathcal{M}$ contains an element $\boldsymbol{f}$ satisfying the "global" positivity constraint $\boldsymbol{f} \in (\mathbb{A}^+)^K$ if and only if it contains elements $\boldsymbol{f}_r, \boldsymbol{f}_v$ that satisfy the "local" positivity constraints 1 and 2. Condition 1 can be seen as a local positivity constraint at the positive reals $r \in \mathbb{R}^n_{>0}$, while condition 2 can be seen as a local positivity constraint at infinity with direction $v \in (\mathbb{R}^n)^*$.

Recall that the constraint (4.12) in Theorem 4.4.8 is motivated by our characterization of face-accessibility and hence cannot be removed. Theorem 4.4.8 generalizes Theorem 4.4.9 by including Property (4.12) in the "global" part of the statement. Correspondingly, a similar Property (4.13) is added to the "local" part of the statement. Note that the sets I' and J' in the "local" Property (4.13) are different from the sets I and J in the "global" Property (4.12). This is the main reason why, as explained in the beginning of this subsection, we need to keep I and J as variables instead of fixing them as $\{1, \ldots, K\}$ and $\emptyset$. The complex interaction between the sets M_v, O_v and $\boldsymbol{f}$ constitutes the main difficulty in generalizing from Theorem 4.4.9 to Theorem 4.4.8.

4.4.4 Decidability

Theorem 4.4.8 is the key to an algorithm that finds $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfying Property (4.11). Indeed, we have:

Theorem 4.4.10. *Fix $n \in \mathbb{N}$. Suppose we are given as input a set of elements $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m \in \mathbb{A}^K$ with integer coefficients, as well as the vectors $a_1, \dots, a_K \in \mathbb{Z}^n$ and two subsets I, J of $\{1, \dots, K\}$. Denote by $\mathcal{M}$ be the $\mathbb{A}$-submodule of $\mathbb{A}^K$ generated by $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$. It is decidable whether there exists $\boldsymbol{f} \in \mathcal{M}$ satisfying $\boldsymbol{f} \in (\mathbb{A}^+)^K$ and*

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^* . \tag{4.14}$$

Here, if $n = 0$ then $\mathbb{A}$ is understood as $\mathbb{R}$, and Property (4.14) is considered trivially true.

The full proof of Theorem 4.4.10 will be given Section 4.7. The main idea is as follows. We use an induction on the number of variables n. Theorem 4.4.8 allows us to reduce the decision problem into verifying two conditions (LocR) and (LocInf). Deciding Condition (LocR) can be done using the first order theory of reals. For Condition (LocInf), the key idea is to show that it suffices to decide it for *countably* many v. For each v we can decide (LocInf) by an appropriate application of the induction hypothesis. We then run two parallel procedures. One procedure enumerates all elements in $\mathcal{M}_{\mathbb{Z}}$ and checks if any one of them is in $(\mathbb{A}^+)^K$ and satisfies Property (4.14); the other procedure enumerates the countably many v and checks if Condition (LocInf) is false. Theorem 4.4.8 guarantees that one of the two procedures must terminate.

Putting together Lemma 2.2.4, Propositions 4.4.4-4.4.5, Lemmas 4.4.6-4.4.7, and Theorem 4.4.10, we obtain our main technical result:

Theorem 4.4.1. *Let $\mathcal{Y}$ be a $\mathbb{Z}[X_1^{\pm}, \dots, X_n^{\pm}]$-module with a given finite presentation. Suppose we are given a finite subset $\mathcal{G}$ of the semidirect product $\mathcal{Y} \rtimes \mathbb{Z}^n$, such that the subgroup $\langle \mathcal{G} \rangle_{grp}$ of $\mathcal{Y} \rtimes \mathbb{Z}^n$ admits the image $\mathbb{Z}^n$ under the canonical projection $\mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$. Then it is decidable whether the semigroup $\langle \mathcal{G} \rangle$ is a group.*

Proof. Recall that elements in $\mathbb{N}[\overline{X}^{\pm}]^K$ have an one-to-one correspondence with $\mathcal{G}$-graphs (Subsection 4.4.2). Therefore, Proposition 4.4.4 and 4.4.5 show that it suffices to decide whether the module $\mathcal{M}_{\mathbb{Z}}$ (defined in (4.9)) contains an element $\boldsymbol{f} \in \left(\mathbb{N}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.8). We use Lemma 4.4.6 to compute a basis of $\mathcal{M}_{\mathbb{Z}}$. Lemma 4.4.7 then shows it suffices to decide whether there exists $\boldsymbol{f} \in \mathcal{M} \cap \left(\mathbb{R}_{\geq 0}[\overline{X}^{\pm}]^*\right)^K$ satisfying Property (4.8). But Property (4.8) is

simply Property (4.14) with $I = \{1, \ldots, K\}, J = \emptyset$. So Theorem 4.4.10 shows this is decidable. □

Our main result follows from Theorem 4.4.1.

Theorem 4.1.2. *The Group Problem (hence also the Identity Problem) is decidable in all finitely generated metabelian groups.*

Proof. Let G be a finitely generated metabelian group. By Lemma 2.2.4, decidability of the Group Problem in G subsumes decidability of the Identity Problem. Given a finite subset $\mathcal{G}$ in G, we use Proposition 4.2.2 to construct a a finitely presented $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module $\mathcal{Y}$ for some $n \in \mathbb{N}$, as well as a subset $\widetilde{\mathcal{G}}$ of the group $\mathcal{Y} \rtimes \mathbb{Z}^n$, such that $\langle \mathcal{G} \rangle$ is a group if and only if $\langle \widetilde{\mathcal{G}} \rangle$ is a group. Furthermore, the constructed set $\widetilde{\mathcal{G}}$ satisfies $\pi(\langle \widetilde{\mathcal{G}} \rangle_{grp}) = \mathbb{Z}^n$ under the canonical projection $\pi \colon \mathcal{Y} \rtimes \mathbb{Z}^n \to \mathbb{Z}^n$. Theorem 4.4.1 shows we can decide whether $\langle \widetilde{\mathcal{G}} \rangle$ is a group. Therefore, it is decidable whether $\langle \mathcal{G} \rangle$ is a group. □

4.5 From face-accessible graphs to to connected graphs

In this section we prove Theorem 4.4.3:

Theorem 4.4.3. *Let Γ be a $\mathcal{G}$-graph that is symmetric, face-accessible and $\mathbb{Z}^n$-generating. Then there exist $z_1, \ldots, z_m \in \mathbb{Z}^n$, such that the union of translations $\widehat{\Gamma} \coloneqq \bigcup_{i=1}^{m} (\Gamma + z_i)$ is an Eulerian graph.*

Let Γ be a $\mathcal{G}$-graph that is symmetric, face-accessible and $\mathbb{Z}^n$-generating. Recall that being Eulerian is equivalent to being symmetric and connected. Since a union of symmetric graphs is automatically symmetric, it suffices to consider the connectivity of $\widehat{\Gamma}$. Note that in a symmetric graph, there exists a path from vertex v to w if and only if there exists a path from w to v. Therefore it will suffice to show connectivity for the undirected version of the graph $\widehat{\Gamma}$.

Let x be an arbitrary point in $\mathbb{R}^n$. Given $c \in \mathbb{R}^n$ and $r \in \mathbb{R}_{>0}$, denote by $scale(x, c, r)$ the *scaling* of x with centre c by the ratio r. That is,

$$scale(x, c, r) \coloneqq c + r \cdot (x - c).$$

Let S be an arbitrary set in $\mathbb{R}^n$, define

$$scale(S, c, r) \coloneqq \{scale(x, c, r) \mid x \in S\}.$$

When the centre of a scale is the origin 0, we simplify the notation by defining

$$rS \coloneqq scale(S, 0, r)$$

Let C be the convex hull of $V(\Gamma)$. Since Γ is $\mathbb{Z}^n$-generating, the polytope C is of dimension n. For any $N \in \mathbb{N}$, let $NC \coloneqq scale(C, 0, N)$. Define

$$S_N \coloneqq \{z \in \mathbb{Z}^n \mid z + C \subset NC\}.$$

That is, S_N is the set of translation vectors z that make $C + z$ stay in NC. Consider the graph

$$\Gamma_N \coloneqq \sum_{z \in S_N} (\Gamma + z).$$

We have $V(\Gamma_N) \subseteq NC$. Intuitively, Γ_N is the union of translations of the graph Γ whose convex hull is contained in NC. See Figure 4.12 and 4.13 for an illustration. Our goal is to prove that for some large N, the graph Γ_N is connected.

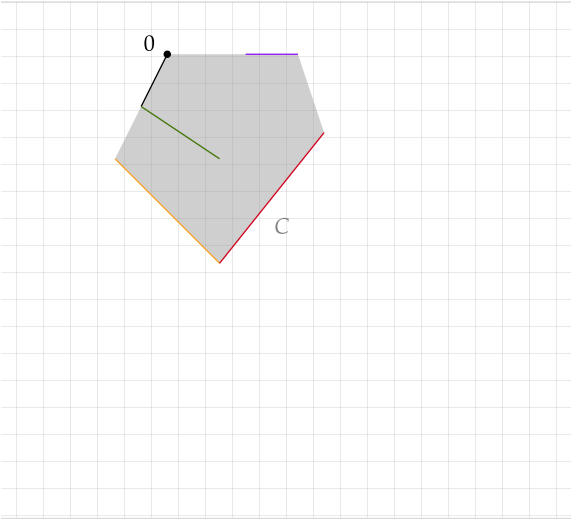

Figure 4.12: A graph Γ. Its convex hull C is covered in grey, and each edge is denoted with a different colour.

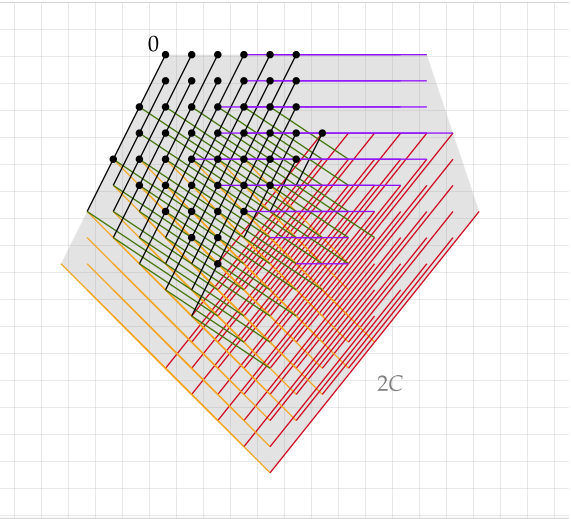

Figure 4.13: The graph Γ_N with $N = 2$, consisting of translations of Γ. The polytope $2C$ is covered in grey.

First we define the (infinite) graph $\Gamma_{\mathbb{Q}}$ as follows. The vertices of $\Gamma_{\mathbb{Q}}$ are $V(\Gamma_{\mathbb{Q}}) \coloneqq C \cap \mathbb{Q}^n$. The edges of $\Gamma_{\mathbb{Q}}$ are

$$E(\Gamma_{\mathbb{Q}}) \coloneqq \{scale(e, c, r) \mid e \in E(\Gamma), c \in C, r \in (0, 1) \cap \mathbb{Q}\}.$$

That is, $\Gamma_{\mathbb{Q}}$ is the union of all scaled versions of Γ that completely falls inside C. Intuitively, $\Gamma_{\mathbb{Q}}$ can be seen as the "limit" of $\frac{1}{N}\Gamma_N$ when N tends towards infinity. See Figure 4.14 for an

illustration.

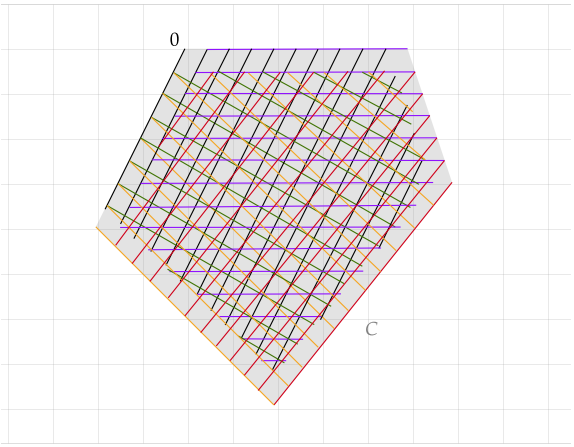

Figure 4.14: Illustration for $\Gamma_{\mathbb{Q}}$, where Γ is as in Figure 4.12.

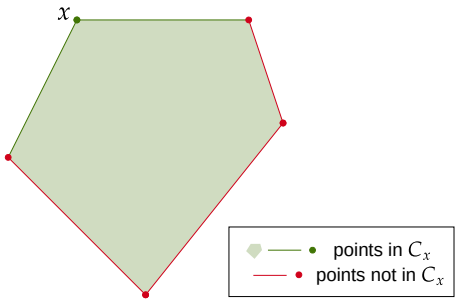

Figure 4.15: Illustration of the set C_x.

Define the *face lattice* $\mathrm{Lat}(C)$ to be the set of all faces of C. For $F \in \mathrm{Lat}(C)$, its *relative interior* $\mathrm{int}(F)$ is the set of points in F that are not contained in any sub-face of F:

$$\mathrm{int}(F) \coloneqq \{x \in F \mid x \notin F' \text{ for all faces } F' \subsetneq F\}.$$

The relative interiors of faces constitute a partition of C:

$$C = \bigcup_{F \in \mathrm{Lat}(C)} \mathrm{int}(F).$$

For any point $x \in C$, define F_x to be the face of C such that $x \in \mathrm{int}(F_x)$. This is the smallest face containing x. For any face F of C, we have $x \in F$ if and only if $F_x \subseteq F$. Define

$$C_x \coloneqq \bigcup_{F \in \mathrm{Lat}(C), F \supseteq F_x} \mathrm{int}(F).$$

That is, the set C_x is the union of the interior of all faces containing F_x. See Figure 4.15 for an illustration. This is also the union of the interior of all faces containing x.

Observation 4.5.1. *Let x, y be points in C. If $y \in C_x$ then $C_y \subseteq C_x$.*

Proof. If $y \in C_x$, then $\mathrm{int}(F_y) \subseteq C_x$, so $F_y \supseteq F_x$. Therefore,

$$C_y = \bigcup_{F \in \mathrm{Lat}(C), F \supseteq F_y} \mathrm{int}(F) \subseteq \bigcup_{F \in \mathrm{Lat}(C), F \supseteq F_x} \mathrm{int}(F) = C_x.$$

□

Lemma 4.5.2. *For any $c \in C_x$ and $r \in (0,1)$, we have $scale(C,c,r) \subset C_x$.*

Proof. Let $v \in C$, we show that $scale(v,c,r) \in C_x$. Since $c \in C_x = \bigcup_{F \in \mathrm{Lat}(C), F \supseteq F_x} \mathrm{int}(F)$, we have $\mathrm{int}(F_c) \subset C_x$. Therefore $F_c \supseteq F_x$ because F_c must appear in the index set $\{F \in \mathrm{Lat}(C), F \supseteq F_x\}$.

Denote by $seg(c,v)$ the closed segment that connects c and v, and denote $\mathrm{int}(seg(c,v)) \coloneqq seg(c,v) \setminus \{c,v\}$. Let F be the smallest face containing the $seg(c,v)$, then $F \supseteq F_c \supseteq F_x$. Hence, $scale(v,c,r) \in \mathrm{int}(seg(c,v)) \subset \mathrm{int}(F) \subset \bigcup_{F \in \mathrm{Lat}(C), F \supseteq F_x} \mathrm{int}(F) = C_x$. □

Since $C = \bigcup_{x \in C} C_x$, we have $E(\Gamma_{\mathbb{Q}}) = \bigcup_{x \in C} E_x(\Gamma_{\mathbb{Q}})$, where

$$E_x(\Gamma_{\mathbb{Q}}) \coloneqq \{scale(e,c,r) \mid e \in E(\Gamma), c \in C_x, r \in (0,1) \cap \mathbb{Q}\}.$$

Every edge in $E_x(\Gamma_{\mathbb{Q}})$ is contained in C_x by Lemma 4.5.2.

Lemma 4.5.3. *Every vertex x of $\Gamma_{\mathbb{Q}}$ is connected to a point in $\mathrm{int}(C)$ by a finite path P_x consisting of edges in $E_x(\Gamma_{\mathbb{Q}})$.*

Proof. See Figure 4.16 for an illustration of the proof. We first show that x is connected by an edge $e_x \in E_x(\Gamma_{\mathbb{Q}})$ to some $x' \in C_x$ where $F_{x'} \supsetneq F_x$.

Since Γ is face-accessible, there exists an edge $e \in E(\Gamma)$ connecting $w \in F_x$ and $w' \in C \setminus F_x$. Since $x \in \mathrm{int}(F_x)$ and $w \in F_x$, there exists $\varepsilon \in \mathbb{Q}_{>0}$ such that $c \coloneqq scale(x,w,1+\varepsilon) \in \mathrm{int}(F_x) \subseteq C_x$. Then $x = scale(w,c,\frac{\varepsilon}{1+\varepsilon})$, and $x' \coloneqq scale(w',c,\frac{\varepsilon}{1+\varepsilon}) \in C_x$ by Lemma 4.5.2. We also have $x' \notin F_x$ since $w' \notin F_x$ and $c \in F_x$. So $F_{x'} \supsetneq F_x$. Therefore, the edge $e_x \coloneqq scale(e,c,\frac{\varepsilon}{1+\varepsilon})$ is in $E_x(\Gamma_{\mathbb{Q}})$ and connects x and x'.

Note that $x' \in C_x$, so $C_{x'} \subseteq C_x$ by Observation 4.5.1, and thus $E_{x'}(\Gamma_{\mathbb{Q}}) \subseteq E_x(\Gamma_{\mathbb{Q}})$. Repeating this process for x', we can find a sequence of edges $e_x, e_{x'}, \ldots$, respectively in $E_x(\Gamma_{\mathbb{Q}}) \subseteq E_{x'}(\Gamma_{\mathbb{Q}}) \subseteq \cdots$, that gradually connects x to the interiors of faces of increasingly higher dimensions. Eventually x is connected to a point in $\mathrm{int}(C)$ by a path P_x. □

Let x be an arbitrary point in C. For each edge e' in P_x (the path defined in Lemma 4.5.3), write $e' = scale(e,c,r)$ where $e \in E(\Gamma), c \in C_x, r \in (0,1) \cap \mathbb{Q}$; define the polytope

$$C(e') \coloneqq scale(C,c,r). \tag{4.15}$$

Then $e' \subseteq C(e') \subseteq C_x$ by Lemma 4.5.2. Therefore, defining the finite union of polytopes

$$U_x \coloneqq \bigcup_{e' \in P_x} C(e'),$$

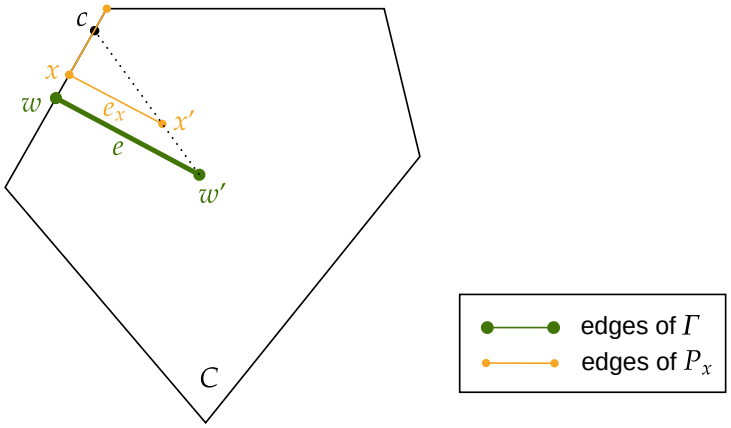

Figure 4.16: Illustration for Lemma 4.5.3.

we have $P_x \subseteq U_x \subseteq C_x$ and U_x is compact. See Figure 4.17 for an illustration.

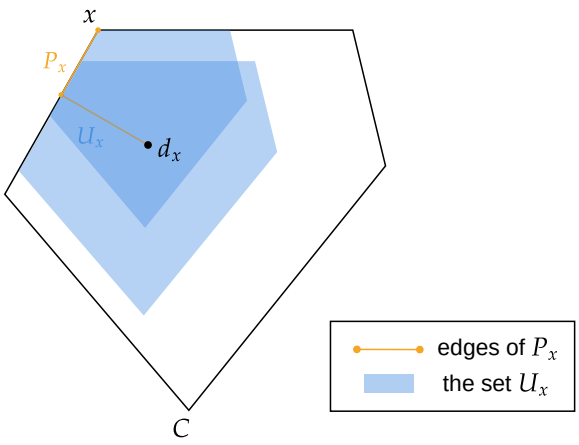

Figure 4.17: Illustration of U_x.

Consider the topology of C inherited from the Euclidean topology of $\mathbb{R}^n$ (that is, the open subsets of C are of the form $C \cap U$ where U is an open subset of $\mathbb{R}^n$). Then C is compact under this topology. Furthermore, U_x is still compact under this topology.

Fact 4.5.4. For each $x \in C$, the set C_x is an open subset of C.

Proof. It suffices to show that $C \setminus C_x$ is closed. Indeed, $C \setminus C_x = \bigcup_{F \in \mathrm{Lat}(C), x \notin F} \mathrm{int}(F)$. For any $F \in \mathrm{Lat}(C), x \notin F$, we have $x \notin F'$ for all faces $F' \subseteq F$. Therefore $\bigcup_{F \in \mathrm{Lat}(C), x \notin F} \mathrm{int}(F) = \bigcup_{F \in \mathrm{Lat}(C), x \notin F} \bigcup_{F' \in \mathrm{Lat}(C), F' \subseteq F} \mathrm{int}(F') = \bigcup_{F \in \mathrm{Lat}(C), x \notin F} F$. So $C \setminus C_x = \bigcup_{F \in \mathrm{Lat}(C), x \notin F} F$ is a finite union of (closed) faces, and is hence closed. □

Denote by $d_x := d(P_x) \in \mathrm{int}(C)$ the destination of P_x. Fix a point c_0 in the interior of C. Then d_x is contained in the interior of $scale(C, c_0, r_x)$ for some rational $0 < r_x < 1$.

Lemma 4.5.5. *Let x be any point in C.*

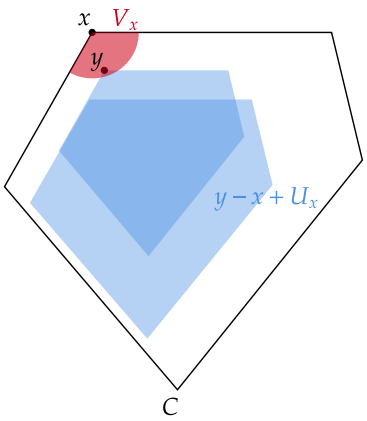

Figure 4.18: Illustration of V_x.

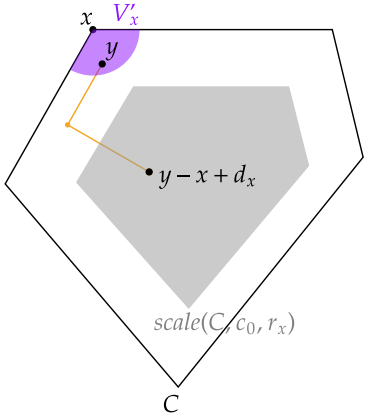

Figure 4.19: Illustration of V'_x.

(1) There exists an open neighbourhood $V_x \subset C$ of x, such that for all $y \in V_x$, we have $(y - x) + U_x \subseteq C_x$. See Figure 4.18 for illustration.

(2) There exists an open neighbourhood $V'_x \subset C$ of x, such that for all $y \in V'_x$, we have $(y - x) + d_x \subseteq scale(C, c_0, r_x)$. See Figure 4.19 for illustration.

Proof. For $y \in C, r \in \mathbb{R}_{>0}$, denote by $B(y, r)$ the open ball centered at y with radius r. Denote $B_C(y, r) \coloneqq C \cap B(y, r)$; it is an open subset of C.

(1) For each $y \in U_x \subset C_x$, let s_y be the supremum of real numbers s such that $B_C(y, s) \subset C_x$. We have $s_y > 0$ since C_x is an open set of C. The function $f : y \mapsto s_y$ is continuous on U_x since $|s_y - s_{y'}| \leq |y - y'|$. Since U_x is compact, f attains a minimum $s_{min} > 0$ on U_x. Then let $V_x \coloneqq B_C(x, s_{min})$, we have $(y - x) + U_x \subseteq C_x$ for all $y \in V_x$.

(2) Since $scale(C, c_0, r_x)$ is an n-dimensional polytope, its interior is an open set of C. Since d_x is in the interior of $scale(C, c_0, r_x)$, there exists $\rho_x > 0$ such that $B(d_x, \rho_x)$ is contained in the interior of $scale(C, c_0, r_x)$. We then simply take $V'_x \coloneqq B_C(x, \rho_x)$. □

For each $x \in C$, denote $W_x \coloneqq V_x \cap V'_x$. The open sets $W_x, x \in C$ cover the compact set C, so we can choose a finite number of representatives $x_1, \ldots, x_m$ such that $W_{x_1} \cup \cdots \cup W_{x_m} = C$. Let $R \coloneqq \max\{r_{x_1}, \ldots, r_{x_m}\}$.

Therefore, for each $y \in C$, there exists $i \in \{1, \ldots, m\}$ such that $y \in W_{x_i}$; the set $U_{x_i} + (y - x_i)$ is contained in C, and the path $P_{x_i} + (y - x_i)$ leads from y to the point $(y - x_i) + d_{x_i}$ in the interior of $scale(C, c_0, R)$.

For each edge $e' = scale(e, c, r), e \in E(\Gamma), c \in C_x, r \in (0, 1) \cap \mathbb{Q}$ in each of the paths $P_{x_1}, \ldots, P_{x_m}$, let $n(e')$ be a positive integer such that $n(e') \cdot r \in \mathbb{N}$ and $n(e') \cdot c \in \mathbb{Z}^n$. Furthermore,

let n_0 be a positive integer such that $n_0 \cdot c_0 \in \mathbb{Z}^n$. Define

$$N_0 \coloneqq n_0 \cdot \prod_{e' \in P_{x_1} \cup \cdots \cup P_{x_m}} n(e').$$

Then in particular, $N_0 x_i$ has only integer entries for all $i = 1, \ldots, m$; and $N_0 \cdot s(e')$ has only integer entries for all $e' \in P_{x_1} \cup \cdots \cup P_{x_1}$.

Lemma 4.5.6. *Let $N \geq 1$ be such that $N_0 \mid N$. Then every vertex in Γ_N is connected to some vertex in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$.*

Proof. See Figure 4.20 and 4.21 for an illustration of the proof.

Take any vertex y in Γ_N, we show it is connected to some vertex in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$. We have $y \in \mathrm{conv}(\Gamma_N) \subseteq NC$, so the point $y' \coloneqq \frac{y}{N} \in \frac{1}{N}C$ is contained in one of $W_{x_1}, \ldots, W_{x_m}$. Without loss of generality suppose $y' \in W_{x_1}$, so y' is connected in $\Gamma_{\mathbb{Q}}$ to the point $(y' - x_1) + d_{x_1} \in scale(C, c_0, R)$ by the path $(y' - x_1) + P_{x_1}$. We will show that $scale((y' - x_1) + P_{x_1}, 0, N)$ is a path in Γ_N. If this is the case, then it connects the vertex $Ny' = y$ to the vertex $N(y' - x_1 + d_{x_1}) \in scale(NC, Nc_0, R)$ and we are done.

In order to show that $scale((y' - x_1) + P_{x_1}, 0, N)$ is a path in Γ_N, it suffices to show that for each edge $e' \in P_{x_1}$, the segment $scale((y' - x_1) + e', 0, N)$ is a concatenation of edges in Γ_N. Again write $e' = scale(e, c, r), e \in E(\Gamma), c \in C_x, r \in (0,1) \cap \mathbb{Q}$. Consider the polytope $C(e') \coloneqq scale(C, c, r)$ as defined in (4.15). Since $y' \in W_{x_1}$, the translation $C'' \coloneqq (y' - x_1) + C(e') \subseteq (y' - x_1) + U_{x_1}$ is contained in C. Therefore, defining $\widetilde{C} \coloneqq NC''$, the polytope $\widetilde{C}$ is contained in NC. Also, the edge $e'' \coloneqq (y' - x_1) + e'$ is contained in C'', so $Ns(e'') \in \widetilde{C}$. Denote $\widetilde{s} \coloneqq Ns(e'') = y - Nx_1 + Ns(e') \in \widetilde{C} \cap \mathbb{Z}^n$.

Given convex polytopes A and A' and points $z, z' \in \mathbb{R}^n$. We denote by $(A, z) \sim (A', z')$, if

(i) either there exists $c \in \mathbb{R}^n$ and $r \in \mathbb{R}_{\geq 0}$, such that $z' = scale(z, c, r)$ and $A' = scale(A, c, r)$.
(ii) or there exists $t \in \mathbb{R}^n$, such that $z' = z + t$ and $A' = A + t$.

In other words, $(A, z) \sim (A', z')$ if the pair (A', z') can be obtained from (A, z) by simultaneous scaling or translation. Obviously, the relation $\sim$ is transitive, meaning $(A, z) \sim (A', z'), (A', z') \sim (A'', z'') \implies (A, z) \sim (A'', z'')$. We have

$$(\widetilde{C}, \widetilde{s}) = (NC'', Ns(e'')) \sim (C'', s(e'')) \sim (C(e'), s(e')) \sim (C, s(e)).$$

We claim that

$$\widetilde{s} - s(e) + C = scale(\widetilde{C}, \widetilde{s}, \frac{1}{rN}). \tag{4.16}$$

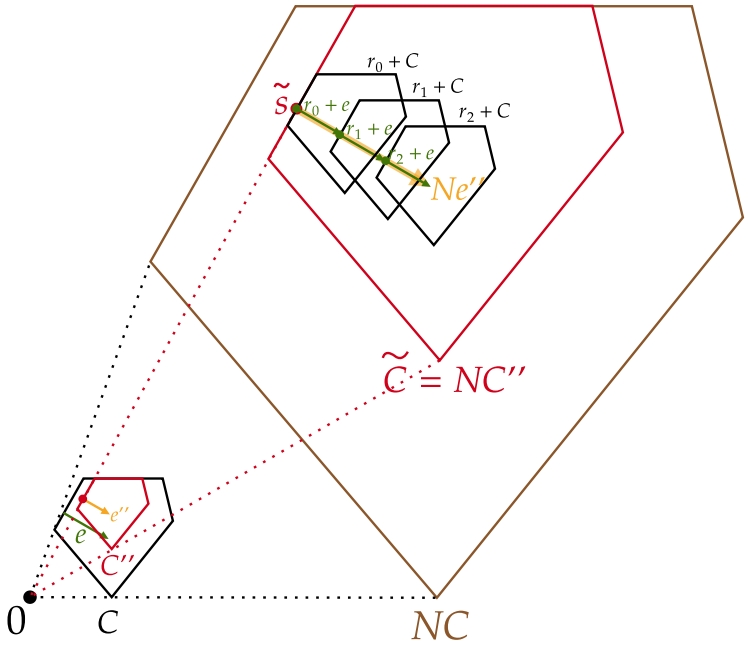

Figure 4.20: Illustration 1 of Lemma 4.5.6.

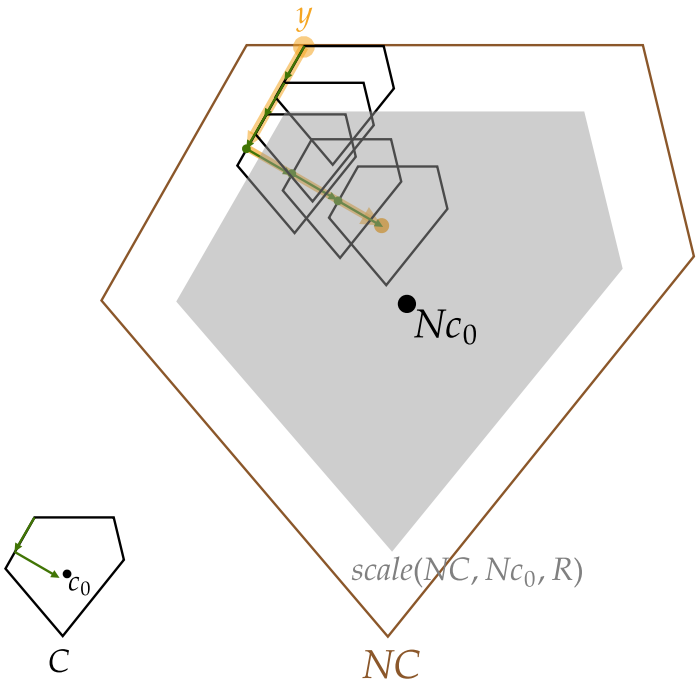

Figure 4.21: Illustration 2 of Lemma 4.5.6.

Indeed,

$$\left(scale(\widetilde{C}, \widetilde{s}, \frac{1}{rN}), \widetilde{s}\right) \sim (\widetilde{C}, \widetilde{s}) \sim (C, s(e)).$$

Since $scale(\widetilde{C}, \widetilde{s}, \frac{1}{rN})$ is a translation of $\frac{1}{rN}\widetilde{C} = C$, the point $s(e)$ becomes $\widetilde{s}$ under the same translation. Thus, we have $\widetilde{s} - s(e) + C = scale(\widetilde{C}, \widetilde{s}, \frac{1}{rN})$.

For every $k = 0, 1, \ldots, rN - 1$, define

$$r_k \coloneqq k(d(e) - s(e)) + \widetilde{s} - s(e) \in \mathbb{Z}^n.$$

Consider the polytopes $r_0 + C, \ldots, r_{rN-1} + C$. For every $k = 0, 1, \ldots, rN - 1$, we have

$$\begin{aligned} r_k + C = k(d(e) - s(e)) + \widetilde{s} - s(e) + C &\overset{(4.16)}{=} k(d(e) - s(e)) + scale(\widetilde{C}, \widetilde{s}, \frac{1}{rN}) \\ &= scale(\widetilde{C}, \widetilde{s} + \frac{krN}{rN-1}(d(e) - s(e)), \frac{1}{rN}). \quad (4.17) \end{aligned}$$

The centre of the scaling in the right hand side of (4.17) satisfies

$$\widetilde{s} + \frac{krN}{rN-1}(d(e) - s(e)) \in \widetilde{s} + rN(e - s(e)) = \widetilde{s} + N(e'' - s(e'')) = Ne'' \subseteq NC'' = \widetilde{C}.$$

Therefore

$$scale(\widetilde{C}, \widetilde{s} + \frac{krN}{rN-1}(d(e) - s(e)), \frac{1}{rN}) \subseteq \widetilde{C},$$

since $rN > 1$. Combined with (4.17), this yields

$$r_k + C \subseteq \widetilde{C} = NC'' \subseteq NC.$$

Thus, $r_k \in S_N$, and $r_k + e$ is an edge of Γ_N for $k = 0, 1, \ldots rN - 1$. Concatenating these rN edges, we obtain the segment from the point $\widetilde{s} = Ns(e'')$ to the point

$$\widetilde{s} + rN(d(e) - s(e)) = \widetilde{s} + N(d(e'') - s(e'')) = N(s(e'') + d(e'') - s(e'')) = Nd(e'').$$

This is exactly the segment $Ne'' = scale((y' - x_1) + e', 0, N)$. We have thus shown that $scale((y' - x_1) + e', 0, N)$ is a concatenation of edges in Γ_N, concluding our proof. □

For $k = 1, \ldots, n$, let $e_k = (0, \ldots, 1, \ldots, 0)$ denote the k-th element in the canonical basis of $\mathbb{Z}^n$. Note that Γ is $\mathbb{Z}^n$-generating. Therefore for each $k = 1, \ldots, n$, there exists a concatenation Q_k of translations of edges in $E(\Gamma)$, such that Q_k connects from 0 to e_k. In other words, one can find a sequence $i_1, \ldots, i_m \in \{1, \ldots, K\}$, such that $a_{i_1} + \cdots + a_{i_m} = e_k$; and Q_k is the concatenation of edges with label $i_1, \ldots, i_m$. *A priori*, Q_k is not a path in Γ_N, since its translations of edges might not appear in Γ_N. Let M_k be the total length of the edges appearing in Q_k; that is, $M_k \coloneqq \|a_{i_1}\| + \cdots + \|a_{i_m}\|$ where $\|\cdot\|$ denotes the Euclidean norm. Let $M \coloneqq \max_{1 \le k \le n} M_k$.

Denote by ∂C the boundary of C; that is, ∂C is the union of strict faces of C. For two sets $S, T \in \mathbb{R}^n$, define their *distance* to be $dist(S, T) \coloneqq \inf\{\|t - s\| \mid s \in S, t \in T\}$. The *diameter* of

C is defined as $diam(C) \coloneqq \sup\{\|t-s\| \mid s,t \in C\}$. Let $N_1 \in \mathbb{N}$ be such that

$$N_1 \cdot dist(scale(C, c_0, R), \partial C) > M + \sqrt{n} + diam(C). \tag{4.18}$$

Such an N_1 exists because $scale(C, c_0, R)$ and ∂C are disjoint compact sets, so their distance is larger than zero.

Lemma 4.5.7. *Let $N \in \mathbb{N}$ be such that $N > N_1$. Then every two vertices in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$ are connected in the graph Γ_N.*

Proof. See Figure 4.22 for an illustration of the proof.

Let v_1, v_2 be two arbitrary vertices in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$. There exists a path $P_{\mathbb{Z}^n}(v_1, v_2)$ in the grid $\mathbb{Z}^n$ from v_1 to v_2, such that each point in $P_{\mathbb{Z}^n}(v_1, v_2)$ is at most of distance $\sqrt{n}$ from the segment $seg(v_1, v_2)$. The path $P_{\mathbb{Z}^n}(v_1, v_2)$ consists of translations of the segments $seg(0, e_k), k = 1, \ldots, n$. For $k = 1, \ldots, n$, replacing each segment $seg(0, e_k) + z$ in $P_{\mathbb{Z}^n}(v_1, v_2)$ by the translation $Q_k + z$ of the path Q_k, we obtain a path $P_\Gamma(v_1, v_2)$. We now show that each edge of $P_\Gamma(v_1, v_2)$ is in $E(\Gamma_N)$.

Each point in $P_\Gamma(v_1, v_2)$ is at most of distance $\sqrt{n} + M$ from the segment $seg(v_1, v_2) \subset scale(NC, Nc_0, R)$, so it is at most of distance $\sqrt{n} + M$ from $scale(NC, Nc_0, R)$. By the definition (4.18) of N_1, we have

$$dist\left(scale(NC, Nc_0, R), \partial\left(NC\right)\right) = N \cdot dist(scale(C, c_0, R), \partial C) > M + \sqrt{n} + diam(C).$$

Therefore each point in $P_\Gamma(v_1, v_2)$ is at least of distance $diam(C)$ from the boundary $\partial\left(NC\right)$. Take an arbitrary edge e in $P_\Gamma(v_1, v_2)$, it comes from some translation $\Gamma + z$ of the graph Γ. Therefore e is contained in $C + z$. Since e is of distance at least $diam(C)$ from $\partial\left(NC\right)$, the polytope $C + z$ must be contained in NC. Hence $z \in S_N$ and so e is an edge of Γ_N. We have thus shown that each edge of $P_\Gamma(v_1, v_2)$ is in $E(\Gamma_N)$. Therefore every two vertices in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$ are connected in Γ_N. □

Theorem 4.4.3. *Let Γ be a $\mathcal{G}$-graph that is symmetric, face-accessible and $\mathbb{Z}^n$-generating. Then there exist $z_1, \ldots, z_m \in \mathbb{Z}^n$, such that the union of translations $\widehat{\Gamma} \coloneqq \bigcup_{i=1}^{m} (\Gamma + z_i)$ is an Eulerian graph.*

Proof. See Figure 4.23 for an illustration of the proof. Let $N \in \mathbb{N}$ be such that $N_0 \mid N$ and $N > N_1$. We show that the graph Γ_N is connected. Take any two vertices v, w of the graph Γ_N. Since $N_0 \mid N$, Lemma 4.5.6 shows that v and w are respectively connected in Γ_N to two vertices

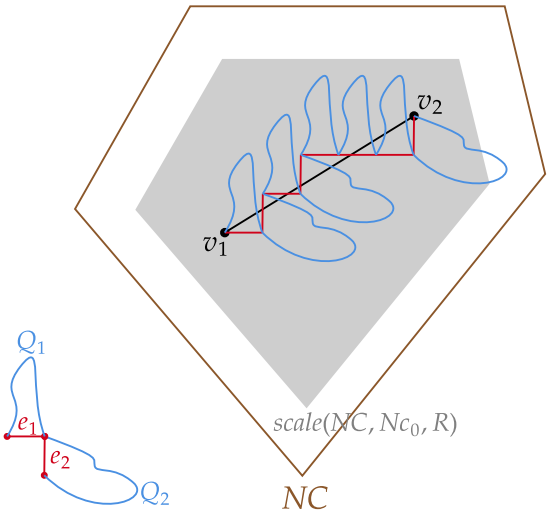

Figure 4.22: Illustration of Lemma 4.5.7.

v_1 and v_2 in $scale(NC, Nc_0, R) \cap V(\Gamma_N)$. Since $N > N_1$, Lemma 4.5.7 shows that v_1 and v_2 are connected in Γ_N. Therefore, v and w are connected in Γ_N. Since a union of symmetric graphs is also symmetric, Γ_N is symmetric and connected, hence Eulerian. □

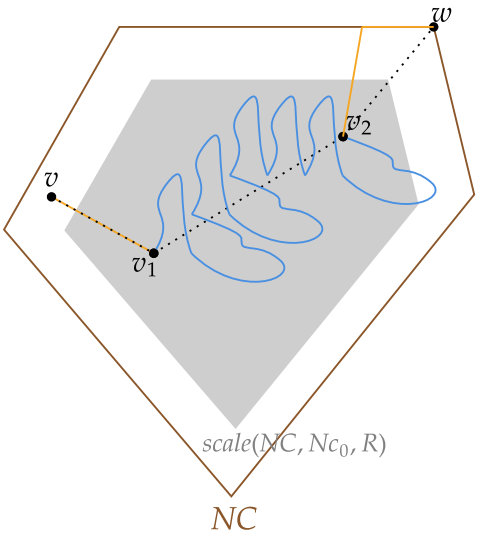

Figure 4.23: Illustration of proof of Theorem 4.4.3.

4.6 A local-global principle

In this section we prove Theorem 4.4.8:

Theorem 4.4.8. *Let $\mathcal{M}$ be a $\mathbb{A}$-submodule of $\mathbb{A}^K$ and I, J be two subsets of $\{1, \ldots, K\}$. There*

exists $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ *satisfying*

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^*, \tag{4.12}$$

if and only if the two following conditions are satisfied:

1. **(LocR):** *For every* $r \in \mathbb{R}^n_{>0}$*, there exists* $\boldsymbol{f}_r \in \mathcal{M}$ *such that* $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$.
2. **(LocInf):** *For every* $v \in (\mathbb{R}^n)^*$*, there exists* $\boldsymbol{f}_v \in \mathcal{M}$*, such that*
   (a) $\mathrm{in}_v(\boldsymbol{f}_v) \in (\mathbb{A}^+)^K$.
   (b) *Denote* $I' \coloneqq M_v(I, \boldsymbol{f}_v), J' \coloneqq O_v \cup J$*. We have*

$$\left(O_w \cup J'\right) \cap M_w(I', \mathrm{in}_v(\boldsymbol{f}_v)) \neq \emptyset \quad \text{for every } w \in (\mathbb{R}^n)^*. \tag{4.13}$$

Theorem 4.4.8 can be considered as a generalization of Einsiedler, Mouat and Tuncel's local-global principle (Theorem 4.4.9), where the additional Property (4.12) (as well as (4.13)) is included. Many components of the original proof [36] of Theorem 4.4.9 fail when trying to integrate Property (4.12), notably [36, Lemma 3.2] and [36, Lemma 5.2]. In order to take into account this extra property, we need to introduce new arguments to rework several parts of the original proof. The key new component will be the following Lemma 4.6.1, which shows a certain "continuity" of Property (4.12) when changing the direction v by a small amount.

Before stating Lemma 4.6.1, we will make a few observations. Define the quotient

$$D_n \coloneqq (\mathbb{R}^n)^* / \mathbb{R}_{>0}.$$

That is, elements of D_n are of the form $v\mathbb{R}_{>0}, v \in (\mathbb{R}^n)^*$, where $v\mathbb{R}_{>0} = v'\mathbb{R}_{>0}$ if and only if $v = r \cdot v'$ for some $r \in \mathbb{R}_{>0}$. The quotient D_n can be identified with the unit sphere of dimension n since every $v\mathbb{R}_{>0}$ is equal to exactly one $v'\mathbb{R}_{>0}$ with $\|v'\| = 1$. We equip D_n with the standard topology of the unit sphere. Note that $\mathrm{in}_v(\cdot), M_v(\cdot)$ and O_v are invariant when scaling v by any positive real number.

Lemma 4.6.1. *Fix* $v \in (\mathbb{R}^n)^*$*, a set* $I \subseteq \{1, \ldots, K\}$ *and* $\boldsymbol{f} \in \mathbb{A}^K$*. There exists an open neighbourhood* $U \subseteq D_n$ *of* $v\mathbb{R}_{>0}$*, such that for every* $w \in (\mathbb{R}^n)^*$ *with* $(v + w)\mathbb{R}_{>0} \in U$*, we have*

$$\mathrm{in}_{v+w}(\boldsymbol{f}) = \mathrm{in}_w(\mathrm{in}_v(\boldsymbol{f})), \quad M_{v+w}(I, \boldsymbol{f}) = M_w(M_v(I, \boldsymbol{f}), \mathrm{in}_v(\boldsymbol{f})), \quad \text{and} \quad O_{v+w} = O_v \cup O_w. \tag{4.19}$$

Proof. See Figures 4.24 and 4.25 for an illustration. For each $i = 1, \ldots, K$ such that $f_i \neq 0$,

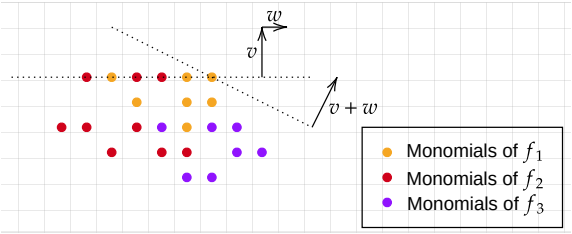

Figure 4.24: Illustration of Lemma 4.6.1. Here, $M_v(\{1,2,3\}, \boldsymbol{f}) = \{1,2\}$.

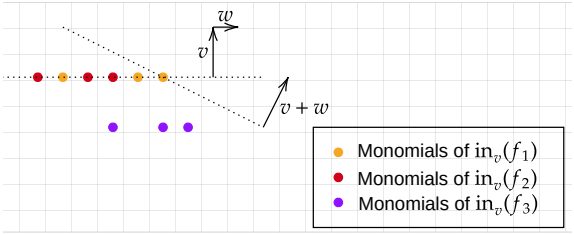

Figure 4.25: Continuation of the example in Figure 4.24. Here, $M_{v+w}(\{1,2,3\}, \boldsymbol{f}) = \{1\} = M_w(\{1,2\}, \mathrm{in}_v(\boldsymbol{f}))$.

write $f_i = g_i + h_i$ where $g_i \coloneqq \mathrm{in}_v(f_i)$ and $\deg_v(h_i) < \deg_v(g_i)$.

Since $\deg_v(h_i)$ and $\deg_v(g_i)$ vary continuously when v varies in $(\mathbb{R}^n)^*$, there exists an open neighbourhood $U_i \subseteq D_n$ of $v\mathbb{R}_{>0}$ such that $\deg_{v'}(h_i) < \deg_{v'}(g_i)$ for every $v'\mathbb{R}_{>0} \in U_i$. Therefore, for $(v+w)\mathbb{R}_{>0} \in U_i$ we have

$$\mathrm{in}_{v+w}(f_i) = \mathrm{in}_{v+w}(g_i) = \mathrm{in}_w(g_i) = \mathrm{in}_w(\mathrm{in}_v(f_i)),$$

where $\mathrm{in}_{v+w}(g_i) = \mathrm{in}_w(g_i)$ can be justified as follows. For every monomials $c\overline{X}^b$ appearing in g_i, we have $\deg_v(c\overline{X}^b) = \deg_v g_i$ since all monomials in $g_i = \mathrm{in}_v(f_i)$ have the same $\deg_v$. Hence, we have $\deg_{v+w}(c\overline{X}^b) = v^\top b + w^\top b = \deg_v g_i + w^\top b = \deg_v g_i + \deg_w(c\overline{X}^b)$; therefore the monomials in g_i with maximal $\deg_{v+w}$ are exactly those with maximal $\deg_w$. This yields $\mathrm{in}_{v+w}(g_i) = \mathrm{in}_w(g_i)$.

Note that for each $i \in M_v(I, \boldsymbol{f})$ and $i' \in I \setminus M_v(I, \boldsymbol{f})$, we have $\deg_v(f_i) > \deg_v(f_{i'})$. Again by the continuity of $\deg_v$ with respect to v, there exists an open neighbourhood U' of $v\mathbb{R}_{>0}$ such that for all $v'\mathbb{R}_{>0} \in U'$, we have $\deg_{v'}(f_i) > \deg_{v'}(f_{i'})$ for all $i \in M_v(I, \boldsymbol{f}), i' \in I \setminus M_v(I, \boldsymbol{f})$. Then for every $(v+w)\mathbb{R}_{>0} \in \cap_{i \in I, f_i \neq 0} U_i \cap U'$, we have

$$\begin{aligned}
&M_{v+w}(I, \boldsymbol{f}) \\
&= \left\{ i \in I \;\middle|\; \deg_{v+w}(f_i) = \max_{i' \in I} \{\deg_{v+w}(f_{i'})\} \right\} \\
&= \left\{ i \in M_v(I, \boldsymbol{f}) \;\middle|\; \deg_{v+w}(f_i) = \max_{i' \in M_v(I, \boldsymbol{f})} \{\deg_{v+w}(f_{i'})\} \right\} \qquad (\text{since } (v+w)\mathbb{R}_{>0} \in U') \\
&= \left\{ i \in M_v(I, \boldsymbol{f}) \;\middle|\; \deg_w(\mathrm{in}_v(f_i)) = \max_{i' \in M_v(I, \boldsymbol{f})} \{\deg_w(\mathrm{in}_v(f_{i'}))\} \right\} \qquad (\text{since } (v+w)\mathbb{R}_{>0} \in U_i, i \in I) \\
&= M_w(M_v(I, \boldsymbol{f}), \mathrm{in}_v(\boldsymbol{f})).
\end{aligned}$$

Finally, take any $i \in \{1, \ldots, K\}$. If $a_i \not\perp v$ then there exists an open neighbourhood $U_i'' \subseteq D_n$ of $v\mathbb{R}_{>0}$ such that for every $v'\mathbb{R}_{>0} \in U_i''$ we have $a_i \not\perp v'$. If $a_i \perp v$ then for every $w \in (\mathbb{R}^n)^*$ we have $a_i \perp (v+w) \iff a_i \perp w$. Take $U'' \coloneqq \cup_{i \in \{1,\ldots,K\}, a_i \not\perp v} U_i''$. For all $(v+w)\mathbb{R}_{>0} \in U''$, we

have

$$\begin{aligned}
O_{v+w} &= \{i \in \{1,\dots,K\} \mid a_i \not\perp (v+w)\} \\
&= \{i \in \{1,\dots,K\} \mid (a_i \not\perp v \text{ and } a_i \not\perp (v+w)) \text{ or } (a_i \perp v \text{ and } a_i \not\perp (v+w))\} \\
&= \{i \in \{1,\dots,K\} \mid (a_i \not\perp v) \text{ or } (a_i \perp v \text{ and } a_i \not\perp (v+w))\} \quad (\text{since } (v+w)\mathbb{R}_{>0} \in U_i'') \\
&= \{i \in \{1,\dots,K\} \mid (a_i \not\perp v) \text{ or } (a_i \perp v \text{ and } a_i \not\perp w)\} \\
&= \{i \in \{1,\dots,K\} \mid (a_i \not\perp v) \text{ or } (a_i \not\perp w)\} \\
&= O_v \cup O_w.
\end{aligned}$$

We conclude the proof by taking $U = \cap_{i\in I, f_i\neq 0} U_i \cap U' \cap U''$. □

As an illustration of how to integrate Property (4.12) into the local-global principle using Lemma 4.6.1, we first give a proof of the "only if" part of Theorem 4.4.8. As a comparison, the "only if" part of the original Theorem 4.4.9 is immediate.

Proof of "only if" part of Theorem 4.4.8. Suppose $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfies Property (4.12). To show (LocR), simply take $\boldsymbol{f}_r \coloneqq \boldsymbol{f}$ for all $r \in \mathbb{R}^n_{>0}$, then $\boldsymbol{f}(r) \in \mathbb{R}^K_{>0}$. As for (LocInf), for every $v \in (\mathbb{R}^n)^*$ we show that $\boldsymbol{f}_v \coloneqq \boldsymbol{f}$ satisfies Properties (LocInf)(a) and (b). Property (LocInf)(a) is satisfied because every monomial in $\boldsymbol{f} \in (\mathbb{A}^+)^K$ has positive coefficient. We now show Property (LocInf)(b).

When $w \in v\mathbb{R}_{>0}$, we have $O_w \cup J' = O_w \cup O_v \cup J = O_v \cup J$ and $M_w(I', \mathrm{in}_v(\boldsymbol{f})) = M_w(M_v(I,\boldsymbol{f}), \mathrm{in}_v(\boldsymbol{f})) = M_v(M_v(I,\boldsymbol{f}), \mathrm{in}_v(\boldsymbol{f})) = M_v(I,\boldsymbol{f})$, so Property (LocInf)(b) is equivalent to $(O_v \cup J) \cap M_v(I,\boldsymbol{f}) \neq \emptyset$. This is exactly the Property (4.12) satisfied by $\boldsymbol{f}$.

When $w \notin v\mathbb{R}_{>0}$, let $U \subseteq D_n$ be the open neighbourhood of $v\mathbb{R}_{>0}$ defined in Lemma 4.6.1. Recall that scaling w by any positive real does not change the Property (4.13). Therefore by scaling w by a small enough positive real we can suppose $(v+w)\mathbb{R}_{>0} \in U$. We have

$$\mathrm{in}_{v+w}(\boldsymbol{f}) = \mathrm{in}_w(\mathrm{in}_v(\boldsymbol{f})), \quad M_{v+w}(I,\boldsymbol{f}) = M_w(I', \mathrm{in}_v(\boldsymbol{f})), \quad \text{and} \quad O_{v+w} = O_v \cup O_w,$$

where $I' = M_v(I,\boldsymbol{f})$. Therefore $(O_{v+w} \cup J) \cap M_{v+w}(I,\boldsymbol{f}) = (O_w \cup O_v \cup J) \cap M_w(I', \mathrm{in}_v(\boldsymbol{f})) = (O_w \cup J') \cap M_w(I', \mathrm{in}_v(\boldsymbol{f}))$. Since $\boldsymbol{f}$ satisfies Property (4.12), we have $(O_{v+w} \cup J) \cap M_{v+w}(I,\boldsymbol{f}) \neq \emptyset$. Therefore we also have $(O_w \cup J') \cap M_w(I', \mathrm{in}_v(\boldsymbol{f})) \neq \emptyset$ for all $w \notin v\mathbb{R}_{>0}$. □

We now start working towards proving the "if" part of Theorem 4.4.8. The main idea is a "gluing" procedure inspired by the original proof [36] of Theorem 4.4.9. The roadmap for

proving the "if" part of Theorem 4.4.8 is illustrated in Figure 4.26, where the case $n = K = 1$ is taken as an example.

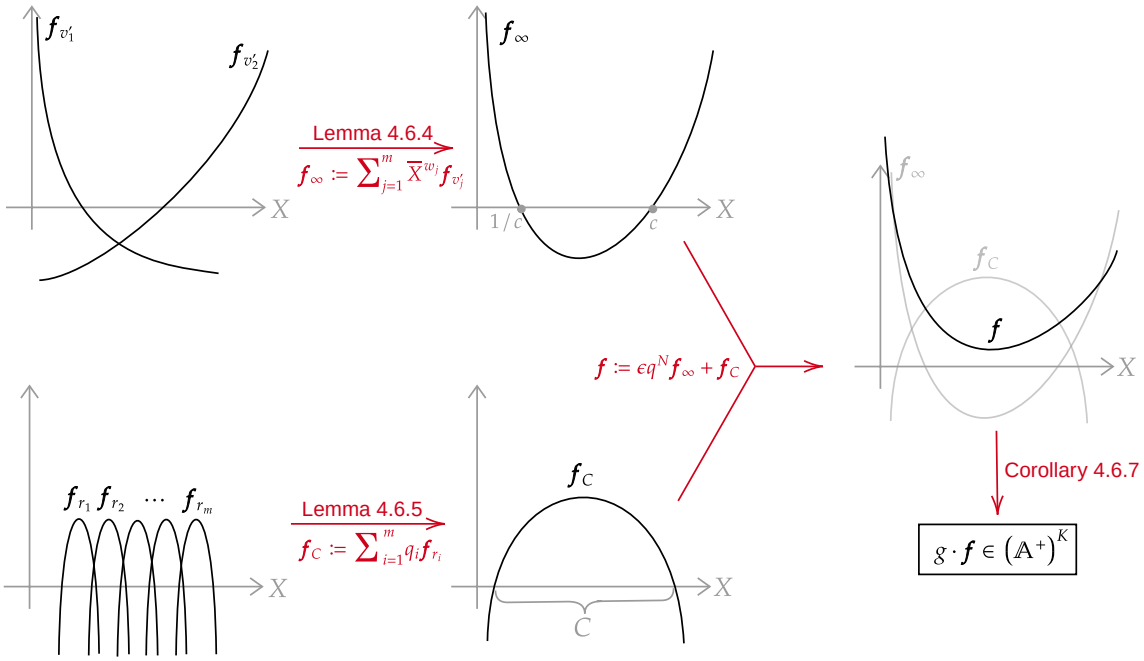

Figure 4.26: Roadmap for proofing the "if" part of Theorem 4.4.8, illustrated for $n = K = 1$.

The following lemma is the foundation of our gluing argument. It shows the "continuity" of Condition (LocInf) when changing the direction v by a small amount.

Lemma 4.6.2. *Suppose $v \in (\mathbb{R}^n)^*$ and $\boldsymbol{f}_v \in \mathcal{M}$ satisfies Properties (LocInf)(a) and (b) of Theorem 4.4.8. Then there exists an open neighbourhood $U_v \subseteq D_n$ of $v\mathbb{R}_{>0}$ such that for every $v'\mathbb{R}_{>0} \in U_v$, we have*

(i) $\mathrm{in}_{v'}(\boldsymbol{f}_v) \in (\mathbb{A}^+)^K$.

(ii) $(O_{v'} \cup J) \cap M_{v'}(I, \boldsymbol{f}_v) \neq \emptyset$.

Proof. We use Lemma 4.6.1 on v, I and $\boldsymbol{f}_v$ to obtain an open neighbourhood $U_v \subseteq D_n$ of $v\mathbb{R}_{>0}$, where for all $v'\mathbb{R}_{>0} = (v+w)\mathbb{R}_{>0} \in U_v$ we have

$$\mathrm{in}_{v+w}(\boldsymbol{f}_v) = \mathrm{in}_w(\mathrm{in}_v(\boldsymbol{f}_v)),\ M_{v+w}(I, \boldsymbol{f}_v) = M_w(M_v(I, \boldsymbol{f}_v), \mathrm{in}_v(\boldsymbol{f}_v)),\ \text{and } O_{v+w} = O_v \cup O_w.$$

Note that $\mathrm{in}_v(\boldsymbol{f}_v) \in (\mathbb{A}^+)^K$ by Property (LocInf)(a) of $\boldsymbol{f}_v$. Since taking the initial polynomial of any polynomial in $\mathbb{A}^+$ yields an element of $\mathbb{A}^+$, we have $\mathrm{in}_{v+w}(\boldsymbol{f}_v) = \mathrm{in}_w(\mathrm{in}_v(\boldsymbol{f}_v)) \in (\mathbb{A}^+)^K$. Furthermore, $(O_{v+w} \cup J) \cap M_{v+w}(I, \boldsymbol{f}_v) = (O_w \cup O_v \cup J) \cap M_w(M_v(I, \boldsymbol{f}_v), \mathrm{in}_v(\boldsymbol{f}_v)) = (O_w \cup J') \cap M_w(I', \mathrm{in}_v(\boldsymbol{f}_v))$, which is non-empty by Property (LocInf)(b) of $\boldsymbol{f}_v$. Therefore, both (i) and (ii) are satisfied for $v'\mathbb{R}_{>0} = (v+w)\mathbb{R}_{>0} \in U_v$. □

Before proceeding further, we recall the definition of the *Lebesgue number* of the open covering of a metric space.

Lemma 4.6.3 (The Lebesgue number lemma [76, Chapter 3, Lemma 7.2])**.** *Let $\{U_\alpha \mid \alpha \in A\}$ be an open covering of the metric space X. (That is, $X = \bigcup_{\alpha\in A} U_\alpha$ where each U_α is an open set.) If X is compact, then there is a $\delta > 0$ such that every subset of X with diameter less than δ is contained in some $U_\alpha, \alpha \in A$.*

The number $\delta > 0$ is called a *Lebesgue number* for the covering $\{U_\alpha \mid \alpha \in A\}$.

The following lemma shows that one can "glue" all different $\boldsymbol{f}_v, v \in (\mathbb{R}^n)^*$ together to obtain a single $\boldsymbol{f}$ that has positive initial polynomial at every direction $v \in (\mathbb{R}^n)^*$.

Lemma 4.6.4. *Suppose Condition (LocInf) of Theorem 4.4.8 is satisfied. Then there exists $\boldsymbol{f} \in \mathcal{M}$ that satisfies*

(i) $\mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ *for all* $v \in (\mathbb{R}^n)^*$.

(ii) $(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset$ *for all* $v \in (\mathbb{R}^n)^*$.

Proof. The main steps of our proof follow that of [36, Lemma 5.2]. See Figure 4.27 for an illustration of our proof.

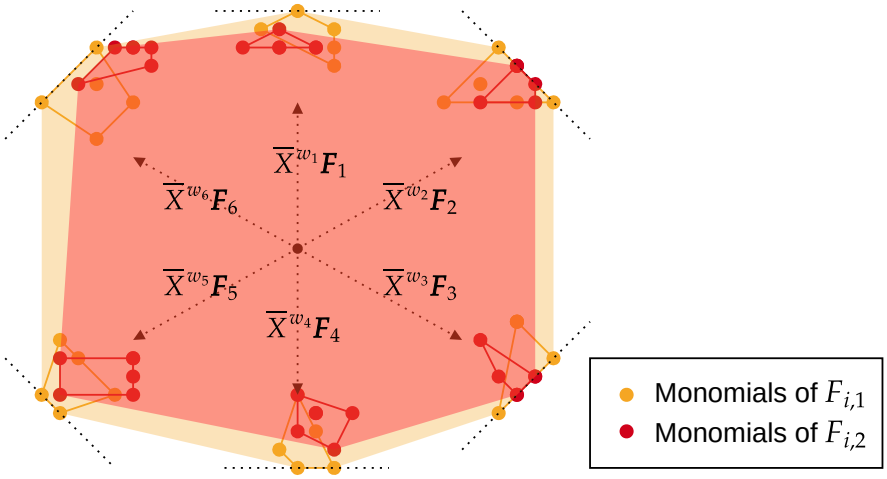

Figure 4.27: Illustration of the proof of Lemma 4.6.4. In this example, $a_1 = a_2 = (0, 0), I = \{1, 2\}, J = \{1\}$. That is, condition (ii) requires $1 \in M_v(\{1, 2\}, \boldsymbol{f})$ for all $v \in (\mathbb{R}^n)^*$.

For each $v \in (\mathbb{R}^n)^*, \|v\| = 1$, let $\boldsymbol{f}_v$ be the element of $\mathcal{M}$ that satisfies Properties (LocInf)(a) and (b). Let $U_v \subseteq D_n$ be an open neighbourhood of $v\mathbb{R}_{>0}$ defined in Lemma 4.6.2. The family of sets $U_v, v \in (\mathbb{R}^n)^*, \|v\| = 1$ forms an open cover of the compact set D_n. We identify D_n with the unit sphere in $\mathbb{R}^n$, and consider the metric on the unit sphere inherited from $\mathbb{R}^n$. In particular, each U_v is now considered as an open subset of the unit sphere.

Let $2\lambda < 1$ be a Lebesgue number of the open covering $\{U_v \mid v \in (\mathbb{R}^n)^*, \|v\| = 1\}$ of the unit sphere D_n. This means that every ball of radius λ in D_n is contained in some U_v. Take a finite collection of balls of radius λ which cover D_n, and label their centers $v_1, \ldots, v_m$: these points on the unit sphere. Note that each ball $B(v_j, \lambda), j = 1, \ldots, m$, is contained in some $U_{v'_j}$. Let $\boldsymbol{F}_j \coloneqq \boldsymbol{f}_{v'_j}$, so that $\mathrm{in}_v(\boldsymbol{F}_j) \in (\mathbb{A}^+)^K$ and $(O_v \cup J) \cap M_v(I, \boldsymbol{F}_j) \neq \emptyset$ for all $v \in B(v_j, \lambda)$ (by Lemma 4.6.2). Let 2κ be a Lebesgue number for the open cover $\{B(v_j, \lambda) \mid j = 1, \ldots, m\}$ of D_n. Then for any $v \in D_n$ there exists $j \in \{1, \ldots, m\}$ such that $B(v, \kappa) \subset B(v_j, \lambda)$ and, in particular, $\|v_j - v\| < \lambda - \kappa$.

Let δ be the infimum of

$$\left\{ v^\top v_j - v^\top v_{j'} \;\middle|\; \|v\| = 1, \|v_j - v\| < \lambda - \kappa, \|v_{j'} - v\| \geq \lambda, j, j' = 1, \ldots, m \right\}.$$

Note that $\delta \geq \kappa(\lambda - \frac{\kappa}{2}) > 0$ since for all $v, v_j, v_{j'}$ of norm one we have

$$v^\top v_j - v^\top v_{j'} = \frac{1}{2}\left(\|v_{j'} - v\|^2 - \|v_j - v\|^2\right) \geq \frac{1}{2}(\lambda^2 - (\lambda - \kappa)^2) = \kappa(\lambda - \frac{\kappa}{2}).$$

Choose r large enough so that $|\deg_v(F_{j,i})| < \frac{\delta}{2} r - \frac{\sqrt{n}}{2}$ for all $\|v\| = 1$, $i = 1, \ldots, K$ and $j = 1, \ldots, m$. Such an r exists because $F_{j,i} \neq 0$ (since $\mathrm{in}_v(\boldsymbol{F}_j) \in (\mathbb{A}^+)^K$), $\deg_v(F_{j,i})$ is continuous with respect to v, and D_n is compact.

For $j = 1, \ldots, m$ pick $w_j \in \mathbb{Z}^n$ such that $\|w_j - r v_j\| \leq \frac{\sqrt{n}}{2}$. Let

$$\boldsymbol{f} \coloneqq \sum_{j=1}^{m} \overline{X}^{w_j} \boldsymbol{F}_j.$$

We show that $\boldsymbol{f}$ satisfies both conditions (i) and (ii). Consider any $v \in (\mathbb{R}^n)^*$ with norm one. Let $j \in \{1, \ldots, m\}$ be such that $\|v - v_j\| < \lambda - \kappa$. For any $j' \in \{1, \ldots, m\}$ with $\|v - v_{j'}\| \geq \lambda$ we have

$$\begin{aligned}
\max_{1 \leq i \leq K} \left\{ \deg_v \left(\overline{X}^{w_{j'}} F_{j',i} \right) \right\} &= v^\top w_{j'} + \max_{1 \leq i \leq K} \left\{ \deg_v \left(F_{j',i} \right) \right\} \\
&< v^\top w_{j'} + \frac{\delta}{2} r - \frac{\sqrt{n}}{2} \\
&\leq r v^\top v_{j'} + \frac{\delta}{2} r \\
&= r v^\top v_j - r(v^\top v_j - v^\top v_{j'}) + \frac{\delta}{2} r \\
&\leq r v^\top v_j - \frac{\delta}{2} r \\
&\leq v^\top w_j + \frac{\sqrt{n}}{2} - \frac{\delta}{2} r
\end{aligned}$$

$$< v^\top w_j + \min_{1\le i\le K}\{\deg_v(F_{j,i})\}$$
$$= \min_{1\le i\le K}\left\{\deg_v\left(\overline{X}^{w_j}F_{j,i}\right)\right\}. \tag{4.20}$$

For the remaining indices j' with $\|v-v_{j'}\|<\lambda$ we already know that $\mathrm{in}_v(\boldsymbol{F}_{j'})\in(\mathbb{A}^+)^K$. Since there can be no cancellation with those initial parts, we get

$$\mathrm{in}_v(\boldsymbol{f}) = \mathrm{in}_v\left(\sum_{j'\colon \|v-v_{j'}\|<\lambda}\overline{X}^{w_{j'}}\boldsymbol{F}_{j'}\right)\in\left(\mathbb{A}^+\right)^K.$$

Therefore $\boldsymbol{f}$ satisfies condition (i).

For condition (ii), take any point v on the unit sphere, let $j'\in\{1,\dots,m\}$ be such that $\max_{i\in I}\{\deg_v(\overline{X}^{w_{j'}}F_{j',i})\} = \max_{1\le j\le m}\max_{i\in I}\{\deg_v(\overline{X}^{w_j}F_{j,i})\}$. We must have $\|v-v_{j'}\|<\lambda$. Indeed, if we had $\|v-v_{j'}\|\ge\lambda$ then there would exist $j\in\{1,\dots,m\}$ such that $\|v-v_j\|<\lambda-\kappa$ and Inequality (4.20) yields $\max_{i\in I}\{\deg_v(\overline{X}^{w_{j'}}F_{j',i})\}<\max_{i\in I}\{\deg_v(\overline{X}^{w_j}F_{j,i})\}$, a contradiction.

We will now show $M_v(I,\boldsymbol{F}_{j'})\subseteq M_v(I,\boldsymbol{f})$. Take any $i'\in M_v(I,\boldsymbol{F}_{j'})$, we show $i'\in M_v(I,\boldsymbol{f})$. On one hand, $\mathrm{in}_v(\boldsymbol{F}_{j'})\in(\mathbb{A}^+)^K$ because $\|v-v_{j'}\|<\lambda$. We have

$$\begin{aligned}\deg_v(\overline{X}^{w_{j'}}F_{j',i'}) &= \max_{i\in I}\{\deg_v(\overline{X}^{w_{j'}}F_{j',i})\} = \max_{1\le j\le m}\max_{i\in I}\{\deg_v(\overline{X}^{w_j}F_{j,i})\}\\ &= \max_{i\in I}\max_{j:\|v-v_j\|<\lambda}\{\deg_v(\overline{X}^{w_j}F_{j,i})\} = \max_{i\in I}\deg_v\left(\sum_{j:\|v-v_j\|<\lambda}\overline{X}^{w_j}F_{j,i}\right)\\ &= \max_{i\in I}\deg_v\left(\sum_{j=1}^{m}\overline{X}^{w_j}F_{j,i}\right) = \max_{i\in I}\deg_v(f_i), \end{aligned}\tag{4.21}$$

since there can be no cancellation when summing $\mathrm{in}_v(\overline{X}^{w_j}F_{j,i})\in\mathbb{A}^+$ for $j,\|v-v_j\|<\lambda$.

On the other hand,

$$\deg_v(\overline{X}^{w_{j'}}F_{j',i'}) = \max_{1\le j\le m}\max_{i\in I}\{\deg_v(\overline{X}^{w_j}F_{j,i})\}\ge\max_{1\le j\le m}\{\deg_v(\overline{X}^{w_j}F_{j,i'})\}.$$

So $\deg_v(\overline{X}^{w_{j'}}F_{j',i'}) = \max_{1\le j\le m}\{\deg_v(\overline{X}^{w_j}F_{j,i'})\}$. Hence,

$$\begin{aligned}\deg_v(\overline{X}^{w_{j'}}F_{j',i'}) &= \max_{1\le j\le m}\{\deg_v(\overline{X}^{w_j}F_{j,i'})\} = \max_{j:\|v-v_j\|<\lambda}\{\deg_v(\overline{X}^{w_j}F_{j,i'})\}\\ &= \deg_v\left(\sum_{j:\|v-v_j\|<\lambda}\overline{X}^{w_j}F_{j,i'}\right) = \deg_v\left(\sum_{j=1}^{m}\overline{X}^{w_j}F_{j,i'}\right) = \deg_v(f_{i'}),\end{aligned}$$

since there can be no cancellation when summing $\mathrm{in}_v(\overline{X}^{w_j}F_{j,i'})\in\mathbb{A}^+$ for $j,\|v-v_j\|<\lambda$.

Hence $\deg_v(f_{i'}) = \deg_v(\overline{X}^{w_{j'}} F_{j',i'}) = \max_{i\in I} \deg_v(f_i)$, which yields $i' \in M_v(I, \boldsymbol{f})$. Since this holds for all $i' \in M_v(I, \boldsymbol{F}_{j'})$, we have shown $M_v(I, \boldsymbol{F}_{j'}) \subseteq M_v(I, \boldsymbol{f})$. Thus, $(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \supseteq (O_v \cup J) \cap M_v(I, \boldsymbol{F}_{j'}) \neq \emptyset$. Therefore $\boldsymbol{f}$ satisfies condition (ii). □

Denote by $\boldsymbol{f}_\infty$ the element $\boldsymbol{f} \in \mathcal{M}$ obtained in Lemma 4.6.4. Since $\mathrm{in}_v(\boldsymbol{f}_\infty) \in (\mathbb{A}^+)^K$ for all $v \in (\mathbb{R}^n)^*$, there exists $c > 1$ such that $\boldsymbol{f}_\infty(x) \in \mathbb{R}^K_{>0}$ for all $x \in \mathbb{R}^n_{>0} \setminus [1/c, c]^n$. Define the following compact set in $\mathbb{R}^n_{>0}$:

$$C \coloneqq [1/(4nc), 4nc]^n \supsetneq [1/c, c]^n.$$

Lemma 4.6.5. *Let $\mathcal{M}$ be an $\mathbb{A}$-submodule of $\mathbb{A}^K$ and $C \subset \mathbb{R}^n_{>0}$ be a compact set. Suppose for all $r \in C$ there exists $\boldsymbol{f}_r \in \mathcal{M}$ with $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$. Then there exists $\boldsymbol{f} \in \mathcal{M}$ such that $\boldsymbol{f}(x) \in \mathbb{R}^n_{>0}$ for all $x \in C$.*

Proof. For each $r \in C$, by the continuity of polynomial functions, there is an open ball $B(r, b_r)$, centered at r, with radius b_r, such that $\boldsymbol{f}_r(x) \in \mathbb{R}^n_{>0}$ for all $x \in B(r, b_r)$.

Consider the open cover $B(r, \frac{b_r}{2}), r \in C$ of the set C. Since C is compact, there is a finite subcover, which we denote by $B(r_1, \frac{b_{r_1}}{2}), \cdots, B(r_m, \frac{b_{r_m}}{2})$. By the compactness of C and closed balls, we have

$$s \coloneqq \min_{1\le i\le K, 1\le j\le m} \inf_{\|x-r_j\|\le \frac{b_j}{2}} f_{r_j,i}(x) > 0,$$

and

$$t \coloneqq \max_{1\le i\le K, 1\le j\le m} \sup_{x\in C} \left|f_{r_j,i}(x)\right| < \infty.$$

Let $\delta > 0$ be such that $\delta < \frac{s}{mt}$. By the Stone-Weierstrass theorem, for each $1 \le j \le m$, there exists a polynomial $q_j \in \mathbb{A}$ such that

(i) $0 < q_j(x)$ for all $x \in C$,

(ii) $q_j(x) < \delta$ for all $x \in C \setminus B(r_j, b_{r_j})$,

(iii) $q_j(x) > 1$ for all $x \in B(r_j, \frac{b_{r_j}}{2})$.

We claim that the sum $\boldsymbol{f} \coloneqq \sum_{j=1}^m q_j \cdot \boldsymbol{f}_{r_j} \in \mathcal{M}$ satisfies $\boldsymbol{f}(x) \in \mathbb{R}^n_{>0}$ for all $x \in C$. Indeed, take any $x \in C$. Without loss of generality suppose $x \in B(r_1, \frac{b_{r_1}}{2})$. Then for $i = 1, \ldots, K$, we have

$$\begin{aligned} f_i(x) = q_1(x) \cdot f_{r_1,i}(x) + \sum_{j>1, x\notin B(r_j,b_{r_j})} q_j(x) \cdot f_{r_j,i}(x) + \sum_{j>1, x\in B(r_j,b_{r_j})} q_j(x) \cdot f_{r_j,i}(x) \\ \ge 1 \cdot s - \sum_{j>1, x\notin B(r_j,b_{r_j})} \delta t + \sum_{j>1, x\in B(r_j,b_{r_j})} 0 \ge s - (m-1)\delta t > 0. \end{aligned}$$

□

Denote by $\boldsymbol{f}_C$ the element $\boldsymbol{f} \in \mathcal{M}$ obtained in Lemma 4.6.5. We also need the following theorem from Handelman:

Theorem 4.6.6 (Handelman's Theorem [31], [44, V.6. Theorem C])**.** *Let $f \in \mathbb{A}$ be a polynomial. There exists $g \in \mathbb{A}^+$ such that $fg \in \mathbb{A}^+$ if and only if the two following conditions are satisfied:*

(i) For all $r \in \mathbb{R}^n_{>0}$, we have $f(r) > 0$.

(ii) For all $v \in (\mathbb{R}^n)^*$ and $r \in \mathbb{R}^n_{>0}$, we have $\mathrm{in}_v(f)(r) > 0$.

Corollary 4.6.7. *Let $\boldsymbol{f} \in \mathbb{A}^K$. There exists $g \in \mathbb{A}^+$ such that $g\boldsymbol{f} \in (\mathbb{A}^+)^K$ if and only if the two following conditions are satisfied:*

(i) For all $r \in \mathbb{R}^n_{>0}$, we have $\boldsymbol{f}(r) \in \mathbb{R}^K_{>0}$.

(ii) For all $v \in (\mathbb{R}^n)^*$ and $r \in \mathbb{R}^n_{>0}$, we have $\mathrm{in}_v(\boldsymbol{f})(r) \in \mathbb{R}^K_{>0}$.

Proof. If there exists $g \in \mathbb{A}^+$ such that $g \cdot \boldsymbol{f} \in (\mathbb{A}^+)^K$, then obviously (i) and (ii) are satisfied.

On the other hand, let $\boldsymbol{f} \in \mathbb{A}^k$ satisfy (i) and (ii). By Handelman's theorem (Theorem 4.6.6), there exist $g_1, \ldots, g_K \in \mathbb{A}^+$ such that $f_1 g_1 \in \mathbb{A}^+, \ldots, f_K g_K \in \mathbb{A}^+$. Let $g \coloneqq g_1 g_2 \cdots g_K$, then $g\boldsymbol{f} \in (\mathbb{A}^+)^K$. □

We are now ready to prove the "if" part of Theorem 4.4.8 by "gluing" together the elements $\boldsymbol{f}_\infty, \boldsymbol{f}_C \in \mathcal{M}$ obtained respectively in Lemma 4.6.4 and 4.6.5.

Proof of "if" part of Theorem 4.4.8. Let $\boldsymbol{f}_\infty, \boldsymbol{f}_C \in \mathcal{M}$ be the elements obtained respectively in Lemma 4.6.4 and 4.6.5. Define the polynomial

$$q \coloneqq \frac{1}{2nc} \sum_{i=1}^{n} (X_i + X_i^{-1}) \in \mathbb{A}^+.$$

It is easy to see that we have $\deg_w(q) > 0$ for all $w \in (\mathbb{R}^n)^*$. By the compactness of the unit sphere, the value $\inf_{\|w\|=1} \deg_w(q)$ is positive.

Let $\epsilon > 0$ be such that

$$\epsilon \cdot \boldsymbol{f}_\infty(x) + \boldsymbol{f}_C(x) \in \mathbb{R}^n_{>0} \tag{4.22}$$

for all $x \in C$. Such an ϵ exists by the compactness of C and because $\boldsymbol{f}_C(x) > 0$ for all $x \in C$. We claim that there exists $N \in \mathbb{N}$ such that the element

$$\boldsymbol{f} \coloneqq \epsilon q^N \cdot \boldsymbol{f}_\infty + \boldsymbol{f}_C \in \mathcal{M}$$

satisfies Conditions (i) and (ii) in Corollary 4.6.7 simultaneously.

Let $M \in \mathbb{N}$ be such that $\deg_v(f_{\infty,i}) + M \cdot \inf_{\|w\|=1} \deg_w(q) > \deg_v(f_{C,i})$ for all $v \in (\mathbb{R}^n)^*, \|v\| = 1$ and $i = 1, \ldots, K$. Such an M exists by the compactness of the unit sphere and because $\inf_{\|w\|=1} \deg_w(q) > 0$. Let

$$\boldsymbol{g} \coloneqq \epsilon q^M \cdot \boldsymbol{f}_\infty + \boldsymbol{f}_C.$$

Then for all $v \in (\mathbb{R}^n)^*, i = 1, \ldots, K$, we have $\deg_v(\epsilon q^M \cdot f_{\infty,i}) = M \cdot \deg_v(q) + \deg_v(f_{\infty,i}) > \deg_v(f_{C,i})$. Therefore $\mathrm{in}_v(\boldsymbol{g}) = \mathrm{in}_v(\epsilon q^M \cdot \boldsymbol{f}_\infty) \in (\mathbb{A}^+)^K$ for all $v \in (\mathbb{R}^n)^*$. Therefore, there exists another compact set $[1/d, d]^n \supset C$ such that $\boldsymbol{g}(x) \in \mathbb{R}^K_{>0}$ for all $x \in \mathbb{R}^n_{>0} \setminus [1/d, d]^n$. Since $[1/d, d]^n \supset C = [1/(4nc), 4nc]^n$, we have $d \geq 4nc$. Since the set $[1/d, d]^n \setminus (1/(4nc), 4nc)^n$ is compact and $\boldsymbol{f}_\infty(x) \in \mathbb{R}^K_{>0}$ for all $x \in [1/d, d]^n \setminus (1/(4nc), 4nc)^n \subseteq \mathbb{R}^n_{>0} \setminus [1/c, c]^n$, there exists $N > M$ such that

$$\epsilon f_{\infty,i}(x) \cdot 2^N + f_{C,i}(x) > 0 \tag{4.23}$$

for all $x \in [1/d, d]^n \setminus (1/(4nc), 4nc)^n$ and all $i = 1, \ldots, K$. We prove that for this N, the element $\boldsymbol{f} \coloneqq \epsilon q^N \cdot \boldsymbol{f}_\infty + \boldsymbol{f}_C$ satisfies Conditions (i) and (ii) in Corollary 4.6.7 simultaneously.

Fix any $i \in \{1, \ldots, K\}$. For every $x \in \mathbb{R}^n_{>0} \setminus [1/d, d]^n$, we have $q(x) > \frac{d}{2nc} > 1$ and $f_{\infty,i}(x) > 0$, so

$$f_i(x) = \epsilon q(x)^N \cdot f_{\infty,i}(x) + f_{C,i}(x) > \epsilon q(x)^M \cdot f_{\infty,i}(x) + f_{C,i}(x) = g_i(x) > 0.$$

For every $x \in [1/d, d]^n \setminus C = [1/d, d]^n \setminus [1/(4nc), 4nc]^n$, we have $x_{i'} \geq 4nc$ for at least one $i' \in \{1, \ldots, K\}$. Since $f_{\infty,i}(x) > 0$ by the definition of C, we have

$$f_i(x) = \epsilon f_{\infty,i}(x) \cdot \left(\sum_{j=1}^n \frac{x_j + x_j^{-1}}{2nc} \right)^N + f_{C,i}(x) \geq \epsilon f_{\infty,i}(x) \cdot 2^N + f_{C,i}(x) > 0$$

by $\sum_{j=1}^n (x_j + x_j^{-1}) > x_{i'} \geq 4nc$ and Inequality (4.23).

For every $x \in C \setminus [1/c, c]^n$, we have

$$f_i(x) = \epsilon q(x)^N \cdot f_{\infty,i}(x) + f_{C,i}(x) > 0$$

since $f_{\infty,i}(x) > 0$ for all $x \notin [1/c, c]^n$ and $f_{C,i}(x) > 0$ for all $x \in C$.

For every $x \in [1/c, c]^n$, we have

$$f_i(x) = \epsilon f_{\infty,i}(x) \cdot \left(\sum_{i=1}^n \frac{x_i + x_i^{-1}}{2nc} \right)^N + f_{C,i}(x) \geq \min\{\epsilon f_{\infty,i}(x), 0\} + f_{C,i}(x) > 0.$$

The last inequality is due to $f_{C,i}(x) > 0$ and Inequality (4.22). The second to last inequality can be justified as follows. If $f_{\infty,i}(x) \geq 0$ then $f_{\infty,i}(x) \cdot \left(\sum_{i=1}^{n} \frac{x_i + x_i^{-1}}{2nc}\right)^N \geq 0$, otherwise $\left(\sum_{i=1}^{n} \frac{x_i + x_i^{-1}}{2nc}\right)^N \leq \left(\sum_{i=1}^{n} \frac{2c}{2nc}\right)^N = 1$ so $f_{\infty,i}(x) \cdot \left(\sum_{i=1}^{n} \frac{x_i + x_i^{-1}}{2nc}\right)^N \geq f_{\infty,i}(x)$.

Therefore, for every $x \in \mathbb{R}^n_{>0}$, we have $f_i(x) > 0$. In other words, $\boldsymbol{f}$ satisfies Conditions (i) in Corollary 4.6.7. Furthermore, since $N > M$ we have $\deg_v(q^N \cdot f_{\infty,i}) > \deg_v(f_{C,i})$ for $i = 1, \ldots, K, v \in (\mathbb{R}^n)^*$. Hence $\mathrm{in}_v(\boldsymbol{f}) = \mathrm{in}_v(\epsilon q^N \cdot \boldsymbol{f}_\infty) \in (\mathbb{A}^+)^K$ and $M_v(I, \boldsymbol{f}) = M_v(I, \boldsymbol{f}_\infty)$ for all $v \in (\mathbb{R}^n)^*$. Therefore, $\boldsymbol{f}$ satisfies Conditions (ii) in Corollary 4.6.7.

Therefore, by Corollary 4.6.7, we can find $g \in \mathbb{A}^+$ such that $g\boldsymbol{f} \in (\mathbb{A}^+)^K$. We have at the same time $g\boldsymbol{f} \in \mathcal{M}$ as well as $(O_v \cup J) \cap M_v(I, g\boldsymbol{f}) = (O_v \cup J) \cap M_v(I, \boldsymbol{f}) = (O_v \cup J) \cap M_v(I, \boldsymbol{f}_\infty) \neq \emptyset$ for all $v \in (\mathbb{R}^n)^*$. We have thus found the required element $g\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfying Property (4.12). □

4.7 Decidability of local conditions

In this section we prove Theorem 4.4.10:

Theorem 4.4.10. *Fix $n \in \mathbb{N}$. Suppose we are given as input a set of elements $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathbb{A}^K$ with integer coefficients, as well as the vectors $a_1, \ldots, a_K \in \mathbb{Z}^n$ and two subsets I, J of $\{1, \ldots, K\}$. Denote by $\mathcal{M}$ be the $\mathbb{A}$-submodule of $\mathbb{A}^K$ generated by $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$. It is decidable whether there exists $\boldsymbol{f} \in \mathcal{M}$ satisfying $\boldsymbol{f} \in (\mathbb{A}^+)^K$ and*

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^*. \tag{4.14}$$

Here, if $n = 0$ then $\mathbb{A}$ is understood as $\mathbb{R}$, and Property (4.14) is considered trivially true.

By the local-global principle (Theorem 4.4.8), this amounts to showing decidability of the two "local" Conditions (LocR) and (LocInf).

4.7.1 Decidability of local condition at positive reals (LocR)

In this subsection we show that the Condition (LocR) of Theorem 4.4.8 is decidable. Let $\mathcal{M}$ be a $\mathbb{A}$-submodule of $\mathbb{A}^K$.

Lemma 4.7.1. *Let $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$ be the generators for $\mathcal{M}$. Condition (LocR) of Theorem 4.4.8 is equivalent to the following:*

1. **(LocRLin)** *For every $r \in \mathbb{R}^n_{>0}$, there exist $x_1, \ldots, x_m \in \mathbb{R}$ such that $\sum_{i=1}^{m} x_i \boldsymbol{g}_i(r) \in \mathbb{R}^K_{>0}$.*

Proof. (LocR) $\Longrightarrow$ (LocRLin): Suppose Condition (LocR) of Theorem 4.4.8 is true. For $r \in \mathbb{R}^n_{>0}$, Condition (LocR) shows there exist $p_1, \ldots, p_m \in \mathbb{A}$ such that $\sum_{i=1}^m p_i \boldsymbol{g}_i = \boldsymbol{f}_r$ with $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$. Then letting $x_1 \coloneqq p_1(r), \ldots, x_m \coloneqq p_m(r)$ we have $\sum_{i=1}^m x_i \boldsymbol{g}_i(r) = \boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$.

(LocRLin) $\Longrightarrow$ (LocR): Suppose Condition (LocRLin) is true. For any $r \in \mathbb{R}^n_{>0}$, Condition (LocRLin) shows there exist $x_1, \ldots, x_m \in \mathbb{R}_{>0}$ such that $\sum_{i=1}^m x_i \boldsymbol{g}_i(r) \in \mathbb{R}^K_{>0}$. Then $\boldsymbol{f}_r \coloneqq \sum_{i=1}^m x_i \boldsymbol{g}_i \in \mathcal{M}$ satisfies $\boldsymbol{f}_r(r) \in \mathbb{R}^K_{>0}$. □

Proposition 4.7.2. *Given the generators $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$ for $\mathcal{M}$, it is decidable whether Condition (LocR) of Theorem 4.4.8 is satisfied.*

Proof. By Lemma 4.7.1, it suffices to decide Condition (LocRLin). This is expressible in the first order theory of the reals:

$$\forall r_1 > 0 \cdots \forall r_n > 0, \exists x_1 \cdots \exists x_m, \left(\sum_{i=1}^m x_i g_{i,1}(r_1, \ldots, r_n) > 0 \right) \wedge \cdots \wedge \left(\sum_{i=1}^m x_i g_{i,K}(r_1, \ldots, r_n) > 0 \right).$$

By Tarski's theorem [94], the truth of this sentence is decidable. □

4.7.2 Local condition at infinity: shifted initials (LocInfShift)

In this subsection we introduce the *shifted initials*, in order to replace Condition (LocInf) of Theorem 4.4.8 with a new Condition (LocInfShift). Our definition follows that of [36, Section 1].

Suppose we are given $\boldsymbol{f} \in \mathbb{A}^K$, $v \in (\mathbb{R}^n)^*$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K$. Then the *shifted initials* $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = (\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_1, \ldots, \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_K)$ are defined as

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i \coloneqq \begin{cases} \mathrm{in}_v(f_i) & \text{if } \deg_v(f_i) + \alpha_i = \max_{1 \le i' \le K} \{\deg_v(f_{i'}) + \alpha_{i'}\}, \\ 0 & \text{if } \deg_v(f_i) + \alpha_i < \max_{1 \le i' \le K} \{\deg_v(f_{i'}) + \alpha_{i'}\}, \end{cases}$$

for $i = 1, \ldots, K$.

Lemma 4.7.3. *Let $\boldsymbol{f} \in \mathbb{A}^K$ and $v \in (\mathbb{R}^n)^*$. Then $\mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ if and only if there exists $\boldsymbol{\alpha} \in \mathbb{R}^K$ such that $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$. Furthermore, in this case we have $\mathrm{in}_v(\boldsymbol{f}) = \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})$ and $\alpha_1 + \deg_v(f_1) = \cdots = \alpha_K + \deg_v(f_K)$.*

Proof. If $\mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$, then $f_1, \ldots, f_K$ are non-zero. Let $\alpha_1 \coloneqq -\deg_v(f_1), \ldots, \alpha_K \coloneqq -\deg_v(f_K)$. We have $\deg_v(f_1) + \alpha_1 = \cdots = \deg_v(f_K) + \alpha_K = \max_{1 \le i \le K} \{\deg_v(f_i) + \alpha_i\}$, so $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = \mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$.

If $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$, then $\mathrm{in}_v(f_i) = \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i \in \mathbb{A}^+$ for $i = 1, \ldots, K$. Therefore $\mathrm{in}_v(\boldsymbol{f}) = \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$.

Furthermore, in this case, since $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i \neq 0$ for $i = 1, \ldots, K$, we have $\deg_v(f_i) + \alpha_i = \max_{1 \leq i' \leq K}\{\deg_v(f_{i'}) + \alpha_{i'}\}$. Hence $\alpha_1 + \deg_v(f_1) = \cdots = \alpha_K + \deg_v(f_K)$. □

Given $v = (v_1, \ldots, v_n)^\top \in (\mathbb{R}^n)^*$, denote by $\sum_{k=1}^n \mathbb{Z}v_k$ the $\mathbb{Z}$-module generated by $v_1, \ldots, v_n$:

$$\sum_{k=1}^n \mathbb{Z}v_k \coloneqq \left\{ \sum_{k=1}^n z_k v_k \;\middle|\; z_1, \ldots, z_n \in \mathbb{Z} \right\}.$$

Then for every $f \in \mathbb{A} \setminus \{0\}$, we have $\deg_v(f) \in \sum_{k=1}^n \mathbb{Z}v_k$.

Proposition 4.7.4. *Condition* (LocInf) *of Theorem 4.4.8 is equivalent to the following:*

2. **(LocInfShift):** *For every $v \in (\mathbb{R}^n)^*$, there exists $\boldsymbol{f} \in \mathcal{M}$ as well as $\boldsymbol{\alpha} \in (\sum_{k=1}^n \mathbb{Z}v_k)^K$ satisfying the following properties:*
   (a) $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$.
   (b) Denote $I' \coloneqq \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\}$, $J' \coloneqq O_v \cup J$. We have

$$\big(O_w \cup J'\big) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) \neq \emptyset \quad \text{for every } w \in (\mathbb{R}^n)^*.$$

Proof. **(LocInf) $\Longrightarrow$ (LocInfShift)**. Suppose Condition (LocInf) of Theorem 4.4.8 is true. Fix a vector $v \in (\mathbb{R}^n)^*$. Then there exists $\boldsymbol{f} \in \mathcal{M}$, such that $\mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ satisfies Property (LocInf)(b). As in Lemma 4.7.3, we can let $\alpha_i \coloneqq -\deg_v(f_i)$ for $i = 1, \ldots, K$. Then $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = \mathrm{in}_v(\boldsymbol{f}) \in (\mathbb{A}^+)^K$, satisfying (LocInfShift)(a). Furthermore, we have $\boldsymbol{\alpha} \in (\sum_{k=1}^n \mathbb{Z}v_k)^K$ by the definition of $\alpha_i = -\deg_v(f_i)$. Finally, $I' = \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\} = \{i \in I \mid \deg_v(f_i) = \max_{i' \in I} \deg_v(f_{i'})\} = M_v(I, \boldsymbol{f})$, so (LocInf)(b) implies (LocInfShift)(b).

(LocInfShift) $\Longrightarrow$ (LocInf). Suppose Condition (LocInfShift) is true. Fix a vector $v \in (\mathbb{R}^n)^*$. Then there exists $\boldsymbol{f} \in \mathcal{M}$ as well as $\boldsymbol{\alpha} \in (\mathbb{R}^n)^*$, such that $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ satisfies Property (LocInfShift)(b). By Lemma 4.7.3, we have $\mathrm{in}_v(\boldsymbol{f}) = \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$, and $\alpha_1 + \deg_v(f_1) = \cdots = \alpha_K + \deg_v(f_K)$. Therefore we have $I' = \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\} = \{i \in I \mid \deg_v(f_i) = \min_{i' \in I} \deg_v(f_{i'})\} = M_v(I, \boldsymbol{f})$, so (LocInfShift)(b) implies (LocInf)(b). □

4.7.3 Dimension reduction: a special case

In this and the following subsections we will further reduce Condition (LocInfShift) to a Condition (LocInfD) (which will be defined in Proposition 4.7.11). In this subsection we first consider the special case where the vector $v \in (\mathbb{R}^n)^*$ in Condition (LocInfShift) is of the form $(0, \ldots, 0, v_{d+1}, \ldots, v_n)^\top$, where $v_{d+1}, \ldots, v_n \in \mathbb{R}$ are $\mathbb{Q}$-linearly independent.

Let $\mathcal{M}$ be any $\mathbb{A}$-submodule of $\mathbb{A}^K$. Similar to [36, Section 2], we now define the notion of a *super Gröbner basis* of $\mathcal{M}$. This is a generalized version of the *universal Gröbner basis* [91]

for modules over a Laurent polynomial ring.[3] Note that our definition differs slightly from [36, Section 2], although the intuition is the same. Let $v \in (\mathbb{R}^n)^*, \boldsymbol{\alpha} \in \mathbb{R}^K$. Define $\mathrm{in}_{v,\boldsymbol{\alpha}}(\mathcal{M})$ to be the $\mathbb{A}$-module generated by the elements $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}), \boldsymbol{f} \in \mathcal{M}$:

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\mathcal{M}) \coloneqq \sum_{\boldsymbol{f}\in\mathcal{M}} \mathbb{A} \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = \left\{ \sum_{j=1}^{q} p_j \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}_j) \;\middle|\; q \in \mathbb{N}, p_1, \dots p_q \in \mathbb{A}, \boldsymbol{f}_1, \dots, \boldsymbol{f}_q \in \mathcal{M} \right\}.$$

In general, the module $\mathrm{in}_{v,\boldsymbol{\alpha}}(\mathcal{M})$ is *not* equal to its generating set $\{\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \mid \boldsymbol{f} \in \mathcal{M}\}$.

Definition 4.7.5 (Super Gröbner basis)**.** A set of generators $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$ for the module $\mathcal{M}$ is called a *super Gröbner basis* if for all $v \in (\mathbb{R}^n)^*, \boldsymbol{\alpha} \in \mathbb{R}^K$, the set $\{\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_1), \dots, \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_m)\}$ generates $\mathrm{in}_{v,\boldsymbol{\alpha}}(\mathcal{M})$ as an $\mathbb{A}$-module.

A super Gröbner basis for a module $\mathcal{M}$ always exists and can be effectively computed as in [36, Lemma 2.1]:

Lemma 4.7.6 (Reformulation of [36, Lemma 2.1])**.** *Suppose we are given a finite set of generators*[4] *for the module $\mathcal{M}$. Then a super Gröbner basis of $\mathcal{M}$ is effectively computable.*

Proof. Let $\mathbb{R}[\overline{X}] \coloneqq \mathbb{R}[X_1, \dots, X_n]$ be the usual polynomial ring over n variables (instead of the Laurent polynomial ring $\mathbb{A} = \mathbb{R}[\overline{X}^{\pm}]$). Let $e_1, \dots, e_K$ be the canonical $\mathbb{R}[\overline{X}]$-basis of $\mathbb{R}[\overline{X}]^K$. A *monomial* of $\mathbb{R}[\overline{X}]^K$ is an element of the form $\overline{X}^u e_i$ for some $u \in \mathbb{Z}^n_{\geq 0}, i \in \{1, \dots, K\}$. A *term order* on the monomials of $\mathbb{R}[\overline{X}]^K$ is a total order $\prec$ satisfying

(i) $e_i \prec \overline{X}^u e_i$,
(ii) $\overline{X}^u e_i \prec \overline{X}^{u'} e_{i'} \implies \overline{X}^{u+w} e_i \prec \overline{X}^{u'+w} e_{i'}$,

for all $i, i' \in \{1, \dots, K\}$ and $u, u', w \in \mathbb{Z}^n_{\geq 0}$.

Let N be an $\mathbb{R}[\overline{X}]$-submodule of $\mathbb{R}[\overline{X}]^K$. An element $\boldsymbol{f} \in \mathbb{R}[\overline{X}]^K$ can be written uniquely as a sum $\sum_{u,i} c_{u,i} \overline{X}^u e_i$ with coefficients $c_{u,i}$ in $\mathbb{R}$. Among the finitely many monomials of $\mathbb{R}[\overline{X}]^K$ that have nonzero coefficients in this sum, the one that is maximal according to the term order $\prec$ is denoted $\mathrm{in}_\prec(\boldsymbol{f})$. Define $\mathrm{in}_\prec(N)$, the *initial module* of N with respect to $\prec$, to be the $\mathbb{R}[\overline{X}]$-module generated by all $\mathrm{in}_\prec(\boldsymbol{f}), \boldsymbol{f} \in N$. We say that the elements $\boldsymbol{f}_1, \dots, \boldsymbol{f}_\ell \in N$ form a *Gröbner basis* for N with respect to $\prec$ if $\mathrm{in}_\prec(N)$ is generated as an $\mathbb{R}[\overline{X}]$-module by $\mathrm{in}_\prec(\boldsymbol{f}_1), \dots, \mathrm{in}_\prec(\boldsymbol{f}_\ell)$.

[3] The usual universal Gröbner basis is defined for ideals over a polynomial ring.

[4] There is a subtlety in how the generators are represented, since element of $\mathcal{M}$ are tuples of polynomials over *real* numbers. However, in our application throughout this section, the generators of $\mathcal{M}$ are always tuples of polynomials with *integer* coefficients. Therefore, they can indeed be effectively represented.

A *universal Gröbner basis* of N is given by elements $\boldsymbol{f}_1, \dots, \boldsymbol{f}_\ell$ that form a Gröbner basis of N with respect to *every* term order. A universal Gröbner basis of N exists and can be effectively computed from a set of generators of N as described in [36, Section 2].

Now let $\boldsymbol{f}_1, \dots, \boldsymbol{f}_\ell$ be a set of generators of $\mathcal{M}$ as an $\mathbb{R}[\overline{X}^{\pm}]$-module. Let $\delta = (\delta_1, \dots, \delta_n) \in \{-1, 1\}^n$. Pick $u \in \mathbb{Z}^n$ such that

$$\overline{X}^u \boldsymbol{f}_1, \dots, \overline{X}^u \boldsymbol{f}_\ell \in \left(\mathbb{R}[X_1^{\delta_1}, \dots, X_n^{\delta_n}]\right)^K,$$

and let $\boldsymbol{f}_{\delta,1}, \dots, \boldsymbol{f}_{\delta,\ell_\delta}$ be a universal Gröbner basis for the $\mathbb{R}[X_1^{\delta_1}, \dots, X_n^{\delta_n}]$-submodule generated by $\overline{X}^u \boldsymbol{f}_1, \dots, \overline{X}^u \boldsymbol{f}_\ell$. List the union of $\{\boldsymbol{f}_{\delta,1}, \dots, \boldsymbol{f}_{\delta,\ell_\delta}\}$ over $\delta \in \{-1, 1\}^n$ as $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$. By [36, Lemma 2.1], this is a super Gröbner basis for $\mathcal{M}$. □

It is easy to see the following from the proof: if the given generators for $\mathcal{M}$ are tuples of polynomials with integer coefficients, then Lemma 4.7.6 computes a super Gröbner basis containing only tuples of polynomials with integer coefficients. From now on we fix a super Gröbner basis $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$.

Let $0 \le d \le n-1$ be an integer. From now on we denote

$$\mathbb{A}_d \coloneqq \mathbb{R}[X_1^{\pm}, \dots, X_d^{\pm}], \quad \mathbb{A}_d^+ \coloneqq \mathbb{R}_{\ge 0}[X_1^{\pm}, \dots, X_d^{\pm}]^*.$$

In particular, $\mathbb{A}_0 = \mathbb{R}, \mathbb{A}_0^+ = \mathbb{R}_{>0}$.

As stated in the beginning of this subsection, we now consider the vectors $v \in (\mathbb{R}^n)^*$ with the special form $(0, \dots, 0, v_{d+1}, \dots, v_n)^\top$ where $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent. The following lemma can be seen as a generalization of [36, Lemma 6.2].

Lemma 4.7.7 (Generalization of [36, Lemma 6.2])**.** *Let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$ be a super Gröbner basis of $\mathcal{M}$. Let $v = (0, \dots, 0, v_{d+1}, \dots, v_n)^\top \in (\mathbb{R}^n)^*$ be such that $0 \le d \le n-1$ and $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent. Let $\boldsymbol{\alpha} \in \mathbb{R}^K$. Then there exist $b_i \in \{0\}^d \times \mathbb{Z}^{n-d}$ and $c_j \in \{0\}^d \times \mathbb{Z}^{n-d}$ such that $\overline{X}^{b_i} \overline{X}^{c_j} \operatorname{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \in \mathbb{A}_d$ for $i = 1, \dots, K$ and $j = 1, \dots, m$. See Figure 4.28 for an illustration.*

Proof. Let $j \in \{1, \dots, m\}, i \in \{1, \dots, K\}$, be such that $\operatorname{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$. Let $c\overline{X}^z$ and $c'\overline{X}^{z'}$ be any two monomials appearing in $\operatorname{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i$, then $v^\top z = v^\top z'$. Since $v = (0, \dots, 0, v_{d+1}, \dots, v_n)^\top$ where $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent, this yields $z - z' \in \mathbb{Z}^d \times \{0\}^{n-d}$. This shows that $\overline{X}^{z_{ij}} \operatorname{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \in \mathbb{A}_d$ for some $z_{ij} \in \{0\}^d \times \mathbb{Z}^{n-d}$.

Letting $F \coloneqq \{(i,j) \mid \operatorname{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0\}$, this defines z_{ij} for all $(i,j) \in F$. Note that for $(i,j) \in F$

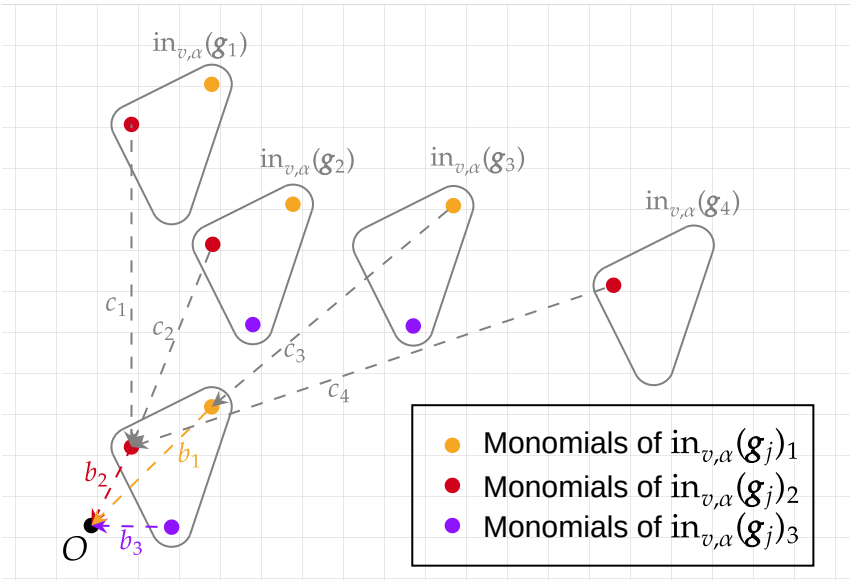

Figure 4.28: Illustration of Lemma 4.7.7 in the case $d = 0$. Note that $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i$ may be zero, in which case it does not contain any monomial.

we have

$$\max_{1\le i'\le K}\left\{\deg_v\left(g_{j,i'}\right)+\alpha_{i'}\right\} = -v^\top z_{ij}+\alpha_i.$$

Considering a sequence

$$(i_0,j_0),(i_1,j_0),(i_1,j_1),\ldots,(i_l,j_{l-1}),(i_l,j_l),(i_0,j_l) \tag{4.24}$$

in F, and writing $i_{l+1}=i_0$, we find that

$$0=\sum_{s=0}^{l}\left(\max_{1\le i'\le K}\{\deg_v(g_{j,i'})+\alpha_{i'}\}-\max_{1\le i'\le K}\{\deg_v(g_{j,i'})+\alpha_{i'}\}\right)=\sum_{s=0}^{l}\left(v^\top z_{i_{s+1}j_s}-v^\top z_{i_sj_s}\right).$$

Since $v=(0,\ldots,0,v_{d+1},\ldots,v_n)^\top$ with $v_{d+1},\ldots,v_n$ being $\mathbb{Q}$-linearly independent and $z_{ij}\in\{0\}^d\times\mathbb{Z}^{n-d}$, the above equation yields

$$\sum_{s=0}^{l}\left(z_{i_{s+1}j_s}-z_{i_sj_s}\right)=0 \tag{4.25}$$

for every allowed sequence (4.24) in F.

We now extend z_{ij} and Equation (4.25) to all pairs $(i,j)\in\{1,\ldots,K\}\times\{1,\ldots,m\}$. Assume z_{ij} is already defined on a set $E\supseteq F$ and (4.25) is valid on E. Pick $(i,j)\notin E$. If there exists a sequence

$$(i_1,j),(i_1,j_1),\ldots,(i_l,j_{l-1}),(i_l,j_l),(i,j_l)\in E,$$

we put $i_{l+1} = i$ and define

$$z_{ij} \coloneqq z_{i_1 j} - \sum_{s=1}^{l} \left(z_{i_s j_s} - z_{i_{s+1} j_s} \right).$$

One easily verifies that Equation (4.25) then holds for every allowed sequence (4.24) in $E \cup \{(i,j)\}$.

If there is no sequence

$$(i_1, j_0), (i_1, j_1), \ldots, (i_l, j_{l-1}), (i_l, j_l), (i_0, j_l)$$

in E, we can take z_{ij} to be any element of $\{0\}^d \times \mathbb{Z}^{n-d}$ and have Equation (4.25) hold for all sequences (4.24) in $E \cup \{(i,j)\}$.

Having thus extended z_{ij} to all pairs $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,m\}$, we define

$$b_i \coloneqq z_{i1},$$

and

$$c_j \coloneqq z_{ij} - z_{i1},$$

which is independent of i thanks to Equation (4.25). Indeed, using Equation (4.25) on the allowed sequence $(i,j), (i',j), (i',1), (i,1)$ we get $z_{ij} - z_{i1} = z_{i'j} - z_{i'1}$. Hence, $z_{ij} = b_i + c_j$ and the lemma follows. □

Suppose $v \in (\mathbb{R}^n)^*$ is such that $v = (0,\ldots,0,v_{d+1},\ldots,v_n)^\top$ with $v_{d+1},\ldots,v_n$ being $\mathbb{Q}$-linearly independent. Let $\boldsymbol{\alpha} \in \mathbb{R}^K$. For each $j = 1,\ldots,m$, define

$$\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) = (\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_1, \ldots, \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_K),$$

where

$$\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \coloneqq \overline{X}^{b_i} \overline{X}^{c_j} \, \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \in \mathbb{A}_d, \quad i = 1,\ldots,K.$$

Here, b_i and c_j are defined as in Lemma 4.7.7. Note that the vectors $b_i, c_j \in \{0\}^d \times \mathbb{Z}^{n-d}$ are not necessarily uniquely determined. However, when $d, v, \boldsymbol{\alpha}$ are fixed, the polynomials $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i, j = 1,\ldots,m, i = 1,\ldots,K$, are uniquely determined by $\boldsymbol{g}_1,\ldots,\boldsymbol{g}_m$. In fact, by Lemma 4.7.7, each $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i$ can be uniquely written as $\overline{X}^s \cdot p$ for some $\overline{X}^s \in \mathbb{R}[X^{\pm}_{d+1},\ldots,X^{\pm}_n]$ and $p \in \mathbb{R}[X^{\pm}_1,\ldots,X^{\pm}_d]$. Therefore $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i$ is uniquely determined as the polynomial p in the decomposition.

Note that if $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$ then $\deg_v(g_{j,i}) = -v^\top(b_i + c_j)$, otherwise $\deg_v(g_{j,i}) < -v^\top(b_i + c_j)$.

In both cases,

$$\deg_v(g_{j,i}) \le -v^\top(b_i + c_j), \tag{4.26}$$

where the equality holds if and only if $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$.

Lemma 4.7.8. *Let $i, i' \in \{1, \dots, K\}$. If there exists $j \in \{1, \dots, m\}$ such that $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0, \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i'} \neq 0$, then $\alpha_i - v^\top b_i = \alpha_{i'} - v^\top b_{i'}$.*

Proof. If $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$ and $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i'} \neq 0$, then $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$ and $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i'} \neq 0$, so $\deg_v\left(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i\right) + \alpha_i = \deg_v\left(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i'}\right) + \alpha_{i'}$.

But $\deg_v\left(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i\right) = \deg_v\left(\overline{X}^{-b_i}\overline{X}^{-c_j} \cdot \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i\right) = -v^\top(b_i + c_j)$. Deriving the same equation for i' we have $-v^\top(b_i + c_j) + \alpha_i = -v^\top(b_{i'} + c_j) + \alpha_{i'}$. This yields $\alpha_i - v^\top b_i = \alpha_{i'} - v^\top b_{i'}$. □

Define by $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M})$ the $\mathbb{A}_d$-module generated by $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_1), \dots, \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_m) \in \mathbb{A}_d^K$:

$$\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}) \coloneqq \sum_{j=1}^m \mathbb{A}_d \cdot \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) = \left\{ \sum_{j=1}^m p_j \cdot \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) \;\middle|\; p_1, \dots p_m \in \mathbb{A}_d \right\}.$$

A key component of proving the original decidability result of Einsiedler et al. [36] is [36, Lemma 3.2], which shows that $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}) \cap \left(\mathbb{A}_d^+\right)^K \neq \emptyset$ implies $\mathcal{M} \cap (\mathbb{A}^+)^K \neq \emptyset$. However, this no longer work when we additionally impose Property (4.12): an element in $\mathcal{M} \cap (\mathbb{A}^+)^K$ satisfying Property (4.12) might not be obtained from an element in $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}) \cap \left(\mathbb{A}_d^+\right)^K$ satisfying a similar property. Indeed, if we directly apply [36, Lemma 3.2] to our situation, the main failure would be in the last paragraph of the proof, where different "levels" of polynomials are combined together to create a positive element. This no longer works if we add in degree constraints. The following lemma shows that [36, Lemma 3.2] can still be made partially compatible with Property (4.12), if we impose the additional constraint $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$.

Lemma 4.7.9. *Let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$ be a super Gröbner basis of $\mathcal{M}$. Let $v = (0, \dots, 0, v_{d+1}, \dots, v_n)^\top$ be such that $0 \le d \le n-1$ and $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent. Let $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$. Denote $I' \coloneqq \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\}$, $J' \coloneqq O_v \cup J$. Denote by $\pi_d \colon \mathbb{Z}^n \to \mathbb{Z}^d$ the projection onto the first d coordinates. For every $u \in \left(\mathbb{R}^d\right)^*$, define $O'_u \coloneqq \{i \in \{1, \dots, K\} \mid \pi_d(a_i) \not\perp u\}$. Then the two following conditions are equivalent:*

(i) (Condition in (LocInfShift)*): There exists $\boldsymbol{f} \in \mathcal{M}$ such that $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ and*

$$\left(O_w \cup J'\right) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) \neq \emptyset \quad \text{for every } w \in (\mathbb{R}^n)^*. \tag{4.27}$$

(ii) We have $J' \cap I' \neq \emptyset$*, and there exists* $\boldsymbol{f}^d \in \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}) \cap \left(\mathbb{A}^+_d\right)^K$*, such that*

$$\left(O'_u \cup J'\right) \cap M_u(I', \boldsymbol{f}^d) \neq \emptyset \quad \text{for every } u \in \left(\mathbb{R}^d\right)^*. \tag{4.28}$$

When $d = 0$*, the Property* (4.28) *is considered trivially true.*

Proof. **(i)** $\Longrightarrow$ **(ii).** Suppose (i) holds. Let $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfy (4.27). We now show (ii).

The property $J' \cap I' \neq \emptyset$ follows from $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ and (4.27) by taking $w \coloneqq v$. Indeed, we have $O_v \cup J' = O_v \cup J$ and

$$\begin{aligned} M_v(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) &= \left\{ i \in I' \;\middle|\; \deg_v(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i) = \max_{i' \in I'} \deg_v(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_{i'}) \right\} \\ &= \left\{ i \in I' \;\middle|\; -\alpha_i = \max_{i' \in I'}(-\alpha_{i'}) \right\} = I', \end{aligned}$$

where the second equality comes from $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$ and Lemma 4.7.3. Therefore, Property (4.27) yields $J' \cap I' \neq \emptyset$ by taking $w \coloneqq v$.

Since $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$ is a super Gröbner basis, we can write

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = \sum_{j=1}^{m} h_j \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) \tag{4.29}$$

for some $h_1, \ldots, h_m \in \mathbb{A}$. Let

$$S \coloneqq \left\{ 1 \leq j \leq m \;\middle|\; \deg_v(h_j) + \max_{1 \leq i \leq K}(\deg_v(\boldsymbol{g}_{j,i}) + \alpha_i) \text{ is maximal} \right\}.$$

Without loss of generality suppose $\sum_{j \in S} \mathrm{in}_v(h_j) \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) \neq 0$, otherwise we can replace each $h_j, j \in S$ by $h_j - \mathrm{in}_v(h_j)$ while (4.29) still holds. We have

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) = \mathrm{in}_{v,\boldsymbol{\alpha}}\left(\sum_{j=1}^{m} h_j \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)\right) = \sum_{j \in S} \mathrm{in}_v(h_j) \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j). \tag{4.30}$$

Indeed, by the definition of the shifted initials $\mathrm{in}_{v,\boldsymbol{\alpha}}$, the right hand side above are the only elements that can contribute to the shifted initials of the sum in the middle. Equation (4.30) shows we can without loss of generality suppose $h_j = \mathrm{in}_v(h_j)$ for all $j \in S$. Denote $D \coloneqq \deg_v(h_j) + \max_{1 \leq i' \leq K}(\deg_v(g_{j,i'}) + \alpha_{i'}), j \in S$. This does not depend on the choice of $j \in S$.

Since $v = (0, \ldots, 0, v_{d+1}, \ldots, v_n)^\top$ such that $v_{d+1}, \ldots, v_n$ are $\mathbb{Q}$-linearly independent and since $h_j = \mathrm{in}_v(h_j)$, we can write $h_j = \overline{X}^{z_j} p_j$, where $p_j \in \mathbb{A}_d$ and $z_j \in \{0\}^d \times \mathbb{Z}^{n-d}$. Note that $D = \deg_v(h_j) + \max_{1 \leq i' \leq K}(\deg_v(g_{j,i'}) + \alpha_{i'}) = v^\top z_j - v^\top(b_i + c_j) + \alpha_i$ for all (i, j) satisfying

$\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0, j \in S$. Since $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$, each $\alpha_i, i = 1, \ldots, K$, can be written as $\alpha_i = v^\top z'_i$ for some $z'_i \in \{0\}^d \times \mathbb{Z}^{n-d}$. So $D = v^\top(z_j - b_i - c_j + z'_i)$ for all (i, j) satisfying $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0, j \in S$.

By the $\mathbb{Q}$-linear independence of the entries in v, there exists a single $z \in \{0\}^d \times \mathbb{Z}^{n-d}$ such that $z = z_j - b_i - c_j + z'_i$ for all (i, j) satisfying $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0, j \in S$. Recall that $\sum_{j\in S} \mathrm{in}_v(h_j) \cdot \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) = \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$. Hence, for each $i \in \{1, \ldots, K\}$ there exists $j \in S$ such that $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$. Then for $i = 1, \ldots, K$, we have

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i = \sum_{j\in S} \overline{X}^{z_j} p_j \,\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i = \sum_{j\in S} p_j \overline{X}^{z_j - b_i - c_j} \,\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i = \overline{X}^{z - z'_i} \sum_{j\in S} p_j \,\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i.$$

Define

$$\boldsymbol{f}^d \coloneqq \sum_{j\in S} p_j \,\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) \in \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}).$$

Then for $i = 1, \ldots, K$,

$$f^d_i = \overline{X}^{z'_i - z} \,\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i \in \mathbb{A}^+ \cap \mathbb{A}_d = \mathbb{A}^+_d.$$

Therefore $\boldsymbol{f}^d \in \left(\mathbb{A}^+_d\right)^K$. It is left to show that $\boldsymbol{f}^d$ satisfies Property (4.28). If $d = 0$ then Property (4.28) is trivially true. Suppose $d \geq 1$. For each $u \in \left(\mathbb{R}^d\right)^*$, let $w \coloneqq (u, 0^{n-d})$ in (4.27). Then

$$\left(O_w \cup J'\right) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) \neq \emptyset. \tag{4.31}$$

For each $i \in \{1, \ldots, K\}$, because $z - z'_i \in \{0\}^d \times \mathbb{Z}^{n-d}$ and $w \in \mathbb{Z}^d \times \{0\}^{n-d}$ we have $\deg_w\left(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i\right) = \deg_w\left(\overline{X}^{z-z'_i} f^d_i\right) = \deg_u\left(f^d_i\right)$. Hence,

$$\begin{aligned} M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) &= \left\{ i \in I' \;\middle|\; \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i) = \max_{i'\in I'} \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_{i'}) \right\} \\ &= \left\{ i \in I' \;\middle|\; \deg_u(f^d_i) = \max_{i'\in I'} \deg_u(f^d_{i'}) \right\} = M_u(I', \boldsymbol{f}^d). \end{aligned} \tag{4.32}$$

Since the last $n - d$ entries of v are $\mathbb{Q}$-linearly independent, we have $a_i \perp v$ if and only if $a_i \in \sum_{k=1}^d \mathbb{Z}e_k$. Furthermore,

$$\begin{aligned} O'_u \cup J' &= O'_u \cup O_v \cup J = \{i \mid \neg(\pi_d(a_i) \perp u \wedge a_i \perp v)\} \cup J \\ &= \left\{ i \;\middle|\; \neg\left(\pi_d(a_i) \perp u \wedge a_i \in \sum_{i=1}^d \mathbb{Z}e_i\right) \right\} \cup J = \left\{ i \;\middle|\; \neg\left(a_i \perp w \wedge a_i \in \sum_{i=1}^d \mathbb{Z}e_i\right) \right\} \cup J \\ &= \{i \mid \neg(a_i \perp w \wedge a_i \perp v)\} \cup J = O_w \cup O_v \cup J = O_w \cup J' \end{aligned} \tag{4.33}$$

Therefore, combining Equations (4.31), (4.32) and (4.33), we obtain

$$\big(O'_u \cup J'\big) \cap M_u(I', \boldsymbol{f}^d) = \big(O_w \cup J'\big) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) \neq \emptyset.$$

We have thus shown that $\boldsymbol{f}^d$ satisfies Property (4.28).

(ii) $\Longrightarrow$ (i). Suppose (ii) holds. Write $\boldsymbol{f}^d = \sum_{j=1}^m p_j \,\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)$ where $p_j \in \mathbb{A}_d$ for $j = 1, \ldots, m$.

For each $i \in \{1, \ldots, K\}$, write $\alpha_i = v^\top z_i$ for some $z_i \in \{0\}^d \times \mathbb{Z}^{n-d}$. By the $\mathbb{Q}$-linear independence of the entries of v, such z_i is unique.

For each $j \in \{1, \ldots, m\}$, take any $i_j \in \{1, \ldots, K\}$ such that $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i_j} \neq 0$, note that the vector $c_j + b_{i_j} - z_{i_j}$ does not depend on the choice of i_j. Indeed, take any other $i'_j \in \{1, \ldots, K\}$ such that $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i'_j} \neq 0$, then by Lemma 4.7.8 we have $\alpha_{i_j} - v^\top b_{i_j} = \alpha_{i'_j} - v^\top b_{i'_j}$. Since $\alpha_{i_j} = v^\top z_{i_j}, \alpha_{i'_j} = v^\top z_{i'_j}$ we have $v^\top(z_{i_j} - b_{i_j}) = v^\top(z_{i'_j} - b_{i'_j})$. By the $\mathbb{Q}$-linear independence of the entries of v, we have $z_{i_j} - b_{i_j} = z_{i'_j} - b_{i'_j}$. So the vector $c_j + b_{i_j} - z_{i_j}$ does not depend on the choice of i_j.

Since $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_{i_j} \neq 0$, for all $i \in \{1, \ldots, K\}$ we have

$$\deg_v(g_{j,i}) + \alpha_i \leq \deg_v(g_{j,i_j}) + \alpha_{i_j}, \tag{4.34}$$

where the equality holds if and only if $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$.

Take

$$\boldsymbol{f} \coloneqq \sum_{j=1}^m \overline{X}^{c_j + b_{i_j} - z_{i_j}} p_j \cdot \boldsymbol{g}_j \in \mathcal{M}.$$

For each $i \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, m\}$, we have

$$\begin{aligned}\deg_v(\overline{X}^{c_j + b_{i_j} - z_{i_j}} p_j g_{j,i}) + \alpha_i &= v^\top(c_j + b_{i_j}) - v^\top z_{i_j} + \deg_v(g_{j,i}) + \alpha_i \\ &\leq -\deg_v(g_{j,i_j}) - \alpha_{i_j} + \deg_v(g_{j,i}) + \alpha_i \leq 0 \end{aligned} \tag{4.35}$$

The first equality follows from $\deg_v(p_j) = 0$ because $p_j \in \mathbb{A}_d$ and $v \in \{0\}^d \times \mathbb{R}^{n-d}$. The first inequality comes from (4.26) and $\alpha_{i_j} = v^\top z_{i_j}$, while the second inequality comes from (4.34). Furthermore, the equality in (4.35) holds if and only if $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0$ by the equality conditions in (4.26) and (4.34).

Hence, for each $i \in \{1, \ldots, K\}$ we have

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i = \sum_{j : \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0} \overline{X}^{c_j + b_{i_j} - z_{i_j}} \,\mathrm{in}_v(p_j g_{j,i}) = \sum_{j : \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0} \overline{X}^{c_j + b_i - z_i} p_j \,\mathrm{in}_v(g_{j,i})$$

$$= \sum_{j:\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i \neq 0} \overline{X}^{c_j+b_i-z_i} p_j \,\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i = \sum_{j=1}^{m} \overline{X}^{c_j+b_i-z_i} p_j \,\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i = \overline{X}^{-z_i} \sum_{j=1}^{m} p_j \,\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j)_i$$
$$= \overline{X}^{-z_i} f_i^d \in \mathbb{A}^+. \quad (4.36)$$

In the first equality, the initial polynomials do not cancel each other because their sum is $\overline{X}^{-z_i} f_i^d \in \mathbb{A}_d^+$. Therefore $\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f}) \in (\mathbb{A}^+)^K$.

We now prove Property (4.27). Recall $I' \coloneqq \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\}$. For $i, i' \in I'$, we have $v^\top z_i = \alpha_i = \alpha_{i'} = v^\top z_{i'}$. By the $\mathbb{Q}$-linear independence of the entries of v, we have

$$z_i = z_{i'} \text{ for all } i, i' \in I'. \quad (4.37)$$

Take any $w \in (\mathbb{R}^n)^*$.

If $w \in \sum_{k=d+1}^n \mathbb{R} e_k$, then

$$\begin{aligned} M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) &= \left\{ i \in I' \;\middle|\; \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i) = \max_{i' \in I'} \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_{i'}) \right\} \\ &= \left\{ i \in I' \;\middle|\; -w^\top z_i = \max_{i' \in I'} \{-w^\top z_{i'}\} \right\} \qquad \text{(by (4.36))} \\ &= I'. \qquad \text{(by (4.37))} \end{aligned}$$

So

$$\big(O_w \cup J'\big) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) = \big(O_w \cup J'\big) \cap I' \supseteq J' \cap I' \neq \emptyset.$$

If $w \notin \sum_{k=d+1}^n \mathbb{R} e_k$ and $d \geq 1$, write $w = w' + u$ where $w' \in \sum_{k=d+1}^n \mathbb{R} e_k$ and $u \in \sum_{k=1}^d \mathbb{R} e_k$. Then

$$\begin{aligned} & M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) \\ &= \left\{ i \in I' \;\middle|\; \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i) = \max_{i' \in I'} \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_{i'}) \right\} \\ &= \left\{ i \in I' \;\middle|\; -w' \cdot z_i + \deg_u(f_i^d) = \max_{i' \in I'} \{-w' \cdot z_{i'} + \deg_u(f_{i'}^d)\} \right\} \qquad \text{(by (4.36))} \\ &= \left\{ i \in I' \;\middle|\; \deg_u(f_i^d) = \max_{i' \in I'} \{\deg_u(f_{i'}^d)\} \right\} \qquad \text{(by (4.37))} \\ &= M_u(I', \boldsymbol{f}^d). \end{aligned}$$

Since the last $n - d$ entries of v are $\mathbb{Q}$-linearly independent, we have $a_i \notin O_v$ if and only if $a_i \in \sum_{k=1}^d \mathbb{Z} e_k$. Hence, $O'_u \cup J' = O_w \cup J'$ as in (4.33). Therefore,

$$\big(O_w \cup J'\big) \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) = \big(O'_u \cup J\big) \cap M_u(I', \boldsymbol{f}^d) \neq \emptyset.$$

If $w \notin \sum_{k=d+1}^{n} \mathbb{R}e_k$ and $d = 0$. We have $f_i^d \in \mathbb{R}$ for $i = 1, \ldots, K$, so

$$\begin{aligned} M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) &= \left\{ i \in I' \;\middle|\; \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_i) = \max_{i' \in I'} \deg_w(\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})_{i'}) \right\} \\ &= \left\{ i \in I' \;\middle|\; -w' \cdot z_i = \max_{i' \in I'} \{-w' \cdot z_{i'}\} \right\} \qquad \text{(by (4.36))} \\ &= I' \qquad \text{(by (4.37))} \end{aligned}$$

So $(O_w \cup J') \cap M_w(I', \mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{f})) = (O_w \cup J') \cap I' \supset J' \cap I' \neq \emptyset$.

This proves Property (4.27). □

4.7.4 Dimension reduction: the general case (LocInfD)

This subsection continues the work of the previous one. Our goal is to reduce Condition (LocInfShift) to a Condition (LocInfD) (which will be defined in Proposition 4.7.11). In the previous subsection we considered the special case where the vector $v \in (\mathbb{R}^n)^*$ in Condition (LocInfShift) is of the form $(0, \ldots, 0, v_{d+1}, \ldots, v_n)^\top$. In this subsection we consider the general case of v. The key idea when dealing with the general case is the following *coordinate change*.

Given a matrix $A = (a_{ij})_{1 \leq i,j \leq n} \in \mathsf{GL}(n, \mathbb{Z})$, define the new variables $X_1', \ldots, X_n'$ where $X_i' \coloneqq X_1^{a_{1i}} X_2^{a_{2i}} \cdots X_n^{a_{ni}}$. Then

$$\mathbb{R}[X_1, \ldots, X_n] = \mathbb{R}[X_1', \ldots, X_n'].$$

In other words, we can define the ring automorphism

$$\varphi_A \colon \mathbb{A} \to \mathbb{A}, \quad X_i \mapsto X_1^{a_{1i}} X_2^{a_{2i}} \cdots X_n^{a_{ni}},$$

such that $\varphi_A(\overline{X}^b) = \overline{X}^{Ab}$. The automorphism φ_A extends entry-wise to $\mathbb{A}^K \to \mathbb{A}^K$.

For each $A \in \mathsf{GL}(n, \mathbb{Z})$, denote by $A^{-\top}$ the inverse of its transpose. Then $(A^{-\top} v)^\top \cdot (Ab) = v^\top b$ for all $v \in (\mathbb{R}^n)^*, b \in \mathbb{Z}^n$. Hence, for any $f \in \mathbb{A}$ we have $\mathrm{in}_{A^{-\top} v}(\varphi_A(f)) = \varphi_A(\mathrm{in}_v(f))$, and for any $\boldsymbol{f} \in \mathbb{A}^K$ we have $M_v(I, \boldsymbol{f}) = M_{A^{-\top} v}(I, \varphi_A(\boldsymbol{f}))$. Furthermore, if we replace the vectors $a_1, \ldots, a_K \in \mathbb{Z}^n$ by the vectors $Aa_1, \ldots, Aa_K \in \mathbb{Z}^n$, then the set O_v becomes $O_{A^{-\top} v}$. It is easy to verify that if $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$ is a super Gröbner basis for $\mathcal{M}$, then $\varphi_A(\boldsymbol{g}_1), \ldots, \varphi_A(\boldsymbol{g}_m)$ is a super Gröbner basis for $\varphi_A(\mathcal{M}) \coloneqq \{\varphi_A(\boldsymbol{f}) \mid \boldsymbol{f} \in \mathcal{M}\}$.

Let $v \in (\mathbb{R}^n)^*$ and let $A \in \mathsf{GL}(n, \mathbb{Z})$ be such that $A^{-\top} v = (0, \ldots, 0, v_{d+1}, \ldots, v_n)^\top$ where $v_{d+1}, \ldots, v_n$ are $\mathbb{Q}$-linearly independent. Then as in the previous section we define the module $\mathrm{in}^d_{A^{-\top} v, \boldsymbol{\alpha}}(\varphi_A(\mathcal{M}))$ to be the module generated by $\mathrm{in}^d_{A^{-\top} v, \boldsymbol{\alpha}}(\varphi_A(\boldsymbol{g}_1)), \ldots, \mathrm{in}^d_{A^{-\top} v, \boldsymbol{\alpha}}(\varphi_A(\boldsymbol{g}_m))$.

The above observation shows the following. Fix $v \in (\mathbb{R}^n)^*$ in (LocInfShift) of Theorem 4.4.8. Given any change of coordinates $A \in \mathsf{GL}(n,\mathbb{Z})$, we can simultaneously multiply $A^{-\top}$ to v and multiply A to all $a_1,\dots,a_K$, while applying φ_A to the super Gröbner basis $\boldsymbol{g}_1,\dots,\boldsymbol{g}_m$ of $\mathcal{M}$. Then the original properties (LocInfShift)(a)(b) are satisfied by $\boldsymbol{f}$ if and only if they are satisfied by $\varphi_A(\boldsymbol{f})$ after the change of coordinates. We will use this observation to reduce the general case for v to the special case considered in the previous subsection.

Fact 4.7.10. For every vector $v \in (\mathbb{R}^n)^*$, there exists a matrix $A \in \mathsf{GL}(n,\mathbb{Z})$ such that $A^{-\top}v = (0,\dots,0,v_{d+1},\dots,v_n)^\top$ with $v_{d+1},\dots,v_n$ being $\mathbb{Q}$-linearly independent.

Proof. It suffices to show the following. If $v \in (\mathbb{R}^n)^*$ is of the form $(0,\dots,0,v_d,\dots,v_n)^\top$, $1 \le d \le n$ where $v_d,\dots,v_n$ are $\mathbb{Q}$-linearly *dependent*, then we can find a matrix $A_d \in \mathsf{GL}(n,\mathbb{Z})$ such that $A_d v = (0,\dots,0,v'_{d+1},\dots,v'_n)^\top$. Indeed, if this is true, then we can find a series of matrices $A_r,\dots,A_{r+s}$ such that $A_{r+s}\cdots A_{r+1}A_r v$ is of the form $(0,\dots,0,v'_{r+s+1},\dots,v'_n)^\top$ with $v'_{r+s+1},\dots,v'_n$ being $\mathbb{Q}$-linearly *independent*. We would then let $A \coloneqq (A_{r+s}\cdots A_{r+1}A_r)^{-\top}$.

Suppose now that $v = (0,\dots,0,v_d,\dots,v_n)^\top$, $1 \le d \le n$ where $v_d,\dots,v_n$ are $\mathbb{Q}$-linearly dependent. Let $z_d,\dots,z_n \in \mathbb{Q}$, not all zero, be such that $z_d v_d + \cdots + z_n v_n = 0$. Multiplying them by a common denominator we can suppose $z_d,\dots,z_n \in \mathbb{Z}$. Using Gaussian pivoting, we can find a matrix $\widetilde{A_d} \in \mathsf{GL}(n-d+1,\mathbb{Z})$ such that $(z_d,\dots,z_n)\widetilde{A_d} = (z,0,\dots,0)$ for some $z \in \mathbb{Z}^*$. Then we have

$$\begin{aligned} 0 = (z_d,\dots,z_n)\cdot(v_d,\dots,v_n)^\top = (z_d,\dots,z_n)\widetilde{A_d}\cdot\widetilde{A_d}^{-1}(v_d,\dots,v_n)^\top \\ = (z,0,\dots,0)\cdot\widetilde{A_d}^{-1}(v_d,\dots,v_n)^\top. \end{aligned}$$

Therefore $\widetilde{A_d}^{-1}(v_d,\dots,v_n)^\top$ is of the form $(0,v'_{d+1},\dots,v'_n)^\top$. We then let $A_d \coloneqq diag(I_{d-1},\widetilde{A_d}^{-1})$. That is, A_d is the block diagonal matrix consisting of the block I_{d-1} of $(d-1)$-dimensional identity matrix and the block $\widetilde{A_d}^{-1}$ of $(n-d+1)$-dimensional matrix. Then $A_d v = (0,\dots,0,v'_{d+1},\dots,v'_n)^\top$. □

Proposition 4.7.11. *Condition* (LocInfShift) *of Proposition 4.7.4 is equivalent to the following:*

2. **(LocInfD):** *For every* $v \in (\mathbb{R}^n)^*$, $A \in \mathsf{GL}(n,\mathbb{Z})$, *such that* $A^{-\top}v = (0,\dots,0,v_{d+1},\dots,v_n)^\top$, $0 \le d \le n-1$ *where* $v_{d+1},\dots,v_n$ *are* $\mathbb{Q}$*-linearly independent, there exist* $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$ *and* $\boldsymbol{f}^d \in \mathrm{in}^d_{A^{-\top}v,\boldsymbol{\alpha}}(\varphi_A(\mathcal{M}))$ *satisfying the following properties:*

   (a) $\boldsymbol{f}^d \in \left(\mathbb{A}^+_d\right)^K$.

(b1) Denote $I' \coloneqq \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\}, J' \coloneqq O_v \cup J$, *we have*

$$J' \cap I' \neq \emptyset. \tag{4.38}$$

(b2) Denote by $\pi_d \coloneqq \mathbb{Z}^n \to \mathbb{Z}^d$ *the projection onto the first* d *coordinates. For* $u \in \left(\mathbb{R}^d\right)^*$, *define* $O'_u \coloneqq \{i \in \{1, \dots, K\} \mid \pi_d(Aa_i) \not\perp u\}$, *we have*

$$\left(O'_u \cup J'\right) \cap M_u(I', \boldsymbol{f}^d) \neq \emptyset \quad \text{for every } u \in \left(\mathbb{R}^d\right)^*. \tag{4.39}$$

Same as in Lemma 4.7.9, the Property (4.39) *is considered trivially true when* $d = 0$.

Proof. Fix a vector $v = (v_1, \dots, v_n)^\top \in (\mathbb{R}^n)^*$ in Condition (LocInfShift). Take any matrix $A \in \mathsf{GL}(n, \mathbb{Z})$ such that $A^{-\top} v = (0, \dots, 0, v'_{d+1}, \dots, v'_n)^\top$ where $v'_{d+1}, \dots, v'_n$ are $\mathbb{Q}$-linearly independent. Note that $\sum_{k=1}^n \mathbb{Z} v_k = \sum_{k=d+1}^n \mathbb{Z} v'_k$ because $A \in \mathsf{GL}(n, \mathbb{Z})$. Therefore, we can apply Lemma 4.7.9 to the super Gröbner basis $\varphi_A(\boldsymbol{g}_1), \dots, \varphi_A(\boldsymbol{g}_m)$, the vector $A^{-\top} v = (0, \dots, 0, v'_{d+1}, \dots, v'_n)^\top$ as well as the vectors $Aa_1, \dots, Aa_K \in \mathbb{Z}^n$. Lemma 4.7.9 shows that there exist $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^n \mathbb{Z} v'_k\right)^K$ and $\boldsymbol{f}^d \in \mathrm{in}^d_{A^{-\top} v, \boldsymbol{\alpha}}(\varphi_A(\mathcal{M}))$ satisfying (LocInfD)(a)(b1)(b2) if and only if there exists $\boldsymbol{\alpha} \in \left(\sum_{k=1}^n \mathbb{Z} v_k\right)^K$ and $\boldsymbol{f} \in \mathcal{M}$ satisfying (LocInfShift)(a)(b). □

4.7.5 Local condition at infinity: computing cells (LocInfCell)

In this subsection we further reduce the Condition (LocInfD) to a Condition (LocInfCell) which consists of verifying a *finite* number of $v \in (\mathbb{R}^n)^*$ for each coordinate-change matrix $A \in \mathsf{GL}(n, \mathbb{Z})$.

Let $v \in (\mathbb{R}^n)^*, \boldsymbol{\alpha} \in \mathbb{R}^K$. Denote by $e_1, \dots, e_K$ the canonical basis of the $\mathbb{A}$-module $\mathbb{A}^K$. We introduce the new variables $T_1, \dots, T_K$ and define an $\mathbb{A}$-module homomorphism

$$\phi : \mathbb{A}^K \to \mathbb{R}[X_1^{\pm}, \dots, X_n^{\pm}, T_1^{\pm}, \dots, T_K^{\pm}], \quad \overline{X}^u e_i \mapsto \overline{X}^u T_i.$$

We have $\phi(\mathrm{in}_{v, \boldsymbol{\alpha}}(\boldsymbol{f})) = \mathrm{in}_{(v, \boldsymbol{\alpha})}(\phi(\boldsymbol{f}))$ for every $\boldsymbol{f} \in \mathbb{A}^K$.

As in the previous subsections, let $\boldsymbol{g}_1, \dots, \boldsymbol{g}_m$ be a super Gröbner basis of $\mathcal{M}$. For each i since $\phi(\boldsymbol{g}_i)$ is a polynomial in $\mathbb{R}[X_1^{\pm}, \dots, X_n^{\pm}, T_1^{\pm}, \dots, T_K^{\pm}]$, there exists a partition of $(\mathbb{R}^n)^* \times \mathbb{R}^K$ such that for any two directions w, w' in the same partition element the initial parts $\mathrm{in}_w(\phi(\boldsymbol{g}_i)), \mathrm{in}_{w'}(\phi(\boldsymbol{g}_i))$ are the same. Let $\mathcal{L}_{\mathcal{M}}$ be the common refinement of the partitions associated to the polynomials $\phi(\boldsymbol{g}_1), \dots, \phi(\boldsymbol{g}_m)$.

From now on we use the term "*cell*" to call an element of a given partition. Fix $I \subseteq \{1, \dots, K\}$. There exists a partition $\mathcal{L}_I$ of $\mathbb{R}^K$ such that for any two vectors $(\alpha_1, \dots, \alpha_K)$,

$(\alpha'_1, \ldots, \alpha'_K)$ in the same cell, we have $\alpha_i > \alpha_j \iff \alpha'_i > \alpha'_j$ and $\alpha_i < \alpha_j \iff \alpha'_i < \alpha'_j$ for all $i, j \in I$. Define the partition $\mathcal{L}'_I \coloneqq (\mathbb{R}^n)^* \times \mathcal{L}_I$ of $(\mathbb{R}^n)^* \times \mathbb{R}^K$ where each cell is of the form $(\mathbb{R}^n)^* \times P, P \in \mathcal{L}_I$.

There exists a partition $\mathcal{L}_O$ of $(\mathbb{R}^n)^*$ such that any two vectors v, v' in the same cell satisfy $v \perp a_i \iff v' \perp a_i$ for all $i \in \{1, \ldots, K\}$. Similar to the definition of $\mathcal{L}'_I$, we define the partition $\mathcal{L}'_O \coloneqq \mathcal{L}_O \times \mathbb{R}^K$ of $(\mathbb{R}^n)^* \times \mathbb{R}^K$.

For any two partition $\mathcal{A}, \mathcal{B}$ of the same set S, define $\mathcal{A} \vee \mathcal{B}$ to be the partition of S whose elements are of the form $A \cap B, A \in \mathcal{A}, B \in \mathcal{B}$. Consider the partition $\mathcal{L}$ of $(\mathbb{R}^n)^* \times \mathbb{R}^K$ defined by

$$\mathcal{L} \coloneqq \mathcal{L}_{\mathcal{M}} \vee \mathcal{L}'_I \vee \mathcal{L}'_O.$$

We point out that from the definition of the partitions $\mathcal{L}_{\mathcal{M}}, \mathcal{L}'_I, \mathcal{L}'_O$, the cells of $\mathcal{L}$ are invariant under scaling by a positive real, meaning $x \in Q \iff r \cdot x \in Q$ for all cells $Q \in \mathcal{L}$ and $r \in \mathbb{R}_{>0}$. By subdividing $\mathcal{L}$ we can suppose that each cell is a convex polyhedron invariant by $\mathbb{R}_{>0}$-scaling.

Let $\pi\colon (\mathbb{R}^n)^* \times \mathbb{R}^K \to (\mathbb{R}^n)^*, (v, \boldsymbol{\alpha}) \mapsto v$ be the canonical projection. For each $Q \in \mathcal{L}$, define the two-element partition $\{\pi(Q), (\mathbb{R}^n)^* \setminus \pi(Q)\}$ of $(\mathbb{R}^n)^*$. Define the following partition of $(\mathbb{R}^n)^*$:

$$\mathcal{P} \coloneqq \bigvee_{Q \in \mathcal{L}} \{\pi(Q), (\mathbb{R}^n)^* \setminus \pi(Q)\}.$$

By this definition, take any $P \in \mathcal{P}$ and $Q \in \mathcal{L}$ with $\pi^{-1}(P) \cap Q \neq \emptyset$, then for $v, v' \in P$, there exists $\boldsymbol{\alpha} \in \mathbb{R}^K$ with $(v, \boldsymbol{\alpha}) \in Q$ if and only if there exists $\boldsymbol{\alpha}' \in \mathbb{R}^K$ with $(v', \boldsymbol{\alpha}') \in Q$. See Figure 4.29 for an illustration.

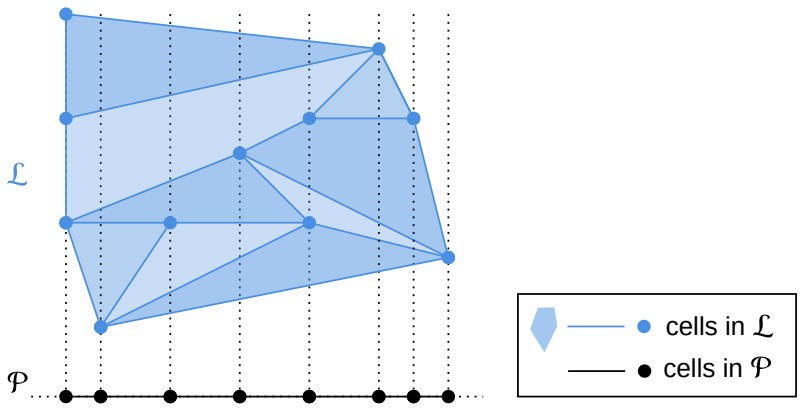

Figure 4.29: Example of the partitions $\mathcal{L}$ and $\mathcal{P}$.

It is important to note that the partitions $\mathcal{L}_{\mathcal{M}}, \mathcal{L}_I, \mathcal{L}_O$ are all defined using equalities and

inequalities with *rational* coefficients. Also, each inequality is strict, so every cell $Q \in \mathcal{L}$ and $P \in \mathcal{P}$ is relatively open (a polyhedron is called relative open if it is open in the smallest linear space containing it). In other words, each cell is defined by a combination of equalities and *strict* inequalities. We also point out that, like the cells of $\mathcal{L}$, the cell of $\mathcal{P}$ are invariant under scaling by a positive real, meaning $x \in P \iff r \cdot x \in P$ for all cells $P \in \mathcal{P}$ and $r \in \mathbb{R}_{>0}$.

For any coordinate change $A \in \mathsf{GL}(n, \mathbb{Z})$, we similarly define the partitions $A^{-\top}\mathcal{L}$ and $A^{-\top}\mathcal{P}$ based on the super Gröbner basis $\varphi_A(\boldsymbol{g}_1), \dots, \varphi_A(\boldsymbol{g}_m)$ and the vectors $Aa_1, \dots, Aa_K$. In particular, each cell of $A^{-\top}\mathcal{L}$ is of the form $diag(A^{-\top}, I_K) \cdot Q, Q \in \mathcal{L}$, and each cell of $A^{-\top}\mathcal{P}$ is of the form $A^{-\top} \cdot P, P \in \mathcal{P}$. By the definition of $\mathcal{L}$ and $A^{-\top}\mathcal{L}$, we immediately obtain the following.

Lemma 4.7.12. *Fix a change of coordinates $A \in \mathsf{GL}(n, \mathbb{Z})$, two sets $I, J \subseteq \{1, \dots, K\}$, and a cell $Q \in A^{-\top}\mathcal{L}$. Then the sets I', J' defined in* (LocInfD)*(b1) are effectively computable and do not depend on $v, \boldsymbol{\alpha}$ as long as $(A^{-\top}v, \boldsymbol{\alpha}) \in Q$. In particular, the Property* (LocInfD)*(b1) is either always true or always false for $v, \boldsymbol{\alpha}, (A^{-\top}v, \boldsymbol{\alpha}) \in Q$.*

Proof. By the definition of the partition $\mathcal{L}_I$, the set $I' = \{i \in I \mid \alpha_i = \min_{i' \in I} \alpha_{i'}\}$ in Property (LocInfD)(b1) only depends on which cell of $\mathcal{L}_I$ contains $\boldsymbol{\alpha}$. Hence I' only depends on the cell of $A^{-\top}\mathcal{L}$ containing $(A^{-\top}v, \boldsymbol{\alpha})$. Similarly, by the definition of the partition $\mathcal{L}_O$, the set O_v only depends on which cell of $\mathcal{L}_O$ contains v. Hence $J' = O_v \cup J$ only depends on the cell of $A^{-\top}\mathcal{L}$ containing $(A^{-\top}v, \boldsymbol{\alpha})$. The sets I', J' can be computed by taking an arbitrary element $(v, \boldsymbol{\alpha}) \in Q$ with rational entries. Therefore, the truth of Property (LocInfD)(b1) only depends on the cell of $A^{-\top}\mathcal{L}$ containing $(A^{-\top}v, \boldsymbol{\alpha})$. □

Let $Q \in \mathcal{L}$. For $(v, \boldsymbol{\alpha}), (v', \boldsymbol{\alpha}') \in Q$, we have

$$\mathrm{in}_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j) = \mathrm{in}_{v',\boldsymbol{\alpha}'}(\boldsymbol{g}_j)$$

for all $j = 1, \dots, m$. Thus, if $v = (0, \dots, 0, v_{d+1}, \dots, v_n)^\top$ is such that $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent, then $\mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M})$ depends only on the cell $Q \in \mathcal{L}$ containing $(v, \boldsymbol{\alpha})$. Hence, we can denote

$$\mathrm{in}^d_Q(\boldsymbol{g}_j) \coloneqq \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\boldsymbol{g}_j), \quad j = 1, \dots, m, \quad \mathrm{in}^d_Q(\mathcal{M}) \coloneqq \mathrm{in}^d_{v,\boldsymbol{\alpha}}(\mathcal{M}), \quad \text{where } (v, \boldsymbol{\alpha}) \in Q.$$

If $A^{-\top}v = (0, \dots, 0, v_{d+1}, \dots, v_n)^\top$ is such that $v_{d+1}, \dots, v_n$ are $\mathbb{Q}$-linearly independent, then $\mathrm{in}^d_{A^{-\top}v,\boldsymbol{\alpha}}(\varphi_A(\mathcal{M}))$ depends only on the cell $Q \in A^{-\top}\mathcal{L}$ containing $(A^{-\top}v, \boldsymbol{\alpha})$. Similarly,

for $j = 1, \ldots, m$, we can denote

$$\mathrm{in}_Q^d(\varphi_A(\boldsymbol{g}_j)) \coloneqq \mathrm{in}_{A^{-\top}v,\boldsymbol{\alpha}}^d(\varphi_A(\boldsymbol{g}_j)),\ \mathrm{in}_Q^d(\varphi_A(\mathcal{M})) \coloneqq \mathrm{in}_{A^{-\top}v,\boldsymbol{\alpha}}^d(\varphi_A(\mathcal{M})),\ \text{ where } (A^{-\top}v, \boldsymbol{\alpha}) \in Q.$$

The inputs in Theorem 4.4.10 are generators for modules $\mathcal{M}$ over $\mathbb{A} = \mathbb{R}[X_1^{\pm}, \ldots, X_n^{\pm}]$, vectors $a_1, \ldots, a_K$ in $\mathbb{Z}^n$ and two sets I, J. Our strategy is to use induction on n to prove Theorem 4.4.10. The base case $n = 0$ reduces to linear programming. Indeed, when $n = 0$, $\mathbb{A} = \mathbb{R}, \mathbb{A}^+ = \mathbb{R}_{>0}$, the Property (4.14) is trivially true; and the problem becomes the following: given an $\mathbb{R}$-submodule $\mathcal{M}$ of $\mathbb{R}^K$, decide whether $\mathcal{M}$ contains an element of $\mathbb{R}^K_{>0}$. Since the given generators of $\mathcal{M}$ all have integer coefficients, this is decidable using linear programming.

The following lemma shows that a decision procedure for Theorem 4.4.10 with smaller n can help us decide for a given cell $Q \in \mathcal{L}$ if the module $\mathrm{in}_Q^d(\varphi_A(\mathcal{M}))$ contains an $\boldsymbol{f}^d$ satisfying the Properties (LocInfD)(a) and (b2).

Lemma 4.7.13. *Fix a change of coordinates $A \in \mathsf{GL}(n, \mathbb{Z})$, two sets $I, J \subseteq \{1, \ldots, K\}$, and a number $0 \le d \le n-1$. Suppose Theorem 4.4.10 is true for all n_0, $0 \le n_0 \le n-1$. Fix a cell $Q \in A^{-\top}\mathcal{L}$, let I', J' be the pair of sets defined in* (LocInfD)*(b1). We can decide whether the module $\mathrm{in}_Q^d(\varphi_A(\mathcal{M}))$ contains an element $\boldsymbol{f}^d$ satisfying the Properties* (LocInfD)*(a) and (b2).*

Proof. Suppose Theorem 4.4.10 is true for all $0 \le n_0 \le n-1$. In particular it is true for $d \le n-1$. Fix a cell $Q \in A^{-\top}\mathcal{L}$. In (LocInfD), the $\mathbb{A}_d$-submodule $\mathrm{in}_{A^{-\top}v,\boldsymbol{\alpha}}^d(\varphi_A(\mathcal{M})) = \mathrm{in}_Q^d(\varphi_A(\mathcal{M}))$ of $\mathbb{A}_d^K$ is generated by the elements $\mathrm{in}_Q^d(\varphi_A(\boldsymbol{g}_j)), j = 1, \ldots, m$.

We then apply Theorem 4.4.10 the following way: replace n by d; replace the elements $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathbb{A}^K$ by the elements $\mathrm{in}_Q^d(\varphi_A(\boldsymbol{g}_1)), \ldots, \mathrm{in}_Q^d(\varphi_A(\boldsymbol{g}_m)) \in \mathbb{A}_d^K$; replace the vectors $a_1, \ldots, a_K \in \mathbb{Z}^n$ by the vectors $\pi_d(Aa_1), \ldots, \pi_d(Aa_K) \in \mathbb{Z}^d$; and replace the sets I, J by the sets I', J'. Then Theorem 4.4.10 shows we can decide whether $\mathrm{in}_Q^d(\varphi_A(\mathcal{M}))$ contains an element $\boldsymbol{f}^d$ satisfying $\boldsymbol{f}^d \in \left(\mathbb{A}_d^+\right)^K$ and

$$\left(O'_u \cup J'\right) \cap M_u(I', \boldsymbol{f}^d) \neq \emptyset, \quad \text{for every } u \in \left(\mathbb{R}^d\right)^*.$$

These are exactly the Properties (LocInfD)(a) and (b2). □

Denote by $Op(A, d)$ the union of all cells $Q \in A^{-\top}\mathcal{L}$ such that the Property (LocInfD)(b1) is true for $(A^{-\top}v, \boldsymbol{\alpha}) \in Q$, and such that $\mathrm{in}_Q^d(\varphi_A(\mathcal{M}))$ contains an element $\boldsymbol{f}^d$ satisfying the Properties (LocInfD)(a)(b2). By Lemma 4.7.12 and 4.7.13, the set $Op(A, d)$ is effectively computable as a finite union of polyhedra defined over rational coefficients (supposing Theorem 4.4.10 is true for all $0 \le n_0 \le n-1$). See Figure 4.30 for an illustration of $Op(A, d)$.

Proposition 4.7.14. *Condition* (LocInfD) *of Proposition 4.7.11 is equivalent to the following:*

2. **(LocInfCell):** *For every* $A \in \mathsf{GL}(n,\mathbb{Z})$ *and every number* $0 \leq d \leq n-1$*, the following is true:*

   (a) *For every* $v = (0,\dots,0,v_{d+1},\dots,v_n)^\top \in \{0\}^d \times (\mathbb{R}^{n-d})^*$ *with* $v_{d+1},\dots,v_n$ *being* $\mathbb{Q}$*-linearly independent, there exists* $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^{n} \mathbb{Z}v_k\right)^K$ *with* $(v,\boldsymbol{\alpha}) \in Op(A,d)$*.*

Proof. This follows directly from the definition of $Op(A,d)$. □

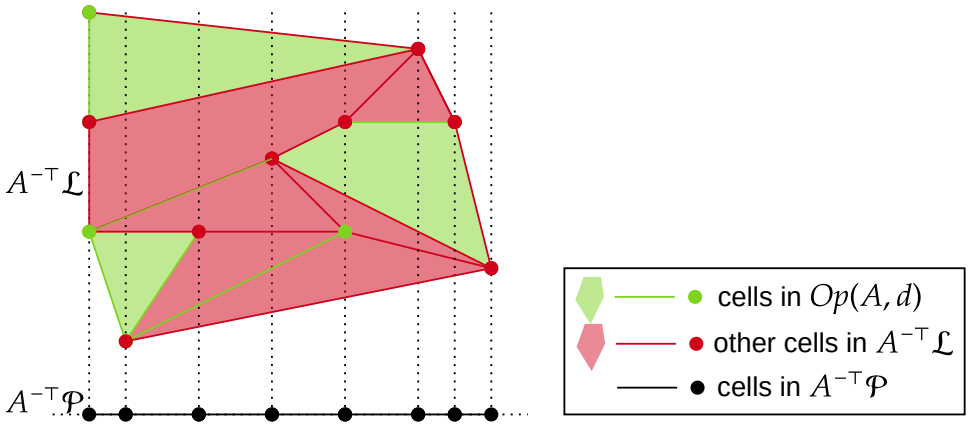

Figure 4.30: Example of $Op(A,d)$.

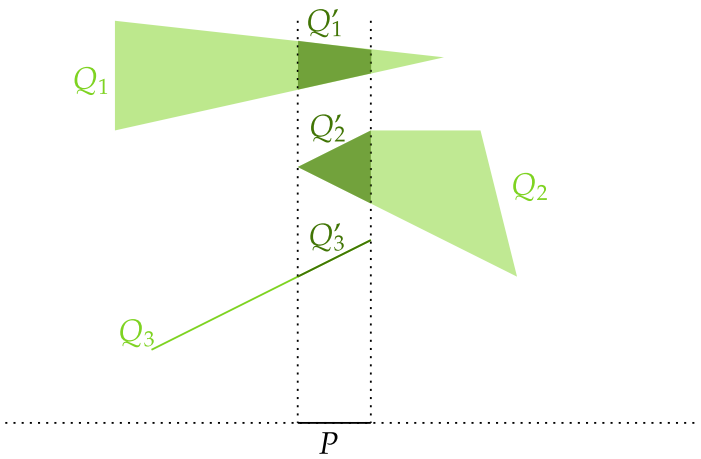

Figure 4.31: Illustration of Lemma 4.7.15

Lemma 4.7.15. *Given* $A \in \mathsf{GL}(n,\mathbb{Z}), d \in \mathbb{N}$ *and given* $Op(A,d)$ *as a finite union of polyhedra defined over rational coefficients, it is decidable whether the statement* (LocInfCell)(a) *is true.*

Proof. Replace each cell Q in $A^{-\top}\mathcal{L}$ with its intersection with $\{0\}^d \times (\mathbb{R}^{n-d})^* \times \mathbb{R}^K$; and replace each cell $P \in A^{-\top}\mathcal{P}$ with its intersection with $\{0\}^d \times (\mathbb{R}^{n-d})^*$. We can suppose $A^{-\top}\mathcal{P}$ is a partition of $(\mathbb{R}^{n-d})^*$, $A^{-\top}\mathcal{L}$ is a partition of $(\mathbb{R}^{n-d})^* \times \mathbb{R}^K$, $Op(A,d) \subseteq (\mathbb{R}^{n-d})^* \times \mathbb{R}^K$ is a union of cells in $A^{-\top}\mathcal{L}$, and that $v \in \left(\mathbb{R}^{n-d}\right)^*$. We separate two cases.

(1) **When $d = n-1$.** In this case, since the partitions are invariant under scaling, we can suppose $v_n \in \{1, -1\}$. Then for each case $v_n = 1$ and $v_n = -1$, decide whether there exists $\boldsymbol{\alpha} \in (\mathbb{Z}v_n)^K = \mathbb{Z}^K$ with $(v, \boldsymbol{\alpha}) \in Op(A,d)$. Since $Op(A,d)$ is a finite union of polyhedra defined using rational coefficients, this is decidable using integer programming.

(2) **When $d \leq n-2$.** See Figure 4.31 for an illustration in this case. Whenever $v = (v_{d+1}, \ldots, v_n)^\top$ with $v_{d+1}, \ldots, v_n$ being $\mathbb{Q}$-linearly independent, v must fall in a cell $P \in A^{-\top}\mathcal{P}$ of full dimension (dimension $n-d$). For each cell $P \in A^{-\top}\mathcal{P}$ of dimension $n-d$, consider all cells $Q \subseteq Op(A,d)$ such that $\pi(Q) \cap P \neq \emptyset$. If there is no such cell Q then statement (LocInfCell)(a) is false. Indeed, in this case, since $P \in A^{-\top}\mathcal{P}$ is of dimension $n-d$, it contains an element $v = (v_{d+1}, \ldots, v_n)^\top$ with $v_{d+1}, \ldots, v_n$ being $\mathbb{Q}$-linearly independent. Then for this v, there does not exist any $\boldsymbol{\alpha} \in \mathbb{R}^K$ such that $(v, \boldsymbol{\alpha}) \in Op(A,d)$, so statement (LocInfCell)(a) is false.

Suppose now that for every cell $P \in \mathcal{P}$ of dimension $n-d$ there exist cells $Q \subseteq Op(A,d)$ such that $\pi(Q) \cap P \neq \emptyset$. Fix a cell $P \in \mathcal{P}$, let $Q_1, \ldots, Q_\ell$ denote all cells in $Op(A,d)$ such that $\pi(Q) \cap P \neq \emptyset$. Define $Q'_t \coloneqq Q_t \cap \pi^{-1}(P)$ for $t = 1, \ldots, \ell$. Each Q'_t is a relatively open polyhedron.

Take an arbitrary Q'_t, it is defined by the following equations and inequalities:

$$(v_{d+1}, \ldots, v_n)^\top \in P \tag{4.40}$$

$$\beta_{j,1}\alpha_1 + \cdots + \beta_{j,K}\alpha_K = \gamma_{j,d+1}v_{d+1} + \cdots + \gamma_{j,n}v_n, \quad j = 1, \ldots, m_0, \tag{4.41}$$

$$\delta_{j,1}\alpha_1 + \cdots + \delta_{j,K}\alpha_K < \epsilon_{j,d+1}v_{d+1} + \cdots + \epsilon_{j,n}v_n, \quad j = 1, \ldots, m_1. \tag{4.42}$$

Where $\beta_{j,i}, \gamma_{j,i}, \delta_{j,i}, \epsilon_{j,i}$ are all rational numbers. Note that by the definition of P, for $v, v' \in P$, there exists $\boldsymbol{\alpha} \in \mathbb{R}^K$ with $(v, \boldsymbol{\alpha}) \in Q$ if and only if there exists $\boldsymbol{\alpha}' \in \mathbb{R}^K$ with $(v', \boldsymbol{\alpha}') \in Q$. Therefore $\pi(Q'_t) = P$ and we can suppose that the left hand sides of (4.41) and (4.42) do not vanish (so that no extra constraint on $(v_{d+1}, \ldots, v_n)^\top$ other than (4.40) is imposed).

Using Gaussian pivoting and possibly exchanging the orders of $\alpha_i, i = 1, \ldots, K$, we can

rewrite the above equations and inequalities into the "reduced echelon" form

$$(v_{d+1},\dots,v_n)^\top \in P \tag{4.43}$$

$$\alpha_i = \beta_{i,D+1}\alpha_{D+1}+\cdots+\beta_{i,K}\alpha_K+\gamma_{i,d+1}v_{d+1}+\cdots+\gamma_{i,n}v_n,\quad i=1,\dots,D, \tag{4.44}$$

$$\delta_{j,D+1}\alpha_{D+1}+\cdots+\delta_{j,K}\alpha_K < \epsilon_{j,d+1}v_{d+1}+\cdots+\epsilon_{j,n}v_n,\quad j=1,\dots,m_1. \tag{4.45}$$

In particular, the number $D\in\mathbb{N}$ is such that Q'_t is a polyhedron of dimension $n-d+K-D$. Let $M\in\mathbb{N}$ be a common denominator of all $\beta_{i,j},\gamma_{i,k}, i=1,\dots,D, j=D+1,\dots,K, k=d+1,\dots,n$. We multiply both sides of the Equations (4.44) by M, and suppose Q'_t is defined by

$$(v_{d+1},\dots,v_n)^\top \in P \tag{4.46}$$

$$M\alpha_i = \beta_{i,D+1}\alpha_{D+1}+\cdots+\beta_{i,K}\alpha_K+\gamma_{i,d+1}v_{d+1}+\cdots+\gamma_{i,n}v_n,\quad i=1,\dots,D, \tag{4.47}$$

$$\delta_{j,D+1}\alpha_{D+1}+\cdots+\delta_{j,K}\alpha_K < \epsilon_{j,d+1}v_{d+1}+\cdots+\epsilon_{j,n}v_n,\quad j=1,\dots,m_1, \tag{4.48}$$

where $\beta_{i,j},\gamma_{i,k}, i=1,\dots,D, j=D+1,\dots,K, k=d+1,\dots,n$, are *integers*.
Fix any $v=(v_{d+1},\dots,v_n)^\top\in P$ with $v_{d+1},\dots,v_n$ being $\mathbb{Q}$-linearly independent. We claim the following. There exists $\boldsymbol{\alpha}\in\left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$ such that $(v,\boldsymbol{\alpha})\in Q'_t$, if and only if the following system of $(n-d)D$ equations has integer solutions $z_{i,k}, i=1,\dots,D, D+1,\dots,K, k=d+1,\dots,n$.

$$\begin{aligned}
Mz_{i,d+1} &= \beta_{i,D+1}z_{D+1,d+1}+\cdots+\beta_{i,K}z_{K,d+1}+\gamma_{i,d+1},\quad i=1,\dots,D,\\
Mz_{i,d+2} &= \beta_{i,D+1}z_{D+1,d+2}+\cdots+\beta_{i,K}z_{K,d+2}+\gamma_{i,d+2},\quad i=1,\dots,D,\\
&\vdots\\
Mz_{i,n} &= \beta_{i,D+1}z_{D+1,n}+\cdots+\beta_{i,K}z_{K,n}+\gamma_{i,n},\quad i=1,\dots,D.
\end{aligned} \tag{4.49}$$

We now prove this claim. For the "only if" implication, suppose there exists an $\boldsymbol{\alpha}\in\left(\sum_{k=d+1}^n \mathbb{Z}v_k\right)^K$ such that $(v,\boldsymbol{\alpha})\in Q'_t$. For each $i\in\{1,\dots,K\}$, write $\alpha_i=\sum_{k=d+1}^n z_{i,k}v_k$, then the Equations (4.47) become

$$M\sum_{k=d+1}^n z_{i,k}v_k = \beta_{i,D+1}\sum_{k=d+1}^n z_{D+1,k}v_k+\cdots+\beta_{i,K}\sum_{k=d+1}^n z_{K,k}v_k+\gamma_{i,d+1}v_{d+1}+\cdots+\gamma_{i,n}v_n,$$
$$i=1,\dots,D. \tag{4.50}$$

Since $v_{d+1},\dots,v_n$ are $\mathbb{Q}$-linearly independent, Equations (4.50) hold if and only if for all

$k = d+1, \ldots, n$, the coefficients of v_k on both sides are equal. That is,

$$Mz_{i,k} = \beta_{i,D+1} z_{D+1,k} + \cdots + \beta_{i,K} z_{i,k} + \gamma_{i,k}, \quad i = 1, \ldots, D, \quad k = d+1, \ldots, n.$$

This is exactly the system (4.49).

For the "if" implication, suppose the system of equations (4.49) has integer solutions $z_{i,k}, i = 1, \ldots, D, D+1, \ldots, K, k = d+1, \ldots, n$.

For each tuple of integers $c_{i,k}, i = D+1, \ldots, K, k = d+1, \ldots, n$, we can construct a new solution of the system (4.49) by letting $z'_{i,k} \coloneqq z_{i,k} + Mc_{i,k}, i = D+1, \ldots, K, k = d+1, \ldots, n$, and

$$\begin{aligned}
z'_{i,d+1} &\coloneqq z_{i,d+1} + \beta_{i,D+1} c_{D+1,d+1} + \cdots + \beta_{i,K} c_{K,d+1}, \quad i = 1, \ldots, D, \\
z'_{i,d+2} &= z_{i,d+2} + \beta_{i,D+1} c_{D+1,d+2} + \cdots + \beta_{i,K} c_{K,d+2}, \quad i = 1, \ldots, D, \\
&\vdots \\
z'_{i,n} &= z_{i,n} + \beta_{i,D+1} c_{D+1,n} + \cdots + \beta_{i,K} c_{K,n}, \quad i = 1, \ldots, D.
\end{aligned} \tag{4.51}$$

Since $z'_{i,k}, i = 1, \ldots, D, D+1, \ldots, K, k = d+1, \ldots, n$, is a solution for (4.49), it is easy to verify that $\alpha_i \coloneqq \sum_{k=d+1}^{n} z'_{i,k} v_k, i = 1, \ldots, K$ constitute a solution for (4.47). We now show that for every tuple $(v_{d+1}, \ldots, v_n)^\top \in P$, we can actually find integers $c_{i,k}, i = D+1, \ldots, K, k = d+1, \ldots, n$ such that $\alpha_i \coloneqq \sum_{k=d+1}^{n} (z_{i,k} + Mc_{i,k}) v_k, i = D+1, \ldots, n$, satisfy also (4.48).

Since Q'_t is relatively open and non-empty, the set $\pi^{-1}((v_{d+1}, \ldots, v_n)^\top) \cap Q'_t$ is also non-empty. Therefore, the solution set $\mathcal{A}$ for $(\alpha_{D+1}, \ldots, \alpha_K) \in \mathbb{R}^{K-D}$ of the inequalities (4.48) is non-empty. This solution set $\mathcal{A}$ is an open subset of $\mathbb{R}^{K-D}$ since it is define by strict inequalities. Since $v_{d+1}, \ldots, v_n$ are $\mathbb{Q}$-linearly independent, the set

$$\left\{ \alpha_i \coloneqq \sum_{k=d+1}^{n} (z_{i,k} + Mc_{i,k}) v_k \;\middle|\; c_{i,d+1}, \ldots, c_{i,n} \in \mathbb{Z} \right\}$$

is dense in $\mathbb{R}$ for every $i \in \{D+1, \ldots, K\}$. Thus we can find $c_{D+1,d+1}, \ldots, c_{D+1,n} \in \mathbb{Z}$ such that $\alpha_{D+1} \coloneqq \sum_{k=d+1}^{n} (z_{D+1,k} + Mc_{D+1,k}) v_k$ satisfies $(\alpha_{D+1}, x_{D+2}, \ldots, x_n) \in \mathcal{A}$ for some $x_{D+2}, \ldots, x_n \in \mathbb{R}$. Similarly, by the openness of $\mathcal{A}$, we can then find $c_{D+2,d+1}, \ldots, c_{D+2,n} \in \mathbb{Z}$ such that $\alpha_{D+2} \coloneqq \sum_{k=d+1}^{n} (z_{D+2,k} + Mc_{D+2,k}) v_i$ satisfies $(\alpha_{D+1}, \alpha_{D+2}, x_{D+3}, \ldots, x_n) \in \mathcal{A}$ for some $x_{D+3}, \ldots, x_n \in \mathbb{R}$. Continue this way and we will find integers $c_{i,k}, i = D+1, \ldots, K, k = d+1, \ldots, n$ such that $(\alpha_{D+1}, \ldots, \alpha_n) \in \mathcal{A}$. This tuple $(\alpha_{D+1}, \ldots, \alpha_n)$ satisfies (4.48). Since $\alpha_i = \sum_{k=d+1}^{n} z'_{i,k} v_k, i = 1, \ldots, K$, is a solution for (4.47) regardless

of the choice of $c_{i,k}$, both (4.47) and (4.48) are now satisfied. We have proved the "if" implication.

Note that whether the system (4.49) has integer solutions depend only on the coefficients $\beta_{i,k}, \gamma_{i,k}$, $i = 1, \ldots, D, j = D+1, \ldots, K, k = d+1, \ldots, n$. These coefficients are determined by the polyhedron Q'_t, but not on the choice of v. For each $Q'_t, t = 1, \ldots, \ell$, we can decide whether its system (4.49) has integer solutions (by integer programming, or more efficiently, by considering solutions modulo M). If for some $t \in \{1, \ldots, \ell\}$, its system (4.49) has integer solutions, then the above claim shows that for all $v = (v_{d+1}, \ldots, v_n)^\top \in P$ with $v_{d+1}, \ldots, v_n$ being $\mathbb{Q}$-linearly independent, there exists $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^{n} \mathbb{Z} v_k\right)^K$ such that $(v, \boldsymbol{\alpha}) \in Q'_t$. Otherwise, if for all $t \in \{1, \ldots, \ell\}$, its system (4.49) has no integer solutions, then for any $v = (v_{d+1}, \ldots, v_n)^\top \in P$ with $v_{d+1}, \ldots, v_n$ being $\mathbb{Q}$-linearly independent, there does not exist $\boldsymbol{\alpha} \in \left(\sum_{k=d+1}^{n} \mathbb{Z} v_k\right)^K$ such that $(v, \boldsymbol{\alpha}) \in Q'_t$.

To summarize, in order to decide whether the statement (LocInfCell)(a) is true, it suffices to enumerate all cells $P \in A^{-\top}\mathcal{P}$ of dimension $n - d$. For a cell P, if there is no cell $Q \subseteq Op(A, d)$ such that $\pi(Q) \cap P \neq \emptyset$. then statement (LocInfCell)(a) is false. Otherwise, for each $Q'_1, \ldots, Q'_\ell$, check whether system (4.49) has integer solutions; in case an integer solution exists for some Q'_t, we call the cell P "operational". If every cell $P \in \mathcal{P}$ of dimension $n-d$ is operational, then statement (LocInfCell)(a) is true, otherwise it is false.

□

4.7.6 Proving Theorem 4.4.10: induction and a double procedure

In this subsection we give the full proof of Theorem 4.4.10. The overall strategy is to use induction on n, while deciding the Conditions (LocR) and (LocInf) from the local-global principle (Theorem 4.4.8).

Theorem 4.4.10. *Fix $n \in \mathbb{N}$. Suppose we are given as input a set of elements $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathbb{A}^K$ with integer coefficients, as well as the vectors $a_1, \ldots, a_K \in \mathbb{Z}^n$ and two subsets I, J of $\{1, \ldots, K\}$. Denote by $\mathcal{M}$ be the $\mathbb{A}$-submodule of $\mathbb{A}^K$ generated by $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m$. It is decidable whether there exists $\boldsymbol{f} \in \mathcal{M}$ satisfying $\boldsymbol{f} \in (\mathbb{A}^+)^K$ and*

$$(O_v \cup J) \cap M_v(I, \boldsymbol{f}) \neq \emptyset \quad \text{for every } v \in (\mathbb{R}^n)^*. \tag{4.14}$$

Here, if $n = 0$ then $\mathbb{A}$ is understood as $\mathbb{R}$, and Property (4.14) is considered trivially true.

Proof. We use induction on n. As remarked in Subsection 4.7.5, the base case $n = 0$ degenerates into linear programming (given an $\mathbb{R}$-submodule $\mathcal{M}$ of $\mathbb{R}^K$, decide whether $\mathcal{M}$ contains an

element of $\mathbb{R}^K_{>0}$). Suppose we have a decision procedure for all $n_0 < n$, we now construct a procedure for n.

By the local-global principle (Theorem 4.4.8), it suffices to decide whether the two conditions (LocR) and (LocInf) are both satisfied. First we check if (LocR) is true using Proposition 4.7.2. If (LocR) is false then we return False and conclude there is no $\boldsymbol{f} \in \mathcal{M} \cap (\mathbb{A}^+)^K$ satisfying (4.14). If (LocR) is true then we proceed.

We now run the two following procedures *in parallel*:

1. **Procedure A:** We enumerate all elements of the $\mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}]$-module:

$$\mathcal{M}_{\mathbb{Z}} \coloneqq \left\{ \sum_{j=1}^{m} h_j \cdot \boldsymbol{g}_j \;\middle|\; h_1, \ldots, h_m \in \mathbb{Z}[X_1^{\pm}, \ldots, X_n^{\pm}] \right\}.$$

   For each element $\boldsymbol{f} \in \mathcal{M}_{\mathbb{Z}}$, check if $\boldsymbol{f}$ is in $(\mathbb{A}^+)^K$ and if it satisfies Property (4.14). This can be done in the following way: since the entries of $\boldsymbol{f}$ contain finitely many monomials, it suffices to check Property (4.14) for a finite number of v. Indeed, since each of $f_1, \ldots, f_K$ has only finitely many monomials, there exists a partition $\mathcal{L}_{\boldsymbol{f}}$ of $(\mathbb{R}^n)^*$ such that for each cell $L \in \mathcal{L}_{\boldsymbol{f}}$, all directions $v \in L$ yield the same set $M_v(I, \boldsymbol{f})$. Recall the partition $\mathcal{L}_O$ of $(\mathbb{R}^n)^*$ defined in Subsection 4.7.5. For each cell $L \in \mathcal{L}_O$, all directions $v \in L$ yield the same set O_v. Therefore, it suffices to check Property (4.14) for one vector v in each cell of the partition $\mathcal{L}_{\boldsymbol{f}} \vee \mathcal{L}_O$. This can be done in finite time for any given $\boldsymbol{f}$. If some element $\boldsymbol{f} \in \mathcal{M}_{\mathbb{Z}}$ is in $(\mathbb{A}^+)^K$ and satisfies Property (4.14), we stop the procedure and return True.

2. **Procedure B:** We enumerate all $A \in \mathsf{GL}(n, \mathbb{Z})$ and $d \in \{0, 1, \ldots, n-1\}$. For each A and d, compute $Op(A, d)$ using Lemma 4.7.13 and the induction hypothesis on n. Using Lemma 4.7.15, we check if the statement (LocInfCell)(a) from Proposition 4.7.14 is false. If for some A, d, the statement (LocInfCell)(a) is false, then we stop the procedure and return False.

We claim that one of the two above procedures must stop.

Indeed, if $\mathcal{M}$ contains an element of $(\mathbb{A}^+)^K$ satisfying Property (4.14), then there exists an element $\boldsymbol{f} \in \mathcal{M}_{\mathbb{Z}} \cap (\mathbb{A}^+)^K$ satisfying Property (4.14) (see Lemma 4.4.7). In this case, Procedure A terminates by finding an element $\boldsymbol{f}$ of $\mathcal{M}_{\mathbb{Z}} \cap (\mathbb{A}^+)^K$ that satisfies Property (4.14).

If $\mathcal{M}$ does not contain an element of $(\mathbb{A}^+)^K$ satisfying Property (4.14), then by Theorem 4.4.8, Condition (LocInf) must be false (since we have already checked (LocR) to be true). By the chain of Propositions 4.7.4, 4.7.11 and 4.7.14, the statement (LocInfCell)(a) must be false for

some $A \in \mathsf{GL}(n, \mathbb{Z})$ and $d \in \{0, 1, \ldots, n-1\}$. In this case, Procedure B terminates by finding $A \in \mathsf{GL}(n, \mathbb{Z})$ and $d \in \{0, 1, \ldots, n-1\}$ where the statement (LocInfCell)(a) is false.

Therefore, by running Procedure A and Procedure B in parallel, we obtain an algorithm that always terminates for n. □

Chapter 5

The special affine group $\mathsf{SA}(2, \mathbb{Z})$

5.1 Introduction and main result

Many semigroup algorithmic problems were first considered in the context of matrix semigroups. Indeed, the first undecidability result given by Markov [72] showed undecidability of Semigroup Membership for integer matrices of dimension six. Most group and semigroup algorithmic problems remain undecidable even for matrix groups of smaller dimensions. Mikhailova [74] famously showed undecidability of Group Membership (and hence also Semigroup Membership) in the special linear group $\mathsf{SL}(4, \mathbb{Z})$. Later, Bell and Potapov [16] showed undecidability of the Identity Problem and the Group Problem in $\mathsf{SL}(4, \mathbb{Z})$. Both undecidability results stem from the fact that $\mathsf{SL}(4, \mathbb{Z})$ contains as a subgroup a direct product of two free groups over two generators. In $\mathsf{SL}(2, \mathbb{Z})$, Semigroup Membership, the Identity Problem and the Group Problem were shown to be NP-complete by Bell, Hirvensalo and Potapov [14, 15], while Group Membership was shown to be in PTIME by Lohrey [66]. All these complexity results suppose the elements of $\mathsf{SL}(2, \mathbb{Z})$ to be represented by matrices with binary encoded entries. It remains an intricate open problem whether any of these four algorithmic problems is decidable in $\mathsf{SL}(3, \mathbb{Z})$. Nevertheless, Ko, Niskanen and Potapov [58] recently showed that the direct product $\{a, b\}^* \times \{a, b\}^*$ of two free monoids over two generators cannot be embedded into the group $\mathsf{SL}(3, \mathbb{Z})$. This excluded the possibility of directly embedding the Post Correspondence Problem (Theorem 4.1.1), and suggested that Semigroup Membership in $\mathsf{SL}(3, \mathbb{Z})$ might be decidable.

Since most algorithmic problems have been solved for $\mathsf{SL}(2, \mathbb{Z})$, but seem currently out of reach for $\mathsf{SL}(3, \mathbb{Z})$, this chapter will focus on an intermediate group between $\mathsf{SL}(2, \mathbb{Z})$ and $\mathsf{SL}(3, \mathbb{Z})$: the *special affine group* $\mathsf{SA}(2, \mathbb{Z})$. Recall that elements of $\mathsf{SA}(2, \mathbb{Z})$ are 3×3 integer matrices of

the following form.

$$\mathsf{SA}(2,\mathbb{Z}) \coloneqq \left\{ \begin{pmatrix} A & \boldsymbol{a} \\ 0 & 1 \end{pmatrix} \;\middle|\; A \in \mathsf{SL}(2,\mathbb{Z}), \boldsymbol{a} \in \mathbb{Z}^2 \right\} \subsetneq \mathsf{SL}(3,\mathbb{Z}).$$

To be precise, elements of $\mathsf{SA}(2,\mathbb{Z})$ are represented using matrices with *binary encoded* entries. We denote by $(A, \boldsymbol{a})$ the element $\begin{pmatrix} A & \boldsymbol{a} \\ 0 & 1 \end{pmatrix}$, then the neutral element of $\mathsf{SA}(2,\mathbb{Z})$ is $(I, \mathbf{0})$; the group law in $\mathsf{SA}(2,\mathbb{Z})$ is given by

$$(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (AB, A\boldsymbol{b} + \boldsymbol{a}).$$

Naturally, $\mathsf{SA}(2,\mathbb{Z})$ has a subgroup $\{(A, \mathbf{0}) \mid A \in \mathsf{SL}(2,\mathbb{Z})\} \cong \mathsf{SL}(2,\mathbb{Z})$.

Special affine groups are important in the context of many fundamental problems, such as Lie groups [98], polyhedral geometry [75], dynamical systems [24], quadrics [93], computer vision [39, 63] and gauge theory [2]. Apart from the intrinsic interest to study $\mathsf{SA}(2,\mathbb{Z})$, we also point out that the Special affine group has tight connections to various reachability problems. Some of the central questions in automated verification include reachability problems in *Affine Vector Addition Systems* and *Affine Vector Addition Systems with states (Affine VASS)* over the integers [85]. While both problems as well as many of their variations have been shown to be decidable in dimension one and undecidable for dimension three [38, 59], few results are known for dimension two. Since the study of these reachability problems in dimension two necessitates the study of sub-semigroups of $\mathsf{SA}(2,\mathbb{Z})$, the techniques introduced in this chapter might provide insights into these open problems.

Currently, among the decision problems introduced in Section 2.2.2, the only known result in $\mathsf{SA}(2,\mathbb{Z})$ is the decidability of Group Membership. This can be deduced from the recent work of Delgado [32], who showed decidability of Group Membership in the semidirect product $\mathbb{Z}^m \rtimes F$, where F is a free group. Delgado's result relies on generalizing the techniques of *Stallings foldings* [52], and can therefore be extended to the case where F is *virtually free*. This can then be applied to the group $\mathsf{SA}(2,\mathbb{Z}) = \mathbb{Z}^2 \rtimes \mathsf{SL}(2,\mathbb{Z})$, since $\mathsf{SL}(2,\mathbb{Z})$ is virtually free. In this chapter, we step further by considering the Identity Problem and the Group Problem in $\mathsf{SA}(2,\mathbb{Z})$.

Main result

Our main result is decidability and NP-completeness of the Identity Problem and the Group Problem in $\mathsf{SA}(2,\mathbb{Z})$. This extends the NP-completeness result of Bell et al. [14] for the Identity Problem and the Group Problem in $\mathsf{SL}(2,\mathbb{Z})$.

Theorem 5.1.1. *The Group Problem and the Identity Problem in* $\mathsf{SA}(2,\mathbb{Z})$ *are NP-complete.*

The NP-hard lower bounds in $\mathsf{SA}(2,\mathbb{Z})$ directly follows from its embedding of the subgroup $\mathsf{SL}(2,\mathbb{Z})$. Therefore our main contribution is the decidability and the NP upper bounds.

Beyond the Identity Problem and the Group Problem, we will also discuss some obstacles to generalizing our results to Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$. Our results actually show that Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$ is decidable in many cases under additional constraints. We identify one of the remaining difficult cases, namely when the semigroup $\langle\mathcal{G}\rangle$ is isomorphic to a sub-semigroup of the semidirect product $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$, where λ is a quadratic integer. Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ remains an open problem. However, the group $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ bears certain similarities to the *Baumslag-Solitar group* $\mathsf{BS}(1,q) \coloneqq \mathbb{Z}[\frac{1}{q}] \rtimes_q \mathbb{Z}$; and a recent result by Cadilhac, Chistikov and Zetzsche [25] showed decidability of the *rational subset membership problem*[1] in $\mathsf{BS}(1,q)$ by considering rational languages of *base-q expansions*. Despite some visible difficulties, it would be interesting in the future to adapt this approach to study Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$, namely by considering rational languages of *base-*λ *expansions* [20], where λ is an algebraic integer.

Organization of the chapter

The organization of this chapter is as follows. Section 5.2 contains the preliminaries in linear algebra, group theory, as well as a classification of elements in $\mathsf{SL}(2,\mathbb{Z})$. Section 5.3 gives an overview of the decision procedure for the Group Problem (and hence the Identity Problem) in $\mathsf{SA}(2,\mathbb{Z})$. This procedure relies on an effective dichotomy on the structure of subgroups of $\mathsf{SL}(2,\mathbb{Z})$ (the *Tits alternative*, Theorem 5.3.2). We then solve the Group Problem in the two cases of the dichotomy. Section 5.4 deals with the case when the projection of $\langle\mathcal{G}\rangle_{grp}$ on $\mathsf{SL}(2,\mathbb{Z})$ contains a non-abelian free subgroup. Section 5.5 deals with the case when the projection of $\langle\mathcal{G}\rangle_{grp}$ on $\mathsf{SL}(2,\mathbb{Z})$ is virtually solvable. Finally, in Section 5.6 we discuss possible extensions of our result and obstacles to solving Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$.

5.2 Preliminaries

Full-image words

Let G be an arbitrary group and $\mathcal{G} = \{g_1, \ldots, g_K\}$ be a set of elements in G. Recall from Lemma 2.2.1 that the semigroup $\langle\mathcal{G}\rangle$ is a group if and only if the neutral element I of G is represented by a full-image word $w \in \mathcal{G}^*$. The following lemma slightly extends this result.

[1] The rational subset membership problem subsumes Semigroup Membership.

Lemma 5.2.1. *Let $\mathcal{G} = \{g_1, \dots, g_K\}$ be a set of elements in a group G. Suppose the semigroup $\langle \mathcal{G} \rangle$ is a group, then every element $g \in \langle \mathcal{G} \rangle$ is represented by a full-image word over $\mathcal{G}$.*

Proof. If $\langle \mathcal{G} \rangle$ is a group, then by Lemma 2.2.1, the neutral element I is represented by some full-image word $w_I \in \mathcal{G}^*$. Then for any element $g \in \langle \mathcal{G} \rangle$, represented by some word $w_g \in \mathcal{G}^*$, the word $w \coloneqq w_g w_I$ is a full-image word representing g. □

Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \dots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements in $\mathsf{SA}(2, \mathbb{Z})$. Suppose that an element $A \in \langle A_1, \dots, A_K \rangle$ in $\mathsf{SL}(2, \mathbb{Z})$ is represented by a full-image word w in the alphabet $\{A_1, \dots, A_K\}$. Then replacing each letter A_i in w by $(A_i, \boldsymbol{a}_i)$, we obtain a product $(A, \boldsymbol{a}) \in \mathsf{SA}(2, \mathbb{Z})$ for some $\boldsymbol{a} \in \mathbb{Z}^2$, represented by a full-image word over $\mathcal{G}$.

Linear algebra

For an arbitrary matrix $A \in \mathsf{SL}(2, \mathbb{Z})$, an *invariant subspace* of A is a $\mathbb{C}$-linear subspace V of $\mathbb{C}^2$ such that $AV = V$. If the eigenvalues of A are reals, then one can suppose that the invariant spaces of A are subspaces of $\mathbb{R}^2$. Denote by $\mathrm{Lat}(A)$ the set of *dimension one* invariant subspaces of A. If $A \notin \{I, -I\}$, then $\mathrm{Lat}(A)$ has one or two elements.

Two matrices A and B are called *conjugates* over a field $\mathbb{K}$ if $P^{-1}AP = B$ for some invertible matrix P with entries in $\mathbb{K}$. This is denoted as $A \overset{\mathbb{K}}{\sim} B$. Two matrices $A, B \in \mathsf{SL}(2, \mathbb{Z})$ are called *simultaneously triangularizable* if there exists a complex matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both upper-triangular. It is easy to see that if $\mathrm{Lat}(A) \cap \mathrm{Lat}(B) \neq \emptyset$, then A and B are simultaneously triangularizable. Indeed, let v be any non-zero vector in a subspace of $\mathrm{Lat}(A) \cap \mathrm{Lat}(B)$, and let $P = (v, w)$ be such that w is linearly independent from v. Then both $P^{-1}AP$ and $P^{-1}BP$ are upper-triangular.

Group theory

Recall the definition of a solvable group (Definition 2.1.9). Every subgroup of a solvable group is solvable [34, Proposition 13.91]. For any field $\mathbb{K}$ and a number n, denote by $\mathsf{T}(n, \mathbb{K})$ the group of $n \times n$ invertible upper-triangular matrices with entries in $\mathbb{K}$. Then $\mathsf{T}(n, \mathbb{K})$ is a solvable group [12]. In particular, if two matrices $A, B \in \mathsf{SL}(2, \mathbb{Z})$ are simultaneously triangularizable, then the group G they generate is isomorphic to a subgroup of $\mathsf{T}(2, \mathbb{C})$; thus G is solvable.

Recall the definition of a free group (Subsection 2.1.1). Like solvable groups, the class of free groups is stable under taking subgroups:

Theorem 5.2.2 (Nielsen–Schreier [90, Chapter I, Theorem 5])**.** *Every subgroup of a free group is free.*

Given an arbitrary group G with neutral element I. Recall that an element $T \in G$ is called *torsion* if $T^m = I$ for some $m \geq 1$. A group is called *torsion-free* if it does not contain any non-trivial torsion element (i.e. if its only torsion element is the neutral element). In particular, a free group is torsion-free.

A group is called *virtually solvable* if it admits a finite index subgroup that is solvable. Similarly, a group is called *virtually free* if it admits a finite index subgroup that is free. The following is a classic result on the structure of $\mathsf{SL}(2,\mathbb{Z})$.

Theorem 5.2.3 ([80])**.** *The group $\mathsf{SL}(2,\mathbb{Z})$ is virtually free. Moreover, it contains a finite index free subgroup $F(\{S,T\})$ over two generators.*

Based on this fact, Bell, Hirvensalo, and Potapov showed the following complexity result.

Theorem 5.2.4 (Bell, Hirvensalo, Potapov [14])**.** *The Identity Problem and the Group Problem in $\mathsf{SL}(2,\mathbb{Z})$ are NP-complete.*

Classification of elements in $\mathsf{SL}(2,\mathbb{Z})$

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix in $\mathsf{SL}(2,\mathbb{Z})$. The characteristic polynomial of A is $f(X) = X^2 - (a+d)X + (ad-bc) = X^2 - (a+d)X + 1$. Consider the five following cases.

(i) $a + d = 0$.

In this case, $A \overset{\mathbb{Q}(i)}{\sim} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, the eigenvalues of A are i and $-i$. We have $A^4 = I$, so A is a torsion element.

(ii) $|a + d| = 1$.

In this case, $A \overset{\mathbb{Q}(\omega)}{\sim} \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix}$, where $\omega = \frac{1+\sqrt{3}i}{2}$ or $\frac{-1+\sqrt{3}i}{2}$. In both cases, we have $A^6 = I$, so A is a torsion element.

(iii) $a + d = 2$.

In this case, either $A = I$, or $A \overset{\mathbb{Q}}{\sim} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. In the second situation, we call A a *shear*. The only eigenvalue of A is 1. If A is a shear, then $\mathrm{Lat}(A)$ has exactly one element. See Figure 5.2 for an illustration.

(iv) $a + d = -2$.

In this case, either $A = -I$, or $A \overset{\mathbb{Q}}{\sim} \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$. In the second situation, we call A a *twisted inversion*. In particular, if A is a twisted inversion, then $(A+I)^2 = 0$, and A^2 is a shear, and $\mathrm{Lat}(A)$ has exactly one element.

(v) $|a+d| \geq 3$.

In this case, $A \overset{\mathbb{R}}{\sim} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$, where λ is the root of $f(X)$ such that $|\lambda| \geq 1$. Furthermore, λ is real. In this case, we call A a *scale*. If $\lambda > 0$, we call A a *positive scale*; if $\lambda < 0$, we call A an *inverting scale*. In both cases, $\mathrm{Lat}(A)$ has two elements, one element is the invariant space corresponding to the eigenvalue λ, and is called the *stretching direction*; the other element is the invariant space corresponding to the eigenvalue λ^{-1}, and is called the *compressing direction*. See Figure 5.1 for an illustration.

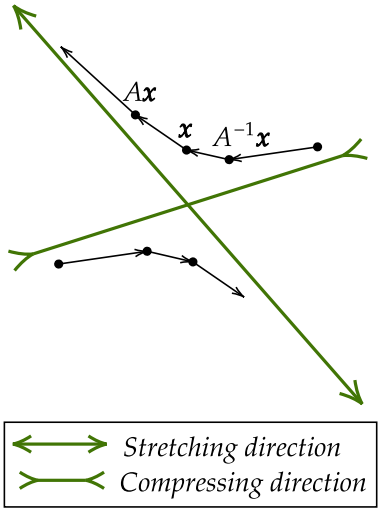

Figure 5.1: Illustration for a positive scale.

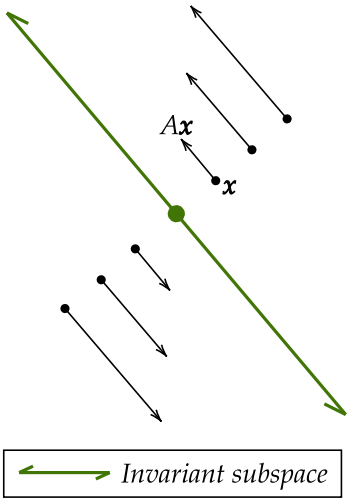

Figure 5.2: Illustration for a shear.

5.3 Overview of decision procedures

In this section we give an overview of the decision procedure for the Group Problem (and hence the Identity Problem) in $\mathsf{SA}(2,\mathbb{Z})$. We state two propositions (Propositions 5.3.3 and 5.3.4) regarding the structure of sub-semigroups of $\mathsf{SA}(2,\mathbb{Z})$. Assuming these propositions, we prove NP-completeness of the Identity Problem and the Group Problem in $\mathsf{SA}(2,\mathbb{Z})$. The proofs of propositions 5.3.3 and 5.3.4 are delayed until Section 5.4 and 5.5.

We focus on solving the Group Problem, because by Lemma 2.2.4, decidability and an NP upper bound of the Group Problem will imply decidability and an NP upper bound of the Identity Problem. Fix a set

$$\mathcal{G} \coloneqq \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$$

of elements in $\mathsf{SA}(2,\mathbb{Z})$. The following lemma shows that, if the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is not a group, then $\langle \mathcal{G} \rangle$ is also not a group. Therefore, to decide whether $\langle \mathcal{G} \rangle$ is a group, we can focus on the case where H is a group.

Lemma 5.3.1. *Let $\mathcal{G} \coloneqq \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$. If the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is not a group, then the semigroup $\langle \mathcal{G} \rangle$ is also not a group.*

Proof. If $\langle \mathcal{G} \rangle$ is a group, then $(A_i^{-1}, -A_i^{-1}\boldsymbol{a}_i) = (A_i, \boldsymbol{a}_i)^{-1} \in \langle \mathcal{G} \rangle$ for all i. Therefore, $A_i^{-1} \in \langle A_1, \ldots, A_K \rangle$ for all i. Thus $\langle A_1, \ldots, A_K \rangle$ is a group. □

Suppose now that H is a group. The key idea of solving the Group Problem is the following dichotomy known as the *Tits alternative*.

Theorem 5.3.2 (Tits alternative [96, Theorem 1], effective version [12, Theorem 1.1])**.** *Fix any $n \in \mathbb{N}$. Given a finitely generated subgroup H of $\mathsf{SL}(n, \mathbb{Z})$, exactly one of the following is true:*

(i) H contains a non-abelian free subgroup.
(ii) H is virtually solvable.

Furthermore, given a set of group generators of H, it is decidable in PTIME which of the two cases is true.

In case of H containing a non-abelian free subgroup, we will prove the following Proposition 5.3.3, which shows that the semigroup $\langle \mathcal{G} \rangle$ must be a group.

Proposition 5.3.3. *Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$, such that the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a group. Suppose H contains a non-abelian free subgroup, then $\langle \mathcal{G} \rangle$ is a group.*

The proof of Proposition 5.3.3 is highly non-trivial and will be given in Section 5.4. The proof is mainly geometric: it consists of analysing the action of $\mathsf{SA}(2, \mathbb{Z})$ on the lattice $\mathbb{Z}^2$.

In case of H being virtually solvable, we will prove the following Proposition 5.3.4 which refines the Tits alternative. In particular, it shows that virtually solvable subgroups of $\mathsf{SL}(2, \mathbb{Z})$ have relatively simple structure: it is either trivial, or it contains a non-trivial torsion element, or it is infinite cyclic.

Proposition 5.3.4. *Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$, such that the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a group. Suppose H is virtually solvable, then exactly one of the following six conditions holds:*

(i) H is the trivial group.
(ii) H contains a non-trivial torsion element.
(iii) $H = \langle A \rangle_{grp}$, where A is a twisted inversion.
(iv) $H = \langle A \rangle_{grp}$, where A is a shear.

(v) $H = \langle A \rangle_{grp}$*, where* A *is an inverting scale.*

(vi) $H = \langle A \rangle_{grp}$*, where* A *is a positive scale.*

Furthermore, in cases (ii), (iii) and (v), the semigroup $\langle \mathcal{G} \rangle$ *is always a group. Overall, it is decidable in PTIME whether* $\langle \mathcal{G} \rangle$ *is a group.*

Proposition 5.3.4 will be proved in Section 5.5. The proof will be mainly algebraic: it consists of analysing the structure of sub-semigroups of a virtually solvable group. Lemma 5.3.1-Proposition 5.3.4 yield the decidability of the Group Problem (and consequently, the Identity Problem) in $\mathsf{SA}(2, \mathbb{Z})$. An overview of the procedure is given in Algorithm 5.1. The justification of each step is given in parentheses with reference to the corresponding lemmas or propositions.

Theorem 5.1.1. *The Group Problem and the Identity Problem in* $\mathsf{SA}(2, \mathbb{Z})$ *are NP-complete.*

Proof. The NP-hard lower bounds come from the NP-completeness of both problems in the subgroup $\mathsf{SL}(2, \mathbb{Z}) \cong \{(A, \mathbf{0}) \mid A \in \mathsf{SL}(2, \mathbb{Z})\} \leq \mathsf{SA}(2, \mathbb{Z})$ (see Theorem 5.2.4).

To show decidability and the NP upper bounds, we first solve the Group Problem. Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set in $\mathsf{SA}(2, \mathbb{Z})$. As a first step, we can check in NP whether $\langle A_1, \ldots, A_K \rangle$ is a group (Theorem 5.2.4). If $\langle A_1, \ldots, A_K \rangle$ is not a group, then $\langle \mathcal{G} \rangle$ is not a group by Lemma 5.3.1.

Suppose now that $\langle A_1, \ldots, A_K \rangle$ is a group. As the second step, we check in PTIME whether $\langle A_1, \ldots, A_K \rangle$ contains a non-abelian free subgroup using the Tits alternative (Theorem 5.3.2). If $\langle A_1, \ldots, A_K \rangle$ contains a non-abelian free subgroup, then the Group Problem has a positive answer by Proposition 5.3.3. Otherwise, $\langle A_1, \ldots, A_K \rangle$ is virtually solvable, and we can decide the Group Problem in PTIME using Proposition 5.3.4. In total, the Group Problem for $\mathcal{G}$ can be decided in NP.

By Lemma 2.2.4, the Identity Problem in $\mathsf{SA}(2, \mathbb{Z})$ can also be decided in NP. □

5.4 Non-abelian free subgroup

In this section we prove Proposition 5.3.3:

Proposition 5.3.3. *Let* $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ *be a set of elements of* $\mathsf{SA}(2, \mathbb{Z})$*, such that the semigroup* $H \coloneqq \langle A_1, \ldots, A_K \rangle$ *is a group. Suppose* H *contains a non-abelian free subgroup, then* $\langle \mathcal{G} \rangle$ *is a group.*

The proof depends on two lemmas concerning the effect of "pumping" a word $(A, \boldsymbol{a})(B, \boldsymbol{b})$, where $(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x})$ for some $\boldsymbol{x} \in \mathbb{Z}^2$. Let A be a scale with $\mathrm{Lat}(A) = \{V, W\}$. Since

Algorithm 5.1 Deciding the Group Problem in $\mathsf{SA}(2,\mathbb{Z})$.

Input: A subset $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ of $\mathsf{SA}(2,\mathbb{Z})$.

Output: **True** ($\langle\mathcal{G}\rangle$ is a group) or **False** ($\langle\mathcal{G}\rangle$ is not a group).

1. Decide whether the semigroup $H \coloneqq \langle A_1, \ldots, A_K\rangle$ is a group by Theorem 5.2.4. If H is not a group, return **False**. (Lemma 5.3.1)
2. Decide for $H \leq \mathsf{SL}(2,\mathbb{Z})$ which case of Theorem 5.3.2 is true.
3. If H contains a non-abelian free subgroup, return **True**. (Proposition 5.3.3)
4. If H is virtually free, decide which case of Proposition 5.3.4 is true using Lemma 5.5.4.
   (i) If H is trivial, decide whether $n_1\boldsymbol{a}_1 + \cdots n_K\boldsymbol{a}_K = \boldsymbol{0}$ has a solution over $\mathbb{Z}^K_{>0}$. If yes, return **True**, otherwise return **False**. (Proposition 5.5.5)
   (ii) If H contains a non-trivial torsion element, return **True**. (Proposition 5.5.6)
   (iii) If $H = \langle A\rangle_{grp}$, where A is a twisted inversion, return **True**. (Proposition 5.5.7)
   (iv) If $H = \langle A\rangle_{grp}$, where A is a shear, compute the set $\varphi(\mathcal{G})$ defined by Equation (5.7). Decide whether $\langle\varphi(\mathcal{G})\rangle$ is a group using Algorithm 3.1 (Theorem 5.5.8). If yes, return **True**; if not, return **False**. (Corollary 5.5.9)
   (v) If $H = \langle A\rangle_{grp}$, where A is a inverting scale, return **True**. (Proposition 5.5.10)
   (vi) If $H = \langle A\rangle_{grp}$, where A is a positive scale, compute the set $\mathcal{S}$ defined by Equation (5.10). Decide whether the condition in Proposition 5.5.14 is satisfied for $\mathcal{S}$. If yes, return **True**; if not, return **False**. (Corollary 5.5.15)

V, W are distinct one dimensional subspaces of $\mathbb{R}^2$, every element $\boldsymbol{x} \in \mathbb{R}^2$ can be written uniquely as $\boldsymbol{x} = \boldsymbol{x}_V + \boldsymbol{x}_W$, where $\boldsymbol{x}_V \in V, \boldsymbol{x}_W \in W$. We will adopt this notation in the following lemma. For an element $\boldsymbol{y} \in \mathbb{R}^2$, $\boldsymbol{y}_V, \boldsymbol{y}_W$ are defined similarly. We use $\|\cdot\|$ to denote the Euclidean norm.

Lemma 5.4.1. *Let $(A, \boldsymbol{a}), (B, \boldsymbol{b})$ be elements of $\mathsf{SA}(2,\mathbb{Z})$ such that $(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x})$ for some $\boldsymbol{x} \in \mathbb{Z}^2$. Suppose A is a scale; denote by V, W the elements of $\mathrm{Lat}(A)$, and suppose $\boldsymbol{x} \notin V \cup W$. Let $\boldsymbol{v}$ be any non-zero vector in the subspace V.*

Then for every $\varepsilon \in (0,1)$, there exists a word $w \in \{(A, \boldsymbol{a}), (B, \boldsymbol{b})\}^*$, such that $(A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{y})$, where $\boldsymbol{y} \in \mathbb{Z}^2$ satisfies

$$1 - \frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon, \quad \boldsymbol{y}_V^\top \boldsymbol{x}_V > 0, \quad \boldsymbol{y}_W^\top \boldsymbol{x}_W > 0. \tag{5.1}$$

In other words, the acute angle θ between $\boldsymbol{y}$ and V satisfies $1 - \cos\theta < \varepsilon$. Also, $\boldsymbol{y}$ and $\boldsymbol{x}$ lie in same cone out of the four cut out by V and W. See Figure 5.3 for an illustration.

Proof. Since $(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x})$, we have $B = A^{-1}$ and $\boldsymbol{x} = A\boldsymbol{b} + \boldsymbol{a}$.

Let λ be the eigenvalue of A associated to the invariant subspace V, then λ^{-1} is the eigenvalue

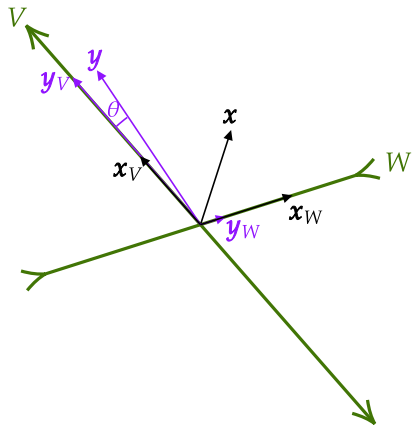

Figure 5.3: Illustration for Lemma 5.4.1.

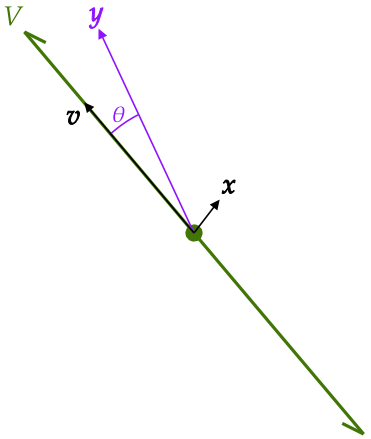

Figure 5.4: Illustration for Lemma 5.4.2.

associated to the invariant subspace W. Then we have $A^n\boldsymbol{x} = \lambda^n \boldsymbol{x}_V + \lambda^{-n}\boldsymbol{x}_W$ for every integer n. Consider two cases.

1. If V is a stretching direction. In this case, $|\lambda| > 1$. Let $m > 0$ be a positive integer, and let $w \coloneqq (A, \boldsymbol{a})^{m-1}(B, \boldsymbol{b})^{m-1}$. Consider the product

$$\begin{aligned}
&(A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) \\
&= (A, \boldsymbol{a})^m (B, \boldsymbol{b})^m \\
&= (I, (I + A + \cdots + A^{(m-1)})(A\boldsymbol{b} + \boldsymbol{a})) \\
&= (I, (I + A + \cdots + A^{(m-1)})\boldsymbol{x}) \\
&= \left(I, \sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W\right) \qquad (5.2)
\end{aligned}$$

   Let $\boldsymbol{y} \coloneqq \sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W$, then $\boldsymbol{y}_V = \sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V$, $\boldsymbol{y}_W = \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W$. Since $\boldsymbol{x} \notin V \cup W$, we have $\boldsymbol{x}_V \neq \boldsymbol{0}, \boldsymbol{x}_W \neq \boldsymbol{0}$. When m is odd, we have $\sum_{i=0}^{m-1} \lambda^i > 0$, $\sum_{i=0}^{m-1} \lambda^{-i} > 0$, so $\boldsymbol{y}_V^\top \boldsymbol{x}_V > 0$ and $\boldsymbol{y}_W^\top \boldsymbol{x}_W > 0$.

   Note that we always have $\frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} \leq 1$ by the Cauchy-Schwarz inequality. We then show that when m tends towards infinity, the value $\frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|}$ will tend to one. Indeed,

$$\begin{aligned}
\frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} &= \frac{\left|\sum_{i=0}^{m-1} \lambda^i \boldsymbol{v}^\top \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i} \boldsymbol{v}^\top \boldsymbol{x}_W\right|}{\|\boldsymbol{v}\| \left\|\sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W\right\|} \\
&\geq \frac{\left|\left(\sum_{i=0}^{m-1} \lambda^i\right) \boldsymbol{v}^\top \boldsymbol{x}_V\right|}{\|\boldsymbol{v}\| \left\|\sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W\right\|} - \frac{\left|\left(\sum_{i=0}^{m-1} \lambda^{-i}\right) \boldsymbol{v}^\top \boldsymbol{x}_W\right|}{\|\boldsymbol{v}\| \left\|\sum_{i=0}^{m-1} \lambda^i \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i}\boldsymbol{x}_W\right\|}
\end{aligned}$$

$$= \frac{\|\boldsymbol{v}\|\|\boldsymbol{x}_V\|}{\|\boldsymbol{v}\|\,\|\boldsymbol{x}_V + \lambda^{1-m}\boldsymbol{x}_W\|} - \frac{\left|\boldsymbol{v}^\top \boldsymbol{x}_W\right|}{\|\boldsymbol{v}\|\,\|\lambda^{m-1}\boldsymbol{x}_V + \boldsymbol{x}_W\|}$$

When $m \to \infty$, the right hand side tends towards $1 - 0 = 1$. This is because $|\lambda| > 1$ and $\boldsymbol{v} \neq \boldsymbol{0}, \boldsymbol{x}_V \neq \boldsymbol{0}$. Hence, for a large enough odd integer m, we have

$$1 - \frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon, \quad \boldsymbol{y}_V^\top \boldsymbol{x}_V > 0, \quad \boldsymbol{y}_W^\top \boldsymbol{x}_W > 0.$$

2. If V is a compressing direction. In this case, $|\lambda| < 1$. Let $m > 0$ be a positive integer, and let $w \coloneqq (B, \boldsymbol{b})^m (A, \boldsymbol{a})^m$. Consider the product

$$\begin{aligned}
&(A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) \\
&= (A, \boldsymbol{a})(B, \boldsymbol{b})(B, \boldsymbol{b})^{m-1}(A, \boldsymbol{a})^{m-1}(A, \boldsymbol{a})(B, \boldsymbol{b}) \\
&= \left(I, (2I + A^{-1} + \cdots + A^{-(m-1)})(A\boldsymbol{b} + \boldsymbol{a})\right) \\
&= \left(I, (2I + A^{-1} + \cdots + A^{-(m-1)})\boldsymbol{x}\right) \\
&= \left(I, \left(1 + \sum_{i=0}^{m-1} \lambda^{-i}\right)\boldsymbol{x}_V + \left(1 + \sum_{i=0}^{m-1} \lambda^{i}\right)\boldsymbol{x}_W\right)
\end{aligned}$$

Let $\boldsymbol{y} \coloneqq \left(1 + \sum_{i=0}^{m-1} \lambda^{-i}\right)\boldsymbol{x}_V + \left(1 + \sum_{i=0}^{m-1} \lambda^{i}\right)\boldsymbol{x}_W$, then $\boldsymbol{y}_V = \left(1 + \sum_{i=0}^{m-1} \lambda^{-i}\right)\boldsymbol{x}_V$, $\boldsymbol{y}_W = \left(1 + \sum_{i=0}^{m-1} \lambda^{i}\right)\boldsymbol{x}_W$. Since $\boldsymbol{x} \notin V \cup W$, we have $\boldsymbol{x}_V \neq \boldsymbol{0}$ and $\boldsymbol{x}_W \neq \boldsymbol{0}$. When m is odd, we have $\sum_{i=0}^{m-1} \lambda^i > 0$, $\sum_{i=0}^{m-1} \lambda^{-i} > 0$, so $\boldsymbol{y}_V^\top \boldsymbol{x}_V > 0, \boldsymbol{y}_W^\top \boldsymbol{x}_W > 0$.

We then show that when m tends towards infinity, the value $\frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} \leq 1$ will tend to one. Indeed,

$$\begin{aligned}
\frac{\left|\boldsymbol{v}^\top \boldsymbol{y}\right|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} &= \frac{\left|\left(1 + \sum_{i=0}^{m-1} \lambda^{-i}\right)\boldsymbol{v}^\top \boldsymbol{x}_V + \left(1 + \sum_{i=0}^{m-1} \lambda^{i}\right)\boldsymbol{v}^\top \boldsymbol{x}_W\right|}{\|\boldsymbol{v}\|\left\|\left(1 + \sum_{i=0}^{m-1} \lambda^{-i}\right)\boldsymbol{x}_V + \left(1 + \sum_{i=0}^{m-1} \lambda^{i}\right)\boldsymbol{x}_W\right\|} \\
&\geq \frac{\|\boldsymbol{v}\|\|\boldsymbol{x}_V\|}{\|\boldsymbol{v}\|\left\|\boldsymbol{x}_V + \frac{1+\sum_{i=0}^{m-1} \lambda^{i}}{1+\sum_{i=0}^{m-1} \lambda^{-i}}\boldsymbol{x}_W\right\|} - \frac{\left|\boldsymbol{v}^\top \boldsymbol{x}_W\right|}{\|\boldsymbol{v}\|\left\|\frac{1+\sum_{i=0}^{m-1} \lambda^{-i}}{1+\sum_{i=0}^{m-1} \lambda^{i}}\boldsymbol{x}_V + \boldsymbol{x}_W\right\|}
\end{aligned}$$

When $m \to \infty$, the right hand side tends towards $1 - 0 = 1$. This is because $|\lambda| < 1$ and $\boldsymbol{v} \neq \boldsymbol{0}, \boldsymbol{x}_V \neq \boldsymbol{0}$, so

$$\lim_{m \to \infty} \frac{1 + \sum_{i=0}^{m-1} \lambda^i}{1 + \sum_{i=0}^{m-1} \lambda^{-i}} = 0.$$

Hence, for a large enough odd integer m, we have

$$1 - \frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon, \quad \boldsymbol{y}_V^\top \boldsymbol{x}_V > 0, \quad \boldsymbol{y}_W^\top \boldsymbol{x}_W > 0.$$

Combining the two cases yields the desired result. □

A similar lemma can be proved for shears. In this case we want the stronger condition $1 - \frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon$ instead of $1 - \frac{|\boldsymbol{v}^\top \boldsymbol{y}|}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon$.

Lemma 5.4.2. *Let $(A, \boldsymbol{a}), (B, \boldsymbol{b})$ be elements of $\mathsf{SA}(2, \mathbb{Z})$ such that $(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x})$ for some $\boldsymbol{x} \in \mathbb{Z}^2$. Suppose A is a shear, $\mathrm{Lat}(A) = \{V\}$, and $\boldsymbol{x} \notin V$. Let $\boldsymbol{v}$ be any non-zero vector in the subspace V.*

Then for every $\varepsilon \in (0, 1)$, there exists a word $w \in \{(A, \boldsymbol{a}), (B, \boldsymbol{b})\}^*$, such that $(A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{y})$, where $\boldsymbol{y} \in \mathbb{Z}^2$ satisfies

$$1 - \frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon, \tag{5.3}$$

and $\boldsymbol{y}$ and $\boldsymbol{x}$ lie in same halfspace cut by V. In other words, the angle θ between $\boldsymbol{y}$ and $\boldsymbol{v}$ satisfies $1 - \cos\theta < \varepsilon$. See Figure 5.4 for an illustration.

Proof. Since $(A, \boldsymbol{a}) \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x})$, we have $B = A^{-1}$ and $\boldsymbol{x} = A\boldsymbol{b} + \boldsymbol{a}$.

Let W be the orthogonal space of V, and $\boldsymbol{w}$ be a non-zero vector in W. Under the basis $\{\boldsymbol{v}, \boldsymbol{w}\}$, the matrix A has the form $\begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$, where $\mu \neq 0$. Then for every integer n, we have $A^n \boldsymbol{x} = \boldsymbol{x} + n\mu c \boldsymbol{v}$, where c is a scalar such that $c\boldsymbol{w} = \boldsymbol{x}_W$. Since $\boldsymbol{x} \notin V$, we have $\boldsymbol{x}_W \neq \boldsymbol{0}$, so $c \neq 0$. Consider two cases.

1. If $\mu c > 0$. Let $m > 0$ be a positive integer, and let $w \coloneqq (A, \boldsymbol{a})^{m-1}(B, \boldsymbol{b})^{m-1}$. Consider the product

$$\begin{aligned} & (A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) \\ = {} & (A, \boldsymbol{a})^m (B, \boldsymbol{b})^m \\ = {} & \left(I, (I + A + \cdots + A^{(m-1)})(A\boldsymbol{b} + \boldsymbol{a})\right) \\ = {} & \left(I, (I + A + \cdots + A^{(m-1)})\boldsymbol{x}\right) \\ = {} & \left(I, m\boldsymbol{x} + \frac{m(m-1)}{2}\mu c \boldsymbol{v}\right). \end{aligned}$$

   Let $\boldsymbol{y} \coloneqq m\boldsymbol{x} + \frac{m(m-1)}{2}\mu c \boldsymbol{v}$, then $\boldsymbol{y}$ and $\boldsymbol{x}$ lie in same halfspace cut by V.

   We then show that when m tends towards infinity, $\frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|}$ will tend to one. Indeed,

$$\frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} = \frac{m\boldsymbol{v}^\top \boldsymbol{x} + \frac{m(m-1)}{2}\mu c \boldsymbol{v}^\top \boldsymbol{v}}{\|\boldsymbol{v}\| \left\| m\boldsymbol{x} + \frac{m(m-1)}{2}\mu c \boldsymbol{v} \right\|}$$

$$= \frac{\boldsymbol{v}^\top \boldsymbol{x}}{\|\boldsymbol{v}\| \left\| \boldsymbol{x} + \frac{(m-1)}{2}\mu c \boldsymbol{v} \right\|} + \frac{\frac{(m-1)}{2}\mu c \|\boldsymbol{v}\|}{\left\| \boldsymbol{x} + \frac{(m-1)}{2}\mu c \boldsymbol{v} \right\|}.$$

When $m \to \infty$, this expression tends towards $0+1=1$, because $\mu c > 0$ and $\boldsymbol{v} \neq \boldsymbol{0}$. Hence, for a large enough odd integer m, we have $1 - \frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon$.

2. If $\mu c < 0$. Let $m > 0$ be a positive integer, and let $w \coloneqq (B, \boldsymbol{b})^m (A, \boldsymbol{a})^m$. Consider the product

$$\begin{aligned}
&(A, \boldsymbol{a}) \cdot w \cdot (B, \boldsymbol{b}) \\
&= (A, \boldsymbol{a})(B, \boldsymbol{b})(B, \boldsymbol{b})^{m-1}(A, \boldsymbol{a})^{m-1}(A, \boldsymbol{a})(B, \boldsymbol{b}) \\
&= \left(I, (2I + A^{-1} + \cdots + A^{-(m-1)})(A\boldsymbol{b} + \boldsymbol{a})\right) \\
&= \left(I, (2I + A^{-1} + \cdots + A^{-(m-1)})\boldsymbol{x}\right) \\
&= \left(I, (m+1)\boldsymbol{x} - \frac{m(m-1)}{2}\mu c \boldsymbol{v}\right).
\end{aligned}$$

Let $\boldsymbol{y} \coloneqq (m+1)\boldsymbol{x} - \frac{m(m-1)}{2}\mu c \boldsymbol{v}$, then $\boldsymbol{y}$ and $\boldsymbol{x}$ lie in same halfspace cut by V.

We then show that when m tends towards infinity, $\frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|}$ will tend to one. Indeed,

$$\begin{aligned}
\frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} &= \frac{(m+1)\boldsymbol{v}^\top \boldsymbol{x} - \frac{m(m-1)}{2}\mu c \boldsymbol{v}^\top \boldsymbol{v}}{\|\boldsymbol{v}\| \left\| (m+1)\boldsymbol{x} - \frac{m(m-1)}{2}\mu c \boldsymbol{v} \right\|} \\
&= \frac{\boldsymbol{v}^\top \boldsymbol{x}}{\|\boldsymbol{v}\| \left\| \boldsymbol{x} - \frac{m(m-1)}{2(m+1)}\mu c \boldsymbol{v} \right\|} + \frac{-\frac{m(m-1)}{2(m+1)}\mu c \|\boldsymbol{v}\|}{\left\| \boldsymbol{x} - \frac{m(m-1)}{2(m+1)}\mu c \boldsymbol{v} \right\|}.
\end{aligned}$$

When $m \to \infty$, this expression tends towards $0+1=1$, because $\mu c < 0$ and $\boldsymbol{v} \neq \boldsymbol{0}$. Hence, for a large enough odd integer m, we have $1 - \frac{\boldsymbol{v}^\top \boldsymbol{y}}{\|\boldsymbol{v}\|\|\boldsymbol{y}\|} < \varepsilon$.

□

We now prove Proposition 5.3.3.

Proposition 5.3.3. *Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$, such that the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a group. Suppose H contains a non-abelian free subgroup, then $\langle \mathcal{G} \rangle$ is a group.*

Proof. Suppose that the group $\langle A_1, \ldots, A_K \rangle$ contains a non-abelian free subgroup. In particular, it contains a subgroup isomorphic to the free group F_2 of two generators. Let A, B be elements of $\langle A_1, \ldots, A_K \rangle$ that generate this free group. Since A, B are non-torsion, they are twisted inversions, scales, or shears. Hence, $A^2, B^2 \in \langle A_1, \ldots, A_K \rangle$ are *positive* scales or shears. Since

A and B generate a non-abelian free group, A^2 and B^2 also generate a non-abelian free group. Therefore, we can replace A, B by A^2, B^2, and without loss of generality suppose A and B to be positive scales or shears.

Since $\langle A_1, \ldots, A_K \rangle$ is a group, A and B can be represented by full-image words over the alphabet $\{A_1, \ldots, A_K\}$ due to Lemma 5.2.1. Hence, let $(A, \boldsymbol{a})$ and $(B, \boldsymbol{b})$ be elements in $\langle \mathcal{G} \rangle$ represented by full-image words.

Note that A, B are not simultaneously triangularizable, otherwise the group they generate is isomorphic to a subgroup of $\mathsf{T}(2, \mathbb{C})$, which is solvable (see Section 5.2). This contradicts that fact that A, B generate a non-abelian free group (see Theorem 5.3.2). Therefore, $\mathrm{Lat}(A) \cap \mathrm{Lat}(B) = \emptyset$. There are three cases to consider:

(1) Both A and B are positive scales.

(2) One of A and B is a shear.

(3) Both A and B are shears.

Case 1. Both A and B are positive scales. Denote $Y = A^{-1}B^{-1} \in \langle A_1, \ldots, A_K \rangle$, we have $AYB = I$. Let $\boldsymbol{y} \in \mathbb{Z}^2$ be such that $(Y, \boldsymbol{y}) \in \langle \mathcal{G} \rangle$, and let $\boldsymbol{x} \in \mathbb{Z}^2$ be such that $(A, \boldsymbol{a})(Y, \boldsymbol{y})(B, \boldsymbol{b}) = (I, \boldsymbol{x})$. If $\boldsymbol{x} = \boldsymbol{0}$, then $(I, \boldsymbol{0})$ can be represented as a full-image word over $\mathcal{G}$, since $(A, \boldsymbol{a})$ can. In this case, $\langle \mathcal{G} \rangle$ is a group by Lemma 5.2.1. Otherwise, there are two subcases. Starting from any one-dimension subspace S, rotate S counter-clockwise and consider its sequence of encounters with $\mathrm{Lat}(A)$ and $\mathrm{Lat}(B)$ before finishing a full cycle. Either the sequence of encounters is a cyclic permutation of $(\mathrm{Lat}(A), \mathrm{Lat}(A), \mathrm{Lat}(B), \mathrm{Lat}(B))$, or it is a cyclic permutation of $(\mathrm{Lat}(A), \mathrm{Lat}(B), \mathrm{Lat}(A), \mathrm{Lat}(B))$.

(a) First consider the case of $(\mathrm{Lat}(A), \mathrm{Lat}(A), \mathrm{Lat}(B), \mathrm{Lat}(B))$. An illustration for the proof of this case is shown in Figure 5.5. Denote by V_A, W_A, V_B, W_B these subspaces in this order. Since $\boldsymbol{x} \neq \boldsymbol{0}$, either $\boldsymbol{x} \notin V_A \cup W_A$, or $\boldsymbol{x} \notin V_B \cup W_B$. Without loss of generality suppose $\boldsymbol{x} \notin V_B \cup W_B$.

Let $\varepsilon > 0$. Apply Lemma 5.4.1 to ε, to the elements $(A', \boldsymbol{a}') \coloneqq (A, \boldsymbol{a})(Y, \boldsymbol{y})$ and $(B, \boldsymbol{b})$, and to the subspaces V_B, W_B of $\mathrm{Lat}(B) = \mathrm{Lat}(A')$. (Note that $A' = B^{-1}$, so $\mathrm{Lat}(A') = \mathrm{Lat}(B)$.) Since $(A', \boldsymbol{a}')(B, \boldsymbol{b}) = (A, \boldsymbol{a})(Y, \boldsymbol{y})(B, \boldsymbol{b}) = (I, \boldsymbol{x})$, Lemma 5.4.1 shows there exists an element $w \in \langle \mathcal{G} \rangle$ such that $(A', \boldsymbol{a}') \cdot w \cdot (B, \boldsymbol{b}) = (I, \boldsymbol{x}_1)$, with

$$1 - \frac{|\boldsymbol{v}_B^\top \boldsymbol{x}_1|}{\|\boldsymbol{v}_B\| \|\boldsymbol{x}_1\|} < \varepsilon,\ (\boldsymbol{x}_1)_{V_B}^\top \boldsymbol{x}_{V_B} > 0,\ (\boldsymbol{x}_1)_{W_B}^\top \boldsymbol{x}_{W_B} > 0. \tag{5.4}$$

Here, $\boldsymbol{v}_B$ is a non-zero vector in V_B. In other words, when ε is sufficiently small, the acute angle between $\boldsymbol{x}_1$ and V_B is also sufficiently small. Also, $\boldsymbol{x}_1$ and $\boldsymbol{x}$ lie in same cone out

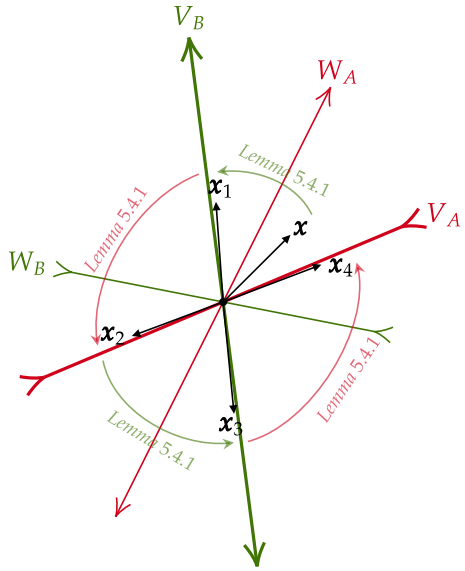

Figure 5.5: Illustration for case 1(a).

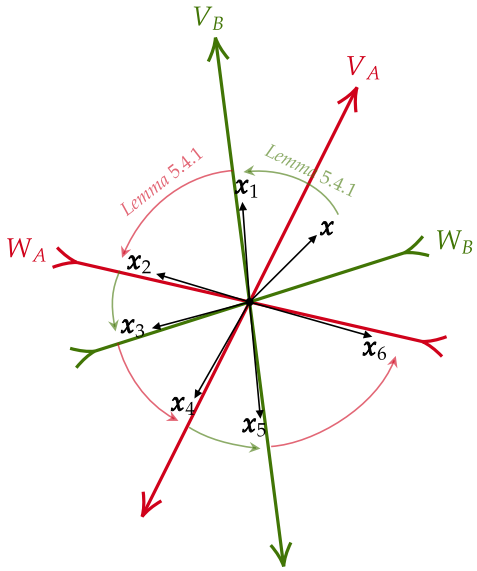

Figure 5.6: Illustration for case 1(b).

of the four cut by V_B and W_B. This gives us an element $(Y_1, \boldsymbol{y}_1) \coloneqq (Y, \boldsymbol{y}) \cdot w \in \langle \mathcal{G} \rangle$ such that $(A, \boldsymbol{a})(Y_1, \boldsymbol{y}_1)(B, \boldsymbol{b}) = (I, \boldsymbol{x}_1)$ with $\boldsymbol{x}_1$ satisfying (5.4), see Figure 5.5.

Next, apply Lemma 5.4.1 to ε, to the elements $(A, \boldsymbol{a})$ and $(B', \boldsymbol{b}') \coloneqq (Y_1, \boldsymbol{y}_1)(B, \boldsymbol{b})$ of $\mathsf{SA}(2, \mathbb{Z})$, and to the subspaces V_A, W_A in $\mathrm{Lat}(A)$. Same as above, we obtain an element $(Y_2, \boldsymbol{y}_2) \in \mathcal{G}$ such that $(A, \boldsymbol{a})(Y_2, \boldsymbol{y}_2)(B, \boldsymbol{b}) = (I, \boldsymbol{x}_2)$ with $\boldsymbol{x}_2$ satisfying

$$1 - \frac{|\boldsymbol{v}_A^\top \boldsymbol{x}_2|}{\|\boldsymbol{v}_A\| \|\boldsymbol{x}_2\|} < \varepsilon,\ (\boldsymbol{x}_1)_{V_A}^\top \boldsymbol{x}_{V_A} > 0,\ (\boldsymbol{x}_1)_{W_A}^\top \cdot \boldsymbol{x}_{W_A} > 0. \tag{5.5}$$

Here, $\boldsymbol{v}_A$ is a non-zero vector in V_A. In other words, when ε is sufficiently small, the acute angle between $\boldsymbol{x}_2$ and V_A is also sufficiently small. Hence, one can take ε such that $\boldsymbol{x}_2$ and $-\boldsymbol{x}_1$ lie in the same cone out of the four cut by V_B and W_B. Furthermore, $\boldsymbol{x}_2$ and $\boldsymbol{x}_1$ lie in same cone out of the four cut by V_A and W_A.

We follow this pattern and apply Lemma 5.4.1 again on $(A', \boldsymbol{a}') \coloneqq (A, \boldsymbol{a})(Y_2, \boldsymbol{y}_2)$ and $(B, \boldsymbol{b})$, and to the subspaces V_B, W_B. This yields an element $(Y_3, \boldsymbol{y}_3) \in \langle \mathcal{G} \rangle$ such that $(A, \boldsymbol{a})(Y_3, \boldsymbol{y}_3)(B, \boldsymbol{b}) = (I, \boldsymbol{x}_3)$ with $\boldsymbol{x}_3$ very close to V_B, but lies in the same cone cut by V_B and W_B as $\boldsymbol{x}_2$.

Finally we apply Lemma 5.4.1 again on $(A, \boldsymbol{a})$ and $(B', \boldsymbol{b}') \coloneqq (Y_3, \boldsymbol{y}_3)(B, \boldsymbol{b})$, and to the subspaces V_A, W_A. This yields $(Y_4, \boldsymbol{y}_4) \in \langle \mathcal{G} \rangle$ such that $(A, \boldsymbol{a})(Y_4, \boldsymbol{y}_4)(B, \boldsymbol{b}) = (I, \boldsymbol{x}_4)$ with $\boldsymbol{x}_4$ very close to V_A, but lies in the same cone cut by V_A and W_A as $\boldsymbol{x}_3$.

When ε is small enough, the angles of the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4$ are sufficiently close to V_B, V_A, V_B, V_A, with opposing directions. Hence, they generate $\mathbb{Q}^2$ as a $\mathbb{Q}_{\geq 0}$-cone. In other

words, there exist *positive* integers n_1, n_2, n_3, n_4 such that $n_1\boldsymbol{x}_1 + n_2\boldsymbol{x}_2 + n_3\boldsymbol{x}_3 + n_4\boldsymbol{x}_4 = \mathbf{0}$. Therefore,

$$(I, \boldsymbol{x}_1)^{n_1}(I, \boldsymbol{x}_2)^{n_2}(I, \boldsymbol{x}_3)^{n_3}(I, \boldsymbol{x}_4)^{n_4} = (I, \mathbf{0}).$$

Since the element $(A, \boldsymbol{a})$ can be represented as a full-image word over $\mathcal{G}$, the element $(I, \boldsymbol{x}_1)$ and hence $(I, \mathbf{0})$ can be represented by a full-image word as well. Therefore, $\langle \mathcal{G} \rangle$ is a group.

(b) Next consider the case (Lat(A), Lat(B), Lat(A), Lat(B)). Denote by V_A, V_B, W_A, W_B these subspaces in this order. Without loss of generality suppose $\boldsymbol{x} \notin V_B \cup W_B$. The strategy is exactly the same as the previous case, see Figure 5.6. However, in the present case, we need to apply Lemma 5.4.1 for a total of six times, where V will be the subspaces $V_B, W_A, W_B, V_A, V_B, W_A$ respectively.

In this way, we obtain $(I, \boldsymbol{x}_1)$, $(I, \boldsymbol{x}_2)$, $(I, \boldsymbol{x}_3)$, $(I, \boldsymbol{x}_4)$, $(I, \boldsymbol{x}_5)$, $(I, \boldsymbol{x}_6)$ with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_6$ generating $\mathbb{Q}^2$ as a $\mathbb{Q}_{\geq 0}$-cone. Hence, there exist positive integers $n_1, \ldots, n_6$ such that

$$(I, \boldsymbol{x}_1)^{n_1} \cdots (I, \boldsymbol{x}_6)^{n_6} = (I, \mathbf{0}).$$

Similarly, since the element $(A, \boldsymbol{a})$ (and hence $(I, \mathbf{0})$) can be represented as a full-image word over $\mathcal{G}$, the semigroup $\langle \mathcal{G} \rangle$ is a group.

This concludes case 1.

Case 2. One of A and B is a shear. The approach is similar to the previous cases, but we have to apply Lemma 5.4.1 and Lemma 5.4.2 alternately. See Figure 5.7 for an illustration.

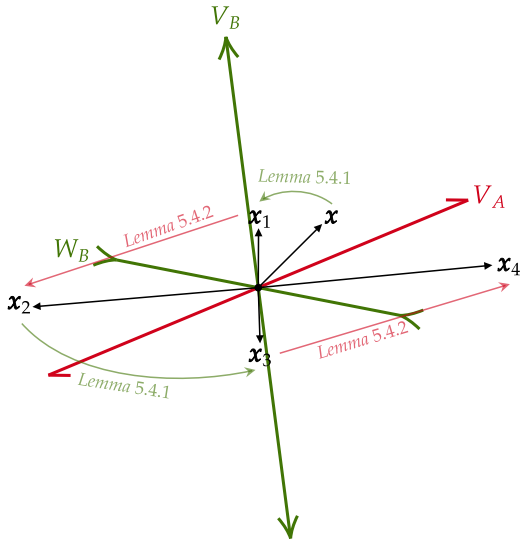

Figure 5.7: Illustration for case 2.

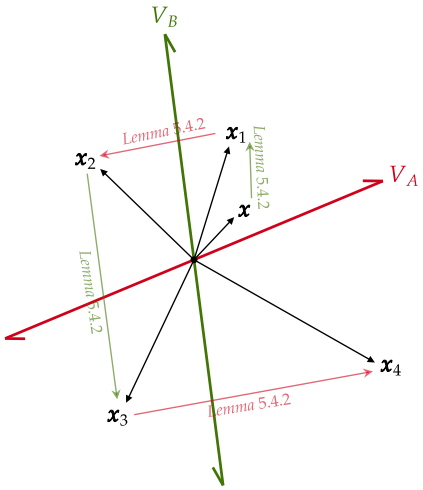

Figure 5.8: Illustration for case 3.

Without loss of generality, let A be the shear and B be the positive scale. Starting from any one-dimension subspace S, rotate S counter-clockwise and consider its sequence of encounters with $\mathrm{Lat}(A)$ and $\mathrm{Lat}(B)$ before finishing a full cycle. The sequence of encounters must a cyclic permutation of $(\mathrm{Lat}(A), \mathrm{Lat}(B), \mathrm{Lat}(B))$. Denote by V_A, V_B, W_B be these subspaces in this order. Without loss of generality suppose $\boldsymbol{x} \notin V_B \cup W_B$, the case where $\boldsymbol{x} \notin V_A$ is analogous.

For a small enough ε, apply Lemma 5.4.1 to $\boldsymbol{x}$ to obtain $\boldsymbol{x}_1$ that is sufficiently close to V_B; then apply Lemma 5.4.2 to $\boldsymbol{x}_1$ to obtain $\boldsymbol{x}_2$ that is sufficiently close to V_A and lies in the same cone cut by V_B and W_B as $-\boldsymbol{x}_1$; then apply Lemma 5.4.1 to $\boldsymbol{x}_2$ to obtain $\boldsymbol{x}_3$ that is sufficiently close to V_B and lies in a different halfspace cut by V_A than $\boldsymbol{x}_2$; finally, apply Lemma 5.4.2 to $\boldsymbol{x}_3$ to obtain $\boldsymbol{x}_4$ that is sufficiently close to V_A and lies in the same cone cut by V_B and W_B as $-\boldsymbol{x}_3$. In this way, we obtain $(I, \boldsymbol{x}_1), (I, \boldsymbol{x}_2), (I, \boldsymbol{x}_3), (I, \boldsymbol{x}_4)$ with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ generating $\mathbb{Q}^2$ as a $\mathbb{Q}_{\geq 0}$-cone. Hence, there exist positive integers $n_1, \ldots, n_4$ such that

$$(I, \boldsymbol{x}_1)^{n_1} \cdots (I, \boldsymbol{x}_4)^{n_4} = (I, \boldsymbol{0}).$$

Since the element $(A, \boldsymbol{a})$ (and hence $(I, \boldsymbol{0})$) can be represented as a full-image word over $\mathcal{G}$, the semigroup $\langle \mathcal{G} \rangle$ is a group.

Case 3. Both A and B are shears. The approach is similar to the previous cases, but we have to apply Lemma 5.4.2 only. See Figure 5.8 for an illustration.

Denote by V_A, V_B respectively the elements of $\mathrm{Lat}(A)$ and $\mathrm{Lat}(B)$. Without loss of generality suppose $\boldsymbol{x} \notin V_B$, the case where $\boldsymbol{x} \notin V_A$ is analogous. For a small enough ε, apply Lemma 5.4.2 to $\boldsymbol{x}$ to obtain $\boldsymbol{x}_1$ that is sufficiently close to V_B; then apply Lemma 5.4.2 again to $\boldsymbol{x}_1$ to obtain $\boldsymbol{x}_2$ that is sufficiently close to V_A and lies in a different halfspace cut by V_B than $\boldsymbol{x}_1$; then apply Lemma 5.4.2 to $\boldsymbol{x}_2$ to obtain $\boldsymbol{x}_3$ that is sufficiently close to V_B and lies in a different halfspace cut by V_A than $\boldsymbol{x}_2$; finally, apply Lemma 5.4.2 to $\boldsymbol{x}_3$ to obtain $\boldsymbol{x}_4$ that is sufficiently close to V_A and lies in a different halfspace cut by V_B than $\boldsymbol{x}_3$. In this way, we obtain $(I, \boldsymbol{x}_1), (I, \boldsymbol{x}_2), (I, \boldsymbol{x}_3), (I, \boldsymbol{x}_4)$ with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ generating $\mathbb{Q}^2$ as a $\mathbb{Q}_{\geq 0}$-cone. Hence, there exist positive integers $n_1, \ldots, n_4$ such that

$$(I, \boldsymbol{x}_1)^{n_1} \cdots (I, \boldsymbol{x}_4)^{n_4} = (I, \boldsymbol{0}).$$

Since the element $(A, \boldsymbol{a})$ (and hence $(I, \boldsymbol{0})$) can be represented as a full-image word over $\mathcal{G}$, the semigroup $\langle \mathcal{G} \rangle$ is a group. □

5.5 Virtual solvability

In this section we prove Proposition 5.3.4:

Proposition 5.3.4. *Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$, such that the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a group. Suppose H is virtually solvable, then exactly one of the following six conditions holds:*

(i) H is the trivial group.
(ii) H contains a non-trivial torsion element.
(iii) $H = \langle A \rangle_{grp}$, where A is a twisted inversion.
(iv) $H = \langle A \rangle_{grp}$, where A is a shear.
(v) $H = \langle A \rangle_{grp}$, where A is an inverting scale.
(vi) $H = \langle A \rangle_{grp}$, where A is a positive scale.

Furthermore, in cases (ii), (iii) and (v), the semigroup $\langle \mathcal{G} \rangle$ is always a group. Overall, it is decidable in PTIME whether $\langle \mathcal{G} \rangle$ is a group.

As in the statement of the proposition, we fix a set $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ of elements in $\mathsf{SA}(2, \mathbb{Z})$, such that $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a virtually solvable group.

Lemma 5.5.1. *Let H be a virtually solvable subgroup of $\mathsf{SL}(2, \mathbb{Z})$. Then H is either finite, or it contains a finite index subgroup H' isomorphic to $\mathbb{Z}$.*

Proof. By Theorem 5.2.3, let $F \leq \mathsf{SL}(2, \mathbb{Z})$ be a free subgroup over two generators, such that the index $[\mathsf{SL}(2, \mathbb{Z}) : F]$ is finite. Then $[H : F \cap H] \leq [\mathsf{SL}(2, \mathbb{Z}) : F]$, so the group $H' \coloneqq F \cap H$ is of finite index in H. But H' is a subgroup of the free group F, so it must be free by Theorem 5.2.2. On the other hand, H' is a subgroup of the virtually solvable group H, so H' is virtually solvable. Since $H' \leq \mathsf{SL}(2, \mathbb{Z})$ is virtually solvable and free, it must be abelian by Theorem 5.3.2. Therefore either $H' \cong \mathbb{Z}$ or $H' = \{I\}$; in the second case, H is finite. □

In case H is finite, it is either trivial or it contains a non-trivial torsion element. We further analyse the case where H contains a finite index subgroup H' isomorphic to $\mathbb{Z}$. We need a deep result from Swan:

Lemma 5.5.2 (Swan [92, Theorem B])**.** *A group that is torsion-free and virtually free is free.*

Lemma 5.5.3. *Let $H \leq \mathsf{SL}(2, \mathbb{Z})$ be a group which contains a finite index subgroup H' isomorphic to $\mathbb{Z}$. Then either H contains a non-trivial torsion element, or it is also isomorphic to $\mathbb{Z}$.*

Proof. Suppose H is torsion-free. Since H is virtually free and torsion-free, by Lemma 5.5.2, it is free. Hence either $H \cong \mathbb{Z}$, or H is a non-abelian free group. But a non-abelian free group is not virtually solvable (Theorem 5.3.2), so it cannot contain a finite index subgroup H' isomorphic to $\mathbb{Z}$. Therefore, we must have $H \cong \mathbb{Z}$. □

By Lemma 5.5.1 and 5.5.3, we can already prove the first part of Proposition 5.3.4. In particular, by Lemma 5.5.1, H is either finite or contains a finite index subgroup isomorphic to $\mathbb{Z}$. If H is finite, then either it is trivial or it contains a non-trivial torsion element. If H contains a finite index subgroup isomorphic to $\mathbb{Z}$, then by Lemma 5.5.3, either it contains a torsion element, or it is isomorphic to $\mathbb{Z}$. Therefore, we have proved that exactly one of the following holds.

(1) H is the trivial group.

(2) H contains a non-trivial torsion element.

(3) H is isomorphic to $\mathbb{Z}$.

In case (3), the generator of H is either a twisted inversion, a shear, an inverting scale, or a positive scale. This corresponds to the cases (iii)-(vi) of Proposition 5.3.4. We have thus proved the first part of Proposition 5.3.4. The following lemma shows that one can decide which of the six cases in Proposition 5.3.4 is true.

Lemma 5.5.4. *Let $A_1, \ldots, A_K$ be matrices in $\mathsf{SL}(2, \mathbb{Z})$. The following can be done in PTIME:*

(i) decide whether the group *$H \coloneqq \langle A_1, \ldots, A_K \rangle_{grp}$ is trivial.*

(ii) decide whether H is isomorphic to $\mathbb{Z}$.

(iii) compute a generator A of H in case $H \cong \mathbb{Z}$.

Proof. (i). H is trivial if and only if $A_i = I$ for all i.

(ii). First we check whether H is abelian, this is done simply by checking whether $A_i A_j = A_j A_i$ for all $1 \leq i, j \leq K$. If H is not abelian, then it is not isomorphic to $\mathbb{Z}$.

Suppose H is abelian. Then the group homomorphism

$$\begin{aligned} \varphi \colon \mathbb{Z}^K &\longrightarrow H \\ (n_1, \ldots, n_K) &\longmapsto A_1^{n_1} \cdots A_K^{n_K} \end{aligned}$$

is surjective. The kernel

$$\Lambda \coloneqq \ker(\varphi) = \{(n_1, \ldots, n_K) \mid A_1^{n_1} \cdots A_K^{n_K} = I\}$$

is a finitely generated subgroup of $\mathbb{Z}^K$. A $\mathbb{Z}$-basis of Λ is computable in PTIME by a classic result of Babai et al. [4].

Let $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_m$ be a $\mathbb{Z}$-basis of Λ. Define

$$\overline{\Lambda} \coloneqq \left\{ \boldsymbol{n} \in \mathbb{Z}^K \;\middle|\; \boldsymbol{n} = q_1\boldsymbol{\ell}_1 + \cdots + q_m\boldsymbol{\ell}_m, \text{ where } q_1, \ldots, q_m \in \mathbb{Q} \right\},$$

which is the lattice of integer points in the $\mathbb{Q}$-linear space spanned by Λ. A basis of $\overline{\Lambda}$ is effectively computable in PTIME using the *Hermite Normal Form* [23, Chapter 14]. It is then decidable in PTIME whether $\overline{\Lambda} = \Lambda$, again using the Hermite Normal Form.

We claim that H is torsion-free if and only if $\overline{\Lambda} = \Lambda$. Indeed, let $T = A_1^{n_1} \cdots A_K^{n_K}$ be a non-trivial torsion element of H, then $T^m = I$ for some $m > 1$. Therefore $A_1^{mn_1} \cdots A_K^{mn_K} = I$, so $(mn_1, \ldots, mn_K) \in \Lambda$. This shows $(n_1, \ldots, n_K) \in \overline{\Lambda}$. But $T \neq I$, so $(n_1, \ldots, n_K) \notin \Lambda$. Hence, $T = A_1^{n_1} \cdots A_K^{n_K}$ is a non-trivial torsion element if and only if $(n_1, \ldots, n_K) \in \overline{\Lambda} \setminus \Lambda$.

This proves the claim. Therefore it is decidable in PTIME whether H contains a non-trivial torsion element. By Lemma 5.5.3, it is decidable in PTIME whether $H \cong \mathbb{Z}$.

(iii). If H is isomorphic to $\mathbb{Z}$, then the quotient group $\mathbb{Z}^K/\Lambda \cong H$ is isomorphic to $\mathbb{Z}$. Using the Hermite Normal Form, one can in PTIME compute an element $\boldsymbol{x} = (x_1, \ldots, x_K) \in \mathbb{Z}^K$ such that $\boldsymbol{x}, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_m$ form a $\mathbb{Z}$-basis of $\mathbb{Z}^K$. (Recall that $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_m$ are a $\mathbb{Z}$-basis of Λ.) Then $\boldsymbol{x} + \Lambda$ generates $\mathbb{Z}^K/\Lambda$. Therefore $A \coloneqq \varphi(\boldsymbol{x}) = A_1^{x_1} \cdots A_K^{x_K}$ generates H. □

We now proceed to prove the PTIME decidability claim of Proposition 5.3.4 for all six cases. We may note that in the cases (iii)-(vi), the group $\langle \mathcal{G} \rangle_{grp}$ is actually metabelian, hence the Group Problem for $\mathcal{G}$ is already decidable by Theorem 4.1.2. However, the difficult part is to obtain an algorithm that runs in PTIME. Let $H \coloneqq \langle A_1, \ldots, A_K \rangle$.

5.5.1 H is trivial

In this case, $\mathcal{G} = \{(I, \boldsymbol{a}_1), \cdots, (I, \boldsymbol{a}_K)\}$.

Proposition 5.5.5. *Let $\mathcal{G} = \{(I, \boldsymbol{a}_1), \cdots, (I, \boldsymbol{a}_K)\}$. Then $\langle \mathcal{G} \rangle$ is a group if and only if the equation*

$$n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + \cdots n_K\boldsymbol{a}_K = \boldsymbol{0} \tag{5.6}$$

has a solution $(n_1, \ldots, n_K) \in \mathbb{Z}_{>0}^K$. In particular, this is decidable in PTIME.

Proof. Note that the matrices in $\mathcal{G}$ commute. Therefore by Lemma 2.2.1, $\langle \mathcal{G} \rangle$ is a group if and only if Equation (5.6) has a solution $(n_1, \ldots, n_K) \in \mathbb{Z}_{>0}^K$. By the homogeneity of Equation (5.6),

it has a solution in $\mathbb{Z}_{>0}^K$ if and only if it has a solution in $\mathbb{Q}_{>0}^K$. This is decidable in PTIME by linear programming. □

5.5.2 H contains a torsion element

We show that $\langle \mathcal{G} \rangle$ is always a group in case H contains a non-trivial torsion element.

Proposition 5.5.6. *Let $\mathcal{G} \coloneqq \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$. If the semigroup $\langle A_1, \ldots, A_K \rangle$ is a group containing a non-trivial torsion element, then $\langle \mathcal{G} \rangle$ is a group.*

Proof. Suppose $\langle A_1, \ldots, A_K \rangle$ contains a non-trivial torsion element T. Let $m > 1$ be such that $T^m = I$. Let $\boldsymbol{t} \in \mathbb{Z}^2$ be a vector such that $(T, \boldsymbol{t})$ is an element in $\langle \mathcal{G} \rangle$ represented by a full-image word (such a word exists by Lemma 5.2.1). Then

$$(T, \boldsymbol{t})^m = (T^m, (I + T + \cdots + T^{m-1})\boldsymbol{t}) = (I, (I - T)^{-1}(I - T^m)\boldsymbol{t}) = (I, \boldsymbol{0}).$$

Here, $I - T$ is invertible because $T \in \mathsf{SL}(2, \mathbb{Z})$ is non-trivial torsion, so the eigenvalues of T are all different from one (see the classification of element of $\mathsf{SL}(2, \mathbb{Z})$ in Section 5.2). We conclude that $(I, \boldsymbol{0})$ can be represented by a full-image word over the alphabet $\mathcal{G}$. Hence, $\langle \mathcal{G} \rangle$ is a group by Lemma 2.2.1. □

In the next four cases, H is isomorphic to $\mathbb{Z}$. Let A be a generator of H, computable in PTIME by Lemma 5.5.4. The *Jordan Normal Form* of A can be computed in PTIME [40]. One can directly obtain from its Jordan Normal Form whether A is a twisted inversion, a shear or a scale.

5.5.3 H is generated by a twisted inversion A

We show that if A is a twisted inversion, then $\langle \mathcal{G} \rangle$ is always a group.

Proposition 5.5.7. *If the generator A of the group $\langle A_1, \ldots, A_K \rangle \cong \mathbb{Z}$ is a twisted inversion, then $\langle \mathcal{G} \rangle$ is a group.*

Proof. Since $\langle A_1, \ldots, A_K \rangle$ is isomorphic to $\mathbb{Z}$, we have $A, A^{-1} \in \langle A_1, \ldots, A_K \rangle$. As $\langle A_1, \ldots, A_K \rangle$ is a group, let $(A, \boldsymbol{a})$ and $(A^{-1}, \boldsymbol{b})$ be elements of $\langle \mathcal{G} \rangle$ represented by full-image words over $\mathcal{G}$.

We claim that

$$(A, \boldsymbol{a})^2 \cdot (A^{-1}, \boldsymbol{b})^3 \cdot (A, \boldsymbol{a})^2 \cdot (A^{-1}, \boldsymbol{b}) = (I, \boldsymbol{0}).$$

By direct computation, $(A, \boldsymbol{a})^2 \cdot (A^{-1}, \boldsymbol{b})^3 \cdot (A, \boldsymbol{a})^2 \cdot (A^{-1}, \boldsymbol{b}) = (I, A^{-1}(A + I)^2\boldsymbol{a} + (A + I)^2\boldsymbol{b})$. But since A is a twisted inversion, we have $(A + I)^2 = 0$. Therefore, $(A, \boldsymbol{a})^2 \cdot (A^{-1}, \boldsymbol{b})^3 \cdot (A, \boldsymbol{a})^2 \cdot$

$(A^{-1}, \boldsymbol{b}) = (I, \mathbf{0})$. We conclude that $(I, \mathbf{0})$ can be represented as a full-image word over $\mathcal{G}$. Hence, $\langle \mathcal{G} \rangle$ is a group. □

5.5.4 H is generated by a shear A

If A is a shear, the main idea in this case is that the semigroup generated by $\mathcal{G}$ can be embedded as a subsemigroup of the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$. Recall from Section 3.1 that the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$ is defined as:

$$\mathrm{H}_3(\mathbb{Q}) \coloneqq \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \;\middle|\; a, b, c \in \mathbb{Q} \right\}.$$

Since A is a shear, it is triangularizable over $\mathbb{Q}$. Since $A_1, \ldots, A_K$ all commute, they all have the same eigenvalue one, and are simultaneously triangularizable. Let P be a matrix (with entries in $\mathbb{Q}$) such that $P^{-1}A_iP$ are all of the form $\begin{pmatrix} 1 & \lambda_i \\ 0 & 1 \end{pmatrix}$ with $\lambda_i \in \mathbb{Q}$. Then, we have the following conjugation:

$$\begin{pmatrix} P^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_i & \boldsymbol{a}_i \\ 0 & 1 \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} P^{-1}A_iP & P^{-1}\boldsymbol{a}_i \\ 0 & 1 \end{pmatrix} \in \mathrm{H}_3(\mathbb{Q}).$$

This shows that the map

$$\varphi : \mathcal{G} \to \mathrm{H}_3(\mathbb{Q}), \quad (A_i, \boldsymbol{a}_i) \mapsto \begin{pmatrix} P^{-1}A_iP & P^{-1}\boldsymbol{a}_i \\ 0 & 1 \end{pmatrix}, \tag{5.7}$$

extends to an injective semigroup homomorphism from $\langle \mathcal{G} \rangle$ to $\mathrm{H}_3(\mathbb{Q})$. Therefore, $\langle \mathcal{G} \rangle$ is a group if and only if $\langle \varphi(\mathcal{G}) \rangle$ is a group. The Group Problem in $\mathrm{H}_3(\mathbb{Q})$ is decidable in PTIME as a special case of Theorem 3.1.1:

Theorem 5.5.8 (Corollary of Theorem 3.1.1)**.** *The Group Problem in* $\mathrm{H}_3(\mathbb{Q})$ *is decidable in PTIME.*

Proof. Recall that $\mathrm{H}_3(\mathbb{Q})$ is $\mathsf{UT}(3, \mathbb{Q})$, where the Invertible Subset can be computed in PTIME by Algorithm 3.1 (Theorem 3.1.1). By Lemma 2.2.3(iii), the Group Problem can be decided in PTIME. □

The elements $\varphi(A_1), \ldots, \varphi(A_K)$ can be computed in PTIME. By Theorem 5.5.8, we immediately obtain:

Corollary 5.5.9. *If the generator A of the group $\langle A_1, \ldots, A_K \rangle \cong \mathbb{Z}$ is a shear, then the Group Problem for $\langle \mathcal{G} \rangle$ is equivalent to the Group Problem for $\langle \varphi(\mathcal{G}) \rangle$, which is decidable in PTIME.*

5.5.5 H is generated by an inverting scale A

We show that if A is an inverting scale, then $\langle \mathcal{G} \rangle$ is a group.

Proposition 5.5.10. *If the generator A of the group $\langle A_1, \dots, A_K \rangle \cong \mathbb{Z}$ is an inverting scale, then $\langle \mathcal{G} \rangle$ is a group.*

Proof. Since $\langle A_1, \dots, A_K \rangle$ is a group, we have $A, A^{-1} \in \langle A_1, \dots, A_K \rangle$. Recall that $\lambda < -1$ is the smaller eigenvalue of A. Let V and W be the invariant spaces of A corresponding to the eigenvalues λ and λ^{-1}, and let $\boldsymbol{x}_V, \boldsymbol{x}_W$ be non-zero vectors respectively in V and W. In particular we have $A\boldsymbol{x}_V = \lambda \boldsymbol{x}_V, A\boldsymbol{x}_W = \lambda^{-1}\boldsymbol{x}_W$. As $\langle A_1, \dots, A_K \rangle$ is a group, let $(A, a_+\boldsymbol{x}_V + b_+\boldsymbol{x}_W)$ and $(A^{-1}, a_-\boldsymbol{x}_V + b_-\boldsymbol{x}_W)$ be elements of $\langle \mathcal{G} \rangle$ represented by full-image words.

Then for any $m > 0$, the elements

$$\begin{aligned}
(I, \boldsymbol{v}_m) &\coloneqq (A, a_+\boldsymbol{x}_V + b_+\boldsymbol{x}_W)^m \cdot (A^{-1}, a_-\boldsymbol{x}_V + b_-\boldsymbol{x}_W)^m \\
&= \left(A^m, \sum_{i=0}^{m-1} \lambda^i a_+ \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i} b_+ \boldsymbol{x}_W \right) \cdot \left(A^{-m}, \sum_{i=0}^{m-1} \lambda^{-i} a_- \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^i b_- \boldsymbol{x}_W \right) \\
&= \left(I, \left(\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}} \right)(1-\lambda^m)\boldsymbol{x}_V + \left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} \right)(1-\lambda^{-m})\boldsymbol{x}_W \right)
\end{aligned}$$

and

$$\begin{aligned}
(I, \boldsymbol{w}_m) &\coloneqq (A^{-1}, a_-\boldsymbol{x}_V + b_-\boldsymbol{x}_W)^m \cdot (A, a_+\boldsymbol{x}_V + b_+\boldsymbol{x}_W)^m \\
&= \left(A^{-m}, \sum_{i=0}^{m-1} \lambda^{-i} a_- \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^i b_- \boldsymbol{x}_W \right) \cdot \left(A^m, \sum_{i=0}^{m-1} \lambda^i a_+ \boldsymbol{x}_V + \sum_{i=0}^{m-1} \lambda^{-i} b_+ \boldsymbol{x}_W \right) \\
&= \left(I, \left(\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}} \right)(\lambda^{-m}-1)\boldsymbol{x}_V + \left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} \right)(\lambda^m-1)\boldsymbol{x}_W \right)
\end{aligned}$$

are in $\langle \mathcal{G} \rangle$ and represented by full-image words.

Consider four cases:

1. $\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}} = \frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} = 0$. In this case we have directly $\boldsymbol{v}_m = \boldsymbol{0}$ for all m. So $(I, \boldsymbol{0})$ can be represented by a full-image word.

2. $\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}} = 0$, but $\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} \neq 0$. When m is even, $\boldsymbol{w}_m$ is a positive multiple of $\left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} \right)\boldsymbol{x}_W$, and when n is odd, $\boldsymbol{w}_n$ is a negative multiple of $\left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} \right)\boldsymbol{x}_W$. Therefore, there exists positive real numbers r_1, r_2 such that $r_1\boldsymbol{w}_m + r_2\boldsymbol{w}_n = \boldsymbol{0}$. Since $\boldsymbol{w}_m, \boldsymbol{w}_n$ have integer entries, there even exist positive *integers* n_1, n_2 such that $n_1\boldsymbol{w}_m + n_2\boldsymbol{w}_n = \boldsymbol{0}$. Therefore $(I, \boldsymbol{0}) = (I, \boldsymbol{w}_m)^{n_1}(I, \boldsymbol{w}_n)^{n_2}$ can be represented by a full-image word.

3. $\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda} = 0$, but $\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}} \neq 0$. This case is the exact symmetry of the previous case.

4. Both $\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}}$ and $\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda}$ are non-zero. In this case, consider the vectors $\boldsymbol{v}_{2m}$, $\boldsymbol{v}_{2m+1}$, $\boldsymbol{w}_{2m}$, $\boldsymbol{w}_{2m+1}$ when m tends to infinity. Since $\lambda < -1$, we have $\lim_{m\to\infty}(1-\lambda^{2m}) = -\infty$, $\lim_{m\to\infty}(1-\lambda^{2m+1}) = +\infty$, $\lim_{m\to\infty}(1-\lambda^{-2m}) = 1$, $\lim_{m\to\infty}(1-\lambda^{-2m-1}) = 1$. Therefore, the *direction* of $\boldsymbol{v}_{2m}$ tends towards $-\left(\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}}\right)\boldsymbol{x}_V$, the direction of $\boldsymbol{v}_{2m+1}$ tends towards $\left(\frac{a_+}{1-\lambda} - \frac{a_-}{1-\lambda^{-1}}\right)\boldsymbol{x}_V$, the direction of $\boldsymbol{w}_{2m}$ tends towards $\left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda}\right)\boldsymbol{x}_W$, the direction of $\boldsymbol{w}_{2m+1}$ tends towards $-\left(\frac{b_+}{1-\lambda^{-1}} - \frac{b_-}{1-\lambda}\right)\boldsymbol{x}_W$. Hence, when m is large enough, the four vectors $\boldsymbol{v}_{2m}$, $\boldsymbol{v}_{2m+1}$, $\boldsymbol{w}_{2m}$, $\boldsymbol{w}_{2m+1}$ in $\mathbb{Z}^2$ generate $\mathbb{Q}^2$ as a $\mathbb{Q}_{\geq 0}$-cone. Hence, there exist positive integers $t_1, \ldots, t_4$ such that

$$(I, \boldsymbol{v}_{2m})^{t_1} \cdots (I, \boldsymbol{w}_{2m+1})^{t_4} = (I, \boldsymbol{0}).$$

   Therefore $(I, \boldsymbol{0})$ can be represented by a full-image word.

In all cases, $(I, \boldsymbol{0})$ can be represented by a full-image word, so $\langle \mathcal{G} \rangle$ is a group. □

5.5.6 H is generated by a positive scale A

This is the most technical case in this section. We show that if A is a positive scale, then we can decide whether $\langle \mathcal{G} \rangle$ is a group in PTIME. Let $P = (\boldsymbol{x}_V, \boldsymbol{x}_W) \in \mathsf{SL}(2, \mathbb{R})$ be a change of basis matrix such that $P^{-1}AP = A'$, where A' is diagonal and can be written as

$$A' = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

with $\lambda > 1$. For $i = 1, \ldots, K$, let $A_i' \coloneqq P^{-1}A_iP$ and $\boldsymbol{a}_i' \coloneqq P^{-1}\boldsymbol{a}_i$, with

$$A_i' = \begin{pmatrix} \lambda^{z_i} & 0 \\ 0 & \lambda^{-z_i} \end{pmatrix}, z_i \in \mathbb{Z}, \quad \boldsymbol{a}_i' = \begin{pmatrix} a_i \\ b_i \end{pmatrix}.$$

These are the forms of A_i and $\boldsymbol{a}_i$ under the new basis $(\boldsymbol{x}_V, \boldsymbol{x}_W)$. In particular we have $A\boldsymbol{x}_V = \lambda\boldsymbol{x}_V, A\boldsymbol{x}_W = \lambda^{-1}\boldsymbol{x}_W$.

Define the sets

$$J_+ \coloneqq \{i \mid z_i > 0\},\ J_- \coloneqq \{i \mid z_i < 0\},\ J_0 \coloneqq \{i \mid z_i = 0\}.$$

Then $J_+ \cup J_- \cup J_0 = \{1, \ldots, K\}$. Since $\langle A_1, \ldots, A_K \rangle$ is isomorphic to $\mathbb{Z}$, we have $J_+ \neq \emptyset, J_- \neq \emptyset$.

Divide the set $\mathbb{R}^2$ into nine parts:

$$\mathcal{R}_{++} \coloneqq \{(x,y) \mid x>0, y>0\}, \mathcal{R}_{+0} \coloneqq \{(x,y) \mid x>0, y=0\}, \mathcal{R}_{+-} \coloneqq \{(x,y) \mid x>0, y<0\},$$
$$\mathcal{R}_{0+} \coloneqq \{(x,y) \mid x=0, y>0\}, \mathcal{R}_{00} \coloneqq \{(x,y) \mid x=0, y=0\}, \mathcal{R}_{0-} \coloneqq \{(x,y) \mid x=0, y<0\},$$
$$\mathcal{R}_{-+} \coloneqq \{(x,y) \mid x<0, y>0\}, \mathcal{R}_{-0} \coloneqq \{(x,y) \mid x<0, y=0\}, \mathcal{R}_{--} \coloneqq \{(x,y) \mid x<0, y<0\}.$$

Each part is called a *cell*. See Figure 5.9.

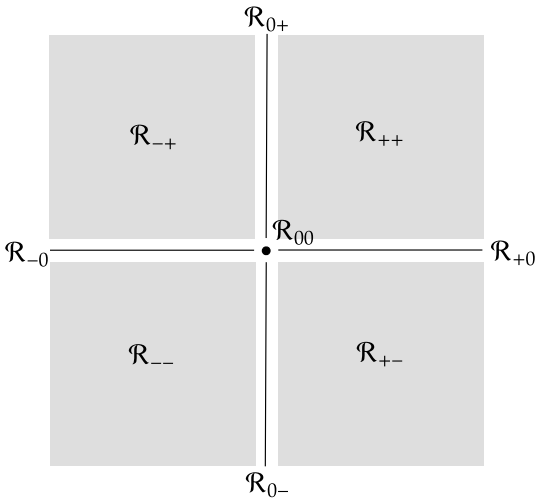

Figure 5.9: Cells.

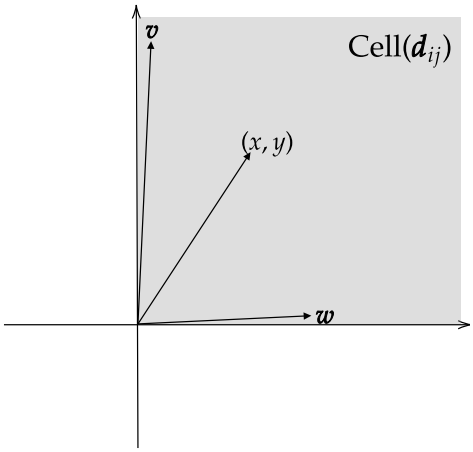

Figure 5.10: Illustration for Lemma 5.5.11.

For an element $\boldsymbol{x} \in \mathbb{R}^2$, denote by $\mathrm{Cell}(\boldsymbol{x})$ the cell which it belongs to. Writing $\boldsymbol{x} = (x_1, x_2)^\top$, it is easy to see that

$$\mathrm{Cell}(\boldsymbol{x}) = \left\{ (r_1 x_1, r_2 x_2)^\top \;\middle|\; r_1, r_2 \in \mathbb{R}_{>0} \right\}. \tag{5.8}$$

In particular, $\boldsymbol{x}$ and $A'^z \boldsymbol{x}$ are in the same cell for all integers z.

For each $i \in J_+ \cup J_- \cup J_0$, denote

$$n_i \coloneqq \begin{cases} |z_i| & z_i \neq 0 \\ 1 & z_i = 0. \end{cases}$$

For each tuple $(i,j) \in J_+ \times J_-$, define the vector

$$\boldsymbol{d}_{ij} \coloneqq (d_{ija}, d_{ijb})^\top \in \mathbb{R}^2$$

where

$$d_{ija} \coloneqq -\frac{a_i}{1-\lambda^{n_i}} + \frac{a_j}{1-\lambda^{-n_j}}, \;\; d_{ijb} \coloneqq \frac{b_i}{1-\lambda^{n_i}} - \frac{b_j}{1-\lambda^{-n_j}}.$$

These expressions are defined so that

$$(A'_i, \boldsymbol{a}'_i)^{n_j}(A'_j, \boldsymbol{a}'_j)^{n_i} = \left(I, (1-\lambda^{-n_i n_j})\cdot(\lambda^{n_i n_j} d_{ija}, d_{ijb})^\top\right).$$

For each element $k \in J_0$, define the vector

$$\boldsymbol{e}_k \coloneqq (a_k, b_k)^\top \in \mathbb{R}^2.$$

Similarly, this expression is defined so that

$$(A'_k, \boldsymbol{a}'_k)^{n_k} = (I, \boldsymbol{e}_k).$$

Denote by $\mathcal{G}'$ the new alphabet $\{(A'_1, \boldsymbol{a}'_1), \ldots, (A'_K, \boldsymbol{a}'_K)\}$. The following lemma shows that the rays $\boldsymbol{v}\mathbb{R}_{>0}$ such that $(I, \boldsymbol{v}) \in \langle \mathcal{G}' \rangle$ can "fill up" the cells $\mathrm{Cell}(\boldsymbol{d}_{ij}), (i,j) \in J_+ \times J_-$.

Lemma 5.5.11. *Let (i,j) be a pair in $J_+ \times J_-$ and $(x,y)^\top \in \mathbb{R}^2$ be a vector in $\mathrm{Cell}(\boldsymbol{d}_{ij})$. If $\lambda > 1$, then there exist elements $(I, \boldsymbol{v})$ and $(I, \boldsymbol{w})$ in $\langle \mathcal{G}' \rangle$, where the vector $(x,y)^\top$ can be written as $r_1\boldsymbol{v} + r_2\boldsymbol{w}$ for some $r_1, r_2 \in \mathbb{R}_{>0}$. Furthermore, $(I, \boldsymbol{v})$ and $(I, \boldsymbol{w})$ are represented by words over the alphabet $\mathcal{G}'$, such that the letters $(A'_i, \boldsymbol{a}'_i)$ and $(A'_j, \boldsymbol{a}'_j)$ both occur. See Figure 5.10 for an illustration.*

Proof. For any $p \in \mathbb{Z}_{>0}$, denote

$$(I, \boldsymbol{v}_p) \coloneqq (A'_i, \boldsymbol{a}'_i)^{pn_j}(A'_j, \boldsymbol{a}'_j)^{pn_i}, \quad (I, \boldsymbol{w}_p) \coloneqq (A'_j, \boldsymbol{a}'_j)^{pn_i}(A'_i, \boldsymbol{a}'_i)^{pn_j}.$$

By direct computation, we have

$$\boldsymbol{v}_p = (1-\lambda^{-pn_i n_j})\cdot(\lambda^{pn_i n_j} d_{ija}, d_{ijb})^\top, \quad \boldsymbol{w}_p = (1-\lambda^{-pn_i n_j})\cdot(d_{ija}, \lambda^{pn_i n_j} d_{ijb})^\top. \tag{5.9}$$

Since $\lambda > 1$, we have $1 - \lambda^{-pn_i n_j} > 0$.

If $\mathrm{Cell}(\boldsymbol{d}_{ij})$ is of dimension one or zero, then either $d_{ija} = 0$ or $d_{ijb} = 0$. In both cases, $(x,y)^\top, \boldsymbol{v}_p, \boldsymbol{w}_p$ are linearly dependant and have the same direction, so $(x,y)^\top$ can be written as $r_1\boldsymbol{v}_p + r_2\boldsymbol{w}_p$ for some $r_1, r_2 \in \mathbb{R}_{>0}$.

If $\mathrm{Cell}(\boldsymbol{d}_{ij})$ is of dimension two, then $d_{ija} \neq 0$ and $d_{ijb} \neq 0$. Let p be large enough so that $\frac{\lambda^{pn_i n_j} d_{ija}}{d_{ijb}} > \frac{x}{y} > \frac{d_{ija}}{\lambda^{pn_i n_j} d_{ijb}}$. Then $(x,y)^\top$ can be written as $r_1\boldsymbol{v}_p + r_2\boldsymbol{w}_p$ for some $r_1, r_2 \in \mathbb{R}_{>0}$. □

A similar lemma can be shown for $\mathrm{Cell}(\boldsymbol{e}_k)$: the rays $\boldsymbol{v}\mathbb{R}_{>0}$ such that $(I, \boldsymbol{v}) \in \langle \mathcal{G}' \rangle$ can fill up the cells $\mathrm{Cell}(\boldsymbol{e}_k), k \in J_0$.

Lemma 5.5.12. *Suppose J_+ and J_- are non-empty. Let k be an element in J_0, and $(x,y)^\top \in \mathbb{R}^2$ be a vector in* $\mathrm{Cell}(\boldsymbol{e}_k)$*. If $\lambda > 1$, then there exist elements $(I,\boldsymbol{v})$ and $(I,\boldsymbol{w})$ in $\langle \mathcal{G}' \rangle$, such that the vector $(x,y)^\top$ can be written as $r_1\boldsymbol{v} + r_2\boldsymbol{w}$ for some $r_1, r_2 \in \mathbb{R}_{>0}$. Furthermore, $(I,\boldsymbol{v})$ and $(I,\boldsymbol{w})$ are represented as words over the alphabet $\mathcal{G}'$, such that the letter $(A'_k, \boldsymbol{a}'_k)$ occurs.*

Proof. Let (i,j) be any pair in $J_+ \times J_-$. For any $p \in \mathbb{Z}_{\geq 0}$, $q \in \mathbb{Z}_{>0}$, denote

$$(I,\boldsymbol{v}_{pq}) \coloneqq A_i'^{pn_j} A_k'^{\,q} A_j'^{pn_i}, \quad (I,\boldsymbol{w}_{pq}) \coloneqq A_j'^{pn_i} A_k'^{\,q} A_i'^{pn_j},$$

By direct computation, we have

$$\boldsymbol{v}_{pq} = (1-\lambda^{-pn_in_j}) \cdot (\lambda^{pn_in_j} d_{ija}, d_{ijb})^\top + q \cdot (\lambda^{pn_in_j} a_k, \lambda^{-pn_in_j} b_k)^\top,$$

$$\boldsymbol{w}_{pq} = (1-\lambda^{-pn_in_j}) \cdot (d_{ija}, \lambda^{pn_in_j} d_{ijb})^\top + q \cdot (\lambda^{-pn_in_j} a_k, \lambda^{pn_in_j} b_k)^\top.$$

Since $\lambda > 1$, we have $1 - \lambda^{-pn_in_j} > 0$.

If $\mathrm{Cell}(\boldsymbol{e}_k)$ has dimension one or zero, then take $p = 0$ and the statement is trivial. Suppose that $\mathrm{Cell}(\boldsymbol{e}_k)$ has dimension two, then we have $d_{ija} \neq 0$ and $d_{ijb} \neq 0$. Fix a large enough p so that $\frac{\lambda^{pn_in_j} a_k}{\lambda^{-pn_in_j} b_k} > \frac{x}{y} > \frac{\lambda^{-pn_in_j} a_k}{\lambda^{pn_in_j} b_k}$. Then $(x,y)^\top$ can be written as $r_1(\lambda^{pn_in_j} a_k, \lambda^{-pn_in_j} b_k)^\top + r_2(\lambda^{-pn_in_j} a_k, \lambda^{pn_in_j} b_k)^\top$ for some $r_1, r_2 \in \mathbb{R}_{>0}$. When q tends towards infinity, the direction of the vectors $\boldsymbol{v}_{pq}, \boldsymbol{w}_{pq}$ tends respectively to $(\lambda^{pn_in_j} a_k, \lambda^{-pn_in_j} b_k)^\top$ and $(\lambda^{-pn_in_j} a_k, \lambda^{pn_in_j} b_k)^\top$. Therefore, for large enough q, the vector $(x,y)^\top$ can be written as $r'_1\boldsymbol{v}_{pq} + r'_2\boldsymbol{w}_{pq}$ for some $r'_1, r'_2 \in \mathbb{R}_{>0}$. □

Define the *radical* $\widehat{(A'_i, \boldsymbol{a}'_i)}$ of $(A'_i, \boldsymbol{a}'_i)$ as

$$\widehat{(A'_i, \boldsymbol{a}'_i)} \coloneqq \left(\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} a_i \cdot \frac{1-\lambda}{1-\lambda^{n_i}} \\ b_i \cdot \frac{1-\lambda^{-1}}{1-\lambda^{-n_i}} \end{pmatrix} \right)$$

if $z_i = n_i > 0$, and

$$\widehat{(A'_i, \boldsymbol{a}'_i)} \coloneqq \left(\begin{pmatrix} \lambda^{-1} & 0 \\ 0 & \lambda \end{pmatrix}, \begin{pmatrix} a_i \cdot \frac{1-\lambda^{-1}}{1-\lambda^{-n_i}} \\ b_i \cdot \frac{1-\lambda}{1-\lambda^{n_i}} \end{pmatrix} \right)$$

if $z_i = -n_i < 0$, and

$$\widehat{(A'_i, \boldsymbol{a}'_i)} \coloneqq (A'_i, \boldsymbol{a}'_i)$$

if $z_i = 0$. This is defined so that $\widehat{(A'_i, \boldsymbol{a}'_i)}^{\,n_i} = (A'_i, \boldsymbol{a}'_i)$ in all cases.

Define the alphabet

$$\widehat{\mathcal{G}} \coloneqq \left\{ \widehat{(A'_1, \boldsymbol{a}'_1)}, \ldots, \widehat{(A'_K, \boldsymbol{a}'_K)} \right\}.$$

Define the following union of cells:

$$\mathcal{S} \coloneqq \left(\bigcup_{i \in J_+, j \in J_-} \mathrm{Cell}(\boldsymbol{d}_{ij}) \right) \bigcup \left(\bigcup_{k \in J_0} \mathrm{Cell}(\boldsymbol{e}_k) \right). \tag{5.10}$$

Lemma 5.5.13. *Let $w \coloneqq (C_1, \boldsymbol{c}_1) \cdots (C_M, \boldsymbol{c}_M)$ be a full-image word over the alphabet $\widehat{\mathcal{G}}$, such that $(C_1, \boldsymbol{c}_1) \cdot \cdots (C_M, \boldsymbol{c}_M) = (I, \boldsymbol{x})$. Then there exists a finite non-empty set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\} \subseteq \mathcal{S}$ such that the following conditions are satisfied:*

(i) $r_1 \boldsymbol{s}_1 + \cdots + r_m \boldsymbol{s}_m = \boldsymbol{x}$ for some strictly positive reals $r_1, \ldots, r_m$.

(ii) For each $i \in J_+$, there exist $j \in J_-$ and $\ell \in \{1, \ldots, m\}$, such that $\boldsymbol{s}_\ell \in \mathrm{Cell}(\boldsymbol{d}_{ij})$.

(iii) For each $j \in J_-$, there exist $i \in J_+$ and $\ell \in \{1, \ldots, m\}$, such that $\boldsymbol{s}_\ell \in \mathrm{Cell}(\boldsymbol{d}_{ij})$.

(iv) For each $k \in J_0$, there exist $\ell \in \{1, \ldots, m\}$ such that $\boldsymbol{s}_\ell \in \mathrm{Cell}(\boldsymbol{e}_k)$.

Proof. We call a vector of the form $r_1 \boldsymbol{s}_1 + \cdots + r_m \boldsymbol{s}_m, m \geq 1, r_i \in \mathbb{R}_{>0}, \boldsymbol{s}_i \in \mathcal{S}$, an $\mathbb{R}_{>0}$-*linear combination* of elements in $\mathcal{S}$. For $i = 1, \ldots, M$, write

$$(C_i, \boldsymbol{c}_i) = \left(\begin{pmatrix} \lambda^{t_i} & 0 \\ 0 & \lambda^{-t_i} \end{pmatrix}, \begin{pmatrix} c_i \\ d_i \end{pmatrix} \right),$$

where $t_i \in \{-1, 0, 1\}$. We show that if $(C_1, \boldsymbol{c}_1) \cdot \cdots (C_M, \boldsymbol{c}_M) = (I, \boldsymbol{x})$ then $\boldsymbol{x}$ can be written as an $\mathbb{R}_{>0}$-linear combination of elements in $\mathcal{S}$. We use induction on M. When $M = 1$, $(C_1, \boldsymbol{c}_1) = (I, \boldsymbol{e}_k)$ for some $k \in J_0$, and the statement is obvious. When $M \geq 2$, distinguish the following three cases.

1. **If $t_1 t_M = 1$.** Suppose $t_1 = t_M = 1$, the case where $t_1 = t_M = -1$ can be done analogously.

   Since $t_1 = 1 > 0$ and $t_1 + \cdots + t_{M-1} = -1 < 0$, there must exist $2 \leq i \leq M - 2$ such that $t_1 + \cdots + t_i = 0$. By induction hypothesis,

   $$(C_1, \boldsymbol{c}_1) \cdots (C_i, \boldsymbol{c}_i) = (I, \boldsymbol{y}), \quad (C_{i+1}, \boldsymbol{c}_{i+1}) \cdots (C_M, \boldsymbol{c}_M) = (I, \boldsymbol{y}'),$$

   where $\boldsymbol{y}$ and $\boldsymbol{y}'$ can be written as a $\mathbb{R}_{>0}$-linear combination of elements in $\mathcal{S}$. Therefore $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{y}'$ also satisfies this claim.

2. **If $t_1 t_M = -1$.** In this case, $C_2 \cdots C_{M-1} = C_1 C_2 \cdots C_M = I$. Define

   $$w' \coloneqq (C_2, \boldsymbol{c}_2) \cdots (C_{M-1}, \boldsymbol{c}_{M-1}).$$

   Then the product of w' is of the form $(I, \boldsymbol{x}')$, where $\boldsymbol{x}'$ is either $\mathbf{0}$ (when w' is empty), or

$\boldsymbol{x}'$ is a $\mathbb{R}_{>0}$-combination of elements in $\mathcal{S}$ by the induction hypothesis. Hence

$$(C_1, \boldsymbol{c}_1) \cdots (C_M, \boldsymbol{c}_M) = (C_1, \boldsymbol{c}_1) \cdot (I, \boldsymbol{x}') \cdot (C_M, \boldsymbol{c}_M) = (I, \boldsymbol{c}_1 + C_1\boldsymbol{c}_M + C_1\boldsymbol{x}').$$

We claim that $\boldsymbol{c}_1 + C_1\boldsymbol{c}_M \in \mathrm{Cell}(\boldsymbol{d}_{ij})$ for some indices $i \in J_+, j \in J_-$. First suppose $t_1 = 1, t_M = -1$. Let $i \in J_+, j \in J_-$ be indices such that $\widehat{(A'_i, \boldsymbol{a}'_i)} = (C_1, \boldsymbol{c}_1)$ and $\widehat{(A'_j, \boldsymbol{a}'_j)} = (C_M, \boldsymbol{c}_M)$, then

$$\begin{aligned}
\boldsymbol{c}_1 + C_1\boldsymbol{c}_M &= \left(a_j \cdot \frac{\lambda - 1}{1 - \lambda^{-n_j}} + a_i \cdot \frac{1 - \lambda}{1 - \lambda^{n_i}}, b_j \cdot \frac{\lambda^{-1} - 1}{1 - \lambda^{n_j}} + b_i \cdot \frac{1 - \lambda^{-1}}{1 - \lambda^{-n_i}} \right)^\top \\
&= \left((\lambda - 1) d_{ija}, (1 - \lambda^{-1}) d_{ijb} \right)^\top \\
&\in \mathrm{Cell}(\boldsymbol{d}_{ij}). \qquad \text{(by Equation (5.8))}
\end{aligned}$$

Next suppose $t_1 = -1, t_M = 1$. Let $i \in J_+, j \in J_-$ be indices such that $\widehat{(A'_j, \boldsymbol{a}'_j)} = (C_1, \boldsymbol{c}_1)$ and $\widehat{(A'_i, \boldsymbol{a}'_i)} = (C_M, \boldsymbol{c}_M)$, then

$$\begin{aligned}
\boldsymbol{c}_1 + C_1\boldsymbol{c}_M &= \left(a_i \cdot \frac{\lambda^{-1} - 1}{1 - \lambda^{n_i}} + a_j \cdot \frac{1 - \lambda^{-1}}{1 - \lambda^{-n_j}}, b_i \cdot \frac{\lambda - 1}{1 - \lambda^{-n_i}} + b_j \cdot \frac{1 - \lambda}{1 - \lambda^{n_j}} \right)^\top \\
&= \left((1 - \lambda^{-1}) d_{ija}, (\lambda - 1) d_{ijb} \right)^\top \\
&\in \mathrm{Cell}(\boldsymbol{d}_{ij}).
\end{aligned}$$

Hence, in both cases, $\boldsymbol{c}_1 + C_1\boldsymbol{c}_M \in \mathrm{Cell}(\boldsymbol{d}_{ij}) \subseteq \mathcal{S}$. We then show that $\boldsymbol{c}_1 + C_1\boldsymbol{c}_M + C_1\boldsymbol{x}'$ is a $\mathbb{R}_{>0}$-combination of elements in $\mathcal{S}$. Since $\boldsymbol{x}'$ is either zero or a $\mathbb{R}_{>0}$-combination of elements in $\mathcal{S}$, write $\boldsymbol{x}' = \sum_{i=1}^m r_i \boldsymbol{s}_i$, where $m \geq 0, r_i > 0, \boldsymbol{s}_i \in \mathcal{S}$. Then $C_1\boldsymbol{x}' = \sum_{i=1}^m r_i C_1 \boldsymbol{s}_i$ is still a $\mathbb{R}_{>0}$-combination of elements in $\mathcal{S}$ by Equation (5.8). Hence, $\boldsymbol{c}_1 + C_1\boldsymbol{c}_M + C_1\boldsymbol{x}'$ is a $\mathbb{R}_{>0}$-combination of elements in $\mathcal{S}$.

3. **If $t_1 t_M = 0$.** Suppose $t_1 = 0$, the case where $t_M = 0$ can be done analogously.

   By induction hypothesis,

   $$(C_1, \boldsymbol{c}_1) = (I, \boldsymbol{y}), \quad (C_2, \boldsymbol{c}_2) \cdots (C_M, \boldsymbol{c}_M) = (I, \boldsymbol{y}'),$$

   where $\boldsymbol{y}$ and $\boldsymbol{y}'$ can be written as a $\mathbb{R}_{>0}$-linear combination of elements in $\mathcal{S}$. Therefore $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{y}'$ also satisfies this claim.

Therefore we have found an $\mathbb{R}_{>0}$-linear combination that satisfies (i). The following can be easily seen from the above induction procedure: if for some $i \in J_+$ the letter $\widehat{(A'_i, \boldsymbol{a}'_i)}$ appears in

w, then the $\mathbb{R}_{>0}$-linear combination contains a vector $\boldsymbol{s}_\ell$ in the cell $\mathrm{Cell}(\boldsymbol{d}_{ij})$ for some $j \in J_-$. Since w is full-image, the condition (ii) in the statement of the Lemma must hold. Similarly, the conditions (iii) and (iv) must also hold. □

Proposition 5.5.14. *The semigroup $\langle \mathcal{G} \rangle$ is a group if and only if there exists a finite set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\} \subseteq \mathcal{S}$ that satisfies the four conditions (i)-(iv) in Lemma 5.5.13.*

Proof. First suppose there exists a finite set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\} \subseteq \mathcal{S}$ satisfying (i)-(iv), we want to find a full-image word w over the alphabet $\mathcal{G}$, representing $(I, \mathbf{0})$.

For any $t \in \{1, \ldots, m\}$, if $\boldsymbol{s}_t$ is an element of $\mathrm{Cell}(\boldsymbol{d}_{ij})$, then by Lemma 5.5.11, there exist $(I, \boldsymbol{v}), (I, \boldsymbol{w}) \in \langle \mathcal{G}' \rangle$ such that $\boldsymbol{s}_t = r_1\boldsymbol{v} + r_2\boldsymbol{w}$ for some $r_1, r_2 > 0$. Also, the letters $(A'_i, \boldsymbol{a}'_i)$ and $(A'_j, \boldsymbol{a}'_j)$ appear in some words representing $(I, \boldsymbol{v})$ and $(I, \boldsymbol{w})$. If $\boldsymbol{s}_t$ is an element of $\mathrm{Cell}(\boldsymbol{e}_k)$, then by Lemma 5.5.12, there exist $(I, \boldsymbol{v}), (I, \boldsymbol{w}) \in \langle \mathcal{G}' \rangle$ such that $\boldsymbol{s}_t = r_1\boldsymbol{v} + r_2\boldsymbol{w}$ for some $r_1, r_2 > 0$. Also, the letter $(A'_k, \boldsymbol{a}'_k)$ appears in some words representing $(I, \boldsymbol{v})$ and $(I, \boldsymbol{w})$.

Therefore, $\boldsymbol{s}_t$ can always be written as a strictly positive linear combination of vectors $\boldsymbol{v}$ with $(I, \boldsymbol{v}) \in \langle \mathcal{G}' \rangle$. Hence, by condition (i), there exist strictly positive reals $r_1, \ldots, r_M$ such that $r_1\boldsymbol{v}'_1 + \cdots + r_M\boldsymbol{v}'_M = \mathbf{0}$, where $(I, \boldsymbol{v}'_t) \in \langle \mathcal{G}' \rangle, t = 1, \ldots, M$. Furthermore, conditions (ii), (iii) and (iv) show that every letter $(A'_i, \boldsymbol{a}'_i)$ in $\mathcal{G}'$ appears at least once in a word representing $(I, \boldsymbol{v}'_t)$ for some $t \in \{1, \ldots, M\}$.

Changing back to the original basis, this shows that $r_1\boldsymbol{v}_1 + \cdots + r_M\boldsymbol{v}_M = \mathbf{0}$, where $(I, \boldsymbol{v}_t) \in \langle \mathcal{G} \rangle$ for $t = 1, \ldots, M$. Since the entries of $\boldsymbol{v}_t$ are all integers, there exist strictly positive *integers* $n_1, \ldots, n_M$, such that $n_1\boldsymbol{v}_1 + \cdots + n_M\boldsymbol{v}_M = \mathbf{0}$. Hence

$$(I, \boldsymbol{v}_1)^{n_1} \cdots (I, \boldsymbol{v}_M)^{n_M} = (I, \mathbf{0}).$$

Every letter $(A_i, \boldsymbol{a}_i)$ in $\mathcal{G}$ appears at least once in a word representing $(I, \boldsymbol{v}_t)$ for some $t \in \{1, \ldots, M\}$. Therefore, $(I, \mathbf{0})$ can be represented as a full-image word. This shows that $\langle \mathcal{G} \rangle$ is a group by Lemma 2.2.1.

Next, suppose $\langle \mathcal{G} \rangle$ is a group, we show that there exists an $\mathbb{R}_{>0}$-linear combination of elements in $\mathcal{S}$ equal to $\mathbf{0}$, that satisfies the conditions (ii), (iii) and (iv).

By Lemma 2.2.1, there exists a full-image word $w \coloneqq (B'_1, \boldsymbol{b}'_1) \cdots (B'_m, \boldsymbol{b}'_m)$ over the alphabet $\mathcal{G}'$, representing $(I, \mathbf{0})$. Replacing each letter $(B'_i, \boldsymbol{b}'_i)$ in w with n_i consecutive letters $\widehat{(B'_i, \boldsymbol{b}'_i)}$, we obtain a full-image word

$$\widehat{w} = (C_1, \boldsymbol{c}_1) \cdots (C_M, \boldsymbol{c}_M)$$

over the alphabet $\widehat{\mathcal{G}}$, representing $(I, \mathbf{0})$.

Then by Lemma 5.5.13, the vector $\mathbf{0}$ can be written as a $\mathbb{R}_{>0}$-linear combination of a finite number of elements in $\mathcal{S}$, satisfying the conditions (ii), (iii) and (iv). □

Corollary 5.5.15. *If the generator A of the group $\langle A_1, \ldots, A_K \rangle \cong \mathbb{Z}$ is a positive scale, it is the decidable in PTIME whether $\langle \mathcal{G} \rangle$ is a group.*

Proof. Since $\langle A_1, \ldots, A_K \rangle \cong \mathbb{Z}$, we have $J_+ \neq \emptyset, J_- \neq \emptyset$. Given $\mathcal{S}$ as a set of cells, it is decidable whether there exist a finite non-empty set of vectors $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\} \subseteq \mathcal{S}$ satisfying conditions (ii), (iii) and (iv), as well as strictly positive reals $r_1, \ldots, r_m$ such that $r_1\boldsymbol{s}_1 + \cdots + r_m\boldsymbol{s}_m = \mathbf{0}$. Indeed, this is true if and only if the largest linear subspace $\mathcal{L}$ of the $\mathbb{R}_{\geq 0}$-cone $\langle \mathcal{S} \rangle_{\mathbb{R}_{\geq 0}}$ contains some cell $\mathrm{Cell}(\boldsymbol{d}_{i*})$ for each $i \in J_+$, some cell $\mathrm{Cell}(\boldsymbol{d}_{*j})$ for each $j \in J_-$, and some cell $\mathrm{Cell}(\boldsymbol{e}_k)$ for each $k \in J_0$. This is decidable in PTIME since the number of cells in $\mathcal{L}$ is finite (at most nine).

Given the input set $\mathcal{G}$, we can compute the set of cells in $\mathcal{S}$ in PTIME. Therefore, by Proposition 5.5.14, it is decidable in PTIME whether $\langle \mathcal{G} \rangle$ is a group. □

Combining all cases, we can now prove the main result of this section.

Proposition 5.3.4. *Let $\mathcal{G} = \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2, \mathbb{Z})$, such that the semigroup $H \coloneqq \langle A_1, \ldots, A_K \rangle$ is a group. Suppose H is virtually solvable, then exactly one of the following six conditions holds:*

(i) H is the trivial group.
(ii) H contains a non-trivial torsion element.
(iii) $H = \langle A \rangle_{grp}$, where A is a twisted inversion.
(iv) $H = \langle A \rangle_{grp}$, where A is a shear.
(v) $H = \langle A \rangle_{grp}$, where A is an inverting scale.
(vi) $H = \langle A \rangle_{grp}$, where A is a positive scale.

Furthermore, in cases (ii), (iii) and (v), the semigroup $\langle \mathcal{G} \rangle$ is always a group. Overall, it is decidable in PTIME whether $\langle \mathcal{G} \rangle$ is a group.

Proof. Division into six cases has already been proved by Lemma 5.5.1 and 5.5.3 as well as the discussion that follows. In cases (ii), (iii) and (v), Proposition 5.5.6, 5.5.7 and 5.5.10 show that $\langle \mathcal{G} \rangle$ is a group. We now show PTIME decidability.

First, we decide in PTIME which of the six cases is true for H, using Lemma 5.5.4. In cases (ii), (iii) and (v), the Group Problem for $\langle \mathcal{G} \rangle$ has positive answer. In cases (i), (iv) and (vi), Proposition 5.5.5, Corollary 5.5.9 and Corollary 5.5.15 show the required PTIME decidability result. □

5.6 Extensions and obstacles to Semigroup Membership

In previous sections we showed decidability and NP-completeness of the Identity Problem and the Group Problem in $\mathsf{SA}(2,\mathbb{Z})$. In this section we discuss possible extensions of our result and obstacles to solving Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$.

Let $\mathcal{G} \coloneqq \{(A_1, \boldsymbol{a}_1), \ldots, (A_K, \boldsymbol{a}_K)\}$ be a set of elements of $\mathsf{SA}(2,\mathbb{Z})$. Compared to the Identity Problem and the Group Problem, the first obvious obstacle for deciding Semigroup Membership for $\langle\mathcal{G}\rangle$ is that we can no longer suppose $H \coloneqq \langle A_1, \ldots, A_K\rangle$ to be a group. However, if we restrict the target to elements of the form $(I, \boldsymbol{a})$, that is, if we want to deciding whether $(I, \boldsymbol{a}) \in \langle\mathcal{G}\rangle$, then we can still suppose H to be a group.

Indeed, if H is not a group, then at least one of the A_i is not invertible in H. Therefore, a word over $\mathcal{G}$ representing $(I, \boldsymbol{a})$ cannot contain the letter $(A_i, \boldsymbol{a}_i)$. We can thus delete $(A_i, \boldsymbol{a}_i)$ from the alphabet $\mathcal{G}$ without changing whether $(I, \boldsymbol{a}) \in \langle\mathcal{G}\rangle$. One can repeat this process until H becomes a group.

Under the additional assumption that H is a group, we can decide Semigroup Membership for $\langle\mathcal{G}\rangle$ in all except one cases. Recall Theorem 5.3.2, if H contains a non-abelian free subgroup, then $\langle\mathcal{G}\rangle$ is a group by Proposition 5.3.3. Hence Semigroup Membership reduces to Group Membership, and is decidable by the result of Delgado [32]. If H is virtually solvable, then consider the six cases in Proposition 5.3.4. Case (ii), (iii) and (v) are easy since $\langle\mathcal{G}\rangle$ becomes a group. In case (i), deciding whether $(I, \boldsymbol{a}) \in \langle\mathcal{G}\rangle$ reduces to solving the linear equation $n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + \cdots n_K\boldsymbol{a}_K = \boldsymbol{a}$ for $(n_1, \ldots, n_K) \in \mathbb{Z}^K_{\geq 0} \setminus \{\boldsymbol{0}\}$, and is decidable by integer programming. In case (iv), Semigroup Membership for $\langle\mathcal{G}\rangle$ reduces to Semigroup Membership in the Heisenberg group $\mathrm{H}_3(\mathbb{Q})$, which is decidable by the result of Colcombet, Ouaknine, Semukhin and Worrell [30, Corollary 8]. Hence, only case (vi) remains unsolved.

We now show that deciding Semigroup Membership in case (vi) is equivalent to deciding Semigroup Membership in a semidirect product $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$. Given $\lambda > 1$ that satisfies a quadratic equation $\lambda^2 - a\lambda + 1 = 0$ for some $a \geq 3$, we define the following semidirect product:

$$\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z} \coloneqq \left\{ \begin{pmatrix} \lambda^k & x \\ 0 & 1 \end{pmatrix} \;\middle|\; k \in \mathbb{Z}, x \in \mathbb{Z}[\lambda] \right\}. \tag{5.11}$$

Here, $\mathbb{Z}[\lambda]$ is the ring generated by 1 and λ. There exist an embedding of $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ as a subgroup of $\mathsf{SA}(2,\mathbb{Z})$ in the following way. For a given λ satisfying $\lambda^2 - a\lambda + 1 = 0$, define the matrix $A_\lambda \coloneqq \begin{pmatrix} 0 & -1 \\ 1 & a \end{pmatrix}$ in $\mathsf{SL}(2,\mathbb{Z})$. Since λ is a quadratic integer, every element $x \in \mathbb{Z}[\lambda]$ can be

written uniquely as $x = x_1 + \lambda x_\lambda$ for some $x_1, x_\lambda \in \mathbb{Z}$. Define the map

$$\begin{aligned} \phi\colon \mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z} &\longrightarrow \mathsf{SA}(2,\mathbb{Z}) \\ \begin{pmatrix} \lambda^k & x \\ 0 & 1 \end{pmatrix} &\longmapsto \left(A_\lambda^k, (x_1, x_\lambda)^\top \right). \end{aligned}$$

It is easy to verify that ϕ is an injective group homomorphism. Furthermore, the image under ϕ of a subsemigroup[2] of $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ satisfies case (vi) of Proposition 5.3.4. Therefore, solving the hard case of Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$ necessitates solving Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$.

On the other hand, given any positive scale $A \in \mathsf{SL}(2,\mathbb{Z})$ with eigenvalue $\lambda > 1$, one can find a change of basis matrix $P \in \mathsf{SL}(2,\mathbb{Z})$ such that $P^{-1}AP = A_\lambda$ (see [27]). Therefore, any (finitely generated) semigroup $\langle \mathcal{G} \rangle$ satisfying case (vi) of Proposition 5.3.4 must be conjugate[3] to a (finitely generated) sub-semigroup of $\phi(\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}) \leq \mathsf{SA}(2,\mathbb{Z})$. Hence, solving Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ is sufficient for solving Semigroup Membership in the case (vi) of Proposition 5.3.4.

Although Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ remains an open problem, the group bears certain similarities to the better studied *Baumslag-Solitar group* $\mathsf{BS}(1,q)$.

$$\mathsf{BS}(1,q) \coloneqq \mathbb{Z}[1/q] \rtimes_q \mathbb{Z} = \left\{ \begin{pmatrix} q^k & x \\ 0 & 1 \end{pmatrix} \;\middle|\; k \in \mathbb{Z}, x \in \mathbb{Z}[1/q] \right\}.$$

Here, $q \geq 2$ is an integer. A recent result by Cadilhac, Chistikov and Zetzsche [25] showed decidability of the *rational subset membership problem* in $\mathsf{BS}(1,q)$ by considering rational languages of *base-q expansions*. This result subsumes decidability of Semigroup Membership in $\mathsf{BS}(1,q)$. Therefore, it would be interesting to adapt this approach to study Semigroup Membership in $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ by considering rational languages of *base-λ expansions* [20], where λ is an algebraic integer. Nevertheless, adaptation of the previous result to a non-integer base of numeration poses various additional difficulties that we have not been able to surmount.

[2]We suppose that the upper-left entries of elements of the semigroup contain both positive and negative exponents of λ, otherwise deciding Semigroup Membership is easy.

[3]The conjugation is realized by the change-of-basis matrix $\mathrm{diag}(P, 1)$.

Chapter 6

Conclusion and outlook

In this chapter, we discuss some possible extensions of our work and pose several open problems that may be interesting for future research in this direction.

6.1 Identity Problem and Group Problem

In this thesis we have obtained decidability of the Identity Problem and the Group Problem in the following groups:

(i) finitely generated nilpotent groups of class at most ten,

(ii) finitely generated metabelian groups,

(iii) the semidirect product $\mathsf{SA}(2,\mathbb{Z}) = \mathbb{Z}^2 \rtimes \mathsf{SL}(2,\mathbb{Z})$. (Note that $\mathsf{SL}(2,\mathbb{Z})$ is virtually free.)

A natural follow-up for (i) is whether it can be extended to groups of arbitrary nilpotency class. The natural next step after (ii) is to consider *centre-by-metabelian* groups. These are groups G that commute with $[[G,G],[G,G]]$. In other words, G is centre-by-metabelian if $[[[G,G],[G,G]],G] = \{I\}$. Centre-by-metabelian groups are solvable of derived length three, but their algorithmic problems remain partially tractable and may still admit decidable Identity Problem. In general, the algorithmic theory of solvable groups of derived length three is highly intractable. Notably, their Word Problem is undecidable [54]. For (iii), a natural generalization is the group $\mathbb{Z}^n \rtimes F$ where F is virtually free and $n \geq 3$. Another possible follow-up is the group $\mathsf{SL}(3,\mathbb{Z})$, but its algorithmic problems seem currently out of reach. To sum up, we have the following natural open problems.

Problem 6.1.1. Are the Identity Problem and the Group Problem decidable in the following groups:

(i) finitely generated nilpotent groups of arbitrary nilpotency class,

(ii) finitely generated centre-by-metabelian groups,

(iii) the semidirect product $\mathbb{Z}^n \rtimes F$ where F is virtually free and $n \geq 3$?

Problem 6.1.2. Is there a solvable group of derived length three[1] where the Identity Problem is undecidable?

Recall that in a fixed group G, decidability of the Group Problem implies decidability of the Identity Problem (Lemma 2.2.4). However, no evidence suggest the converse to be true. Hence, we have the following open problem.

Problem 6.1.3. Is there a group G with decidable Identity Problem but undecidable Group Problem?

6.2 Open problems in specific metabelian groups

In this thesis we focused most effort on the Identity Problem and the Group Problem. This is because Semigroup Membership is known to be undecidable in most natural classes of groups. Nevertheless, certain decidability results [25, 67] for Semigroup Membership have been shown for specific classes of metabelian groups such as the Baumslag-Solitar groups $\mathrm{BS}(1,q), q \geq 2$ and the wreath product $(\mathbb{Z}/p\mathbb{Z}) \wr \mathbb{Z}, p \geq 2$. Both can be seen as quotients of the wreath product $\mathbb{Z} \wr \mathbb{Z}$:

$$\mathrm{BS}(1,q) \coloneqq (\mathbb{Z}[1/q]) \rtimes \mathbb{Z} = \left\{ \begin{pmatrix} X^a & y \\ 0 & 1 \end{pmatrix} \;\middle|\; a \in \mathbb{Z}, y \in \mathbb{Z}[X]/(qX-1) \right\}.$$
$$(\mathbb{Z}/p\mathbb{Z}) \wr \mathbb{Z} \coloneqq \Big((\mathbb{Z}/p\mathbb{Z})[X] \Big) \rtimes \mathbb{Z} = \left\{ \begin{pmatrix} X^a & y \\ 0 & 1 \end{pmatrix} \;\middle|\; a \in \mathbb{Z}, y \in \mathbb{Z}[X]/(p) \right\}.$$

Here, $(qX-1)$ and (p) respectively represent the ideals of $\mathbb{Z}[X]$ generated by $qX-1$ and p. This leads to the question of whether Semigroup Membership is decidable in other quotients of $\mathbb{Z} \wr \mathbb{Z}$:

Problem 6.2.1. For which $f \in \mathbb{Z}[X]$ is Semigroup Membership decidable in the group:

$$\Big(\mathbb{Z}[X]/(f) \Big) \rtimes \mathbb{Z} \coloneqq \left\{ \begin{pmatrix} X^a & y \\ 0 & 1 \end{pmatrix} \;\middle|\; a \in \mathbb{Z}, y \in \mathbb{Z}[X]/(f) \right\} \; ?$$

Here, (f) denotes the ideal generated by the element f. The only known *undecidability* result is the distinguished case $f = 0$. That is, Semigroup Membership is undecidable in the wreath product $\mathbb{Z} \wr \mathbb{Z}$ itself [67].

Recall that in Section 5.6 we have identified the group $\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z}$ as one of the main obstacles to solving Semigroup Membership in $\mathsf{SA}(2,\mathbb{Z})$. This can be considered as a special case of

[1] As a convention, we only consider groups that are finitely presented in the class of solvable groups.

Problem 6.2.1:

$$\mathbb{Z}[\lambda] \rtimes_\lambda \mathbb{Z} \cong \Big(\mathbb{Z}[X]/(X^2 - aX + 1)\Big) \rtimes \mathbb{Z}.$$

Here, $X^2 - aX + 1$ is the minimal polynomial of λ.

One possible way to approach Problem 6.2.1 is to adapt the tools introduced in Chapter 4 and try to reduce it to solving (possibly non-homogeneous) linear equations over polynomial semirings. While solving *a system of* non-homogeneous linear equations over the polynomial semiring $\mathbb{N}[X]$ is undecidable [78], it is unclear whether solving a *single* non-homogeneous linear equation is decidable. This yields the following open problem.

Problem 6.2.2. Given as input the polynomials $y_1, \ldots, y_K, g \in \mathbb{Z}[X^{\pm}]$, can we decide whether the equation

$$f_1 y_1 + f_2 y_2 + \cdots + f_K y_K = g$$

has solution $f_1, \ldots, f_K$ in the semiring $\mathbb{N}[X^{\pm}]$?

Finally, recall that in class two nilpotent groups, Semigroup *Intersection* is decidable (Corollary 3.1.2), despite Semigroup *Membership* being undecidable [86]. We ask the same question for the wreath product $\mathbb{Z} \wr \mathbb{Z}$. The wreath product $\mathbb{Z} \wr \mathbb{Z}$ has undecidable Semigroup *Membership* [67]. However, the Identity Problem, which can be considered as a special case of Semigroup *Intersection*, is decidable by Theorem 4.1.2. Furthermore, $\mathbb{Z} \wr \mathbb{Z}$ cannot embed a direct product of two non-abelian free monoids, making it unlikely to show undecidability of Semigroup Intersection using the Post Correspondence Problem.

Fact 6.2.3. There is no embedding of the monoid $\{a, b\}^* \times \{a, b\}^*$ into the group $\mathbb{Z} \wr \mathbb{Z}$.

Proof. Suppose on the contrary that such an embedding

$$\varphi \colon \{a, b\}^* \times \{a, b\}^* \hookrightarrow \mathbb{Z} \wr \mathbb{Z}$$

exists. Let ϵ denote the empty word of $\{a, b\}^*$, write

$$\varphi((a, \epsilon)) = \begin{pmatrix} X^A & f_1 \\ 0 & 1 \end{pmatrix}, \varphi((b, \epsilon)) = \begin{pmatrix} X^B & f_2 \\ 0 & 1 \end{pmatrix},$$
$$\varphi((\epsilon, a)) = \begin{pmatrix} X^C & f_3 \\ 0 & 1 \end{pmatrix}, \varphi((\epsilon, b)) = \begin{pmatrix} X^D & f_4 \\ 0 & 1 \end{pmatrix}.$$

We have the following equations and inequalities in $\{a, b\}^* \times \{a, b\}^*$:

$$(a, \epsilon) \cdot (\epsilon, a) = (\epsilon, a) \cdot (a, \epsilon), \quad (a, \epsilon) \cdot (\epsilon, b) = (\epsilon, b) \cdot (a, \epsilon),$$

$$(b,\epsilon)\cdot(\epsilon,a)=(\epsilon,a)\cdot(b,\epsilon),\quad (b,\epsilon)\cdot(\epsilon,b)=(\epsilon,b)\cdot(b,\epsilon),$$
$$(a,\epsilon)\cdot(b,\epsilon)\neq(b,\epsilon)\cdot(a,\epsilon),\quad (b,\epsilon)\cdot(a,\epsilon)\neq(a,\epsilon)\cdot(b,\epsilon).$$

Applying φ, we obtain

$$(X^A-1)f_3=(X^C-1)f_1,\quad (X^A-1)f_4=(X^D-1)f_1, \tag{6.1}$$
$$(X^B-1)f_3=(X^C-1)f_2,\quad (X^B-1)f_4=(X^D-1)f_2, \tag{6.2}$$
$$(X^A-1)f_2\neq(X^B-1)f_1,\quad (X^C-1)f_4\neq(X^D-1)f_3. \tag{6.3}$$

Consider three cases.

1. If A,B,C,D are all non-zero. Then (6.1) and (6.2) yield

$$\frac{f_1}{X^A-1}=\frac{f_2}{X^B-1}=\frac{f_3}{X^C-1}=\frac{f_4}{X^D-1}.$$

   This contradicts (6.3).

2. If exactly one of the two sets $\{A,B\}$, $\{C,D\}$ contains zero. Without loss of generality suppose $A=0$ and $0\notin\{C,D\}$. Since $A=0$ and $(X^A-1)f_2\neq(X^B-1)f_1$, we have $B\neq 0$. Then (6.2) yields

$$\frac{f_2}{X^B-1}=\frac{f_3}{X^C-1}=\frac{f_4}{X^D-1},$$

   contradicting $(X^C-1)f_4\neq(X^D-1)f_3$.

3. If both sets $\{A,B\}$ and $\{C,D\}$ contain zero. Without loss of generality suppose $A=C=0$. Since $C=0$ and $(X^C-1)f_4\neq(X^D-1)f_3$, we have $D\neq 0$. Then, $(X^A-1)f_4=(X^D-1)f_1$ and $A=0,D\neq 0$ yield $f_1=0$. Then $(X^A-1)f_2=0=(X^B-1)f_1$, contradicting (6.3).

Therefore, such an embedding φ cannot exist. □

Consequently, decidability of Semigroup Intersection in the group $\mathbb{Z}\wr\mathbb{Z}$ becomes a natural open problem.

Problem 6.2.4. Is Semigroup Intersection decidable in the group $\mathbb{Z}\wr\mathbb{Z}$?

Bibliography

[1] S. I. Adyan. Algorithmic unsolvability of problems of recognition of certain properties of groups. In *Dokl. Akad. Nauk SSSR (N.S.)*, volume 103, pages 533–535, 1955.

[2] L. F. Alday and Y. Tachikawa. Affine SL(2) conformal blocks from 4d gauge theories. *Letters in Mathematical Physics*, 94(1):87–114, 2010.

[3] A. D. Alexandrov. *Convex polyhedra*, volume 109. Springer, 2005.

[4] L. Babai, R. Beals, J.-y. Cai, G. Ivanyos, and E. M. Luks. Multiplicative equations over commuting matrices. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 498–507, 1996.

[5] H. F. Baker. Alternants and continuous groups. *Proceedings of the London Mathematical Society*, 2(1):24–47, 1905.

[6] G. Baumslag. Wreath products and finitely presented groups. *Mathematische Zeitschrift*, 75(1):22–28, 1961.

[7] G. Baumslag. Subgroups of finitely presented metabelian groups. *Journal of the Australian Mathematical Society*, 16(1):98–110, 1973.

[8] G. Baumslag. *Lecture notes on nilpotent groups*. American Mathematical Society, 2007.

[9] G. Baumslag, F. B. Cannonito, and D. J. Robinson. The algorithmic theory of finitely generated metabelian groups. *Transactions of the American Mathematical Society*, 344(2):629–648, 1994.

[10] G. Baumslag, R. Mikhailov, and K. E. Orr. Localization, metabelian groups, and the isomorphism problem. *Transactions of the American Mathematical Society*, 369(10):6823–6852, 2017.

[11] G. Baumslag and Y. Roitberg. Groups with free 2-generator subsemigroups. In *Semigroup forum*, volume 25, pages 135–143. Springer, 1982.

[12] R. Beals. Algorithms for matrix groups and the Tits alternative. *Journal of computer and system sciences*, 58(2):260–279, 1999.

[13] R. Beals and L. Babai. Las Vegas algorithms for matrix groups. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 427–436. IEEE, 1993.

[14] P. C. Bell, M. Hirvensalo, and I. Potapov. The Identity Problem for matrix semigroups in $\mathrm{SL}_2(\mathbb{Z})$ is NP-complete. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 187–206. SIAM, 2017.

[15] P. C. Bell, M. Hirvensalo, and I. Potapov. Private communication, 2023.

[16] P. C. Bell and I. Potapov. On the undecidability of the identity correspondence problem and its applications for word and matrix semigroups. *International Journal of Foundations of Computer Science*, 21(06):963–978, 2010.

[17] C. Berkesch and F.-O. Schreyer. Syzygies, finite length modules, and random curves. *Commutative algebra and noncommutative algebraic geometry*, 1:25–52, 2015.

[18] M. Bhattacharjee, R. G. Möller, D. Macpherson, and P. M. Neumann. *Notes on infinite permutation groups*. Springer, 2006.

[19] N. Blackburn. Conjugacy in nilpotent groups. *Proceedings of the American Mathematical Society*, 16(1):143–148, 1965.

[20] F. Blanchard. β-expansions and symbolic dynamics. *Theoretical Computer Science*, 65(2):131–141, 1989.

[21] V. D. Blondel, E. Jeandel, P. Koiran, and N. Portier. Decidable and undecidable problems about quantum automata. *SIAM Journal on Computing*, 34(6):1464–1473, 2005.

[22] L. A. Bokut'. A basis for free polynilpotent Lie algebras. *Algebra i logika*, 2(4):13–19, 1963.

[23] M. Bremner. *Lattice Basis Reduction: An Introduction to the LLL Algorithm and Its Applications*. CRC Press New York, 2011.

[24] L. M. Cabrer and D. Mundici. Classifying orbits of the affine group over the integers. *Ergodic Theory and Dynamical Systems*, 37(2):440–453, 2017.

[25] M. Cadilhac, D. Chistikov, and G. Zetzsche. Rational subsets of Baumslag-Solitar groups. In *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020*, volume 168 of *LIPIcs*, pages 116:1–116:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[26] J. E. Campbell. On a law of combination of operators (second paper). *Proceedings of the London Mathematical Society*, 1(1):14–32, 1897.

[27] J. T. Campbell and E. C. Trouy. When are two elements of GL(2,Z) similar? *Linear Algebra and its Applications*, 157:175–184, 1991.

[28] F. Casas and A. Murua. An efficient algorithm for computing the Baker-Campbell-Hausdorff series and some of its applications. *Journal of Mathematical Physics*, 50(3):033513, 2009.

[29] C. Choffrut and J. Karhumäki. Some decision problems on integer matrices. *RAIRO-Theoretical Informatics and Applications-Informatique Théorique et Applications*, 39(1):125–131, 2005.

[30] T. Colcombet, J. Ouaknine, P. Semukhin, and J. Worrell. On reachability problems for low-dimensional matrix semigroups. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019*, volume 132 of *LIPIcs*, pages 44:1–44:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[31] V. De Angelis and S. Tuncel. Handelman's theorem on polynomials with positive multiples. *Codes, systems, and graphical models (Minneapolis, MN, 1999)*, pages 439–445, 2001.

[32] J. Delgado Rodríguez. *Extensions of free groups: algebraic, geometric, and algorithmic aspects*. PhD thesis, Universitat Politècnica de Catalunya, 2017.

[33] H. Derksen, E. Jeandel, and P. Koiran. Quantum automata and algebraic groups. *Journal of Symbolic Computation*, 39(3-4):357–371, 2005.

[34] C. Druţu and M. Kapovich. *Geometric group theory*, volume 63. American Mathematical Soc., 2018.

[35] E. B. Dynkin. Calculation of the coefficients in the Campbell-Hausdorff formula. *Selected Papers of E. B. Dynkin with Commentary. Ed. by Yushkenich, A. A.*, pages 31–35, 2000.

[36] M. Einsiedler, R. Mouat, and S. Tuncel. When does a submodule of $(\mathbb{R}[x_1, \ldots, x_k])^n$ contain a positive element? *Monatshefte für Mathematik*, 140(4):267–283, 2003.

[37] D. Eisenbud. *Commutative algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.

[38] A. Finkel, S. Göller, and C. Haase. Reachability in register machines with polynomial updates. In *International Symposium on Mathematical Foundations of Computer Science*, pages 409–420. Springer, 2013.

[39] L. A. Giefer, J. Clemens, and K. Schill. Extended object tracking on the affine group Aff(2). In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2020.

[40] M. Giesbrecht. Nearly optimal algorithms for canonical matrix forms. *SIAM Journal on Computing*, 24(5):948–969, 1995.

[41] F. Grunewald and D. Segal. Some general algorithms. II: Nilpotent groups. *Annals of Mathematics*, 112(3):585–617, 1980.

[42] M. Hall. A basis for free lie rings and higher commutators in free groups. *Proceedings of the American Mathematical Society*, 1(5):575–581, 1950.

[43] P. Hall. Finiteness conditions for soluble groups. *Proceedings of the London Mathematical Society*, s3-4(1):419–436, 01 1954.

[44] D. Handelman. *Positive polynomials and product type actions of compact groups*, volume 320. American Mathematical Soc., 1985.

[45] F. Hausdorff. Die symbolische Exponentialformel in der Gruppentheorie. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 58:19–48, 1906.

[46] D. F. Holt and S. Rees. Solving the word problem in real time. *Journal of the London Mathematical Society*, 63(3):623–639, 2001.

[47] R. Howe. On the role of the Heisenberg group in harmonic analysis. *Bulletin (New Series) of the American Mathematical Society*, 3(2):821–843, 1980.

[48] J. M. Howie. *Fundamentals of Semigroup Theory*. Oxford University Press, 1995.

[49] E. Hrushovski, J. Ouaknine, A. Pouly, and J. Worrell. Polynomial invariants for affine programs. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 530–539, 2018.

[50] J.-i. Igusa. *Theta functions*, volume 194. Springer Science & Business Media, 2012.

[51] D. L. Johnson. *Presentations of groups*. Number 15 in London Mathematical Society Student Texts. Cambridge university press, 1997.

[52] I. Kapovich, R. Weidmann, and A. Myasnikov. Foldings, graphs of groups and the membership problem. *International Journal of Algebra and Computation*, 15(01):95–128, 2005.

[53] M. I. Kargapolov and J. I. Merzljakov. *Fundamentals of the Theory of Groups*, volume 62. Springer, 1979.

[54] O. G. Kharlampovich. A finitely presented solvable group with unsolvable word problem. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 45(4):852–873, 1981.

[55] O. G. Kharlampovich and M. V. Sapir. Algorithmic problems in varieties. *International Journal of Algebra and Computation*, 5(04n05):379–602, 1995.

[56] E. I. Khukhro. *p-Automorphisms of Finite p-Groups*, volume 246. Cambridge University Press, 1998.

[57] A. A. Kirillov. *Lectures on the orbit method*, volume 64. American Mathematical Soc., 2004.

[58] S. Ko, R. Niskanen, and I. Potapov. On the identity problem for the special linear group and the Heisenberg group. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018*, volume 107 of *LIPIcs*, pages 132:1–132:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[59] S.-K. Ko, R. Niskanen, and I. Potapov. Reachability problems in nondeterministic polynomial maps on the integers. In *International Conference on Developments in Language Theory*, pages 465–477. Springer, 2018.

[60] S. Kobayashi and T. Nagano. On filtered Lie algebras and geometric structures I. *Journal of Mathematics and Mechanics*, pages 875–907, 1964.

[61] V. M. Kopytov. Solvability of the problem of occurrence in finitely generated soluble groups of matrices over the field of algebraic numbers. *Algebra and Logic*, 7(6):388–393, 1968.

[62] H. Krovi and M. Rötteler. An efficient quantum algorithm for the hidden subgroup problem over Weyl-Heisenberg groups. In *Mathematical Methods in Computer Science*, pages 70–88. Springer, 2008.

[63] J. Kwon and F. C. Park. Visual tracking via particle filtering on the affine group. *The International Journal of Robotics Research*, 29(2-3):198–217, 2010.

[64] J. C. Lennox and D. J. Robinson. *The theory of infinite soluble groups*. Clarendon press, 2004.

[65] J.-L. Loday. Série de Hausdorff, idempotents Eulériens et algèbres de Hopf. *Expositiones Mathematicae*, 12, 1994.

[66] M. Lohrey. Subgroup membership in GL(2,Z). In *38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[67] M. Lohrey, B. Steinberg, and G. Zetzsche. Rational subsets and submonoids of wreath products. *Information and Computation*, 243:191–204, 2015.

[68] J. Macdonald, A. Miasnikov, and D. Ovchinnikov. Low-complexity computations for nilpotent subgroup problems. *International Journal of Algebra and Computation*, 29(04):639–661, 2019.

[69] W. Magnus. On a theorem for Marshall Hall. *Annals of Mathematics*, pages 764–768, 1939.

[70] A. I. Mal'cev. Nilpotent semigroups. In *Ivanov. Gos Ped. Inst. Ucen Zap. Fiz.-Mat. Nauki*, volume 4, pages 107–111, 1953.

[71] A. I. Mal'cev. On homomorphisms onto finite groups. *Uchen. Zap. Ivanovsk. Ped. Inst.*, 18:49–60, 1958.

[72] A. Markov. On certain insoluble problems concerning matrices. *Doklady Akad. Nauk SSSR*, 57(6):539–542, 1947.

[73] Y. V. Matiyasevich. *Hilbert's Tenth Problem*. MIT press, 1993.

[74] K. A. Mikhailova. The occurrence problem for direct products of groups. *Matematicheskii Sbornik*, 112(2):241–251, 1966.

[75] D. Mundici. Invariant measure under the affine group over $\mathbb{Z}$. *Combinatorics, Probability and Computing*, 23(2):248–268, 2014.

[76] J. R. Munkres. *Topology: A first course*. Prentice-Hall, 1974.

[77] A. Myasnikov and A. Weiß. Parallel complexity for nilpotent groups. *International Journal of Algebra and Computation*, 32(05):895–928, 2022.

[78] P. Narendran. Solving linear equations over polynomial semirings. In *Proceedings 11th Annual IEEE Symposium on Logic in Computer Science*, pages 466–472. IEEE, 1996.

[79] J. v. Neumann. Die Eindeutigkeit der Schrödingerschen Operatoren. *Mathematische Annalen*, 104:570–578, 1931.

[80] M. Newman. The structure of some subgroups of the modular group. *Illinois Journal of Mathematics*, 6(3):480–487, 1962.

[81] G. A. Noskov. Conjugacy problem in metabelian groups. *Mathematical notes of the Academy of Sciences of the USSR*, 31:252–258, 1982.

[82] P. S. Novikov. On the algorithmic unsolvability of the word problem in group theory. *Trudy Matematicheskogo Instituta imeni VA Steklova*, 44:3–143, 1955.

[83] J. Okniński and A. Salwa. Generalised Tits alternative for linear semigroups. *Journal of Pure and Applied Algebra*, 103(2):211–220, 1995.

[84] E. L. Post. A variant of a recursively unsolvable problem. *Bulletin of the American Mathematical Society*, 52(4):264 – 268, 1946.

[85] M. Raskin, F. Mazowiecki, C. Haase, and M. Blondin. Affine extensions of integer vector addition systems with states. *Logical Methods in Computer Science*, 17, 2021.

[86] V. Roman'kov. Undecidability of the submonoid membership problem for a sufficiently large finite direct power of the Heisenberg group. *arXiv preprint arXiv:2209.14786*, 2022.

[87] N. S. Romanovskii. Some algorithmic problems for solvable groups. *Algebra and Logic*, 13(1):13–16, 1974.

[88] F.-O. Schreyer. Die Berechnung von Syzygien mit dem verallgemeinerten Weierstraßschen Divisionssatz. Master's thesis, Fakultät für Mathematik, Universität Hamburg, 1980.

[89] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

[90] J.-P. Serre. *Trees*. Springer Science & Business Media, 2002.

[91] B. Sturmfels. *Gröbner Bases and Convex Polytopes*, volume 8. American Mathematical Soc., 1996.

[92] R. G. Swan. Groups of cohomological dimension one. *Journal of Algebra*, 12(4):585–610, 1969.

[93] A. R. Tarrida. *Affine Maps, Euclidean Motions and Quadrics*. Springer, 2011.

[94] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. second ed., rev., Univ. of California Press, Berkeley, 1951.

[95] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.0)*, 2020. `https://www.sagemath.org`.

[96] J. Tits. Free subgroups in linear groups. *Journal of Algebra*, 20(2):250–270, 1972.

[97] H. Weyl. *The theory of groups and quantum mechanics*. Courier Corporation, 1950.

[98] J. A. Wolf. The affine group of a Lie group. In *Proc. Amer. Math. Soc.*, volume 14, pages 352–353, 1963.

[99] J.-H. Yang. Harmonic analysis on the quotient spaces of Heisenberg groups. *Nagoya mathematical journal*, 123:103–117, 1991.