Aleksi Saarela

Word Equations and Related Topics

Independence, Decidability and Characterizations

Turku Centre *for* Computer Science

TUCS Dissertations
No 145, May 2012

Word Equations and Related Topics

Independence, Decidability and Characterizations

Aleksi Saarela

To be presented, with the permission of the Faculty of Mathematics and Natural Sciences of the University of Turku, for public criticism in Auditorium XXI on May 18, 2012, at 12 noon.

University of Turku
Department of Mathematics and Statistics
FI-20014 Turku, Finland

2012

Supervisor

Professor Juhani Karhumäki
Department of Mathematics and Statistics
University of Turku
Finland

Reviewers

Professor Christian Choffrut
LIAFA
Universite Paris Diderot
France

Professor Wojciech Rytter
Instytut Informatyki
Uniwersytet Warszawski
Poland

Opponent

Professor Dominique Perrin
Laboratoire d'informatique Gaspard-Monge
University of Paris-Est Marne-la-Vallee
France

ISBN 978-952-12-2737-0
ISSN 1239-1883

Abstract

The three main topics of this work are independent systems and chains of word equations, parametric solutions of word equations on three unknowns, and unique decipherability in the monoid of regular languages.

The most important result about independent systems is a new method giving an upper bound for their sizes in the case of three unknowns. The bound depends on the length of the shortest equation. This result has generalizations for decreasing chains and for more than three unknowns. The method also leads to shorter proofs and generalizations of some old results.

Hmelevksii's theorem states that every word equation on three unknowns has a parametric solution. We give a significantly simplified proof for this theorem. As a new result we estimate the lengths of parametric solutions and get a bound for the length of the minimal nontrivial solution and for the complexity of deciding whether such a solution exists.

The unique decipherability problem asks whether given elements of some monoid form a code, that is, whether they satisfy a nontrivial equation. We give characterizations for when a collection of unary regular languages is a code. We also prove that it is undecidable whether a collection of binary regular languages is a code.

Acknowledgements

First of all, I thank my advisor professor Juhani Karhumäki. He introduced me to this wonderful area of research and gave many great problems to work with. His advice has always guided me to the right direction but he has still given me enough freedom.

I am thankful for professors Christian Choffrut and Wojciech Rytter for reviewing my thesis. Their comments were extremely helpful and made the manuscript much better. It was an honor for me that professor Dominique Perrin agreed to act as the opponent.

The Department of Mathematics has provided excellent working conditions. I also gratefully acknowledge the financial support from Turku Centre for Computer Science (TUCS), Turku University Foundation and Academy of Finland. I thank the staff of the department and of TUCS for taking care of all the practical matters.

I thank my colleagues at the department for being a part of my journey to become a mathematician, starting with the courses and competitions and continuing with research, conference trips, nonmathematical activities like badminton, and the everyday life at the department.

Finally, I thank my parents for everything.

Turku, April 2012 Aleksi Saarela

Contents

Chapter 1

Introduction

In this work we are mostly interested in word equations, particularly in questions related to independent systems and parametric solutions, but also in unique decipherability. We start this introduction by giving some background information about these topics, and then we outline the structure and results of this work.

Combinatorics on words is a part of discrete mathematics. It studies the properties of strings of symbols and has applications in many areas from pure mathematics to computer science. The history of combinatorics on words can be said to begin with the works of Thue in the early 20th century [63, 64]. For a long time the research was scattered, but the book *Combinatorics on words* [43] by Lothaire was a sign that the area had become more mature. After that development has been fast, as can be seen in the later books of Lothaire [44, 45]. Some other general references are [9], [4] and [3].

Algebraically words form a free monoid, which is one of the most fundamental algebraic structures. Thus the connections to algebra are natural, but combinatorics on words is also related to other areas of mathematics, like number theory and dynamical systems.

From the point of view of computer science, combinatorics on words is strongly related to automata and formal languages. Algorithmic questions are often studied. There are also connections to some more applied areas, like bioinformatics and pattern recognition.

Theory of word equations is an important part of combinatorics on words. It plays an essential role in many areas of mathematics, such as in representation results of algebra, theory of algorithms and pattern matching. One of the historically important papers on word equations is the article by Lyndon and Schützenberger [46]. During the last decades the area has provided several challenging problems as well as fundamental, or even breakthrough, results in discrete mathematics.

One remarkable result of this topic is the decidability of the satisfiability

problem for word equations. This was proved by Makanin [47] and is in contrast to the same problem on Diophantine equations, which is undecidable [48]. For a presentation of Makanin' result, see [16]. This result was improved to a PSPACE algorithm by Plandowski [52]. The satisfiability problem is known to be NP-hard and has been conjectured to be in NP [53].

In the case of word equations with only three unknowns important results have also been achieved. In one direction Hmelevskii [26] proved that any such constant-free equation is finitely parameterizable, that is the general solution can be expressed as a finite formula on word and numerical parameters. On other direction Spehner [61, 62] classified all sets of relations a given solution, that is a triple of words, can satisfy (see also the paper by Budkina and Markov [7]). A remarkable thing is that both of these results have only very complicated proofs. This is a splendid example of a challenging nature of word problems.

It should be noted that Makanin's result concerns word equations with constants, while Hmelevskii's result is about constant-free equations. When talking about word equations in this work, we mean constant-free equations.

Another remarkable property of word equations is the so-called *Ehrenfeucht compactness property*. It guarantees that any system of word equations is equivalent to some of its finite subsystems. The proofs (see [2] and [20]) are based on a transformation of word equations into Diophantine equations and an application of Hilbert's basis theorem. Although we have this finiteness property, we do not know any upper bound, if it exists, for the size of an equivalent subsystem in terms of the number of unknowns. This holds even in the case of three unknowns.

One of the basic results in the theory of word equations is that a nontrivial equation causes a defect effect. In other words, if n words satisfy a nontrivial relation, then they can be represented as products of $n-1$ words. Not much is known about the additional restrictions caused by several independent relations [22].

In fact, even the following simple question, formulated already in [12], is still unanswered: how large can an independent system of word equations on three unknowns be? The largest known examples consist of three equations. The only known upper bound comes from the Ehrenfeucht compactness property: an independent system cannot be infinite. This question can be obviously asked also in the case of $n > 3$ unknowns. Some results concerning independent systems on three unknowns can be found in [24], [14] and [15], but the open problem seems to be very difficult to approach with current techniques.

There are many variations of the above question. As a related problem we define the notion of *decreasing chains* of word equations. This asks how long chains of word equations exist such that the set of solutions always properly diminishes when a new element of the chain is taken into the sys-

tem. Or more intuitively, how many proper constraints we can define such that each constraint reduces the set of words satisfying these constraints. It is essentially the above compactness property which guarantees that these chains are finite.

Equations can be studied not only for words, that is in a free monoid, but also in other semigroups, see e.g. [23].

Instead of taking a word equation and looking for its solutions, we could also take some words and look for the equations they satisfy. If the words do not satisfy any nontrivial relation, then it is said that they form a code, or have the unique decipherability property. Codes have been studied a lot, see e.g. [5], and they are fundamentally important in message transmission.

The problem of determining whether a set is a code was probably first encountered when asking whether or not a finite encoding $i \mapsto w_i$ can be uniquely decoded, that is whether a finite set of words $\{w_i \mid i \in I\}$ is a free generating set of a submonoid of a free monoid. An affirmative answer to this problem is given by the classical *Sardinas-Patterson algorithm* [59]. This algorithm extends straightforwardly to regular languages, see, for example, Section I.3 in [5].

There are several options to try to extend the above problem. One such direction is to consider instead of the freeness of a finitely generated subsemigroup of Σ^*, the isomorphism of two such semigroups. This reveals some interesting phenomena. First of all the problem remains decidable, see [8], but the proof relies on something surprising, namely systems of equations over free semigroups and their compactness properties. Even more interestingly this approach does not extend to subsemigroups generated by regular languages – in fact, the decidability of their isomorphism is an open problem. Another interesting feature here is that when moving from subsemigroups of a free semigroup to more general semigroups the isomorphism problem becomes undecidable. For example, for finitely generated multiplicative semigroups of 3×3 matrices over natural numbers the isomorphism, or even the freeness, problem is undecidable, see [37] or [21] as a survey.

Like the notion of an equation, also the notion of a code can be extended to other monoids. The unique decipherability problem in a monoid $\mathcal{M}$ asks whether a given finite subset M of $\mathcal{M}$ is a free generating set of the submonoid of $\mathcal{M}$ it generates.

We are interested in the monoid of languages. The product of two languages A and B is defined as the language containing all words uv, where $u \in A$ and $v \in B$. Then the set of all languages is a monoid. Some problems that are easy for words are very hard in this monoid. For example, if $xy = yx$ for two words x, y, then x and y are powers of a common word, but no similar result holds for languages. In fact, the maximal language commuting with a given finite language is not necessarily even recursively enumerable [41]. As another example, it is undecidable whether $AB^iC = DE^iF$ for all

i, where A, B, C, D, E, F are given finite sets [32].

Recently an attempt to study codes in the monoid of languages was made in [10]. It was shown that the unique decipherability in the monoid of unary languages is decidable in the case of finite languages, as well as in the case of regular, that is ultimately periodic languages. Also a simple case of non-unary languages was settled affirmatively in [10]: if there is a letter that appears exactly once in every word of every language, then the problem is decidable. It is also known that the set of finite prefix sets is a free monoid, i.e. generated by a code [50].

Let us now give a brief overview of this work.

We start in Chapter 2 by giving some basic definitions, notation and theorems that will be used later.

In Chapter 3, based on [35], we analyze maximal independent systems of equations and maximal decreasing chains of equations, as well as search for their relations. We also survey known results and open problems. The most fundamental problem asks whether the maximal size of an independent system of word equations on n unknowns is bounded by some function of n. Amazingly, the same problem is open for three unknown equations, although we do not know independent systems of more than three equations in this case. The question about maximal sizes of decreasing chains of equations is equally open. We give new lower bounds in the cases of three and four unknowns: a chain of seven equations on three unknowns and a chain of twelve equations on four unknowns.

In Chapter 4, based on [56], we use polynomials to study some questions related to systems of word equations. Algebraic techniques have been used before, most notably in the proof of Ehrenfeucht's conjecture, which is based on Hilbert's basis theorem. However, the way in which we use polynomials is quite different and allows us to apply linear algebra to the problems. One of the main contributions of this chapter is the development of new methods for attacking problems on word equations. Other contributions include simplified proofs and generalizations for old results and studying maximal sizes of independent systems of equations. In particular, we get the first nontrivial lower bound for the size of independent systems of word equations on three unknowns. This bound depends on the length of the shortest equation in the system. Thus the connection between word equations and linear algebra is not only theoretically interesting, but is also shown to be very useful at establishing simple-looking results that have been previously unknown, or that have had only very complicated proofs. In addition to the results of this section, we believe that the techniques may be useful in further analysis of word equations.

In Chapter 5, based on [34] and [55], we analyze the proof of Hmelevskii's theorem. The result itself is, of course, very well known, see e.g. [43]. However, a compact and readable presentation of it seems to be lacking. We

hope to fill this gap. In other words, we search for a self-contained proof using achievements and tools of combinatorics on words obtained over the last decades. In addition, we conclude an upper bound for the size of the formula giving the general solution of a constant-free equation on three unknowns. Our bound is exponential in terms of the length of the equation. Based on the bound for the parametric solution, we prove that the length of the shortest nontrivial solution is also exponential (if such a solution exists). This connects our work to the satisfiability problem mentioned above, because Plandowski and Rytter proved [53] that there is a nondeterministic algorithm solving the problem in time polynomial in $n \log N$, where n is the length of the equation and N is the length of the shortest solution. From this and our result it follows that the problem of deciding if a constant-free equation on three unknowns has a nontrivial solution is in NP.

In Chapter 6, based on [57] and [36], we study unique decipherability in the monoid of languages. The monoid of unary languages is isomorphic to the additive monoid of sets of natural numbers, so in the unary case we will actually formulate everything in terms of sets of numbers. We will extend the result of [10] by giving a complete characterization of codes in the monoid of unary regular languages. We will also study the power equality problem, that is the problem of determining whether some powers of two sets are equal. As far as we know, the nonunary case is very much untouched. We show that, given a finite collection of regular languages, it is undecidable whether it is a free generating set in the monoid of languages. Our result is based on another undecidability result in [11] which states that the unique decipherability problem is undecidable in the trace monoid $\{a, b\}^* \times \{c, d\}^*$.

Finally, in Chapter 7 we give a summary of the results in this work.

Chapter 2

Preliminaries

We start this chapter with definitions and theorems about words, about equations in semigroups and about word equations. Then we determine the solutions of many simple equations and give some definitions related to parametric words. Finally we talk about unique decipherability. Proofs and more information on combinatorics on words can be found in [43] and [9].

2.1 Words

The set of nonnegative integers is denoted by $\mathbb{N}_0$ and the set of positive integers by $\mathbb{N}_1$.

An *alphabet* Σ is a set of symbols. The elements of Σ are also called *letters*. A *word* w over Σ is a finite sequence of these symbols: $w = a_1 \dots a_n$, where $a_1, \dots, a_n \in \Sigma$ and $n \in \mathbb{N}_0$. If $n = 0$, we get the *empty word*, denoted by ε. The *product* (or *catenation* or *concatenation*) of two words $u = a_1 \dots a_m$ and $v = b_1 \dots b_n$ is $uv = a_1 \dots a_m b_1 \dots b_n$.

If Σ is an alphabet, then Σ^* is the set of all words over Σ, and Σ^+ is the set of all nonempty words. Now Σ^* is a free monoid and Σ^+ is a free semigroup. The empty word ε is the neutral element in Σ^*.

The size of the alphabet is usually not important, as long as there are at least two letters. However, when talking about the ranks of solutions of equations, we have to assume that there are at least as many letters as unknowns.

The *length* of a word $w = a_1 \dots a_n \in \Sigma^*$ is $|w| = n$. The number of occurrences of a letter $a \in \Sigma$ in w is denoted by $|w|_a$.

A word $w \in \Sigma^*$ is a *factor* of a word $t \in \Sigma^*$ if there are $u, v \in \Sigma^*$ such that $t = uwv$. If $u = \varepsilon$, then w is a *prefix* of t. This is denoted by $w \leq t$. If also $w \neq t$, then w is a *proper prefix* and the notation $w < t$ is used.

A word $w \in \Sigma^+$ is *primitive* if it is not of the form u^k for any $u \in \Sigma^+$ and $k > 1$. Every word $w \in \Sigma^+$ can be represented uniquely as u^n with u

primitive; then u is the *primitive root* of w and the notation $u = \rho(w)$ is used.

The *reverse* of a word $w = a_1 \dots a_n$ is $w^R = a_n \dots a_1$.

2.2 Equations in Semigroups

We are mostly interested in word equations, but equations can be defined in any semigroup. In this section we give definitions related to equations in an arbitrary semigroup S. These will be used in Section 3.1, and also in many other sections in the case $S = \Sigma^*$.

Let S be a semigroup and Ξ be a finite nonempty alphabet of unknowns. A (coefficient-free) *equation* $u = v$ consists of two words $u, v \in \Xi^+$. A morphism $h : \Xi^+ \to S$ is a *solution* of this equation if $h(u) = h(v)$. The set of all solutions is denoted by $\mathrm{Sol}(u = v)$.

If S is a monoid, we can use Ξ^* instead of Ξ^+, so that equations like $u = \varepsilon$ are allowed. An equation $u = \varepsilon$ can also be written $u = 1$ if 1 is the neutral element of S.

A set of equations A is a *system of equations*. A morphism is a solution of this system if it a solution of every equation in A. The set of all solutions of A is denoted by $\mathrm{Sol}(A)$. We have

$$\mathrm{Sol}(A) = \bigcap_{E \in A} \mathrm{Sol}(E).$$

As a trivial boundary case, $\mathrm{Sol}(\varnothing)$ is the set of all morphisms $\Xi^+ \to S$.

Two equations (or systems of equations) are *equivalent* if they have the same solutions.

When writing systems of equations, we can omit the braces, so that a system $\{E_1, \dots, E_m\}$ can be written as $E_1, \dots, E_m$, and the set of its solutions as $\mathrm{Sol}(E_1, \dots, E_m)$.

If $\Xi = \{x_1, \dots, x_n\}$ and $w_1, \dots, w_n \in \Sigma^*$, then we can informally talk of the solution

$$x_1 = w_1, \dots, x_n = w_n.$$

This means the morphism h determined by

$$h(x_1) = w_1, \dots, h(x_n) = w_n.$$

An equation is *trivial* if every morphism $S \to \Xi^+$ is a solution; otherwise it is *nontrivial*.

An equation $u = v$ is *balanced* if $|u|_x = |v|_x$ for every unknown x; otherwise it is *unbalanced*.

Equations of the form $u = u$ are always trivial. In some semigroups there are also other trivial equations. For example, in a commutative semigroup every balanced equation is trivial.

A system of equations is *independent* if it is not equivalent to any of its proper subsystems. Another formulation, which is useful when showing that a specific system is independent, is that a system A is independent if for every $E \in A$ there is a morphism that is not a solution of E, but is a solution of all the other equations in A.

Solution sets of systems of equations form a partially ordered set where the order is given by set inclusion. It is natural to consider maximal sizes of chains in this partially ordered set. So if $A_0, \dots, A_m$ are systems of equations and

$$\mathrm{Sol}(A_0) \supsetneq \mathrm{Sol}(A_1) \supsetneq \mathrm{Sol}(A_2) \supsetneq \cdots \supsetneq \mathrm{Sol}(A_m),$$

then how large can m be? If m can be unboundedly large, can there be infinite decreasing or increasing chains?

It will be shown in Theorem 3.1.2 that it is sufficient to consider the case where the systems are of the form

$$A_0 = \varnothing, \quad A_1 = \{E_1\}, \quad A_2 = \{E_1, E_2\}, \quad \dots, \quad A_m = \{E_1, \dots, E_m\}.$$

This justifies the following definitions.

We define *decreasing chains* of equations. A finite sequence of equations $E_1, \dots, E_m$ is a decreasing chain if

$$\mathrm{Sol}(\varnothing) \supsetneq \mathrm{Sol}(E_1) \supsetneq \mathrm{Sol}(E_1, E_2) \supsetneq \cdots \supsetneq \mathrm{Sol}(E_1, \dots, E_m). \tag{2.1}$$

An infinite sequence of equations $E_1, E_2, \dots$ is a decreasing chain if

$$\mathrm{Sol}(\varnothing) \supsetneq \mathrm{Sol}(E_1) \supsetneq \mathrm{Sol}(E_1, E_2) \supsetneq \dots.$$

Similarly we define *increasing chains* of equations. A finite sequence of equations $E_1, \dots, E_m$ is an increasing chain if

$$\mathrm{Sol}(E_1, \dots, E_m) \subsetneq \mathrm{Sol}(E_2, \dots, E_m) \subsetneq \cdots \subsetneq \mathrm{Sol}(E_m) \subsetneq \mathrm{Sol}(\varnothing).$$

An infinite sequence of equations $E_1, E_2, \dots$ is an increasing chain if

$$\mathrm{Sol}(E_1, E_2, \dots) \subsetneq \mathrm{Sol}(E_2, E_3, \dots) \subsetneq \mathrm{Sol}(E_3, E_4, \dots) \subsetneq \dots.$$

Now $E_1, \dots, E_m$ is an increasing chain if and only if $E_m, \dots, E_1$ is a decreasing chain. However, for infinite chains these concepts are essentially different. Note that a chain can be both decreasing and increasing, for example, if the equations form an independent system.

In the definitions the inclusions between the solution sets are trivial, so we could as well write

$$\mathrm{Sol}(\varnothing) \neq \mathrm{Sol}(E_1) \neq \mathrm{Sol}(E_1, E_2) \neq \cdots \neq \mathrm{Sol}(E_1, \dots, E_m)$$

in place of (2.1), and so on. This also means that $E_1, \ldots, E_m$ is a decreasing chain if and only if for every $i \in \{1, \ldots, m\}$ there is a morphism that is not a solution of E_i, but is a solution of the system $E_1, \ldots, E_{i-1}$.

A semigroup has the *compactness property* if every system of equations has an equivalent finite subsystem. Many results on the compactness property are collected in [23]. In terms of chains, the compactness property turns out to be equivalent to the property that every decreasing chain is finite. This is proved in Theorem 3.1.3 and gives further justification for the importance of decreasing chains.

2.3 Word Equations

In this section we state some well known theorems related to word equations, that is equations in the free monoid Σ^*.

Let Ξ be a finite nonempty alphabet of unknowns. A (coefficient-free) *word equation* $u = v$ consists of two words $u, v \in \Xi^+$. A morphism $h : \Xi^* \to \Sigma^*$ is a *solution* of this equation if $h(u) = h(v)$.

This is of course just the definition of equations in the previous section written for free monoids. Similarly we can denote the set of solutions by Sol and extend all the definitions for systems of equations.

A solution h is *periodic* if there exists a $t \in \Sigma^*$ such that every $h(x)$, where $x \in \Xi$, is a power of t. Otherwise h is *nonperiodic*.

Now we state some basic auxiliary results that are needed later, for more see [9].

Theorem 2.3.1 (Commutation)**.** *A nontrivial equation on two unknowns has only periodic solutions.*

Theorem 2.3.2 (Periodic solutions)**.** *Let $\Xi = \{x_1, \ldots, x_n\}$. The periodic solutions of an equation $u = v$ are*

$$x_1 = t^{i_1}, \ldots, x_n = t^{i_n},$$

where $t \in \Sigma^*$ and $i_1, \ldots, i_n$ are numbers satisfying the linear relation

$$|u|_{x_1} i_1 + \cdots + |u|_{x_n} i_n = |v|_{x_1} i_1 + \cdots + |v|_{x_n} i_n.$$

Theorem 2.3.3 (Conjugation)**.** *The solutions of the equation $xz = zy$ are*

$$x = pq,\ y = qp,\ z = p(qp)^i \qquad or \qquad x = y = \varepsilon,\ z = p,$$

where $p, q \in \Sigma^*$ and $i \geq 0$.

Theorem 2.3.4 (Fine and Wilf)**.** *Let $u, v \in \Sigma^+$, $u' < u$, $v' < v$ and*

$$|u^m u'| = |v^n v'| \geq |u| + |v| - \gcd(|u|, |v|).$$

Now $u^m u' = v^n v'$ if and only if $uv = vu$.

The (combinatorial) *rank* of a morphism h is the smallest number r for which there is a set A of r words such that $h(x) \in A^*$ for every unknown x.

A morphism is periodic if and only if its rank is at most one.

It is well known that any nontrivial equation on n variables forces a defect effect; that is, the values of the variables in any solution can be expressed as products of $n-1$ words (see [22] for a survey on the defect effect).

Theorem 2.3.5 (Defect theorem)**.** *Every solution of a nontrivial equation on n variables has rank at most $n-1$.*

The *graph* of a system of word equations is the graph where Ξ is the set of vertices and there is an edge between x and y if one of the equations in the system is of the form $x \cdots = y \cdots$.

Theorem 2.3.6 (Graph lemma)**.** *Consider a system of equations whose graph has r connected components. If h is a solution of this system and $h(x) \neq \varepsilon$ for all $x \in \Xi$, then the rank of h is at most r.*

The above theorem has the following corollary, which is particularly useful in Chapter 5.

Corollary 2.3.7. *Let $A, B, C, D \in \{x, y, z\}^*$. If h is a solution of the pair of equations $xA = yB, xC = zD$ and if $h(x), h(y), h(z) \neq \varepsilon$, then h is periodic.*

The following theorem, proved in [24], is sometimes useful.

Theorem 2.3.8. *If an independent pair of equations on three unknowns has a nonperiodic solution, then the equations must be balanced.*

We will generalize and reproof this result in Section 4.5.

If we need to replace a system of equations with a single equation, the following theorem can be used.

Theorem 2.3.9. *Let E_1, E_2 be a pair of equations. There is an equation that has the same nonperiodic solutions as the pair. If at least one of E_1, E_2 is balanced, then there is an equation that is equivalent to the pair.*

Proof. The first claim was proved in [26]. If E_i is the equation $u_i = v_i$ and E_1 is balanced, then the pair E_1, E_2 is equivalent to the equation $u_1u_2 = v_1v_2$, so the second claim holds. □

Two unbalanced equations can't necessarily be combined. For example, the pair $x = y$, $x = z$ is not equivalent to any single equation.

It was conjectured by Ehrenfeucht in a language theoretic setting that the compactness property holds for free monoids. This conjecture was reformulated in terms of equations in [12], and it was proved independently by Albert and Lawrence [2] and by Guba [20].

Theorem 2.3.10 (Ehrenfeucht's compactness property)**.** *Every infinite system of word equations has an equivalent finite subsystem.*

The proofs are based on Hilbert's basis theorem.

2.4 Examples of Word Equations

In this section we solve some simple word equations. These work as examples, but they are also used in Chapter 5.

The following lemma is needed in Lemma 2.4.8.

Lemma 2.4.1. *If $ww = uwv$, where $u \neq \varepsilon$ and $v \neq \varepsilon$, then w is not primitive.*

We continue by solving a few examples of word equations that are needed later.

Lemma 2.4.2. *The nonperiodic solutions of the equation $xyz = zyx$ are*

$$x = (pq)^i p,\ y = q(pq)^j,\ z = (pq)^k p,$$

where $p, q \in \Sigma^*$, $i, j, k \geq 0$, $pq \neq qp$ and pq can be assumed to be primitive.

Proof. The claimed solutions satisfy the equation and are nonperiodic. If h is a nonperiodic solution, then $h(xyzy) = h(zyxy)$, so $h(xy) = t^m$ and $h(zy) = t^n$, where t is primitive and $m, n > 0$. Now $h(y) = q(pq)^j$, where $pq = t$ and $0 \leq j < m, n$, so $h(x) = (pq)^i p$ and $h(z) = (pq)^k p$, where $i = m - j - 1$ and $k = n - j - 1$. If p and q would commute, the solution would be periodic. □

Lemma 2.4.3. *The nonperiodic solutions of the equation $xyz = zxy$ are*

$$x = (pq)^i p,\ y = q(pq)^j,\ z = (pq)^k,$$

where $p, q \in \Sigma^*$, $i, j, k \geq 0$ and $pq \neq qp$.

Proof. The claimed solutions satisfy the equation and are nonperiodic. If h is a nonperiodic solution, then $h(xy) = t^m$ and $h(z) = t^k$, where $m > 0, k \geq 0$. Now $h(y) = q(pq)^j$, where $pq = t$ and $0 \leq j < m$, so $h(x) = (pq)^i p$ and $h(z) = (pq)^k$, where $i = m - j - 1$. If p and q would commute, the solution would be periodic. □

Lemma 2.4.4. *Let $a \geq 2$. The nonperiodic solutions of the equation $xzx = y^a$ are*

$$x = (pq)^i p,\ y = (pq)^{i+1} p,\ z = qp((pq)^{i+1} p)^{a-2} pq,$$

where $p, q \in \Sigma^*$, $i \geq 0$ and $pq \neq qp$.

Proof. The claimed solutions satisfy the equation and are nonperiodic. Let h be a nonperiodic solution. If it would be $|h(x)| \geq |h(y)|$, then $h(xz)$ and $h(y)$ would be powers of a common word by Theorem 2.3.4, and h would be periodic. Thus $|h(x)| < |h(y)|$. Now $h(y) = uh(x) = h(x)v$, where $u, v \neq \varepsilon$, and $h(z) = vh(y)^{a-2}u$. By Theorem 2.3.3, $u = pq$, $v = qp$, $h(x) = (pq)^i p$, $h(y) = (pq)^{i+1}p$ and $h(z) = qp((pq)^{i+1}p)^{a-2}pq$. If p and q would commute, the solution would be periodic. □

Lemma 2.4.5. *Let $a \geq 2$. The nonperiodic solutions of the equation $xy^a z = zy^a x$ are*

$$x = (pq^a)^i p,\ y = q,\ z = (pq^a)^j p \qquad \text{or}$$
$$\begin{cases} x &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^i, \\ y &= (pq)^{k+1}p, \\ z &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^j, \end{cases}$$

where $p, q \in \Sigma^*$, $i, j, k \geq 0$ and $pq \neq qp$.

Proof. The claimed solutions satisfy the equation and are nonperiodic. If h is a nonperiodic solution, then, by Lemma 2.4.2,

$$h(x) = u(vu)^i,\ h(y^a) = v(uv)^b,\ h(z) = u(vu)^j,$$

where uv is primitive. If $b = 0$, this gives a solution of the first form. If $b > 1$, then, by Theorem 2.3.4, $h(y)$ and vu commute. Then $u = \varepsilon$ or $v = \varepsilon$ and $h(x), h(y), h(z) \in (uv)^*$, which is a contradiction. If $b = 1$, then, by Lemma 2.4.4, $v = (pq)^k p$, $h(y) = (pq)^{k+1}p$ and $u = qp((pq)^{k+1}p)^{a-2}pq$. This gives a solution of the second form. If p and q would commute, the solution would be periodic. □

Lemma 2.4.6. *Those nonperiodic solutions of the equation $xyxz = zx^2y$ that satisfy $|x| \geq |z|$ are*

$$x = (pq)^i p,\ y = qp((pq)^{i+1}p)^j pq,\ z = pq,$$

where $p, q \in \Sigma^*$, $i \geq 1$, $j \geq 0$ and $pq \neq qp$.

Proof. The claimed solutions satisfy the equation and are nonperiodic. If h is a nonperiodic solution, then, by Lemma 2.4.2,

$$h(xy) = (uv)^b u,\ h(x) = v(uv)^c,\ h(z) = (uv)^d u.$$

Because $h(z) \leq h(x) \leq h(xy)$ and $uv \neq vu$, it must be $h(z) = u \leq h(x) = v \leq uv$. Now $h(z) = pq$ and $h(x) = (pq)^i p$, so $y = qp((pq)^{i+1}p)^j pq$. If p and q would commute, the solution would be periodic. □

Lemma 2.4.7. *Let $a, b \geq 1$ and $U, V \in \Xi^*$. If h is a solution of the equation $x^a yU = y^b xV$, then $h(x)$ and $h(y)$ commute.*

Proof. Assume that $h(x) \leq h(y)$. Then $h(y) = h(x)^c t$, where $h(x) \not\leq t$. Because $h(x)^{a+c} \cdots = h(x)^c t \cdots$, it must be $t \leq h(x)$. Now $h(x)^{a+c} t \cdots = h(y)^b h(x) \cdots$ and $|h(x)^{a+c} t|, |h(y)^b h(x)| \geq |h(x)h(y)|$. The claim follows by Theorem 2.3.4. □

Lemma 2.4.8. *The nonperiodic solutions of the equation $xyxzyz = zxzyxy$ are*

$$x = p,\ y = q,\ z = \varepsilon \qquad or \qquad x = p,\ y = q,\ z = pq,$$

where $p, q \in \Sigma^*$ and $pq \neq qp$.

Proof. The claimed solutions satisfy the equation and are nonperiodic. If h is a nonperiodic solution, then, by Lemma 2.4.2,

$$h(xy) = (uv)^i u,\ h(xzy) = v(uv)^j,\ h(z) = (uv)^k u,$$

where uv is primitive. If $|h(x)| \geq |uv|$, then uv and vu are both prefixes of $h(x)$, $uv = vu$, and the solution is periodic. Thus $|h(x)| < |uv|$. Symmetrically $|h(y)| < |uv|$, so $i = 0$ or $i = 1$. If $k > 0$ and $h(x) \neq v$, then uv is a factor of $uvuv$ in a nontrivial way, which contradicts the primitivity of uv by Lemma 2.4.1. If $h(x) = v$, then $h(y) = v(uv)^l$ for some l, u and v satisfy a nontrivial relation and the solution is periodic. Thus $k = 0$ and $h(z) = u$. If $i = 0$, then $h(xy) = h(z)$. If $i = 1$, then either $h(z) = \varepsilon$ or $j = 1$ or $j = 2$ or $v = \varepsilon$. If $j = 1$, then $|v| = 2|u|$, u is a prefix and a suffix of v, and u and v commute. If $j = 2$, then $|u| = 2|v|$, v is a prefix and a suffix of u, and u and v commute. If $v = \varepsilon$, then $|h(x)|, |h(y)| < |uv|$ is not possible. This proves that the claimed solutions are all nonperiodic solutions. If p and q would commute, the solution would be periodic. □

We will use the equation of Lemma 2.4.8 several times. Essentially the same equation was also used in [42] to prove a result about one unknown equations with constants.

2.5 Parametric Words

In this section we define parametric words, parameterizability and parametric solutions. These definitions are used in Chapter 5.

We fix the alphabet of *word parameters* Δ and the set of *numerical parameters* Λ. Now *parametric words* are defined inductively as follows:

(i) if $a \in \Delta \cup \{\varepsilon\}$, then (a) is a parametric word,

(ii) if α and β are parametric words, then so is $(\alpha\beta)$,

(iii) if α is a parametric word and $i \in \Lambda$, then (α^i) is a parametric word.

The set of parametric words is denoted by $\mathcal{P}(\Delta, \Lambda)$. The sets of parameters are always denoted by Δ and Λ.

When there is no danger of confusion, unnecessary parenthesis can be omitted and notations like $\alpha^i\alpha^j = \alpha^{i+j}$ and $(\alpha^i)^j = \alpha^{ij}$ can be used. Then parametric words form a monoid if the product of α and β is defined to be $\alpha\beta$.

If f is a function $\Lambda \to \mathbb{N}_0 = \{0, 1, 2, \dots\}$, we can abuse the notation and use the same symbol for the function that maps parametric words by giving values for the numerical parameters with f: if $a \in \Delta \cup \{\varepsilon\}$, then $f((a)) = a$; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then $f((\alpha\beta)) = f(\alpha)f(\beta)$; if $\alpha \in \mathcal{P}(\Delta, \Lambda)$ and $i \in \Lambda$, then $f((\alpha^i)) = f(\alpha)^{f(i)}$. A parametric word is thus mapped by f to a word of Δ^*. This can be further mapped by a morphism $h : \Delta^* \to \Sigma^*$ to a word of Σ^*. The mapping $h \circ f$ is a *valuation* of a parametric word into Σ^*, and f is its valuation to the set Δ^*.

We define the *length* of a parametric word: the length of ε is zero; if $a \in \Delta$, then the length of a is one; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then the length of $\alpha\beta$ is the sum of the lengths of α and β; if $\alpha \in \mathcal{P}(\Delta, \Lambda) \setminus \{\varepsilon\}$ and $i \in \Lambda$, then the length of α^i is the length of α plus one. The length of α is denoted by $|\alpha|$.

Next we define the *height* of a parametric word: if $a \in \Delta \cup \{\varepsilon\}$, then the height of a is zero; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then the height of $\alpha\beta$ is the maximum of the heights of α and β; if $\alpha \in \mathcal{P}(\Delta, \Lambda) \setminus \{\varepsilon\}$ and $i \in \Lambda$, then the height of α^i is the height of α plus one. Parametric words of height zero can be considered to be words of Δ^*.

A *linear Diophantine relation* R is a disjunction of systems of linear Diophantine equations with lower bounds for the unknowns. For example,

$$((x + y - z = 0) \wedge (x \geq 2)) \vee ((x + y = 3) \wedge (x + z = 4))$$

is a linear Diophantine relation over the unknowns x, y and z. We are only interested in the nonnegative values of the unknowns. If $\Lambda = \{i_1, \dots, i_k\}$, f is a function $\Lambda \to \mathbb{N}_0$, and $f(i_1), \dots, f(i_k)$ satisfy R, then the notation $f \in R$ can be used.

Let S be a set of morphisms $\Xi^* \to \Sigma^*$, $\Lambda = \{i_1, \dots, i_k\}$, h_j a morphism from the monoid Ξ^* to parametric words and R_j a linear Diophantine relation, when $j = 1, \dots, m$. The set $\{(h_j, R_j) \mid 1 \leq j \leq m\}$ is a *parametric representation* of S if

$$S = \{h \circ f \circ h_j \mid 1 \leq j \leq m, f \in R_j\},$$

where $h \circ f$ runs over all valuations to Σ^*. The linear Diophantine relations are not strictly necessary, but they make some proofs easier. A set can be *parameterized* if it has a parametric representation.

It follows immediately that if two sets can be parameterized, then also their union can be parameterized.

The *length* of the parametric representation is the sum of the lengths of all $h_j(x)$, where $j = 1, \ldots, m$ and $x \in \Xi$. This definition does not take into account the linear Diophantine relations. For example, if a is a number, then the length of p^a is a, but if i is a numerical parameter appearing nowhere in the representation and if the equality $i = a$ is added to the linear Diophantine relation, then p^a can be replaced with p^i, whose length is only two.

Let $S, S_1, \ldots, S_n$ be sets of morphisms $\Xi^* \to \Sigma^*$. The set S can be *parameterized in terms of the sets* $S_1, \ldots, S_n$ if there exists morphisms $h_1, \ldots, h_n$ from Ξ^* to $\mathcal{P}(\Xi, \Lambda)$ such that

$$S = \{g \circ f \circ h_j \mid 1 \leq j \leq n, g \in S_j\},$$

where f runs over functions $\Lambda \to \mathbb{N}_0$.

Again it is a direct consequence of the definitions that the parameterizability is preserved in compositions. Namely, if S can be parameterized in terms of the sets $S_1, \ldots, S_n$ and every S_i can be parameterized in terms of the sets $S_{i1}, \ldots, S_{in_i}$, then S can be parameterized in terms of the sets S_{ij}.

We conclude these definitions by saying that solutions of an equation can be *parameterized* if the set of its all solutions can be parameterized. A parametric representation of this set is a *parametric solution* of the equation.

These definitions can be generalized in an obvious way for systems of equations. Theorem 2.3.3 and Lemmas 2.4.2 – 2.4.8 give parametric solutions for some equations. For example, the conjugacy equation $xz = zy$ has a parametric solution $\{(h_1, R), (h_2, R)\}$, where $\Delta = \{p, q\}$, $\Lambda = \{i\}$, $h_1(x) = pq$, $h_1(y) = qp$, $h_1(z) = p(qp)^i$, $h_2(x) = h_2(y) = \varepsilon$, $h_2(z) = p$ and R is the trivial relation satisfied by all functions $f : \Lambda \to \mathbb{N}_0$.

Hmelevskii proved [26] that every equation on three unknowns has a parametric solution. Giving a new proof for this theorem is the topic of Chapter 5.

2.6 Unique Decipherability

This section gives preliminaries for Chapter 6.

A set of words $\{w_1, \ldots, w_n\}$ is a *code*, or has the *unique decipherability property*, if

$$x_1 = w_1, \ldots, x_n = w_n$$

is not a solution of any nontrivial word equation.

A good reference on the theory of codes is [5]. There is a well-known algorithm for determining whether a given set of words is a code [59].

Codes can also be defined in other semigroups than Σ^*: a subset of a semigroup is a code if the elements of the subset do not satisfy a nontrivial equation.

We are interested in the case where the semigroup is the set of all regular languages over Σ. The product of two languages is defined in a usual way: $AB = \{uv \mid u \in A, v \in B\}$.

There are two essentially different cases depending on whether Σ has just one letter or at least two.

If $\Sigma = \{a\}$, then the semigroup of languages is isomorphic to the additive semigroup of sets of nonnegative integers, where the product of two sets is

$$AB = \{x + y \mid x \in A, y \in B\}.$$

The isomorphism is given by $\{a^{k_1}, a^{k_2}, \dots\} \mapsto \{k_1, k_2, \dots\}$. Thus we can consider sets of numbers instead of unary languages.

Chapter 3

Chains and Systems of Equations

In this chapter we study equations, first in an arbitrary semigroup and then in a free monoid. We survey known results and give examples. We prove some elementary theorems that, as far as we know, have not been explicitly stated before. We also give some new lower bounds for the sizes of decreasing chains of equations.

Section 3.1 is devoted to equations in semigroups.

In Section 3.2 we make basic observations about word equations.

In Section 3.3 we make a remark about the trivial cases of one and two unknowns and give lower bounds for the maximal sizes of systems and chains in the cases of three and four unknowns.

In Section 3.4 we consider lower bounds in the general case of n unknowns.

In Section 3.5 we mention some related question.

This chapter is based on the article [35].

3.1 Systems and Chains in Semigroups

In this section we consider independent systems and chains of equations in semigroups. The definitions were given in Section 2.2.

More precisely, we will consider the *maximal* sizes of independent systems of equations and *chains* of equations. If the number of unknowns is n, then the maximal size of an independent system is denoted by $\mathrm{IS}(n)$. We use two special symbols UB and ∞ for the infinite cases: if there are infinite independent systems, then $\mathrm{IS}(n) = \infty$, and if there are only finite but unboundedly large independent systems, then $\mathrm{IS}(n) = \mathrm{UB}$. We extend the order relation of numbers to these symbols: $k < \mathrm{UB} < \infty$ for every integer

k. Similarly the maximal size of a decreasing chain is denoted by $\mathrm{DC}(n)$, and the maximal size of an increasing chain by $\mathrm{IC}(n)$.

Independent systems of equations are a well-known topic (see, e.g., [23]). Chains of equations have been studied less, so we prove here some elementary results about them. The following theorem states the most basic relations between IS, DC and IC.

Theorem 3.1.1. *For every n, $\mathrm{IS}(n) \leq \mathrm{DC}(n), \mathrm{IC}(n)$. If $\mathrm{DC}(n) < \mathrm{UB}$ or $\mathrm{IC}(n) < \mathrm{UB}$, then $\mathrm{DC}(n) = \mathrm{IC}(n)$.*

Proof. Every independent system of equations is also a decreasing and increasing chain of equations, regardless of the order of the equations. This means that $\mathrm{IS}(n) \leq \mathrm{DC}(n), \mathrm{IC}(n)$.

A finite sequence of equations is a decreasing chain if and only if the reverse of this sequence is an increasing chain. Thus $\mathrm{DC}(n) = \mathrm{IC}(n)$ if $\mathrm{DC}(n) < \mathrm{UB}$ or $\mathrm{IC}(n) < \mathrm{UB}$. □

Now we prove two theorems mentioned in Section 2.2.

Theorem 3.1.2. *If there are systems of equations $A_0, \ldots, A_m$ such that*

$$\mathrm{Sol}(A_0) \supsetneq \cdots \supsetneq \mathrm{Sol}(A_m), \tag{3.1}$$

then $\mathrm{DC}(n) \geq m$. If there are systems of equations $A_0, A_1, \ldots$ such that

$$\mathrm{Sol}(A_0) \supsetneq \mathrm{Sol}(A_1) \supsetneq \ldots, \tag{3.2}$$

then $\mathrm{DC}(n) = \infty$. If there are systems of equations $A_1, A_2, \ldots$ such that

$$\mathrm{Sol}(A_1) \subsetneq \mathrm{Sol}(A_2) \subsetneq \ldots, \tag{3.3}$$

then $\mathrm{IC}(n) = \infty$.

Proof. First, assume that (3.1) holds. If we replace every A_i with $A_0 \cup \cdots \cup A_i$, then (3.1) still holds, so we can assume that $A_0 \subsetneq \cdots \subsetneq A_m$. For every $i \in \{1, \ldots, m\}$, there is a solution $h_i \in \mathrm{Sol}(A_{i-1}) \setminus \mathrm{Sol}(A_i)$ and an equation $E_i \in A_i \setminus A_{i-1}$ such that $h_i \notin \mathrm{Sol}(E_i)$. Now $E_1, \ldots, E_m$ is a decreasing chain.

Second, assume that (3.2) holds. If we replace every A_i with $A_0 \cup \cdots \cup A_i$, then (3.2) still holds, so we can assume that $A_0 \subsetneq A_1 \subsetneq \ldots$. For every $i \in \{1, 2, \ldots\}$, there is a solution $h_i \in \mathrm{Sol}(A_{i-1}) \setminus \mathrm{Sol}(A_i)$ and an equation $E_i \in A_i \setminus A_{i-1}$ such that $h_i \notin \mathrm{Sol}(E_i)$. Now $E_1, E_2, \ldots$ is an infinite decreasing chain.

Third, assume that (3.3) holds. If we replace every A_i with $A_i \cup A_{i+1} \cup \ldots$, then (3.3) still holds, so we can assume that $A_1 \supsetneq A_2 \supsetneq \ldots$. For every $i \in \{1, 2, \ldots\}$, there is a solution $h_i \in \mathrm{Sol}(A_{i+1}) \setminus \mathrm{Sol}(A_i)$ and an equation $E_i \in A_i \setminus A_{i+1}$ such that $h_i \notin \mathrm{Sol}(E_i)$. Now $E_1, E_2, \ldots$ is an infinite increasing chain. □

Theorem 3.1.3. *A semigroup has the compactness property if and only if* $\mathrm{DC}(n) \leq \mathrm{UB}$ *for every* n.

Proof. Assume first that the compactness property holds. Let $E_1, E_2, \ldots$ be an infinite decreasing chain of equations. As a system of equations, it is equivalent to some finite subsystem $E_{i_1}, \ldots, E_{i_k}$, where $i_1 < \cdots < i_k$. But now the system $E_1, \ldots, E_{i_k}$ is equivalent to $E_1, \ldots, E_{i_k+1}$. This is a contradiction.

Assume then that $\mathrm{DC}(n) \leq \mathrm{UB}$ for every n. Let $E_1, E_2, \ldots$ be an infinite system of equations. If there is an index N such that $E_1, \ldots, E_i$ is equivalent to $E_1, \ldots, E_{i+1}$ for all $i \geq N$, then the whole system is equivalent to $E_1, \ldots, E_N$. If there is no such index, then let $i_1 < i_2 < \ldots$ be all indexes such that $E_1, \ldots E_{i_k}$ is not equivalent to $E_1, \ldots, E_{i_k+1}$. But then $E_{i_1}, E_{i_2}, \ldots$ is an infinite decreasing chain, which is a contradiction. □

This theorem is one reason why decreasing chains seem to be more interesting than increasing ones. Another reason is that we do not know an example where $\mathrm{IC}(n) = \mathrm{UB}$. If there would not be such examples, then $\mathrm{IC}(n)$ would be completely determined by $\mathrm{DC}(n)$.

Now we examine what kind of combinations of values IS, DC and IC can have. The first two examples give the extreme cases: all three values can be finite, and all three values can be infinite.

Example 3.1.4. In any finite semigroup, $\mathrm{IS}(n)$, $\mathrm{DC}(n)$ and $\mathrm{IC}(n)$ are finite for all n.

Example 3.1.5. In the monoid of functions $\mathbb{Z} \to \mathbb{Z}$, $\mathrm{IS}(2) = \mathrm{DC}(2) = \mathrm{IC}(2) = \infty$. For $j > 0$, let $f_j(a) = 1$ if $a = 2j$ and $f_j(a) = 0$ otherwise. Let $g(a) = a + 2$ for all a. Now $f_j \circ g^i \circ f_j = f_j \circ f_j$ if and only if $i \neq j$, so the infinite system of equations $xy^i x = xx$ $(i = 1, 2, \ldots)$ is independent.

The next example shows that the values of IS, DC and IC can differ significantly.

Example 3.1.6. We give an example of a monoid where $\mathrm{IS}(1) = 1$, $\mathrm{DC}(1) = \mathrm{UB}$ and $\mathrm{IC}(1) = \infty$. The monoid is

$$\langle a_1, a_2, \ldots \mid a_i a_j = a_j a_i,\ a_i^{i+1} = a_i^i \rangle.$$

Every equation on one unknown is of the form $x^i = x^j$. If $i < j$, then this is equivalent to $x^i = x^{i+1}$. So all nontrivial equations are, up to equivalence,

$$x = 1,\ x^2 = x,\ x^3 = x^2,\ \ldots$$

and these have strictly increasing solution sets. Thus $\mathrm{IC}(1) = \infty$, $\mathrm{DC}(1) = \mathrm{UB}$ and $\mathrm{IS}(1) = 1$.

In the next example we need the fact that every commutative monoid has the compactness property, see [23].

Example 3.1.7. In the monoid of complex roots of unity, $\mathrm{IS}(1) = \mathrm{DC}(1) = \mathrm{UB}$ and $\mathrm{IC}(1) = \infty$. Let $p_1, p_2, \ldots$ be distinct primes. If $N = p_1 \ldots p_n$, then the equations $x^{N/p_i} = 1$, where $i = 1, \ldots, n$, form an independent system, so $\mathrm{DC}(1) \geq \mathrm{IS}(1) \geq \mathrm{UB}$. On the other hand, $\mathrm{IS}(1) \leq \mathrm{DC}(1) \leq \mathrm{UB}$, because the monoid is commutative. The equations

$$x^{p_1} = 1,\ x^{p_1 p_2} = 1,\ x^{p_1 p_2 p_3} = 1,\ \ldots$$

form an increasing chain of equations, so $\mathrm{IC}(1) = \infty$.

In the next section we will move to free monoids. They satisfy the compactness property, but the finiteness of $\mathrm{IS}(n)$ is open for $n \geq 3$. This should be compared to the related case of free groups. It was proved by Albert and Lawrence [1] that the compactness property holds for free groups but there are unboundedly large independent systems. Thus the situation is similar as in Example 3.1.7.

Example 3.1.8. Let $\Xi = \{x, y, z\}$ and let S be the free group generated by a and b. In this example we consider *group equations* $u = 1$, where u is an element of the free group generated by Ξ.

Let $[u, v] = u^{-1}v^{-1}uv$ be the *commutator* of u and v and, for $m > 2$, let $[u_1, \ldots, u_m] = [[u_1, \ldots, u_{m-1}], u_m]$ be the *generalized commutator*.

Let $v_i = z^{-i}x^i y^{-1} z^i$. Now the group equations

$$E_i : [v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m] = 1,$$

where $i \in \{1, \ldots, m\}$, form an independent system, because

$$x = a,\ y = a^i,\ z = b$$

is not a solution of E_i but is a solution of every other E_j. Thus there are arbitrarily large independent systems of group equations on three unknowns.

In general, group equations $u = 1$ are not semigroup equations or monoid equations as defined in Section 2.2, because u can contain inverses of the unknowns. However, we can transform a group equation into a monoid equation by replacing x^{-1}, y^{-1} and z^{-1} with new unknowns $\bar{x}$, $\bar{y}$ and $\bar{z}$. Then the above system $E_1, \ldots, E_m$ is transformed into an independent system on six unknowns. This shows that $\mathrm{IS}(6) = \mathrm{UB}$ in the free group S. If we would alter the definition of IS to allow group equations, then we would get $\mathrm{IS}(3) = \mathrm{UB}$.

3.2 Systems and Chains of Word Equations

For the rest of this chapter, we will study independent systems and decreasing chains in the case of free monoids and free semigroups. The symbols IS and DC will always refer to the maximal sizes of independent systems and decreasing chains in the free monoid. In the free semigroup we will use the symbols IS_+ and DC_+.

Periodic solutions are not very interesting, so sometimes it makes sense to consider only systems that have a nonperiodic solution. The maximal size of an independent system in a free monoid having a nonperiodic solution is denoted by IS'. Similarly the maximal size of a decreasing chain having a nonperiodic solution is denoted by DC'. Similar notation can be used for free semigroups.

We can study independent systems or decreasing chains, we can do this in the free monoid or in the free semigroup, and we can choose whether to require that there is a nonperiodic solution. This gives eight related questions, namely what are the values of $\mathrm{IS}, \mathrm{DC}, \mathrm{IS}_+, \mathrm{DC}_+, \mathrm{IS}', \mathrm{DC}', \mathrm{IS}'_+, \mathrm{DC}'_+$.

It is clear that we have

$$\begin{aligned} \mathrm{IS}_+(n) &\le \mathrm{IS}(n), & \mathrm{DC}_+(n) &\le \mathrm{DC}(n), \\ \mathrm{IS}'_+(n) &\le \mathrm{IS}'(n), & \mathrm{DC}'_+(n) &\le \mathrm{DC}'(n) \end{aligned}$$

and

$$\begin{aligned} \mathrm{IS}(n) &\le \mathrm{DC}(n), & \mathrm{IS}_+(n) &\le \mathrm{DC}_+(n), \\ \mathrm{IS}'(n) &\le \mathrm{DC}'(n), & \mathrm{IS}'_+(n) &\le \mathrm{DC}'_+(n) \end{aligned}$$

for all n. We can say a little bit more about the relation between $\mathrm{DC}(n)$ and $\mathrm{DC}'(n)$ and between the other corresponding numbers. If a decreasing chain has a nonperiodic solution, then we can add an equation that forces all unknowns to commute. After that we can add at least one equation, for example $x_1 \dots x_n = 1$, and at most n equations by Theorem 2.3.2 and basic linear algebra. Thus

$$\begin{aligned} \mathrm{DC}'(n) + 2 \le \mathrm{DC}(n) \le \mathrm{DC}'(n) + n + 1, \\ \mathrm{DC}'_+(n) + 2 \le \mathrm{DC}_+(n) \le \mathrm{DC}'_+(n) + n + 1. \end{aligned}$$

for $n \ge 2$. In the case of independent systems we get even tighter bounds. To see this, let $E_1, \dots, E_m$ be an independent system and $h_1, \dots, h_m$ be solutions such that h_i satisfies all of the equations except E_i. If one of the solutions, say h_1, is not periodic, then $E_2, \dots, E_m$ is an independent system with a nonperiodic solution. On the other hand, if every h_i is periodic, then $m \le n$ by linear algebra, and $\mathrm{IS}'_+(n) \ge n - 1$ for $n \ge 3$. Thus

$$\begin{aligned} \mathrm{IS}'(n) \le \mathrm{IS}(n) \le \mathrm{IS}'(n) + 1, \\ \mathrm{IS}'_+(n) \le \mathrm{IS}_+(n) \le \mathrm{IS}'_+(n) + 1, \end{aligned}$$

for $n \geq 3$. This means that IS' and DC' are basically the same as IS and DC if we are only interested in their finiteness or asymptotic growth.

The next theorem follows from Theorems 2.3.10 and 3.1.3.

Theorem 3.2.1. *For all n, we have* $\mathrm{DC}(n) \leq \mathrm{UB}$*, and hence also* $\mathrm{IS}(n) \leq \mathrm{UB}$*.*

No better upper bounds than $\mathrm{DC}(n) \leq \mathrm{UB}$ are known for $n > 2$. Even the seemingly simple question about the size of $\mathrm{IS}'(3)$ is still completely open; the only thing that is known is that $2 \leq \mathrm{IS}'(3) \leq \mathrm{UB}$. The lower bound is given by the example $xyz = zyx, xyyz = zyyx$.

3.3 Three and Four Unknowns

Word equations on one unknown are completely trivial. They are all equivalent to either the equation $x = x$ or the equation $x = \varepsilon$, so

$$\begin{aligned} \mathrm{IS}(1) = \mathrm{DC}(1) = \mathrm{IS}_+(1) = \mathrm{DC}_+(1) = 1, \\ \mathrm{IS}'(1) = \mathrm{DC}'(1) = \mathrm{IS}'_+(1) = \mathrm{DC}'_+(1) = 0. \end{aligned}$$

Equations on two unknowns are not much more interesting. By Theorems 2.3.1 and 2.3.2,

$$\begin{aligned} \mathrm{IS}(2) = \mathrm{IS}_+(2) = 2, \\ \mathrm{DC}(2) = \mathrm{DC}_+(2) = 3, \\ \mathrm{IS}'(2) = \mathrm{DC}'(2) = \mathrm{IS}'_+(2) = \mathrm{DC}'_+(2) = 0. \end{aligned}$$

As soon as there are at least three unknowns, the questions become much more difficult. The cases of three and four variables were studied in [13]. The article gives examples showing that $\mathrm{IS}'_+(3) \geq 2$, $\mathrm{DC}_+(3) \geq 6$, $\mathrm{IS}'_+(4) \geq 3$ and $\mathrm{DC}_+(4) \geq 9$. We are able to give better bounds for $\mathrm{DC}_+(3)$ and $\mathrm{DC}(4)$.

First we assume that there are three unknowns x, y, z. There are trivial examples of independent systems of three equations, for example, $x^2 = y, y^2 = z, z^2 = x$, so $\mathrm{IS}_+(3) \geq 3$. There are also easy examples of independent pairs of equations having a nonperiodic solution, like $xyz = zyx, xyyz = zyyx$, so $\mathrm{IS}'_+(3) \geq 2$. Amazingly, no other bounds are known for $\mathrm{IS}_+(3)$, $\mathrm{IS}'_+(3)$, $\mathrm{IS}(3)$ or $\mathrm{IS}'(3)$.

The following chain of equations shows that $\mathrm{DC}(3) \geq 7$:

$$\begin{aligned} xyz = zxy, &\qquad x = a,\ y = b,\ z = abab \\ xyxzyz = zxzyxy, &\qquad x = a,\ y = b,\ z = ab \\ xz = zx, &\qquad x = a,\ y = b,\ z = \varepsilon \end{aligned}$$

$$
\begin{array}{rl}
xy = yx, & x = a,\ y = a,\ z = a \\
x = \varepsilon, & x = \varepsilon,\ y = a,\ z = a \\
y = \varepsilon, & x = \varepsilon,\ y = \varepsilon,\ z = a \\
z = \varepsilon, & x = \varepsilon,\ y = \varepsilon,\ z = \varepsilon.
\end{array}
$$

Here the second column gives a solution that is not a solution of the equation on the next row, but is a solution of all the preceding equations. The first two equations are those from Lemmas 2.4.3 and 2.4.8. This chain uses the empty word, and thus does not work in the free semigroup as such. However, a slightly more complicated example shows that also $\mathrm{DC}_+(3) \geq 7$:

$$
\begin{array}{rl}
xxyz = zxyx, & x = a,\ y = b,\ z = aabaaba \\
xxyxzyz = zzyxxyx, & x = a,\ y = b,\ z = aaba \\
xz = zx, & x = a,\ y = b,\ z = a \\
xy = yx, & x = a,\ y = aa,\ z = a \\
x = y, & x = a,\ y = a,\ z = aa \\
x = z, & x = a,\ y = a,\ z = a \\
xx = x, & \text{no solutions.}
\end{array}
$$

For some results about the structure of equations in independent systems on three unknowns see [14] and [15].

If we add a fourth unknown t, then we can trivially extend any independent system by adding the equation $t = x$. This gives $\mathrm{IS}_+(4) \geq 4$ and $\mathrm{IS}'_+(4) \geq 3$. For chains the improvements are nontrivial. The following chain of equations shows that $\mathrm{DC}(4) \geq 12$:

$$
\begin{array}{rl}
xyz = zxy, & x = a,\ y = b,\ z = abab,\ t = a \\
xyt = txy, & x = a,\ y = b,\ z = abab,\ t = abab \\
xyxzyz = zxzyxy, & x = a,\ y = b,\ z = ab,\ t = abab \\
xyxtyt = txtyxy, & x = a,\ y = b,\ z = ab,\ t = ab \\
xyxztyzt = ztxztyxy, & x = a,\ y = b,\ z = ab,\ t = \varepsilon \\
xz = zx, & x = a,\ y = b,\ z = \varepsilon,\ t = ab \\
xt = tx, & x = a,\ y = b,\ z = \varepsilon,\ t = \varepsilon \\
xy = yx, & x = a,\ y = a,\ z = a,\ t = a \\
x = \varepsilon, & x = \varepsilon,\ y = a,\ z = a,\ t = a \\
y = \varepsilon, & x = \varepsilon,\ y = \varepsilon,\ z = a,\ t = a \\
z = \varepsilon, & x = \varepsilon,\ y = \varepsilon,\ z = \varepsilon,\ t = a \\
t = \varepsilon, & x = \varepsilon,\ y = \varepsilon,\ z = \varepsilon,\ t = \varepsilon.
\end{array}
$$

The next theorem sums up the new bounds given in this section.

Theorem 3.3.1. $\mathrm{DC}_+(3) \geq 7$ *and* $\mathrm{DC}(4) \geq 12$.

3.4 Lower Bounds

In [33] it is proved that $\mathrm{IS}(n) = \Omega(n^4)$ and $\mathrm{IS}_+(n) = \Omega(n^3)$. The former is proved by a construction that uses $n = 10m$ variables and gives a system of m^4 equations. Thus $\mathrm{IS}(n)$ is asymptotically at least $n^4/10000$. We present here a slightly modified version of this construction. By "reusing" some of the unknowns we get a bound that is asymptotically $n^4/1536$.

Theorem 3.4.1. *If* $n = 4m$, *then* $\mathrm{IS}'(n) \geq m^2(m-1)(m-2)/6$.

Proof. We use unknowns x_i, y_i, z_i, t_i, where $1 \leq i \leq m$. The equations in the system are

$$E(i,j,k,l) : x_i x_j x_k y_i y_j y_k z_i z_j z_k t_l = t_l x_i x_j x_k y_i y_j y_k z_i z_j z_k,$$

where $i, j, k, l \in \{1, \ldots, m\}$ and $i < j < k$. If $i, j, k, l \in \{1, \ldots, m\}$ and $i < j < k$, then

$$x_r = \begin{cases} ab, & \text{if } r \in \{i,j,k\} \\ \varepsilon, & \text{otherwise} \end{cases} \qquad y_r = \begin{cases} a, & \text{if } r \in \{i,j,k\} \\ \varepsilon, & \text{otherwise} \end{cases}$$

$$z_r = \begin{cases} ba, & \text{if } r \in \{i,j,k\} \\ \varepsilon, & \text{otherwise} \end{cases} \qquad t_r = \begin{cases} ababa, & \text{if } r = l \\ \varepsilon, & \text{otherwise} \end{cases}$$

is not a solution of $E(i,j,k,l)$, but is a solution of all the other equations. Thus the system is independent. □

The idea behind this construction (both the original and the modified) is that $(ababa)^k = (ab)^k a^k (ba)^k$ holds for $k < 3$, but not for $k = 3$. It was noted in [51] that if we could find words u_i such that $(u_1 \ldots u_m)^k = u_1^k \ldots u_m^k$ holds for $k < K$, but not for $k = K$, then we could prove that $\mathrm{IS}(n) = \Omega(n^{K+1})$. However, it has been proved that such words do not exist for $K \geq 5$ (see [28]), and conjectured that such words do not exist for $K = 4$.

3.5 Related Questions

We can use the notion of rank to present variations of the question of the maximal size of independent systems.

If a system has only periodic solutions, then the system can be said to force a maximal defect effect, so $\mathrm{IS}'(n)$ is the maximal size of an independent

system not doing that. But how large can an independent system be if it forces only the minimal defect effect, that is, the system has a solution in which the variables cannot be expressed as products of $n-2$ words? In [33] it is proved that there are such systems of size $\Omega(n^3)$ in free monoids and of size $\Omega(n^2)$ in free semigroups. Again, no upper bounds are known.

Another variation would be to consider chains that satisfy

$$\mathrm{Sol}_{n-1}(\varnothing) \supsetneq \mathrm{Sol}_{n-1}(E_1) \supsetneq \mathrm{Sol}_{n-1}(E_1, E_2) \supsetneq \cdots \supsetneq \mathrm{Sol}_{n-1}(E_1, \ldots, E_m),$$

where Sol_{n-1} is the set of solutions of rank $n-1$. These kinds of chains will be studied in Section 4.6.

Instead of trying to find a fixed bound for $\mathrm{IS}(n)$ or for some of the other related numbers, we could also try to find a bound that depends on the lengths of the equations. If $E_1, \ldots, E_m$ is an independent system or chain, then trivially m is exponential with respect to the maximum of the lengths $|E_i|$, because there are only exponentially many equations of certain size. For some questions, we can give bounds that depend only quadratically on the length of the shortest equation of a system, or on the length of the first equation of a chain. This is done in Section 4.6.

Chapter 4

Word Equations, Polynomials and Linear Algebra

In this chapter we introduce a new approach for proving results about word equations with the help of polynomials and linear algebra. This allows us to give new shorter proofs and generalizations for some old theorems.

First, in Section 4.1 we define a way to transform words into polynomials and prove some basic results using these polynomials.

In Section 4.2 we prove that if the lengths of the unknowns are fixed, then there is a connection between the ranks of solutions of a system of equations and the rank of a certain polynomial matrix. This theorem is very important for all the later results.

In Section 4.3 we analyze the results of Section 4.2 when the lengths of the unknowns are not fixed. For every solution these lengths form an n-dimensional vector, called the *length type* of the solution. We prove that the set of length types of all solutions of rank $n-1$ of a pair of equations is covered by a finite union of $(n-1)$-dimensional subspaces if the equations are not equivalent on solutions of rank $n-1$. This means that the solution sets of pairs of equations are in some sense more structured than the solution sets of single equations. This theorem is the key to proving the remaining results.

Section 4.4 contains small generalizations of two earlier results. These are nice examples of the methods developed in Section 4.2 and have some independent interest.

In Section 4.5 we prove a theorem about unbalanced equations. This gives a considerably simpler reproof and a generalization of a result in [24].

Finally, in Section 4.6 we return to the question about sizes of independent systems. There is a trivial bound for the size of a system depending

on the length of the longest equation, because there are only exponentially many equations of a fixed length. We prove that if the system is independent even when considering only solutions of rank $n-1$, then there is an upper bound for the size of the system depending quadratically on the length of the shortest equation. Even though it does not give a fixed bound even in the case of three unknowns, it is a first result of its type – hence opening, we hope, a new avenue for future research.

This chapter is based on the article [56].

4.1 Words and Polynomials

In this section we give proofs for some well-known results that have been mentioned in Chapter 2. These serve as examples of the polynomial methods used. Even though the standard proofs of these are simple, we hope that the proofs given here illustrate how properties of words can be formulated and proved in terms of polynomials.

Let $\Sigma \subset \mathbb{N}_1$ be an alphabet of numbers. For a word $w = a_0 \dots a_{n-1} \in \Sigma^n$ we define a polynomial

$$P_w = a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1}$$

and, if $n = |w| > 0$, a rational function

$$R_w = \frac{P_w}{X^n - 1}.$$

Now $w \mapsto P_w$ is an injective mapping from words to polynomials. Here we need the assumption $0 \notin \Sigma$; if injectivity of P_w would not be needed, then also 0 could be a letter. If $w_1, \dots, w_m \in \Sigma^*$, then

$$P_{w_1 \dots w_m} = P_{w_1} + P_{w_2} X^{|w_1|} + \cdots + P_{w_m} X^{|w_1 \dots w_{m-1}|}, \tag{4.1}$$

and if $w_1, \dots, w_m \in \Sigma^+$, then

$$\begin{aligned} P_{w_1 \dots w_m} =& R_{w_1}(X^{|w_1|} - 1) + R_{w_2}(X^{|w_1 w_2|} - X^{|w_1|}) \\ &+ \cdots + R_{w_m}(X^{|w_1 \dots w_m|} - X^{|w_1 \dots w_{m-1}|}). \end{aligned}$$

If $w \in \Sigma^+$ and $k \in \mathbb{N}_0$, then

$$P_{w^k} = P_w \frac{X^{k|w|} - 1}{X^{|w|} - 1} = R_w(X^{k|w|} - 1).$$

The polynomial P_w can be viewed as a characteristic polynomial of the word w. We could also replace X with a suitable number b and get a number whose reverse b-ary representation is w. Or we could let the coefficients of P_w be from some other commutative ring than $\mathbb{Z}$. Similar ideas have been used to analyze words in many places, see e.g. [39] and [58].

Example 4.1.1. If $w = 1212$, then $P_w = 1 + 2X + X^2 + 2X^3$ and

$$R_w = \frac{1 + 2X + X^2 + 2X^3}{X^4 - 1} = \frac{(1 + X^2)(1 + 2X)}{(X^2 + 1)(X^2 - 1)} = \frac{1 + 2X^2}{X^2 - 1}.$$

Recall that a word w is called primitive if it is not of the form u^k for any $k > 1$. If $w = u^k$ and u is primitive, then u is a *primitive root* of w.

Lemma 4.1.2. *If w is primitive, then P_w is not divisible by any polynomial of the form $(X^{|w|} - 1)/(X^n - 1)$, where $n < |w|$ is a divisor of $|w|$.*

Proof. If P_w is divisible by $(X^{|w|} - 1)/(X^n - 1)$, then there are numbers $a_0, \ldots, a_{n-1}$ such that

$$\begin{aligned} P_w &= (a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1}) \frac{X^{|w|} - 1}{X^n - 1} \\ &= (a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1})(1 + X^n + \cdots + X^{|w|-n}), \end{aligned}$$

so $w = (a_0 \ldots a_{n-1})^{|w|/n}$. □

The next two theorems are among the most basic and well-known results in combinatorics on words (except for item (4) of Theorem 4.1.4, which, however, appeared in [30] in a slightly different form).

Theorem 4.1.3. *Every nonempty word has a unique primitive root.*

Proof. Let $u^m = v^n$, where u and v are primitive. We need to show that $u = v$. We have

$$P_u \frac{X^{m|u|} - 1}{X^{|u|} - 1} = P_{u^m} = P_{v^n} = P_v \frac{X^{n|v|} - 1}{X^{|v|} - 1}.$$

Because $m|u| = n|v|$, we get $P_u(X^{|v|} - 1) = P_v(X^{|u|} - 1)$. If $d = \gcd(|u|, |v|)$, then $\gcd(X^{|u|} - 1, X^{|v|} - 1) = X^d - 1$. Thus P_u must be divisible by $(X^{|u|} - 1)/(X^d - 1)$ and P_v must be divisible by $(X^{|v|} - 1)/(X^d - 1)$. By Lemma 4.1.2, both u and v can be primitive only if $|u| = d = |v|$. □

The primitive root of a word $w \in \Sigma^+$ is denoted by $\rho(w)$.

Theorem 4.1.4. *For $u, v \in \Sigma^+$, the following are equivalent:*

1. *$\rho(u) = \rho(v)$,*
2. *if $U, V \in \{u, v\}^*$ and $|U| = |V|$, then $U = V$,*
3. *u and v satisfy a nontrivial relation,*
4. *$R_u = R_v$.*

Proof. (1) $\Rightarrow$ (2): $U = \rho(u)^{|U|/|\rho(u)|} = \rho(u)^{|V|/|\rho(u)|} = V$.

(2) $\Rightarrow$ (3): Clear.

(3) $\Rightarrow$ (4): Let $u_1 \dots u_m = v_1 \dots v_n$, where $u_i, v_j \in \{u, v\}$. Now

$$\begin{aligned} 0 =& P_{u_1 \dots u_m} - P_{v_1 \dots v_n} \\ =& R_{u_1}(X^{|u_1|} - 1) + \cdots + R_{u_m}(X^{|u_1 \dots u_m|} - X^{|u_1 \dots u_{m-1}|}) \\ & - R_{v_1}(X^{|v_1|} - 1) - \cdots - R_{v_n}(X^{|v_1 \dots v_n|} - X^{|v_1 \dots v_{n-1}|}) \\ =& R_u p - R_v p \end{aligned}$$

for some polynomial p. If $m \neq n$ or $u_i \neq v_i$ for some i, then $p \neq 0$, and thus $R_u = R_v$.

(4) $\Rightarrow$ (1): We have $P_{u^{|v|}} = R_u(X^{|u||v|} - 1) = R_v(X^{|u||v|} - 1) = P_{v^{|u|}}$, so $u^{|v|} = v^{|u|}$ and $\rho(u) = \rho(u^{|v|}) = \rho(v^{|u|}) = \rho(v)$. □

Similarly, polynomials can be used to give a simple proof for Theorem 2.3.4. In fact, one of the original proofs in [18] uses power series. The proof we give here is essentially this original proof formulated in terms of our polynomials. Algebraic techniques have also been used to prove variations of this theorem [49].

Theorem 4.1.5 (Fine and Wilf)**.** *If u^i and v^j have a common prefix of length $|u| + |v| - \gcd(|u|, |v|)$, then $\rho(u) = \rho(v)$.*

Proof. Let $\gcd(|u|, |v|) = d$, $\operatorname{lcm}(|u|, |v|) = m$, $m/|u| = r$ and $m/|v| = s$. If $\rho(u) \neq \rho(v)$, then $u^r \neq v^s$, so u^r and v^s have a maximal common prefix of length $k < m$. Now

$$\begin{aligned} P_{u^r} - P_{v^s} &= \frac{X^{r|u|} - 1}{X^{|u|} - 1} P_u - \frac{X^{s|v|} - 1}{X^{|v|} - 1} P_v \\ &= \frac{(X^m - 1)(X^d - 1)}{(X^{|u|} - 1)(X^{|v|} - 1)} \left(\frac{X^{|v|} - 1}{X^d - 1} P_u - \frac{X^{|u|} - 1}{X^d - 1} P_v \right) \end{aligned}$$

is divisible by X^k, but not by X^{k+1}, so also the polynomial

$$\frac{X^{|v|} - 1}{X^d - 1} P_u - \frac{X^{|u|} - 1}{X^d - 1} P_v$$

is divisible by X^k, but not by X^{k+1}. Thus k can be at most the degree of this polynomial, which is at most $|u| + |v| - d - 1$. □

4.2 Solutions of Fixed Length

In this section we apply polynomial techniques to word equations. From now on, we will assume that there are n unknowns, they are ordered as $x_1, \dots, x_n$ and Ξ is the set of these unknowns.

Recall that the rank of a morphism h is the smallest number r for which there is a set A of r words such that $h(x) \in A^*$ for every unknown x.

Let $h : \Xi^* \to \Sigma^*$ be a morphism. The *length type* of h is the vector

$$L = (|h(x_1)|, \ldots, |h(x_n)|) \in \mathbb{N}_0^n.$$

This length type L determines a morphism

$$\mathrm{len}_L : \Xi^* \to \mathbb{N}_0, \mathrm{len}_L(w) = |h(w)|.$$

It is important that len_L depends only on L and not on h.

If E is a word equation, the set of its solutions is denoted by $\mathrm{Sol}(E)$, the set of solutions of rank r by $\mathrm{Sol}_r(E)$, the set of solutions of length type L by $\mathrm{Sol}^L(E)$ and the set of solutions of rank r and length type L by $\mathrm{Sol}_r^L(E)$. These can be naturally generalized for systems of equations.

For a word equation $E : y_1 \ldots y_k = z_1 \ldots z_l$ (where $y_i, z_i \in \Xi$), a variable $x \in \Xi$ and a length type L, let

$$Q_{E,x,L} = \sum_{y_i = x} X^{\mathrm{len}_L(y_1 \ldots y_{i-1})} - \sum_{z_i = x} X^{\mathrm{len}_L(z_1 \ldots z_{i-1})}.$$

Informally, this polynomial encodes the positions of x in the equation E.

Theorem 4.2.1. *A morphism $h : \Xi^* \to \Sigma^*$ of length type L is a solution of an equation $E : u = v$ if and only if*

$$\sum_{x \in \Xi} Q_{E,x,L} P_{h(x)} = 0.$$

Proof. Now $h(u) = h(v)$ if and only if $P_{h(u)} = P_{h(v)}$, and we can write the polynomial $P_{h(u)} - P_{h(v)}$ as $\sum_{x \in \Xi} Q_{E,x,L} P_{h(x)}$ by (4.1). □

Theorem 4.2.1 means that if we fix a length type L, then we can turn a word equation into a linear equation where the polynomials $Q_{E,x,L}$ are the coefficients. A solution for this linear equation is an n-dimensional vector over the field of rational functions, and $h \in \mathrm{Sol}^L(E)$ corresponds to a solution $(P_{h(x_1)}, \ldots P_{h(x_n)})$ of the linear equation.

Example 4.2.2. Let $\Xi = \{x, y, z\}$, $E : xyz = zxy$ and $L = (1, 1, 2)$. Now

$$Q_{E,x,L} = 1 - X^2, \qquad Q_{E,y,L} = X - X^3, \qquad Q_{E,z,L} = X^2 - 1.$$

If h is the morphism defined by $h(x) = 1$, $h(y) = 2$ and $h(z) = 12$, then h is a solution of E and

$$\begin{aligned} &Q_{E,x,L} P_{h(x)} + Q_{E,y,L} P_{h(y)} + Q_{E,z,L} P_{h(z)} \\ =&(1 - X^2) \cdot 1 + (X - X^3) \cdot 2 + (X^2 - 1)(1 + 2X) = 0. \end{aligned}$$

At this point we start using linear algebra. We will do this over two fields: the field of rational numbers (for the first time in Lemma 4.2.5) and the field of rational functions (for the first time in Lemma 4.2.6). We start with an example.

Example 4.2.3. Consider the morphism $h : \{x_1, x_2, x_3\}^* \to \{1, 2\}^*$ of rank 2 defined by $h(x_1) = 1, h(x_2) = 2, h(x_3) = 12$. If h is a solution of an equation E, then so is $g \circ h$ for every morphism $g : \{1, 2\}^* \to \{1, 2\}^*$. The length type of $g \circ h$ is

$$(|g(1)|, |g(2)|, |g(12)|) = |g(1)| \cdot (1, 0, 1) + |g(2)| \cdot (0, 1, 1).$$

Because the vectors $(1, 0, 1)$ and $(0, 1, 1)$ are linearly independent, these length types essentially form a two-dimensional space (of course $|g(1)|$ and $|g(2)|$ are nonnegative integers, so the length types don't form the whole space). This observation is formalized and generalized in Lemma 4.2.5.

A morphism $\phi : \Xi^* \to \Xi^*$ is an *elementary transformation* if there are two unknowns $x, y \in \Xi$ so that $\phi(y) \in \{xy, x\}$ and $\phi(z) = z$ for $z \in \Xi \setminus \{y\}$. If $\phi(y) = xy$, then ϕ is *regular*, and if $\phi(y) = x$, then ϕ is *singular*. The next lemma follows immediately from results in [43].

Lemma 4.2.4. *Every solution h of an equation E has a factorization $h = \theta \circ \phi \circ \alpha$, where $\alpha(x) \in \{x, \varepsilon\}$ for all $x \in \Xi$, $\phi = \phi_m \circ \cdots \circ \phi_1$, every ϕ_i is an elementary transformation, $\phi \circ \alpha$ is a solution of E and $\theta(x) \neq \varepsilon$ for all $x \in \Xi$. If $\alpha(x) = \varepsilon$ for s unknowns x and t of the ϕ_i are singular, then the rank of $\phi \circ \alpha$ is $n - s - t$.*

Lemma 4.2.5. *Let E be an equation on n unknowns and let $h \in \mathrm{Sol}_r^L(E)$. There is an r-dimensional subspace V of $\mathbb{Q}^n$ containing L such that the set of those length types of morphisms in $\mathrm{Sol}_r(E)$ that are in V is not covered by any finite union of $(r-1)$-dimensional spaces.*

Proof. For arbitrary morphisms $F : \Xi^* \to \Xi^*$ and $G : \Xi^* \to \Sigma^*$, let $L_G = (|G(x_1)|, \ldots, |G(x_n)|)^T$ be the length type of G as a column vector and let $A_F = (|F(x_i)|_{x_j})$ be an $n \times n$ matrix. Now $L_{G \circ F} = A_F L_G$. More generally, if $F_1, \ldots, F_m$ are morphisms $\Xi^* \to \Xi^*$, then

$$L_{G \circ F_m \circ \cdots \circ F_1} = A_{F_1} \ldots A_{F_m} L_G.$$

Let $h = \theta \circ \phi_m \circ \cdots \circ \phi_1 \circ \alpha$ as in Lemma 4.2.4. Let $f = \phi_m \circ \cdots \circ \phi_1 \circ \alpha$. The rank of f is $n - s - t \geq r$, if s and t are as in Lemma 4.2.4. Now $g \circ f$ is a solution of E for every morphism $g : \Xi^* \to \Sigma^*$. The length type of $g \circ f$ is $L_{g \circ f} = L_{g \circ \phi_m \circ \cdots \circ \phi_1 \circ \alpha} = A L_g$, where $A = A_\alpha A_{\phi_1} \ldots A_{\phi_m}$. To prove the theorem, we need to show that the rank of A is at least r. This can be done by determining the ranks of the matrices A_α and A_{ϕ_k}.

The matrix A_α is a diagonal matrix and the ith element on the diagonal is 0 if $\alpha(x_i) = \varepsilon$ and 1 otherwise. Thus the rank of A_α is $n - s$.

If ϕ is the elementary transformation defined by $\phi(x_1) = x_2x_1$, then

$$A_\phi = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

is a matrix of rank n (this is an identity matrix except for the second element on the first row). In general, the rank of A_ϕ is n for every regular elementary transformation ϕ.

If ϕ is the elementary transformation defined by $\phi(x_1) = x_2$, then

$$A_\phi = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

is a matrix of rank $n - 1$ (this is an identity matrix except for the first two elements on the first row). In general, the rank of A_ϕ is $n - 1$ for every singular elementary transformation ϕ.

The rank of A_α is $n - s$, t of the matrices A_{ϕ_k} have rank $n - 1$ and the rest have rank n. Thus the rank of A is at least $n - s - t$, which is at least r. □

Lemma 4.2.6. *Let E be an equation on n unknowns and let $h \in \mathrm{Sol}_r^L(E)$. There are morphisms $f : \Xi^* \to \Xi^*$ and $\theta : \Xi^* \to \Sigma^*$ and polynomials p_{ij} such that the following conditions hold:*

1. *$h = \theta \circ f$,*
2. *f is a solution of E,*
3. *$\theta(x) \neq \varepsilon$ for every $x \in \Xi$,*
4. *$P_{(g\circ f)(x_i)} = \sum p_{ij} P_{g(x_j)}$ for all i, j if $g : \Xi^* \to \Sigma^*$ is a morphism of the same length type as θ,*
5. *r of the vectors $(p_{1j}, \dots, p_{nj}) \in \mathbb{Q}(X)^n$, where $j = 1, \dots, n$, are linearly independent.*

Proof. The proof is quite similar to the proof of Lemma 4.2.5.

For arbitrary morphisms $F : \Xi^* \to \Xi^*$ and $G : \Xi^* \to \Sigma^*$ and length type L, define an n-dimensional column vector $P_G = (P_{G(x_1)}, \ldots, P_{G(x_n)})^T$ and an $n \times n$ polynomial matrix $B_{F,L} = (b_{ij})$, where

$$b_{ij} = \sum_{ux_j \leq F(x_i)} X^{\mathrm{len}_L(u)}.$$

If L is the length type of G, then $P_{G \circ F} = B_{F,L} P_G$. More generally, if $F_1, \ldots, F_m$ are morphisms $\Xi^* \to \Xi^*$ and L_k is the length type of $G \circ F_m \circ \cdots \circ F_{k+1}$, then

$$P_{G \circ F_m \circ \cdots \circ F_1} = B_{F_1,L_1} \ldots B_{F_m,L_m} P_G.$$

The matrices $B_{F,L}$ will be used to define the polynomials p_{ij}.

Let $h = \theta \circ \phi_m \circ \cdots \circ \phi_1 \circ \alpha$ as in Lemma 4.2.4. Let $f = \phi_m \circ \cdots \circ \phi_1 \circ \alpha$. The first three conditions are satisfied by θ and f. The rank of f is $n - s - t \geq r$, if s and t are as in Lemma 4.2.4.

Let L be the length type of θ and let g be a morphism of length type L. Now $P_{g \circ f} = P_{g \circ \phi_m \circ \cdots \circ \phi_1 \circ \alpha} = B P_g$, where $B = B_{\alpha,L_0} B_{\phi_1,L_1} \ldots B_{\phi_m,L_m}$ and L_k is the length type of $g \circ \phi_m \circ \cdots \circ \phi_{k+1}$. Let $B = (p_{ij})$. Now the fourth condition holds, because $P_{g \circ f} = B P_g$.

To prove that the last condition holds, it must be proved that the rank of the matrix B is at least r. This can be done by determining the ranks of the matrices $B_{\alpha,L}$ and $B_{\phi_k,L}$.

The matrix $B_{\alpha,L}$ is a diagonal matrix and the ith element on the diagonal is 0 if $\alpha(x_i) = \varepsilon$ and 1 otherwise. Thus the rank of $B_{\alpha,L}$ is $n - s$.

If ϕ is the elementary transformation defined by $\phi(x_1) = x_2 x_1$, then

$$B_{\phi,L} = \begin{pmatrix} X^{\mathrm{len}_L(x_2)} & 1 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & & & & \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

is a matrix of rank n (this is an identity matrix except for the first two elements on the first row). In general, the rank of $B_{\phi,L}$ is n for every regular elementary transformation ϕ.

If ϕ is the elementary transformation defined by $\phi(x_1) = x_2$, then

$$B_{\phi,L} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & & & & \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

is a matrix of rank $n-1$ (again, this is an identity matrix except for the first two elements on the first row). In general, the rank of $B_{\phi,L}$ is $n-1$ for every singular elementary transformation ϕ.

The rank of B_{α,L_0} is $n-s$, t of the matrices B_{ϕ_k,L_k} have rank $n-1$ and the rest have rank n. Thus the rank of B is at least $n-s-t$, which is at least r. □

With the help of these lemmas, we are going to analyze solutions of some fixed length type. Principal (or fundamental) solutions, which were implicitly present in the previous lemmas (see [43]), have been used in connection with fixed lengths also in [27] and [28].

Theorem 4.2.7. *Let $E_1, \ldots, E_m$ be a system of equations on n unknowns and let $L \in \mathbb{N}_0^n$. Let $q_{ij} = Q_{E_i,x_j,L}$. If $\mathrm{Sol}_r^L(E_1, \ldots, E_m) \neq \varnothing$, then the rank of the $m \times n$ polynomial matrix (q_{ij}) is at most $n-r$. If the rank of the matrix is 1, at most one component of L is zero and the equations are nontrivial, then $\mathrm{Sol}^L(E_1) = \cdots = \mathrm{Sol}^L(E_m)$.*

Proof. Let $h \in \mathrm{Sol}_r^L(E_1, \ldots, E_m)$. If $r = 1$, the first claim follows from Theorem 4.2.1, so assume that $r > 1$. Let E be an equation that has the same nonperiodic solutions as the system. We will use Lemma 4.2.6 for this equation. Fix k and let $g_1 : \Xi^* \to \Sigma^*$ be the morphism determined by $g_1(x_i) = 1^{|\theta(x_i)|}$ for all i and let $g_2 : \Xi^* \to \Sigma^*$ be the morphism determined by $g_2(x_k) = 21^{|\theta(x_k)|-1}$ and $g_2(x_i) = 1^{|\theta(x_i)|}$ for all $i \neq k$. Now $g_1 \circ f$ and $g_2 \circ f$ are solutions of every E_l, so

$$\sum_{i=1}^{n} Q_{E_l,x_i,L} P_{(g_1 \circ f)(x_i)} = 0 \qquad \text{and} \qquad \sum_{i=1}^{n} Q_{E_l,x_i,L} P_{(g_2 \circ f)(x_i)} = 0$$

for all l by Theorem 4.2.1. Because also $P_{(g_1 \circ f)(x_i)} = \sum_{j=1}^{n} p_{ij} P_{g_1(x_j)}$ and $P_{(g_2 \circ f)(x_i)} = \sum_{j=1}^{n} p_{ij} P_{g_2(x_j)}$, we get

$$\begin{aligned} 0 &= \sum_{i=1}^{n} Q_{E_l,x_i,L} (P_{(g_2 \circ f)(x_i)} - P_{(g_1 \circ f)(x_i)}) \\ &= \sum_{i=1}^{n} Q_{E_l,x_i,L} \sum_{j=1}^{n} p_{ij} (P_{g_2(x_j)} - P_{g_1(x_j)}) = \sum_{i=1}^{n} Q_{E_l,x_i,L} p_{ik} \end{aligned}$$

for all l. Thus the vectors $(p_{1j}, \ldots, p_{nj})$ are solutions of the linear system of equations determined by the matrix (q_{ij}). Because at least r of these vectors are linearly independent, the rank of the matrix is at most $n-r$.

If at most one component of L is zero and the equations are nontrivial, then all rows of the matrix are nonzero. If also the rank of the matrix is 1, then all rows are multiples of each other and the second claim follows by Theorem 4.2.1. □

4.3 Sets of Solutions

Now we analyze how the polynomials $Q_{E,x,L}$ behave when L is not fixed. Let

$$\mathcal{M} = \{a_1X_1 + \cdots + a_nX_n \mid a_1, \ldots, a_n \in \mathbb{N}_0\} \subset \mathbb{Z}[X_1, \ldots, X_n]$$

be the additive monoid of linear homogeneous polynomials with nonnegative integer coefficients on the variables $X_1, \ldots, X_n$. The *monoid ring* of $\mathcal{M}$ over $\mathbb{Z}$ is the ring formed by expressions of the form

$$a_1X^{p_1} + \cdots + a_kX^{p_k},$$

where $a_i \in \mathbb{Z}$ and $p_i \in \mathcal{M}$, and the addition and multiplication of these generalized polynomials is defined in a natural way. This ring is denoted by $\mathbb{Z}[X; \mathcal{M}]$. If $L \in \mathbb{Z}^n$, then the value of a polynomial $p \in \mathcal{M}$ at the point $(X_1, \ldots, X_n) = L$ is denoted by $p(L)$, and the polynomial we get by making this substitution in $s \in \mathbb{Z}[X; \mathcal{M}]$ is denoted by $s(L)$.

The ring $\mathbb{Z}[X; \mathcal{M}]$ is isomorphic to the ring $\mathbb{Z}[Y_1, \ldots, Y_n]$ of polynomials on n variables. The isomorphism is given by $X^{X_i} \mapsto Y_i$. However, the generalized polynomials where the exponents are in $\mathcal{M}$ are suitable for our purposes.

If $a_i \leq b_i$ for $i = 1, \ldots, n$, then we use the notation

$$a_1X_1 + \cdots + a_nX_n \preceq b_1X_1 + \cdots + b_nX_n.$$

If $p, q \in \mathcal{M}$ and $p \preceq q$, then $p(L) \leq q(L)$ for all $L \in \mathbb{N}_0^n$.

For an equation $E : x_{i_1} \ldots x_{i_r} = x_{j_1} \ldots x_{j_s}$ we define

$$S_{E,x} = \sum_{x_{i_k}=x} X^{X_{i_1}+\cdots+X_{i_{k-1}}} - \sum_{x_{j_k}=x} X^{X_{j_1}+\cdots+X_{j_{k-1}}} \in \mathbb{Z}[X; \mathcal{M}].$$

Now $S_{E,x}(L) = Q_{E,x,L}$. Theorem 4.2.1 can be formulated in terms of these generalized polynomials.

Theorem 4.3.1. *A morphism $h : \Xi^* \to \Sigma^*$ of length type L is a solution of an equation E if and only if*

$$\sum_{x \in \Xi} S_{E,x}(L) P_{h(x)} = 0.$$

Example 4.3.2. Let $E : x_1x_2x_3 = x_3x_1x_2$. Now

$$S_{E,x_1} = 1 - X^{X_3}, \qquad S_{E,x_2} = X^{X_1} - X^{X_1+X_3}, \qquad S_{E,x_3} = X^{X_1+X_2} - 1.$$

The *length* of an equation $E : u = v$ is $|E| = |uv|$. The number of occurrences of an unknown x in E is $|E|_x = |uv|_x$.

Theorem 4.3.3. *Let E_1, E_2 be a pair of nontrivial equations on n unknowns. Let $\mathrm{Sol}_{n-1}(E_1) \neq \mathrm{Sol}_{n-1}(E_2)$. For some unknowns x_k, x_l, the set of length types of solutions of the pair of rank $n-1$ is covered by a union of $(|E_1|_{x_k} + |E_1|_{x_l})^2$ $(n-1)$-dimensional subspaces of $\mathbb{Q}^n$. If $V_1, \ldots, V_m$ is a minimal such cover and $L \in V_i$ for some i, then $\mathrm{Sol}^L_{n-1}(E_1) = \mathrm{Sol}^L_{n-1}(E_2)$.*

Proof. Let $s_{ij} = S_{E_i,x_j}$ for $i = 1, 2$ and $j = 1, \ldots, n$. If all 2×2 minors of the $2 \times n$ matrix (s_{ij}) are zero, then for all length types L of solutions of rank $n-1$ the rank of the matrix (q_{ij}) in Theorem 4.2.7 is 1 and E_1 and E_2 are equivalent, which is a contradiction. Thus there are k, l such that

$$t_{kl} = s_{1k}s_{2l} - s_{1l}s_{2k} \neq 0.$$

The generalized polynomial t_{kl} can be written as

$$t_{kl} = \sum_{i=1}^{M} X^{p_i} - \sum_{i=1}^{N} X^{q_i},$$

where $p_i, q_i \in \mathcal{M}$ and $p_i \neq q_j$ for all i, j. If L is a length type of a solution of rank $n-1$, then $M = N$ and L must be a solution of the system of equations

$$p_i = q_{\sigma(i)} \qquad (i = 1, \ldots, M) \tag{4.2}$$

for some permutation σ. For every σ the equations determine an at most $(n-1)$-dimensional space.

Let the equations be $E_1 : u_1 = v_1$ and $E_2 : u_2 = v_2$. Let

$$\begin{array}{llll} |u_1|_{x_k} = A, & |v_1|_{x_k} = A', & |u_2|_{x_k} = B, & |v_2|_{x_k} = B', \\ |u_1|_{x_l} = C, & |v_1|_{x_l} = C', & |u_2|_{x_l} = D, & |v_2|_{x_l} = D'. \end{array}$$

Now $s_{1k}, s_{2l}, s_{1l}, s_{2k}$ can be written as

$$\begin{aligned} s_{1k} &= \sum_{i=1}^{A} X^{a_i} - \sum_{i=1}^{A'} X^{a'_i}, \quad s_{2l} = \sum_{i=1}^{B} X^{b_i} - \sum_{i=1}^{B'} X^{b'_i}, \\ s_{1l} &= \sum_{i=1}^{C} X^{c_i} - \sum_{i=1}^{C'} X^{c'_i}, \quad s_{2k} = \sum_{i=1}^{D} X^{d_i} - \sum_{i=1}^{D'} X^{d'_i}, \end{aligned}$$

where $a_i \preceq a_{i+1}$, $a'_i \preceq a'_{i+1}$, and so on. The polynomials p_i form a subset of the polynomials $a_i + b_j$, $a'_i + b'_j$, $c_i + d'_j$ and $c'_i + d_j$ (the reason that they form just a subset is that we assumed $p_i \neq q_j$ for all i, j). For any i, let j_i be the smallest index j such that $a_i + b_j = p_m$ for some m. Now for every i, j, m such that $a_i + b_j = p_m$ we have $a_i + b_{j_i} \preceq p_m$. We can do a similar thing for the polynomials a'_i, b'_i and c_i, d'_i and c'_i, d_i. In this way we obtain at most $A + A' + C + C'$ polynomials p_i such that for any L the value of one of

these polynomials is minimal among the values $p_i(L)$. Similarly we obtain at most $A + A' + C + C'$ "minimal" polynomials q_i. If L satisfies one of the systems (4.2), then the smallest of the values $p_i(L)$ must be the same as the smallest of the values $q_i(L)$. Thus L must satisfy some equation $p_i = q_j$, where p_i and q_j are some of the "minimal" polynomials. There are at most

$$(A + A' + C + C')^2 = (|E_1|_{x_k} + |E_1|_{x_l})^2$$

possible pairs of such polynomials, and each of them determines an $(n-1)$-dimensional space.

Consider the second claim. Because the cover is minimal, there is a solution of rank $n-1$ whose length type is in V_i, but not in any other V_j. By Lemma 4.2.5, the set of length types of solutions of rank $n-1$ in this space cannot be covered by a finite union of $(n-2)$-dimensional spaces. Thus one of the systems (4.2) must determine the space V_i. The same holds for systems coming from all other nonzero 2×2 minors of the matrix (s_{ij}), so E_1 and E_2 have the same solutions of rank $n-1$ and length type L for all $L \in V_i$ by Theorem 4.2.7. □

The following example illustrates the proof of Theorem 4.3.3. It gives a pair of equations on three unknowns where the required number of subspaces is two. The equations are those from Lemmas 2.4.3 and 2.4.8. We do not know any example where more spaces would be necessary.

Example 4.3.4. Consider the equations

$$E_1 : x_1x_2x_3 = x_3x_1x_2 \qquad \text{and} \qquad E_2 : x_1x_2x_1x_3x_2x_3 = x_3x_1x_3x_2x_1x_2$$

and the generalized polynomial

$$\begin{aligned} s =& S_{E_1,x_1}S_{E_2,x_3} - S_{E_1,x_3}S_{E_2,x_1} \\ =& X^{2X_1+X_2} + X^{2X_1+2X_2+X_3} + X^{X_1+2X_3} + X^{X_1+X_2+X_3} \\ & - X^{2X_1+X_2+X_3} - X^{X_1+X_3} - X^{2X_1+2X_2} - X^{X_1+X_2+2X_3}. \end{aligned}$$

If L is a length type of a nontrivial solution of the pair E_1, E_2, then $s(L) = 0$. If $s(L) = 0$, then L must satisfy an equation $p = q$, where

$$p \in \{2X_1 + X_2, X_1 + 2X_3, X_1 + X_2 + X_3\} \quad \text{and} \quad q \in \{X_1 + X_3, 2X_1 + 2X_2\}.$$

The possible relations are

$$X_3 = 0, \qquad X_1 + X_2 = X_3, \qquad X_2 = 0, \qquad X_1 + 2X_2 = 2X_3.$$

If L satisfies one of the first three, then $s(L) = 0$. If L satisfies the last one, then $s(L) \neq 0$, except if $L = 0$. So if h is a nonperiodic solution, then

$$|h(x_3)| = 0 \qquad \text{or} \qquad |h(x_1x_2)| = |h(x_3)| \qquad \text{or} \qquad |h(x_2)| = 0.$$

There are no nonperiodic solutions with $h(x_2) = \varepsilon$, but every h with $h(x_3) = \varepsilon$ or $h(x_1x_2) = h(x_3)$ is a solution.

4.4 Minor Applications

Theorem 2.3.6 can be proved with the help of Theorem 4.2.7.

Theorem 4.4.1 (Graph lemma)**.** *Consider a system of equations whose graph has r connected components. If h is a solution of this system and $h(x_i) \neq \varepsilon$ for all i, then the rank of h is at most r.*

Proof. We can assume that the connected components are

$$\{x_1, \ldots, x_{i_2-1}\}, \{x_{i_2}, \ldots, x_{i_3-1}\}, \ldots, \{x_{i_r}, \ldots, x_n\}$$

and the equations are

$$x_j \cdots = x_{k_j} \cdots ,$$

where $j \in \{1, \ldots, n\} \setminus \{1, i_2, \ldots, i_r\}$ and $k_j < j$. Let q_{ij} be as in Theorem 4.2.7. If we remove the columns $1, i_2, \ldots, i_r$ from the $(n-r) \times n$ matrix (q_{ij}), we obtain a square matrix M where the diagonal elements are not divisible by X, but all elements above the diagonal are divisible by X. This means that $\det(M)$ is not divisible by X, so $\det(M) \neq 0$. Thus the rank of the matrix (q_{ij}) is $n - r$ and h has rank at most r by Theorem 4.2.7. □

The next theorem generalizes a result from [14] for more than three unknowns.

Theorem 4.4.2. *If a pair of nontrivial equations on n unknowns has a solution h of rank $n-1$ where no two of the unknowns commute, then there is a number $k \geq 1$ such that the equations are of the form $x_1 \cdots = x_2^k x_3 \cdots$.*

Proof. By Theorem 4.4.1, the equations must be of the form $x_1 \cdots = x_2 \cdots$. Let them be

$$x_1 u y \cdots = x_2 v z \cdots \qquad \text{and} \qquad x_1 u' y' \cdots = x_2 v' z' \cdots ,$$

where $u, v, u', v' \in \{x_1, x_2\}^*$ and $y, z, y', z' \in \{x_3, \ldots, x_n\}$. We can assume that $z = x_3$ and

$$|h(x_2 v)| \leq |h(x_1 u)|, |h(x_1 u')|, |h(x_2 v')|.$$

If it would be $|h(x_1 u)| = |h(x_2 v)|$, then $h(x_1)$ and $h(x_2)$ would commute, so $|h(x_1 u)| > |h(x_2 v)|$. If v would contain x_1, then $h(x_1)$ and $h(x_2)$ would commute by Theorem 4.1.5, so $v = x_2^{k-1}$ for some $k \geq 1$.

Let L be the length type of h and let q_{ij} be as in Theorem 4.2.7. By Theorem 4.2.7, the rank of the matrix (q_{ij}) must be 1 and thus $q_{12}q_{23} - q_{13}q_{22} = 0$. The term of $q_{13}q_{22}$ of the lowest degree is $X^{|h(x_2^k)|}$. The same must hold for $q_{12}q_{23}$, and thus the term of q_{23} of the lowest degree must be $-X^{|h(x_2^k)|}$. We know that $x_2 v = x_2^k$ and assumed that $|h(x_2 v)| \leq |h(x_2 v')|$.

If it would be $|h(x_2v)| < |h(x_2v')|$, then $h(x_3)$ would start in $h(x_2v'z'\dots)$ before the end $h(x_2v')$, which is not possible. This means that $|h(x_2v')| = |h(x_2^k)| \leq |h(x_1u')|$ and $z' = x_3$. As above, we conclude that $|h(x_2v')| < |h(x_1u')|$, v' cannot contain x_1 and $v' = x_2^{k-1}$. □

It was proved in [38] that if

$$s_0u_1^is_1\dots u_m^is_m = t_0v_1^it_1\dots v_n^it_n$$

holds for $m+n+3$ consecutive values of i, then it holds for all i. By using similar ideas as in Theorem 4.2.7, we improve this bound to $m+n$ and prove that the values do not need to be consecutive. In [38] it was also stated that the arithmetization and matrix techniques in [65] would give a simpler proof of a weaker result. Similar questions have been studied in [29] and there are relations to independent systems [51]; see also the comment in the end of Section 3.4.

Theorem 4.4.3. *Let $m, n \geq 1$, $s_j, t_j \in \Sigma^*$ and $u_j, v_j \in \Sigma^+$. Let*

$$U_i = s_0u_1^is_1\dots u_m^is_m \qquad \text{and} \qquad V_i = t_0v_1^it_1\dots v_n^it_n.$$

If $U_i = V_i$ holds for $m+n$ values of i, then it holds for all i.

Proof. The equation $U_i = V_i$ is equivalent to $P_{U_i} - P_{V_i} = 0$. Because

$$\begin{aligned} P_{U_i} = &\sum_{j=1}^{m} \left(P_{s_{j-1}} + P_{u_j}\frac{X^{i|u_j|}-1}{X^{|u_j|}-1}X^{|s_{j-1}|} \right) X^{i|u_1\dots u_{j-1}|+|s_0\dots s_{j-2}|} \\ &+ P_{s_m}X^{i|u_1\dots u_m|+|s_0\dots s_{m-1}|} \end{aligned}$$

and P_{V_i} is of a similar form, this equation can be written as

$$\sum_{j=0}^{m} y_jX^{i|u_1\dots u_j|} + \sum_{k\in K} z_kX^{i|v_1\dots v_k|} = 0, \tag{4.3}$$

where y_j, z_k are some polynomials that do not depend on i and K is the set of those $k \in \{0,\dots n\}$ for which $|v_1\dots v_k|$ is not any of the numbers $|u_1\dots u_j|$ $(j = 0,\dots,m)$. If $U_{i_1} = V_{i_1}$ and $U_{i_2} = V_{i_2}$, then

$$(i_1-i_2)|u_1\dots u_m| = |U_{i_1}| - |U_{i_2}| = |V_{i_1}| - |V_{i_2}| = (i_1-i_2)|v_1\dots v_n|.$$

Thus $|u_1\dots u_m| = |v_1\dots v_n|$ and the size of K is at most $n-1$. If (4.3) holds for $m+1+\#K \leq m+n$ values of i, it can be viewed as a system of equations where y_j, z_k are unknowns. The coefficients of this system form a generalized Vandermonde matrix whose determinant is nonzero, so the system has a unique solution $y_j = z_k = 0$ for all j, k. Now (4.3) holds for all i and $U_i = V_i$ for all i. □

4.5 Unbalanced Equations

Recall that an equation $u = v$ is *balanced* if $|u|_x = |v|_x$ for every unknown x. With the help of Theorem 4.3.3 we can generalize Theorem 2.3.8. Our proof is also significantly simpler than the original proof of Theorem 2.3.8 in [24].

Theorem 4.5.1. *Let E_1, E_2 be a pair of equations on n unknowns having a solution of rank $n-1$. If E_1 is not balanced, then $\mathrm{Sol}_{n-1}(E_1) \subseteq \mathrm{Sol}_{n-1}(E_2)$.*

Proof. If E_1 is the equation $u = v$ and h is a solution of E_1, then

$$\sum_{i=1}^{n} |u|_{x_i} |h(x_i)| = \sum_{i=1}^{n} |v|_{x_i} |h(x_i)|$$

and $|u|_{x_i} \neq |v|_{x_i}$ for at least one i. Thus the set of length types of solutions of E_1 is covered by a single $(n-1)$-dimensional space V. Because the pair E_1, E_2 has a solution of rank $n-1$, V is a minimal cover for the length types of the solutions of the pair of rank $n-1$. By Theorem 4.3.3, E_1 and E_2 have the same solutions of length type L and rank $n-1$ for all $L \in V$. □

Another way to think of this result is that if E_1 is not balanced but has a solution of rank $n-1$ that is not a solution of E_2, then the pair E_1, E_2 causes a larger than minimal defect effect.

If $h : \Xi^* \to \Sigma^*$ is a morphism, then the *entire system* generated by h is the set of all equations satisfied by h. It is denoted by K_h. As a consequence of Theorem 4.5.1, we get the following result about entire systems. The case of three unknowns was proved in [24].

Corollary 4.5.2. *If $g, h : \Xi^* \to \Sigma^*$ are morphisms of rank $n-1$ and $K_g \neq K_h$, then $K_g \cap K_h$ contains only balanced equations.*

Proof. We can assume that there is an equation $E_2 \in K_g \setminus K_h$. For any equation $E_1 \in K_g \cap K_h$, g is a solution of the pair E_1, E_2 and h is a solution of E_1 but not of E_2. By Theorem 4.5.1, E_1 must be balanced. □

4.6 Upper Bounds for the Lengths of Chains

We study the following variation of the question about maximal sizes of chains: how long can a sequence of nontrivial equations $E_1, \ldots, E_m$ be if

$$\mathrm{Sol}_{n-1}(E_1) \supsetneq \mathrm{Sol}_{n-1}(E_1, E_2) \supsetneq \cdots \supsetneq \mathrm{Sol}_{n-1}(E_1, \ldots, E_m)?$$

We prove an upper bound depending quadratically on the length of the first equation. For three unknowns we get a similar bound for the size of independent systems and chains.

Theorem 4.6.1. *Let $E_1, \ldots, E_m$ be nontrivial equations on n unknowns and let*

$$\mathrm{Sol}_{n-1}(E_1) \supsetneq \mathrm{Sol}_{n-1}(E_1, E_2) \supsetneq \cdots \supsetneq \mathrm{Sol}_{n-1}(E_1, \ldots, E_m) \neq \varnothing.$$

If the set of length types of solutions of the pair E_1, E_2 of rank $n-1$ is covered by a union of N $(n-1)$-dimensional subspaces, then $m \leq N+1$. There are two unknowns x, y such that $m \leq (|E_1|_x + |E_1|_y)^2 + 1$.

Proof. We can assume that E_i is equivalent to the system $E_1, \ldots, E_i$ for all $i \in \{1, \ldots, m\}$. Let the set of length types of solutions of E_2 of rank $n-1$ be covered by the $(n-1)$-dimensional spaces $V_1, \ldots, V_N$. Some subset of these spaces forms a minimal cover for the length types of solutions of E_3 of rank $n-1$. If this minimal cover would be the whole set, then E_2 and E_3 would have the same solutions of rank $n-1$ by the second part of Theorem 4.3.3. Thus the set of length types of solutions of E_3 of rank $n-1$ is covered by some $N-1$ of these spaces. We conclude inductively that the set of length types of solutions of E_i of rank $n-1$ is covered by some $N-i+2$ of these spaces for all $i \in \{2, \ldots, m\}$. It must be $N-m+2 \geq 1$, so $m \leq N+1$. The second claim follows by Theorem 4.3.3. □

In the case of three unknowns, Theorem 4.6.1 gives an upper bound depending on the length of the shortest equation for the size of an independent system of equations, or an upper bound depending on the length of the first equation for the size of a chain of equations. A better bound in Theorem 4.3.3 would immediately give a better bound in the following corollary.

Corollary 4.6.2. *If $E_1, \ldots, E_m$ is an independent system on three unknowns having a nonperiodic solution, then $m \leq (|E_1|_x + |E_1|_y)^2 + 1$ for some $x, y \in \Xi$. If $E_1, \ldots, E_m$ is a decreasing chain of equations on three unknowns, then $m \leq (|E_1|_x + |E_1|_y)^2 + 5$ for some $x, y \in \Xi$.*

Corollary 4.6.2 means that as soon as we take one equation on three unknowns, we get a fixed bound for the size of independent systems containing that equation.

It is worth noting that the bounds in Theorem 4.6.1 and Corollary 4.6.2 do not depend on the number of unknowns, only on the length of one equation.

Getting a similar bound for the sizes of independent systems or decreasing chains in the case of more than three unknowns remains an open problem. Such a bound would have to depend on the number of unknowns. Indeed, in Theorem 4.6.1 it is not enough to assume that the equations are independent and have a common solution of rank $n-1$. If the number of unknowns is not fixed, then there are arbitrarily large such systems where the length of every equation is 10 [33].

Chapter 5

Parametric Solutions

In this chapter we reprove Hmelevskii's theorem: every equation on three unknowns has a parametric solution. We also analyze this proof to obtain an exponential bound for the size of the parametric solution. This gives an exponential bound for the size of the shortest nontrivial solution, which, in turn, proves that the existence of such a solution can be solved in nondeterministic polynomial time.

The general structure of our new proof is quite similar to the original proof of Hmelevskii, and many of the lemmas in this chapter can be found in [26] in some form. However, many parts of the proofs are different. For example, exponential equations are treated in a totally different way. The use of more modern results such as Theorem 2.3.6 simplifies the proofs of many lemmas. Some large parts of the original proof become unnecessary. The results about sizes of parametric solutions and about complexity of solving equations are entirely new.

In Section 5.1 we study the form of parametric solutions.

Section 5.2 deals with exponential equations, which are an important tool used in our proof.

In Section 5.3 we are able to prove Hmelevskii's theorem for a large class of equations. All other equations will be reduced to these so called basic equations later on.

The main tools in this process are images and θ-images, which are the topic of Section 5.4.

Finally, in the last three sections a so called basic tree is constructed for an arbitrary equation, and this completes the proof of Hmelevskii's theorem. An upper bound for the height of such a tree gives an upper bound for the length of the parametric solution. This leads to an upper bound for the length of the shortest nontrivial solution, and to a nondeterministic polynomial time algorithm for solving the existence of such a solution.

This chapter is based on the articles [34] and [55].

5.1 Remarks about Parametric Solutions

Parametric words and solutions were defined in Section 2.5. In this section we take a closer look at them.

In addition to ordinary word equations, we also consider *one-sided* equations $xU \rightrightarrows yV$. A morphism $h : \Xi^* \to \Sigma^*$ is a solution of such an equation, if $h(xU) = h(yV)$ and $|h(x)| \geq |h(y)|$.

Periodic solutions are easy to find and represent, so in many cases it is enough to consider nonperiodic solutions.

If $\Sigma = \{a_1, \ldots, a_n\}$, then $U \in \Sigma^*$ can be denoted $U[a_1, \ldots, a_n]$, and its image under a morphism h can be denoted $h(U) = U[h(a_1), \ldots, h(a_n)]$. If $u \in \Sigma^*$, then by the morphism $a_1 \mapsto u$ we mean the morphism that maps $a_1 \mapsto u$ and $a_i \mapsto a_i$ when $i = 2, \ldots, n$.

The following theorem states that the basic tool in solving equations, namely the cancellation of the first variable, preserves the parameterizability of solutions.

Theorem 5.1.1. *Let $U, V \in \Xi^*$, $x, y \in \Xi$ and $x \neq y$. Let $h : \Xi^* \to \Xi^*$ be the morphism $x \mapsto yx$. If the equation $xh(U) = h(V)$ has a parametric solution, then so does the equation $xU \rightrightarrows yV$.*

Proof. If the equation $xh(U) = h(V)$ has a parametric solution

$$\{(h_j, R_j) \mid 1 \leq j \leq m\},$$

then the equation $xU \rightrightarrows yV$ has the parametric solution

$$\{(h_j \circ h, R_j) \mid 1 \leq j \leq m\}. \qquad \square$$

Next we make some remarks about parametric solutions to increase our understanding of them.

A parametric solution was defined as a set $\{(h_j, R_j) \mid 1 \leq j \leq m\}$. This solution can be written less formally as

$$\begin{aligned} &x = h_1(x),\ y = h_1(y),\ z = h_1(z), \quad R_1 \qquad \text{or} \\ &\ \vdots \\ &x = h_m(x),\ y = h_m(y),\ z = h_m(z), \quad R_m, \end{aligned}$$

if the unknowns are x, y, z. Actually, only one pair (h, R) is needed. For example, if we have a parametric solution

$$x = \alpha_1,\ y = \beta_1,\ z = \gamma_1 \qquad \text{or} \qquad x = \alpha_2,\ y = \beta_2,\ z = \gamma_2,$$

we can replace it with

$$x = \alpha_1^i \alpha_2^j,\ y = \beta_1^i \beta_2^j,\ z = \gamma_1^i \gamma_2^j, \quad i + j = 1,$$

where i and j are new parameters.

On the other hand, the linear Diophantine relations are not necessary either. A parametric solution

$$x = \alpha_1,\ y = \beta_1,\ z = \gamma_1, \quad R$$

can be replaced with a parametric solution of the form

$$\begin{aligned} &x = \alpha_1,\ y = \beta_1,\ z = \gamma_1 \qquad \text{or} \\ &\vdots \\ &x = \alpha_m,\ y = \beta_m,\ z = \gamma_m. \end{aligned}$$

This follows from the results in [17]. We will not give the proof here, but present an example.

Example 5.1.2. Consider the periodic solutions of the equation $x^n = yz$. They are

$$x = t^i,\ y = t^j,\ z = t^k, \quad ni = j + k.$$

We can replace j with $nj' + b$ and k with $nk' + c$, where $0 \leq b, c < n$. Then $i = j' + k' + (b + c)/n$. Only those pairs (b, c) for which $b + c$ is divisible by n are possible. Thus we get a representation

$$\begin{aligned} &x = t^{j'+k'}, \quad y = t^{nj'}, \qquad z = t^{nk'} \qquad \text{or} \\ &x = t^{j'+k'+1},\ y = t^{nj'+1}, \quad z = t^{nk'+n-1} \qquad \text{or} \\ &x = t^{j'+k'+1},\ y = t^{nj'+2}, \quad z = t^{nk'+n-2} \qquad \text{or} \\ &\vdots \\ &x = t^{j'+k'+1},\ y = t^{nj'+n-1},\ z = t^{nk'+1}, \end{aligned}$$

where the parameters j', k' can now have any nonnegative values.

The periodic solutions of an equation on three unknowns can be represented with just one morphism and without any Diophantine relations.

Theorem 5.1.3. *The periodic solutions of an equation $U = V$ have a representation*

$$x = t^p,\ y = t^q,\ z = t^r,$$

where p, q, r are polynomials of numerical parameters

Proof. All periodic solutions of an equation $U = V$ are of the form $x = t^i, y = t^j, z = t^k$, and the exponents i, j, k must satisfy the constraint

$$|U|_x i + |U|_y j + |U|_z k = |V|_x i + |V|_y j + |V|_z k.$$

By permuting the unknowns we can assume that this can be written as $ai = bj + ck$, where a, b, c are nonnegative integers and $a > 0$ (except for some trivial cases). Let $(b_n, c_n)_{n=1}^N$ be a sequence of all solutions $(u, v) \in \{0, \ldots, a-1\}^2$ of the congruence $bu+cv \equiv 0 \pmod{a}$. For each pair (b_n, c_n), we could define a corresponding morphism, and these would together form a parametric representation. This was done in Example 5.1.2. However, we can also replace the exponents i, j, k with the polynomials

$$p = bj_0 + ck_0 + \sum_{n=1}^{N} \frac{bb_n + cc_n}{a} i_n,$$
$$q = aj_0 + \sum_{n=1}^{N} b_n i_n,$$
$$r = ak_0 + \sum_{n=1}^{N} c_n i_n,$$

where $j_0, k_0, i_1, \ldots, i_n$ are new parameters, which can now have any values. Thus the solutions can be represented with one parametric word for each unknown. The parametric representation has at most quadratic length with respect to the length of the equation. □

Theorem 5.1.3 does not hold if instead of periodic solutions we consider all solutions. Indeed, we will show that a parametric solution for the equation $xyxzyz = zxzyxy$ consists of at least three morphisms if linear Diophantine relations are not allowed. The solutions of this equation were determined in Lemma 2.4.8.

The number of occurrences of a letter $a \in \Sigma$ in a parametric word after giving values for the parameters can be viewed as a polynomial, where for every $i \in \Lambda$ there is a variable X_i and for every $p \in \Delta$ and $a \in \Sigma$ there is a variable $X_{p,a}$. Formally, we define the polynomial $|\alpha|_a$ as follows:

(i) if $p \in \Delta$, then $|(p)|_a = X_{p,a}$,

(ii) if α and β are parametric words, then $|(\alpha\beta)|_a = |\alpha|_a + |\beta|_a$,

(iii) if α is a parametric word and $i \in \Lambda$, then $|(\alpha^i)|_a = |\alpha|_a X_i$.

For example, $|(p^i q)^j p|_a = X_{p,a}X_iX_j + X_{q,a}X_j + X_{p,a}$. If φ is a valuation, then $|\varphi(\alpha)|_a$ is the value taken by the polynomial $|\alpha|_a$, when X_i is given the value $\varphi(i)$ (for all $i \in \Lambda$) and $X_{p,a}$ is given the value $|\varphi(p)|_a$ (for all $p \in \Delta$).

Theorem 5.1.4. *The equation $xyxzyz = zxzyxy$ does not have a parametric solution of the form*

$$x = \alpha_1,\ y = \beta_1,\ z = \gamma_1 \qquad \text{or} \qquad x = \alpha_2,\ y = \beta_2,\ z = \gamma_2.$$

Proof. The following are examples of the solutions of the equation:

$$x = a,\ y = b,\ z = \varepsilon; \quad x = a,\ y = b,\ z = ab; \quad x = a,\ y = a,\ z = a. \tag{5.1}$$

We will show that if α, β, γ are parametric words such that

$$x = \varphi(\alpha),\ y = \varphi(\beta),\ z = \varphi(\gamma) \tag{5.2}$$

is a solution of the equation for all valuations φ, then we can get at most one of the three above-mentioned solutions from these parametric words.

Consider the three polynomials $p_1 = |\gamma|_a$, $p_2 = |\alpha\beta|_a - |\gamma|_a$ and $p_3 = |\alpha|_a|\beta|_b - |\alpha|_b|\beta|_a$. The values taken by them are

$$|z|_a,\ |xy|_a - |z|_a,\ |x|_a|y|_b - |x|_b|y|_a, \tag{5.3}$$

where (x, y, z) can be any solution of the equation $xyxzyz = zxzyxy$. If $z = \varepsilon$, then the first value is zero, if $xy = z$, then the second value is zero, and if x and y are powers of a common word, then the third value is zero. For every solution one of these holds by Lemma (2.4.8), so the product of the three polynomials $p_1p_2p_3$ vanishes everywhere. Thus $p_1p_2p_3$ is the zero polynomial. But this means that one of the polynomials p_1, p_2, p_3 must be the zero polynomial. For the first solution in (5.1) only the first value in (5.3) is zero, for the second solution only the second value is zero, and for the third solution only the third value is zero. Thus only one of these solutions can be obtained from (5.2). □

5.2 Exponential Equations

Let α and β be parametric words. The pair (α, β) can be viewed as an equation, referred to as an *exponential equation*. The *height* of this equation is the height of $\alpha\beta$. The solutions of this equation are the functions $f : \Lambda \to \mathbb{N}_0$ that satisfy $f(\alpha) = f(\beta)$. If the numerical parameters are in order $i_1, \ldots, i_n$, then we can talk of the solution $(f(i_1), \ldots, f(i_n))$ or of the solution $i_1 = f(i_1), \ldots, i_n = f(i_n)$.

If we know some parametric words which give all solutions of an equation, but which also give some extra solutions, then often the right solutions can be picked by adding some constraints for the numerical parameters. These constraints can be found by exponential equations, and the following theorems prove that they are in our cases equivalent to linear Diophantine relations.

The proofs in this section are connected to the method of Chapter 4. We will transform words into polynomials when studying exponential equations. Alphabet Ξ with k letters can be thought to be the set $\{1, \ldots, k\}$. If $w = a_0 \ldots a_{n-1} \in \Sigma^n$, then we let

$$P_w = a_0 + a_1X^1 + \cdots + a_{n-1}X^{n-1}$$

like in Section 4.1.

Theorem 5.2.1. *Let $E : \alpha = \beta$ be an exponential equation of height one. There exists a linear Diophantine relation R such that a function $f : \Lambda \to \mathbb{N}_0$ is a solution of E if and only if $f \in R$. The coefficients in R are of size $O(|\alpha\beta|)$.*

Proof. Let

$$\alpha = s_0 t_1^{i_1} s_1 \dots t_m^{i_m} s_m, \qquad \beta = u_0 v_1^{j_1} u_1 \dots v_n^{j_n} u_n,$$

where $s_0, \dots, s_m, u_0, \dots, u_n \in \Delta^*$, $t_1, \dots, t_m, v_1, \dots, v_n \in \Delta^+$ and $i_1, \dots, i_m$, $j_1, \dots, j_n \in \Lambda$. Function f is a solution if and only if $P_{f(\alpha)} = P_{f(\beta)}$. Now $P_{f(\alpha)}$ is

$$\begin{aligned} &P_{s_0} + \frac{P_{t_1}(X^{|t_1|f(i_1)} - 1)}{X^{|t_1|} - 1} \cdot X^{|s_0|} \\ &\quad + \dots + \frac{P_{t_m}(X^{|t_m|f(i_m)} - 1)}{X^{|t_m|} - 1} \cdot X^{|s_0|} X^{|t_1|f(i_1)} X^{|s_1|} \dots X^{|s_{m-1}|} \\ &\quad + P_{s_m} X^{|s_0|} X^{|t_1|f(i_1)} X^{|s_1|} \dots X^{|t_m|f(i_m)}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} &P_{s_0} - \frac{P_{t_1} X^{|s_0|}}{X^{|t_1|} - 1} \\ &+ \sum_{k=1}^{m-1} \left(\frac{P_{t_k}}{X^{|t_k|} - 1} + P_{s_k} - \frac{P_{t_{k+1}} X^{|s_k|}}{X^{|t_{k+1}|} - 1} \right) X^{|s_0 \dots s_{k-1}|} X^{|t_1^{f(i_1)} \dots t_k^{f(i_k)}|} \\ &+ \left(\frac{P_{t_m}}{X^{|t_m|} - 1} + P_{s_m} \right) X^{|s_0 \dots s_{m-1}|} X^{|t_1^{f(i_1)} \dots t_m^{f(i_m)}|} \end{aligned}$$

and $P_{f(\beta)}$ is of the corresponding form. Thus the equation

$$\begin{aligned} &(X^{|t_1|} - 1) \dots (X^{|t_m|} - 1)(X^{|v_1|} - 1) \dots (X^{|v_n|} - 1) P_{f(\alpha)} \\ =&(X^{|t_1|} - 1) \dots (X^{|t_m|} - 1)(X^{|v_1|} - 1) \dots (X^{|v_n|} - 1) P_{f(\beta)} \end{aligned}$$

can be rewritten as

$$X^{p_1} + \dots + X^{p_M} = X^{q_1} + \dots + X^{q_N}, \tag{5.4}$$

where every p_k and q_k is a linear polynomial with unknowns $f(i_l)$, $f(j_l)$. The coefficients in these polynomials are of size $O(|\alpha\beta|)$. Equation (5.4) can be satisfied only if $M = N$. Then it is equivalent to the formula

$$\bigvee_{\pi} \left((p_1 = q_{\pi(1)}) \wedge \dots \wedge (p_N = q_{\pi(N)}) \right),$$

where π runs over all permutations of N elements. Hence the claim follows. □

In some cases Theorem 5.2.1 can be generalized for exponential equations of height two.

Theorem 5.2.2. *Let $\Lambda = \{i, j_1, \ldots, j_r\}$ and let $s_0, \ldots, s_m$, $u_0, \ldots, u_n$ and t be parametric words of height at most one, with no occurrences of the parameters j_k. Let $\alpha = s_0 t^{j_{k_1}} s_1 \ldots t^{j_{k_m}} s_m$ and $\beta = u_0 t^{j_{l_1}} u_1 \ldots t^{j_{l_n}} u_n$. Now there exists a linear Diophantine relation R such that a function $f : \Lambda \to \mathbb{N}_0$ is a solution of the exponential equations $E : \alpha = \beta$ if and only if $f \in R$. The coefficients in R are of size $O(|\alpha\beta|)$.*

Proof. Like in the proof of Theorem 5.2.1, the equation $P_{f(\alpha)} = P_{f(\beta)}$ can be turned into the equation

$$X^{p_1} + \cdots + X^{p_M} = X^{q_1} + \cdots + X^{q_N}, \tag{5.5}$$

where every p_k and q_k is of the form

$$\sum_{k=1}^{r} a_k(bf(i)f(j_k) + cf(j_k)) + df(i) + e$$

for some integers $a_1, \ldots, a_r, b, c, d, e$. The coefficients in these polynomials are of size $O(|\alpha\beta|)$. Equation (5.5) can be satisfied only if $M = N$. Then it is equivalent to the formula

$$\bigvee_{\pi} \left((p_1 = q_{\pi(1)}) \wedge \cdots \wedge (p_N = q_{\pi(N)}) \right),$$

where π runs over all permutations of N elements. Consider now the equations $p_k = q_{\pi(k)}$. They are of the form

$$bf(i)\sum_{k=1}^{r} a_k f(j_k) + c\sum_{k=1}^{r} a_k f(j_k) + df(i) + e = 0$$

for some integers $a_1, \ldots, a_r, b, c, d, e$. If $b = 0$, this is a linear equation. If $b \neq 0$, then

$$f(i) \leq \left|\frac{c}{b}\right| + |d| + |e| \qquad \text{or} \qquad \left| b\sum_{k=1}^{r} a_k f(j_k) \right| \leq \left|\frac{c}{b}\right| + |d| + |e|,$$

because otherwise

$$\left| bf(i)\sum_{k=1}^{r} a_k f(j_k) \right| > \left| c\sum_{k=1}^{r} a_k f(j_k) + df(i) + e \right|.$$

If $f(i)$ or $b\sum_{k=1}^{r} a_k f(j_k)$ is fixed, the equation turns into a linear equation or a pair of linear equations. Hence the claim follows. □

We must examine some exponential equations of height one more closely.

Theorem 5.2.3. *Let $\Lambda = \{i\}$. Let $E : s_0 t^i s_1 \dots t^i s_m = u_0 t^i u_1 \dots t^i u_n$ be an exponential equation of height one, $s_0, \dots, s_m$, $u_0, \dots, u_n$, $t \in \Delta^*$ and $|s_0 \dots s_m u_0 \dots u_n| < S|t|$. There exists a number $T = O(S)$ such that f is a solution of E for every $f(i) \geq T$, or f is not a solution for any $f(i) \geq T$.*

Proof. Like in the proof of Theorem 5.2.1, we get the equation (5.4). The polynomials p_j, q_j are of the form $af(i) + b$. On the other hand, they are exponents of terms of products of $X^{|s_k|}$, P_{s_k}, $X^{|u_k|}$, P_{u_k}, $X^{|t|}$, $X^{|t|f(i)}$, P_t. Each of these polynomials can occur in the products at most once, except $X^{|t|}$ and $X^{|t|f(i)}$, which can occur at most $m + n$ times. Thus $|t|$ divides a and $b \leq 2|s_0 \dots s_m t^{m+n}|$. The equation $p_j = q_{\pi(j)}$ can be written as $Af(i) = B$, where $A = 0$ or $|A| \geq |t|$ and $|B| \leq 2|s_0 \dots s_m u_0 \dots u_n t^{m+n}|$. Now there exists the required number T such that the equations $p_j = q_{\pi(j)}$ have no solutions $f(i) \geq T$, unless the equations are trivial. This proves the claim. □

5.3 Basic Equations

From now on we only consider equations on three unknowns. The alphabet of unknowns is $\Xi = \{x, y, z\}$. The left-hand side of an equation can be assumed to begin with x. We can also assume that x occurs on the right-hand side, but not as the first letter.

Periodic solutions and solutions where some unknown has the value ε are called *trivial*. These are easy to parameterize by Theorems 2.3.1 and 2.3.2.

An equation is a *basic equation* if it is a trivial equation $U = U$, where $U \in \Xi^*$, if it has only trivial solutions, or if it is of one of the following forms, where $a, b \geq 1$, $c \geq 2$ and $t \in \{x, z\}$:

B1. $x^a y \cdots = y^b x \cdots$

B2. $x^2 \cdots \rightrightarrows y^a x \cdots$

B3. $xyt \cdots \rightrightarrows zxy \cdots$

B4. $xyt \cdots \rightrightarrows zyx \cdots$

B5. $xyz \cdots = zxy \cdots$

B6. $xyz \cdots = zyx \cdots$

B7. $xy^c z \cdots = zy^c x \cdots$

B8. $xyt \cdots \rightrightarrows z^a xy \cdots$

B9. $xyxz\cdots \rightrightarrows zx^2y\cdots$

The parameterizability of basic equations is easy to prove with the help of previous lemmas and theorems.

Lemma 5.3.1. *Let $S,T,U,V \in \Xi^*$. Assume that the equation $S = T$ has a parametric solution $\{(h_j, R_j) \mid j = 1,\ldots,m\}$, where $\Delta = \{p,q\}$ and $\Lambda = \{i_1,\ldots,i_k\}$. Assume that the exponential equations $h_j(U) = h_j(V)$ are equivalent to linear Diophantine relations. Then the pair of equations $S = T, U = V$ has a parametric solution.*

Proof. Let $h_j(U) = h_j(V)$ be equivalent to the linear Diophantine relation R'_j. We show that the solutions of the equation have a parametric representation

$$\big\{(h_j, R_j \cap R'_j) \mid j = 1,\ldots,m\big\} \cup A,$$

where A is a parametric representation of the periodic solutions

If $\varphi = h \circ f$ is a valuation in $R_j \cap R'_j$, then $\varphi \circ h_j$ is a solution of $S = T$ and f is a solution of $h_j(U) = h_j(V)$. Now $\varphi \circ h_j$ is also a solution of $U = V$.

If g is a nonperiodic solution of the pair of equations $S = T, U = V$, then $g = \varphi \circ h_j$ for some number j and valuation $\varphi = h \circ f$ satisfying $f \in R_j$. It needs to be shown that f is a solution of $h_j(U) = h_j(V)$. The morphism h is a solution of the equation $f(h_j(U)) = f(h_j(V))$, which has two unknowns. But h cannot be periodic, because g is not periodic. Thus $f(h_j(U))$ and $f(h_j(V))$ must be the same word □

Theorem 5.3.2. *Every basic equation has a parametric solution. The solution is of length $O(1)$ and the coefficients in the linear Diophantine relations are of size $O(n)$, where n is the length of the equation.*

Proof. For equations $U = U$ and for equations with only trivial solutions the claim is clear. We prove it for equations B1 – B9. First we reduce equations to other equations by Theorem 5.1.1. The equation B2 is reduced by the substitution $x \mapsto yx$ to the equation $xyx\cdots = y^a x\cdots$, which is of the form B1. The equations B3 and B4 are reduced by the substitution $x \mapsto zx$ to the equations $xyz\cdots = zxy\cdots$ and $xyz\cdots = yzx\cdots$, which are of the form B5. The equation B8 is reduced by the substitution $x \mapsto zx$ to the equation $xyzA = z^a xyB$ for some $A, B \in \Xi^*$. By Lemma 2.4.7, this is equivalent to the equation $xyzxyzA = zxyz^a xyB$, which is of the form B5. Therefore only the cases B1, B5, B6, B7 and B9 have to be considered; if we can prove the claims about the existence and size of parametric solutions for these equations, then the claims hold for all basic equations.

Consider the equations B1, B5, B6, B7 and B9 as the equation $U = V$ of Lemma 5.3.1, and the equations $xy = yx$, $xyz = zxy$, $xyz = zyx$, $xy^cz = zy^cx$ and $xyxz \rightrightarrows zx^2y$ as the equation $S = T$. For B1 this can be done by

Lemma 2.4.7, otherwise by a length argument. By Lemmas 2.4.2 – 2.4.6, the solutions of these equations are obtained from certain parametric words over word parameters p, q and numerical parameters i, j, k. For equations B1, B5 and B6, the exponential equation of Lemma 5.3.1 will be of height one and Theorem 5.2.1 can be used. For B9 and B7, Theorem 5.2.2 can be used. So the exponential equation is in all cases equivalent to a linear Diophantine relation with coefficients of size $O(n)$. The claim follows from Lemma 5.3.1, because the parametric solutions of Lemmas 2.4.2 – 2.4.6 are of bounded length. □

5.4 Images and θ-Images

In this section we define images and θ-images of equations and prove some results about these. If h is a solution of the equation $xU \rightrightarrows yV$, then $h(y) \leq h(x)$. This fact was already behind Theorem 5.1.1. This will be generalized.

Let $t_1, \ldots, t_n \in \{y, z\}$ and $V = t_1 \ldots t_n$. Let $t_{n+1} = t_1$. If a morphism h is a solution of the equation $E : xU \rightrightarrows VxW$, then

$$h(x) = h(V^k t_1 \ldots t_i)u \tag{5.6}$$

for some numbers k, i and word u satisfying $k \geq 0$, $0 < i \leq n$ and $h(t_{i+1}) \not\leq u$.

On the other hand, a morphism h satisfying (5.6) is a solution of E if and only if $uh(U) = h(t_{i+1} \ldots t_n t_1 \ldots t_i)uh(W)$. We can write $h = g \circ f$, where f is the morphism $x \mapsto V^k t_1 \ldots t_i x$ and g is the morphism for which $g(x) = u$, $g(y) = h(y)$ and $g(z) = h(z)$. Now h is a solution of E if and only if g is a solution of

$$xf(U) \leftleftarrows f(t_{i+1} \ldots t_n t_1 \ldots t_i)xf(W). \tag{5.7}$$

An *image* of an equation $xU[x, y, z] \rightrightarrows V[y, z]xW[x, y, z]$ under the morphism $x \mapsto V^k Px$, where $k \geq 0$, $V = PQ$ and $Q \neq \varepsilon$, is

$$xU[V^k Px, y, z] \leftleftarrows QPxW[V^k Px, y, z].$$

If V contains only one of y, z or if $P = \varepsilon$, the image is *degenerated*.

The m first images of an equation of length n are of length $O(mn)$. Images are needed in the most important reduction steps used in the proof of parameterizability of equations with three unknowns. The solutions of an equation are easily obtained from the solutions of its images, so it is enough to consider them. There are infinitely many images, but a finite number is enough if one of them is turned from a one-sided equation to an ordinary equation.

Equation E is *reduced to the equations* $E_1, \ldots, E_n$ *by an* n*-tuple of substitutions* if E is of the form $xU[x,y,z] \rightrightarrows t_1 \ldots t_k xV[x,y,z]$, where $1 \leq n \leq k$ and $t_1, \ldots, t_k \in \{y,z\}$, equation E_i is

$$xU[t_1 \ldots t_i x, y, z] \lhd\!\lhd t_{i+1} \ldots t_k t_1 \ldots t_i xV[t_1 \ldots t_i x, y, z],$$

when $1 \leq i < n$, and equation E_n is

$$xU[t_1 \ldots t_n x, y, z] = t_{n+1} \ldots t_k t_1 \ldots t_n xV[t_1 \ldots t_n x, y, z].$$

By the above, Theorem 5.1.1 can be generalized.

Theorem 5.4.1. *Let E be an equation of length n. If E is reduced to the equations $E_1, \ldots, E_m$ by an m-tuple of substitutions, and if $E_1, \ldots, E_m$ have parametric solutions of length at most c, then E has a parametric solution of length $O(mn)c$.*

Reductions with n-tuples of substitutions are not sufficient. Other ways to restrict the considerations to a finite number of images are needed.

Equation

$$xU[x,y,z] \rightrightarrows V[y,z]xW[x,y,z]$$

is of *type I* if both unknowns y, z occur in V. Equation

$$xy^bU[x,y,z] \rightrightarrows z^cxV[x,z]yW[x,y,z],$$

where $b, c \geq 1$, is of *type II* if $b > 1$ or $V \neq \varepsilon$.

Theorem 5.4.2. *The solutions of an equation of type I of length n can be parameterized in terms of the solutions of $O(n^2)$ of its images of length $O(n^3)$.*

Proof. Consider the equation $E : xU[x,y,z] \rightrightarrows V[y,z]xW[x,y,z]$, where both y and z occur in V, and its images

$$E_{P,i} : xU[V^iPx, y, z] \lhd\!\lhd QPxW[V^iPx, y, z], \tag{5.8}$$

where $i \geq 0$, $V = PQ$ and $Q \neq \varepsilon$. We show that there exists a number T such that if P and Q are fixed, then the equations (5.8) are equivalent for all $i \geq T$.

Let h be a solution of $E_{P,i}$. Then Theorem 5.2.3 can be used for the exponential equation

$$h(x)U[h(V)^ih(Px), h(y), h(z)] = h(QPx)W[h(V)^ih(Px), h(y), h(z)],$$

where i is considered to be unknown. The bound S in the theorem does not depend on h and is of size $O(n)$, because both y and z occur in V and

$h(x) \leq h(y)$ or $h(x) \leq h(z)$. So there exists a number $T = O(n)$ such that h is a solution for all $i \geq T$ or for no $i \geq T$. Thus the equations $E_{P,T}, E_{P,T+1}, E_{P,T+2}, \ldots$ are equivalent if P is fixed.

Now the images of this theorem can be taken to be $E_{P,j}$, where $P < V$ and $j \leq T$. The solutions of E are $g \circ f \circ h'$, where either h' is the morphism $x \mapsto V^j Px$, g runs over the solutions of the corresponding image $E_{P,j}$, f does nothing and $j < T$, or h' is the morphism $x \mapsto V^{T+i} Px$, i is a numerical parameter, g runs over the solutions of $E_{P,T}$ and f gives values for i. Because $T = O(n)$ and $|V| = O(n)$, there are $O(n^2)$ of these images, and because $|V^j Px| = O(n^2)$, they are of length $O(n^3)$. □

Like in the proof of Theorem 5.4.2, we will often use a variation of the following reasoning: if the images of E are $E_1, E_2, \ldots$, and if $E_m, E_{m+1}, \ldots$ are equivalent, then the solutions of E can be parameterized in terms of the solutions of $E_1, \ldots, E_m$. It is also easy to see that if each of these images has a parametric solution of length at most c, then E has a parametric solution of length $O(m^2)c$. This also holds for θ-images, which are defined later.

The next example shows that the claim in the proof of Theorem 5.4.2 about the existence of a number T does not hold for equations of type II.

Example 5.4.3. The images of the equation $xy \rightrightarrows zx^2$ are $xy \leftleftarrows zxz^i x$. Now $x = a$, $y = ba(ab)^N a$, $z = ab$ is a solution of an image if and only if $i = N$. So all images have different solution sets.

Consider an equation of type II

$$xy^b A[x, y, z] \rightrightarrows z^c x B[x, z] y C[x, y, z], \tag{5.9}$$

where $b, c \geq 1$ and $b > 1$ or $B \neq \varepsilon$. Its images are degenerated and of the form

$$xy^b A[z^i x, y, z] \leftleftarrows z^c x B[z^i x, z] y C[z^i x, y, z]. \tag{5.10}$$

Theorem 5.4.2 holds for some of the equations (5.9).

Theorem 5.4.4. *If $B = z^d$, where $d \geq 1$, then the solutions of* (5.9) *can be parameterized in terms of the solutions of $O(n^2)$ of its images of length $O(n^3)$, where n is the length of the equation.*

Proof. Equation (5.10) is reduced by the mapping $z \mapsto xz$ to the equation

$$y^b A[(xz)^i x, y, xz] = (zx)^c (xz)^d y C[(xz)^i x, y, xz]. \tag{5.11}$$

Let h be its solution. Let $D = h((zx)^c (xz)^d)$ and $h(y) = D^j Y$, where $Y < D$. Then we get the equality

$$\begin{aligned} &Y(D^j Y)^{b-1} A[h((xz)^i x), D^j Y, h(xz)] \\ =&DYC[h((xz)^i x), D^j Y, h(xz)]. \end{aligned} \tag{5.12}$$

On the other hand, if (5.12) holds, then h is a solution of (5.11) and it gives a solution of (5.10). It needs to be shown that there exists a bound $T = O(n^2)$ not depending on h such that if (5.12) holds for some $i \geq T$, then it holds for all $i \geq T$. Then the images (5.10) with $i \leq T$ are sufficient, like in the proof of Theorem 5.4.2.

If $j < c + d + 1 = O(n)$ is fixed, then (5.12) can be considered to be an exponential equation with unknown i, and Theorem 5.2.3 can be used. The bound $S = O(n^2)$ not depending on h exists, because $|D| = (d + c)|h(xz)|$.

For the rest of the proof we consider the case $j \geq c + d + 1$. Let t and v be the primitive roots of $h(xz) = t^{a_1}$ and $D = v^{a_2}$. If these have equal length, then $t = v$ and $h(xz) = h(zx)$, which leads to a periodic solution. Assume that $|t| \neq |v|$. In (5.12), starting from the left, move the powers of t and v as far to the left as possible by changing them to their suitable conjugates, and then combine as much as possible from the right to these powers. This may require replacing i and j with $i' = i - b_1$ and $j' = j - b_2$ for some b_1, b_2. By Theorem 2.3.4, powers of conjugates of $h(xz)$ and D can overlap for at most $|h(xz)D|$ letters, so we can select $b_1, b_2 \leq c + d + 1$, and i' and j' can be used if i and j are large enough. This way (5.12) can be written as

$$s_0 t_1^{p_1} s_1 \dots t_M^{p_M} s_M = u_0 v_1^{q_1} u_1 \dots v_N^{q_N} u_N, \tag{5.13}$$

where $s_0, \dots, s_M, u_0, \dots, u_N \in \Delta^*$, every t_k and v_k is either a conjugate of t or a conjugate of v, and every p_k and q_k is a polynomial of first degree with unknowns i', j'. The coefficients in these polynomials cannot be negative. Also the last letter of t_k is different from the last letter of s_{k-1}, and $t_k \not\leq s_k t_{k+1}^a$ for all a. The same holds for words u_k and v_k and for polynomials q_k. Because the words $h(xz)$ and D consist of $h(x)$ and $h(z)$ and $Y < D$, there exists a bound $S = O(n^2)$ such that

$$|s_0 \dots s_M u_0 \dots u_N (tv)^2| < S|t^a| \qquad \text{and} \qquad b < a_1 S \tag{5.14}$$

when $ai' + b$ is in $\{p_1, \dots, p_M, q_1, \dots, q_N\}$. The same holds with v in place of t, j' in place of i', and a_2 in place of a_1.

We prove by induction with respect to $M+N$ that if (5.13) has a solution f with $f(i'), f(j') \geq S+2$, then $s_k = u_k$, $t_k = v_k$ and $f(p_k) = f(q_k)$ for all k. If $M+N = 0$, then the claim is clear (although the equation is of height zero). If $M = 0, N > 0$, or the other way around, then the exponent occurring in the equation can get only small values. Assume that $M, N > 0$. From (5.14) it follows that $|t_k^{f(p_k)}| > |s_0 \dots s_M u_0 \dots u_N (tv)^2|$ for all k, and similarly for v_k. It can be assumed that $u_0 \leq s_0$, so $v_1 = BA$ and $s_0 = u_0 (BA)^k B$ for some A, B. Now $|v_1^{f(q_1)}| \geq |s_0| + |t_1^2|$ and $|t_1^{f(p_1)}| \geq |(AB)^2|$. Thus the powers of t_1 and AB have a common prefix of length $|t_1 AB|$ and, by Theorem 2.3.4, $t_1 = AB$. Now $t_1 = v_1$, $B = \varepsilon$, $k = 0$ and $s_0 = u_0$. We prove that $f(p_1) = f(q_1)$. From $f(p_1) > f(q_1)$ it would follow that $v_1 = t_1 \leq u_1 v_2^{f(q_2)}$

(or $v_1 = t_1 \leq u_1$ if $N = 1$), which is a contradiction. The case $f(p_1) < f(q_1)$ is symmetric. It follows inductively that $s_k = u_k$, $t_k = v_k$ and $f(p_k) = f(q_k)$ for all k.

Now it can be seen that $p_k - q_k$ contains only one of i' and j', because $t_k = v_k$, and the coefficient of i' or j' is divisible by a_1 or a_2. So if $f(p_k) = f(q_k)$ and $f(i'), f(j') \geq S$, then it must be $p_k = q_k$ because of (5.14). The claim follows. □

Theorem 5.4.2 can be generalized by defining θ-images.

A sequence of equations $E_0, \ldots, E_n$ is a *chain of images* if E_i is an image of E_{i-1} for all i, $1 \leq i \leq n$. Then E_n is an *image of order* n of E_0. If every E_i is a degenerated image, then the chain is degenerated and E_n is a degenerated image of order n. These chains are not related to the decreasing chains of equations that were considered in the earlier chapters.

We define *θ-images* of equations of type I and II. For equations of type I all images are θ-images. For equations of type II the degenerated images of order 2 and nondegenerated images of order 3 are θ-images.

Lemma 5.4.5. *The solutions h of equation* (5.9) *satisfying $|h(y)| \leq |h(z)|$ can be parameterized in terms of the solutions of $O(n)$ of its images of length $O(n^2)$, where n is the length of the equation.*

Proof. This is proved like Theorem 5.4.2. Let E_i be the equation (5.10) and let h be its solution. Theorem 5.2.3 can be used for the exponential equation

$$h(xy^bA[z^ix, y, z]) = h(z^cxB[z^ix, z]yC[z^ix, y, z]),$$

where i is considered to be the unknown. The bound S does not depend on h and is of size $O(n)$, because $h(x), h(y) \leq h(z)$. So there exists a number $T = O(n)$ such that either h is a solution for all $i \geq T$ or for no $i \geq T$. Thus the equations $E_T, E_{T+1}, E_{T+2}, \ldots$ are equivalent. Like in the proof of Theorem 5.4.2, the images E_j, where $j \leq T$, are sufficient. □

Lemma 5.4.6. *The solutions h of equation* (5.9) *satisfying $|h(y)| \leq |h(z)|$ can be parameterized in terms of the solutions of $O(n^{17})$ of its θ-images of length $O(n^{18})$, where n is the length of the equation.*

Proof. Let these solutions be called τ-solutions. Let E_i be the equation (5.10). By Lemma 5.4.5, the τ-solutions can be parameterized in terms of the τ-solutions of $E_0, \ldots, E_T$ for some T. Let P_i be the set of those τ-solutions h of E_i for which $|h(z)| \geq |h(xy)|$, and let Q_i be the set of those τ-solutions h of E_i for which $|h(y)| \leq |h(z)| \leq |h(xy)|$.

Let E_i' be the image of E_i under the morphism $z \mapsto xz$, and let E_i'' be the image of E_i' under the morphism $y \mapsto zy$. From the length constraint $|h(y)| \leq |h(z)| \leq |h(xy)|$ it follows that the set Q_i can be parameterized in

terms of the solutions of E_i'', which is a nondegenerated image of the third order of (5.9).

Consider the set P_i. The equation (5.10) is of type I, so its solutions can be parameterized in terms of the solutions of a finite number of its images. Because of the condition $|h(z)| \geq |h(xy)|$ in the definition of P_i, the image under the morphism $z \mapsto xz$ can be omitted. Let the set thus obtained be F_i. The set P_i can be parameterized in terms of the solutions of equations of F_i. Partition F_i into the sets G_i and H_i of degenerated and nondegenerated images. The equations of H_i are of type I, so their solutions can be parameterized in terms of the solutions of a finite number of their images. These images are nondegenerated images of the third order of the original equation (5.9). The equations of G_i are degenerated images of the second order. So also P_i can be parameterized in terms of the solutions of a finite number of θ-images of (5.9).

In this construction there are $O(n)$ images of the first order of length $O(n^2)$, $O(n^5)$ images of the second order of length $O(n^6)$, and $O(n^{17})$ images of the third order of length $O(n^{18})$. The claim follows. □

Lemma 5.4.7. *Let $A, B, C \in \Xi^*$ and $i, k, a, p, a_1, \ldots, a_n \geq 0$ and $c, q > 0$. Assume that all letters x, y, z occur in A, $y \not\leq A$, $0 < q \leq n$ and $a_q + c + 2 \leq k \leq i - c - |A|$. Let*

$$D_1[x, z] = (zx)^c((xz)^{i+a_1}x)\ldots((xz)^{i+a_{q-1}}x)xz,$$
$$D_2[x, z] = (xz)^{i-k+a_q}x((xz)^{i+a_{q+1}}x)\ldots((xz)^{i+a_n}x)(xz)^p.$$

Now the equations

$$y(D_1B)^aA[(xz)^ix, D_1B, xz] \leftrightarrows zD_2D_1C[x, y, z]$$
$$y(D_1B)^aA[(xz)^ix, D_1B, xz] \leftrightarrows D_2D_1C[x, y, z]$$

have only trivial solutions.

Proof. The first equation is reduced by the morphism $z \mapsto yz$ to the equation

$$(D_1[x, yz]B')^aA[(xyz)^ix, D_1[x, yz]B', xyz] = zD_2[x, yz]D_1[x, yz]C'.$$

If $a > 0$, then the equation is of the form

$$(yzx)^cx\cdots = (zxy)^{i-k}\cdots.$$

Because $c > 0$ and $i - k \geq c + 1$, this equation has only trivial solutions by Corollary 2.3.7. If $a = 0$ and $z^my \leq A$, $m > 0$, then the equation is of the form

$$(xyz)^my\cdots = (zxy)^{i-k}\cdots.$$

Because $i-k \geq m+1$, this equation has only trivial solutions by Corollary 2.3.7. If $a=0$ and $z^m x \leq A$, $m \geq 0$, then the equation is of the form

$$(xyz)^{m+i} \cdots = (zxy)^{i-k+a_q} zxxyz \cdots ,$$

except if $n=q$ and $p=0$, when it is of the form

$$(xyz)^{m+i} \cdots = (zxy)^{i-k+a_q} zx(yzx)^c xyz \cdots .$$

Because $i-k+a_q > 0$ and $i > i-k+a_q+c+1$, this equation has in both cases only trivial solutions by Corollary 2.3.7.

The second equation is similar. It is reduced by the morphism $x \mapsto yx$ to the equation

$$\begin{aligned}&(D_1[yx,z]B')^a A[(yxz)^i yx, D_1[yx,z]B', yxz]\\ =&x(zyx)^{i-k+a_q}((yxz)^{i+a_{q+1}}yx)\ldots((yxz)^{i+a_n}yx)(yxz)^p D_1[yx,z]C'.\end{aligned}$$

If $a>0$, then the equation is of the form

$$(zyx)^c y \cdots = (xzy)^{i-k} \cdots .$$

Because $c>0$ and $i-k \geq c+1$, this equation has only trivial solutions by Corollary 2.3.7. If $a=0$ and $z^m y \leq A$, $m>0$, then the equation is of the form

$$(yxz)^m z \cdots = (xzy)^{i-k} \cdots .$$

Because $i-k \geq m+1$, this equation has only trivial solutions by Corollary 2.3.7. If $a=0$ and $z^m x \leq A$, $m \geq 0$, then the equation is of the form

$$(yxz)^{m+i} \cdots = (xzy)^{i-k+a_q} xyx \cdots ,$$

except if $n=q$ and $p=0$, when it is of the form

$$(yxz)^{m+i} \cdots = (xzy)^{i-k+a_q} x(zyx)^c yx \cdots .$$

Because $i-k+a_q > 0$ and $i > i-k+a_q+c+1$, this equation has in both cases only trivial solutions by Corollary 2.3.7. □

Lemma 5.4.8. *If x occurs in B, then the nonperiodic solutions h of (5.9) satisfying $|h(y)| \geq |h(z)|$, and some periodic solutions, can be parameterized in terms of the solutions of $O(n^5)$ of its θ-images of length $O(n^6)$, where n is the length of the equation.*

Proof. The images of (5.9) are the equations (5.10). Because of the condition $|h(y)| \geq |h(z)|$, it is enough to consider the image of this under the morphism $z \mapsto xz$:

$$y^b A[(xz)^i x, y, xz] \rightrightarrows (zx)^c B[(xz)^i x, xz] y C[(xz)^i x, y, xz]. \tag{5.15}$$

The length constraint is now $|h(y)| \geq |h(xz)|$. Equation (5.15) is a nondegenerated image of the second order of (5.9). Let $D = (zx)^c B[(xz)^i x, xz]$. Now the image of (5.15) under the morphism $y \mapsto D^j D_1 y$, where $j \geq 0$, $D_1 < D$ and $D^j D_1 \neq \varepsilon$, is

$$\begin{aligned} &y(D^j D_1 y)^{b-1} A[(xz)^i x, D^j D_1 y, xz] \\ \leftrightarrows &D_2 D_1 B[(xz)^i x, xz] y C[(xz)^i x, D^j D_1 y, xz], \end{aligned} \tag{5.16}$$

where $D_1 D_2 = D$.

We can write $D = (zx)^c((xz)^{i+a_1}x)\dots((xz)^{i+a_N}x)(xz)^p$, where $N \geq 1$, $p \geq 0$ and $a_1, \dots, a_N \geq 0$. Let $M = \max\{a_l + c + 1 + |A| \mid 1 \leq l \leq N\}$. If D_1 "cuts" the factor $(xz)^i$ in D, then

$$\begin{aligned} D_1 &= (zx)^c((xz)^{i+a_1}x)\dots((xz)^{i+a_{q-1}}x)(xz)^k, \\ D_2 &= (xz)^{i-k+a_q}x((xz)^{i+a_{q+1}}x)\dots((xz)^{i+a_N}x)(xz)^p \end{aligned}$$

or

$$\begin{aligned} D_1 &= (zx)^c((xz)^{i+a_1}x)\dots((xz)^{i+a_{q-1}}x)(xz)^{k-1}x, \\ D_2 &= z(xz)^{i-k+a_q}x((xz)^{i+a_{q+1}}x)\dots((xz)^{i+a_N}x)(xz)^p, \end{aligned}$$

where $0 < k \leq i$ and $0 < q \leq N$. If $M \leq k \leq i - M$, then, by Lemma 5.4.7, equation (5.16) has only trivial solutions.

All nonperiodic solutions h of (5.9) for which $|h(y)| \geq |h(z)|$ are obtained from the solutions of (5.16). Divide the solutions of the original equation into sets P and Q depending on whether they are obtained from (5.16) when $i \leq 2M$ or when $i \geq 2M$. It needs to be shown that these sets, and some periodic solutions, can be parameterized in terms of the solutions of a finite number of equations (5.16).

Let $U \leftrightarrows V$ be the equation (5.16) and let h be its solution. If $i \leq 2M$ is fixed, then $h(U) = h(V)$ can be viewed as an exponential equation with j as the unknown. We use Theorem 5.2.3. It gives a $T = O(n^2)$ such that h is a solution for all $j \geq T$ or for no $j \geq T$. It can be assumed that the same T is valid for all $i \leq 2M$. Like in the proof of Theorem 5.4.2, the set P, and some periodic solutions, can be parameterized in terms of the equations (5.16) with $i \leq 2M$ and $j \leq T$. There are $O(n^3)$ of those.

Consider the set Q. We can write $i = 2M + m$. Replace $(xz)^i$ with $(xz)^M(xz)^m(xz)^M$ in (5.16). Now D_1 can no longer "cut" $(xz)^m$ if we are interested only in equations with nonperiodic solutions. So there are only $O(n^2)$ possibilities for D_1. Fix D_1 and a solution h. Now $h(U) = h(V)$ can be viewed as an exponential equation with j and m as the unknowns. Fix m so that Theorem 5.2.3 can be used. There exists a bound $L = O(n)$ not depending on m such that either h is a solution for all $j \geq L$ or for no $j \geq L$.

Next, fix j and view $h(U) = h(V)$ as an exponential equation with m as the unknown. Now, by Theorem 5.2.3, there exists a bound $N_j = O(nj)$ such that either h is a solution for all $m \geq N_j$ or for no $m \geq N_j$. The bound N_j can be assumed to be increasing with respect to j. By combining these considerations it can be seen that either h is a solution for all $j \geq L$, $m \geq N_L$ or for no $j \geq L$, $m \geq N_L$. The set Q, and some periodic solutions, can be parameterized in terms of the equations (5.16) with $i \leq 2M + N_L$ and $j \leq L$. There are $O(n^3)$ of those for every D_1. This proves the theorem. □

Lemma 5.4.9. *If x occurs in B, then the nonperiodic solutions of* (5.9)*, and some periodic solutions, can be parameterized in terms of the solutions of $O(n^{17})$ of its θ-images of length $O(n^{18})$, where n is the length of the equation.*

Proof. The required θ-images are obtained by combining the sets of Lemmas 5.4.6 and 5.4.8. □

Lemma 5.4.10. *If $B = z^d$, where $d \geq 1$, then the solutions of* (5.9) *can be parameterized in terms of the solutions of $O(n^{26})$ of its θ-images of length $O(n^{27})$, where n is the length of the equation.*

Proof. All images of the equation are degenerate; $O(n^2)$ of these of length $O(n^3)$ can be chosen by Theorem 5.4.4. These images are of type I, so $O(n^6)$ of their images of length $O(n^9)$ can be chosen by Theorem 5.4.2. Of these images of the second order, the nondegenerated images are of type I, so $O(n^{18})$ of their images of length $O(n^{27})$ can be chosen. These nondegenerated images of the third order with the degenerated images of the second order give the set of required θ-images. □

We define a *complete set of θ-images* of an equation of type I or II. For equations of type I it is the set of Theorem 5.4.2. For equations of the form (5.9) it is the set of Lemma 5.4.6 if $B = \varepsilon$, the set of Lemma 5.4.9 if x occurs in B, and the set of Lemma 5.4.10 if $B = z^d$, $d \geq 1$. The next theorem follows immediately from this definition.

Theorem 5.4.11. *Every equation of type I or II of length n has a complete set of θ-images consisting of $O(n^{26})$ equations of length $O(n^{27})$.*

We assume that every complete set of θ-images satisfies the conditions of Theorem 5.4.11.

Theorem 5.4.12. *Let E be a word equation of length n. If $\{E_1, \ldots, E_m\}$ is a complete set of θ-images of E and every E_i has a parametric solution of length at most c, then E has a parametric solution of length $O(mn^{26})c$.*

Proof. For equations of type I this follows from Theorem 5.4.2. Consider the type II equation (5.9). If $B \neq \varepsilon$, then the claim follows from Lemmas 5.4.9 and 5.4.10. Assume that $B = \varepsilon$. By Lemma 5.4.6, it suffices to show that those solutions h of (5.9) for which $|h(y)| \geq |h(z)|$ can be parameterized. Let h be such a solution. Then $h(x) = h(z)^m u$ for some $m \geq 1$ and $u \leq h(z)$, $h(z) = uv$ for some v and $y = vuw$ for some w. Now $h = g \circ f$, where f and g are morphisms, $f(x) = (xz)^m x$, $f(y) = zxy$, $f(z) = xz$ and g is a solution of

$$yzx \cdots = (zx)^c y \cdots . \tag{5.17}$$

On the other hand, all such morphisms h are solutions of (5.9). By Lemma 2.4.7, g is also a solutions of $yzx = zxy$. Now, by Lemmas 2.4.3 and 5.3.1, the solutions g of (5.17) can be parameterized. This gives a parametric representation for the required solutions h if the exponent m in the morphism f is considered to be a numerical parameter. □

5.5 Neighborhoods and Trees

The proof of the parameterizability of equations with three unknowns consists mainly of reducing equations to other equations. This forms a tree-like structure. The intention is to make all leaf equations in this tree to be basic equations. The possible reduction steps are given in the definition of a neighborhood, which is preceded by two lemmas.

Lemma 5.5.1. *Let $u, v, w \in \Sigma^*$, $0 < |w| \leq |u|$ and $c \geq 1$. If*

$$wu^{c+1}v \cdots = u^{c+1}vu \cdots \qquad \text{or} \qquad w(uv)^c u^2 \cdots = (uv)^c u^2 \cdots ,$$

then $uv = vu$.

Proof. Let $u = wt$. From $wu^{c+1}v \cdots = u^{c+1}vu \cdots$ it follows that

$$(wt)^{c+1}v \cdots = t(wt)^c vwt \ldots \quad \text{and} \quad (wt)^{c+1}v = t(wt)^c vw.$$

From $w(uv)^c u^2 \cdots = (uv)^c u^2 \cdots$ it follows that

$$(wtv)^c wtwt \cdots = tv(wtv)^{c-1} wtwt \ldots \quad \text{and} \quad (wtv)^c wt = tv(wtv)^{c-1} wtw.$$

In both cases the beginnings and ends of the last equation give $wt = tw$ and $wtv = tvw$. So $\rho(w) = \rho(t) = \rho(tv) = \rho(v) = \rho(u)$. □

Lemma 5.5.2. *Let E_0 be the equation $xy^a zy^p s \ldots \rightrightarrows zy^b xy^q t \ldots$, where $s, t \in \{x, z\}$ and $a + p \neq b + q$. Let k be an even number such that $2^{(k-4)/2} \geq 1 + |p - q|$. Let $E_0, \ldots, E_k$ be a degenerated chain of images and let E_k be $xP \rightrightarrows zQ$. Now the solutions of E_k satisfying $y \neq 1$ are also solutions of the equation $xy^a zy^b \rightrightarrows zy^b xy^a$.*

Proof. Assume that E_{i+1} is the image of E_i under the morphism $f_i : x \mapsto (zy^b)^{c_i}x$ when i is even, and under the morphism $f_i : z \mapsto (xy^a)^{c_i}z$ when i is odd. Because $f_0(x)$ and $f_0(z)$ and thus $f_0(s)$ and $f_0(t)$ begin with z, the equation E_k is of the form

$$xy^a zy^p r \cdots \rightrightarrows zy^b xy^q r \cdots , \tag{5.18}$$

where

$$r = (f_k \circ \cdots \circ f_1)(z) = (f_k \circ \cdots \circ f_4)((((xy^a)^{c_3}zy^b)^{c_2}xy^a)^{c_1}(xy^a)^{c_3}).$$

Let $F_m = f_m \circ \cdots \circ f_4$. The words xy^a and zy^b occur as factors of $F_4(xy^a)$ at least once, and if they occur as factors of $F_m(xy^a)$ at least $2^{(m-4)/2}$ times, they occur as factors of $F_{m+2}(xy^a)$ at least $2^{(m-2)/2}$ times. Thus, by induction, they occur as factors of $F_k(xy^a)$ at least $2^{(k-4)/2}$ times. If h is a solution of E_k, then

$$\begin{aligned} ||h(xy^a zy^p)| - |h(zy^b xy^q)|| &\leq |a+p-b-q||h(y)| \\ \leq (a+b)|h(y)| + |p-q||h(y)| &\leq (1+|p-q|)|h(xy^a zy^b)| \\ \leq 2^{(k-4)/2}|h(xy^a zy^b)| &\leq |h(F_k(xy^a))|. \end{aligned}$$

Thus, by (5.18),

$$w((u^{c_3}v)^{c_2}u)^{c_1}u^{c_3}\cdots = ((u^{c_3}v)^{c_2}u)^{c_1}u^{c_3}\cdots ,$$

where $u = h(F_k(xy^a))$, $v = h(F_k(zy^b))$ and $|w| \leq |u|$. If $w = \varepsilon$, then $h(xy^a zy^p) = h(zy^b xy^q)$, which is not possible by the assumptions $h(y) \neq \varepsilon$ and $a + p \neq b + q$. Thus it follows from Lemma 5.5.1 that $uv = vu$. It can be seen that $u, v \in \{h(xy^a), h(zy^b)\}^*$, u ends with $h(xy^a)$ and v ends with $h(zy^b)$. This means that $h(xy^a)$ and $h(zy^b)$ satisfy a nontrivial relation. It follows that they commute, that is $h(xy^a zy^b) = h(zy^b xy^a)$. □

The equations $E_1, \ldots, E_n$ form a *neighborhood* of an equation E if one of the following conditions holds:

N1. $E_1, \ldots, E_n$ form a complete set of θ-images of E,

N2. E reduces to $E_1, \ldots, E_n$ with an n-tuple of substitutions,

N3. E is the equation $U = V$, U and V begin with different letters, $n = 2$, and E_1 and E_2 are equations $U \rightrightarrows V$ and $V \rightrightarrows U$,

N4. $n = 1$ and E is the equation $U = V$ and E_1 is the equation $U^R = V^R$,

N5. E is the equation $SU = TV$, $|S|_t = |T|_t$ for all $t \in \Xi$, $n = 1$ and E_1 is the equation $US = VT$,

N6. $n = 1$ and E_1 is E reduced from the left or multiplied from the right,

N7. $n = 1$ and, with the assumptions of Lemma 5.5.2, E is the equation $xP \rightrightarrows zQ$ and E_1 the equation $xy^a zy^b xP \rightrightarrows zy^b xy^a zQ$.

Rules N1 and N2 will be the most important ones. Rule N3 makes it possible to consider one-sided equations. Because of rule N6, it can be assumed that equations are reduced from the left and continue sufficiently far to the right. The other rules are used in some special cases. The next theorem justifies the definition of a neighborhood.

Theorem 5.5.3. *Let E be a word equation of length n and let $E_1, \ldots, E_m$ be its neighborhood. If each E_i has a parametric solution of length at most c, then E has a parametric solution of length $O(mn^{26})c$.*

Compared to the parametric solutions of the equations E_i, the parametric words in the parametric solution of E contain $O(1)$ new numerical parameters, the height of the parametric words can increase by $O(1)$, and the coefficients of the linear Diophantine relations are of the same size.

Proof. For N1 this follows from Theorem 5.4.12, for N2 from Theorem 5.4.1 and for N7 from Lemma 5.5.2. The other cases are clear.

The second paragraph can be deduced by examining the rules in the definition of a neighborhood and, most importantly, the definition of a complete set of θ-images. □

Directed acyclic graph whose vertices are equations is a *tree* of E if the following conditions hold:

(i) only vertex with no incoming edges is E,

(ii) all other vertices have exactly one incoming edge,

(iii) if there are edges from E_0 to exactly $E_1, \ldots, E_n$, then these equations form a neighborhood of E_0.

Theorem 5.5.4. *Let E be a word equation of length n. If E has a tree of height k, then all equations in the tree are of length $O(n)^{27^k}$. If each leaf equation in this tree has a parametric solution of length at most c, then E has a parametric solution of length $O(n)^{52 \cdot 27^k} c$.*

If the leaf equations are basic equations, then the parametric words in the parametric solution of E contain $O(k)$ numerical parameters, their height is $O(k)$, and the coefficients of the linear Diophantine relations are of size $O(n)^{27^k}$.

Proof. In the case N1 the first claim follows directly from Theorem 5.4.11, and for the other cases the bound $O(n)^{27^k}$ is more than enough. Now,

by Theorem 5.5.3, there exists a constant a such that E has a parametric solution of length

$$\begin{aligned}&a(an)^{52}\cdot a((an)^{27})^{52}\cdot a((an)^{27^2})^{52}\cdot\cdots\cdot a((an)^{27^{k-1}})^{52}\cdot c\\ <&a^k(an)^{52\cdot 27^k}c=O(n)^{52\cdot 27^k}c.\end{aligned}$$

The second paragraph follows from the second paragraph of Theorem 5.5.3 and from Theorem 5.3.2. □

A tree in which all leaves are basic equations is a *basic tree.*

If every θ-image of an equation of type I or II has a basic tree, then the equation has a basic tree, because it has a complete set of θ-images. The rule N1 is used in this way instead of explicitly selecting some complete set of θ-images.

The main theorem is proved by a sequence of lemmas. The lemmas are proved by using the rules of the definition of a neighborhood in various ways.

Lemma 5.5.5. *The equation* $xyz^2A[x,y,z]=yz^2xB[x,y,z]$ *has a basic tree.*

Proof. With N5 we get the equation $Axyz^2=Byz^2x$, and then with N4 the equation $z^2yxA^R=xz^2yB^R$. With N3 we get $z^2yxA^R\rightrightarrows xz^2yB^R$ and $z^2yxA^R\lleftarrow xz^2yB^R$. The former is basic of the form B2. The latter is reduced by the pair of substitutions $x\mapsto zx$, $x\mapsto z^2x$ to the equations $zyzx\cdots\rightrightarrows xz^2y\cdots$ and $yz^2x\cdots=xz^2y\cdots$. These are basic of the form B9 or B7 and we get a basic tree. □

Lemma 5.5.6. *Every nondegenerated θ-image of the equation*

$$xyztA[x,y,z]\rightrightarrows zx^2yB[x,y,z],$$

where $t\neq z$, has a basic tree.

Proof. The equation is of type II. Its nondegenerated images of the second order are

$$yxzg(h(tA))\rightrightarrows zx((xy)^jxz)^ixyg(h(B)),\tag{5.19}$$

where h is the morphism $x\mapsto z^ix$ and g is the morphism $z\mapsto(xy)^jxz$. The nondegenerated θ-images are the images of (5.19). We consider the cases $j=0$ and $j>0$.

First, let $j=0$. The images of (5.19) are

$$yxzC[(xz)^ix,D^kD_1y,xz]\lleftarrow D_2D_1yB[(xz)^ix,D^kD_1y,xz],\tag{5.20}$$

where $C=tA$, $D=D_1D_2=zx(xz)^ix$, $D\neq D_1$ and $D^kD_1\neq\varepsilon$. If D_2D_1 begins with one of x^2, xzx, zxz, then (5.20) is a basic equation. Otherwise D_2D_1 begins with zx^2, $D_1=1$ and $k>0$. Then (5.20) is

$$yxzC[(xz)^ix,D^ky,xz]\lleftarrow zx(xz)^ixyB[(xz)^ix,D^ky,xz].$$

This is reduced by the substitution $z \mapsto yz$ to the equation

$$xyzC[(xyz)^i x, E^k y, xyz] = zx(xyz)^i xyB[(xyz)^i x, E^k y, xyz],$$

where $E = yzx(xyz)^i x$. This is equivalent to one of the following pairs of equations:

(a) $xyzx = zxxy$ and $y\cdots = z\cdots$, if $t = x$,

(b) $xyzyzxx = zxxyzxy$ and $y\cdots = z\cdots$, if $t = y$ and $i > 1$,

(c) $xyzyzxx = zxxyzxy$ and $y\cdots = x\cdots$, if $t = y$, $i = 1$ and $y \not\leq B$,

(d) $xyzyzxx = zxxyzxy$ and $(yzx)y\cdots = (yzx)x\cdots$, if $t = y$, $i = 1$ and $y \leq B$.

By Corollary 2.3.7, there are only trivial solutions in all cases.

Next, let $j > 0$. If $t = x$, then (5.19) is

$$yxz((xy)^j xz)^i xg(h(A)) \rightrightarrows zx((xy)^j xz)^i xyg(h(B)).$$

This is equivalent to the pair of equations $yxzx \rightrightarrows zxxy$, $y\cdots = x\cdots$ and has only trivial solutions by Corollary 2.3.7. If $t = y$, then (5.19) is

$$yxzy\cdots \rightrightarrows zxxy(xy)^{j-1}xzxy\cdots .$$

Every image of this equation is of one of the following forms:

$$yx\cdots \leftleftarrows x^2\cdots , \qquad yxz\cdots \leftleftarrows xzx\cdots , \qquad yxzzx^2s\cdots \leftleftarrows zx^2yxzx\cdots ,$$

where $s \neq x$. The first two equations are basic of the form B2 and B3. The third equation is equivalent to the pair of equations $yxzzx^2 \leftleftarrows zx^2yxz$, $s\cdots = x\cdots$ and has only trivial solutions by Corollary 2.3.7. □

5.6 Supporting Equations

We define supporting equations and prove as an intermediate result that they have basic trees.

Let $1 \leq a, b \leq 2$, $d \geq 1$ and $t \neq y$. A *supporting equation* is an equation of the form

$$x^a y^b t\cdots \rightrightarrows zyx\cdots \qquad \text{or} \qquad x^a y^b t\cdots \rightrightarrows zxy\cdots , \tag{5.21}$$

or of the form

$$x^a y^b t\cdots \rightrightarrows z(yz)^d x\cdots . \tag{5.22}$$

A tree whose leaves are basic equations, supporting equations of the form (5.21) or equations $x^2yt\cdots \rightrightarrows zyzxy\cdots$, where $t \neq y$, is a *supporting tree*.

Lemma 5.6.1. *Let $E_0, \ldots, E_3$ be a chain of images of the equation*

$$E_0 : xy^a tA[x, y, z] \rightrightarrows z^c xB[x, z]yC[x, y, z],$$

where $a, c \geq 1$, $A, C \neq \varepsilon$ and $t \neq y$. Assume first that E_2 is a degenerated image. Now

1. *E_2 is of the form $xy^a z \cdots \rightrightarrows zx \cdots$;*
2. *if $a = 2$, $c = 1$, $B = \varepsilon$ and $y \not\leq C$, then E_2 is of the form $xy^2 z \cdots \rightrightarrows zxyx \cdots$;*
3. *if $a = 2$, $c = 1$ and $B = x$, then E_2 is of the form $xy^2 z \cdots \rightrightarrows zx^2 y \cdots$;*
4. *if $a = 1$, then E_2 is basic equation B3 or of the form $xyzs \cdots \rightrightarrows zx^2 y \cdots$, where $s \neq z$.*

Assume then that E_2 is a nondegenerated image. Now

1. *E_3 is a supporting equation;*
2. *if $a = 2$, $c = 1$, $B = \varepsilon$ and $y \not\leq C$, then E_3 is a basic equation or of the form $yxzy \cdots \rightrightarrows zxzy \cdots$;*
3. *if $a = 2$, $c = 1$ and $B = x$, then E_3 is a supporting equation of the form (5.21) or an equation of the form $x^2 ys \cdots \rightrightarrows zyzxy \cdots$, where $s \neq y$;*
4. *if $a = 1$, then E_3 is a supporting equation of the form (5.21).*

Proof. The equation E_1 is of the form

$$xy^a zA_1[x, y, z] \leftleftarrows z^c xB[z^i x, z]yC[z^i x, y, z],$$

where $i > 0$ and $A_1 \neq \varepsilon$. Its image E_2 is of the form

$$D_2 D_1 zA_2[x, y, z] \rightrightarrows zh(z^{c-1} xB[z^i x, z]yC[z^i x, y, z]),$$

where h is the morphism $z \mapsto (xy^a)^j D_1 z$, $j \geq 0$, $D_1 < xy^a$, $(xy^a)^j D_1 \neq \varepsilon$, $D_1 D_2 = xy^a$ and $z \not\leq A_2 \neq \varepsilon$.

The equation E_2 is a degenerated image if and only if $D_1 = \varepsilon$. Then the first four claims are correct.

If $D_1 \neq \varepsilon$, then E_2 is of the form

$$y^{a-b} xy^b z \cdots \rightrightarrows B_1[x, z]y \cdots,$$

where $0 \leq b < a$, B_1 starts with z and neither z^2 nor x^3 is a factor of B_1. Now E_3 is a supporting equation.

If $D_1 \neq \varepsilon$, $a = 2$, $c = 1$, $B = \varepsilon$ and $y \not\leq C$, then E_2 is of the form

$$y^{2-b}xy^bz\cdots \rightrightarrows zxyx\cdots$$

and E_3 is a basic equation of the form B3 or B4 or an equation of the form $yzyx\cdots \leftleftarrows xzyx\cdots$.

If $D_1 \neq \varepsilon$, $a = 2$, $c = 1$ and $B = x$, then E_2 is of the form

$$y^{2-b}xy^bz\cdots \rightrightarrows zx(xz)^ixy\cdots$$

and E_3 is a supporting equation of the form (5.21) or an equation of the form $yzyxz\cdots \leftleftarrows x^2zs\cdots$, where $s \neq z$.

If $D_1 \neq \varepsilon$ and $a = 1$, then E_2 is of the form

$$yxz\cdots \rightrightarrows B_1[x,z]y\cdots$$

and E_3 is a supporting equation of the form (5.21). □

Lemma 5.6.2. *Let $s, t \neq y$. Every nondegenerated θ-image of the equation $xy^2s\cdots \rightrightarrows zxyt\cdots$ has a basic tree. Every nondegenerated θ-image of the equation $xy^2z\cdots \rightrightarrows zx^2y\cdots$ has a supporting tree.*

Proof. For the latter equation this follows from 3 of Lemma 5.6.1. For the former it follows from 2 of Lemma 5.6.1, because the equation $yxzy\cdots \rightrightarrows zxzy\cdots$ is reduced by the substitution $y \mapsto zy$ to the equation of Lemma 5.5.5. □

Lemma 5.6.3. *Let $s \neq x$ and $t \neq y$. Consider the equations*

(a) $xy^2z\cdots \rightrightarrows zx^2y\cdots$,

(b) $xyzs\cdots \rightrightarrows zx^2y\cdots$,

(c) $xy^2z\cdots \rightrightarrows zxyt\cdots$,

(d) $xyzt\cdots \rightrightarrows zy^2x\cdots$,

(e) $xyz\cdots \rightrightarrows zy^2x\cdots$.

the first has a supporting tree and the others have basic trees.

Proof. Let E_0 be one of (a)–(d). It can be written in the form $xy^azy^pu\cdots \rightrightarrows zy^bxy^qv\cdots$, where $u, v \neq y$. Here always $a + p \neq b + q$. Let

$$l \geq 4 + 2\log_2(1 + |p - q|)$$

be even. Form a complete set of θ-images for E_0, a complete set of θ-images for each of these, and so on l times. These θ-images form chains $E_0, \ldots, E_l$.

We show that each chain has an equation with the required tree. This proves the lemma.

First we consider chains of degenerated θ-images. There is a corresponding degenerated chain of ordinary images. Now, by N7, the equation E_l can be replaced by one of the following:

(a') $xy^2z\cdots \rightrightarrows zxy^2\cdots$

(b') $xyz\cdots \rightrightarrows zxy\cdots$

(c') $xy^2z\cdots \rightrightarrows zxy^2\cdots$

(d') $xyzy\cdots \rightrightarrows zy^2x\cdots$.

Equation (b') is basic of the form B3. Equations (a') and (c') are reduced by the substitution $x \mapsto zx$ to equations of Lemma 5.5.5. Equation (d') is reduced to the equation $xyzyP = y^2zxQ$, which can be transformed to $yzyx\cdots = xzy^2\cdots$ by N5 and N4. This has a basic tree by Lemma 5.5.5.

Then we consider nondegenerated chains. We assume that the part $E_0, \ldots, E_{j-1}$ of the chain is degenerated and that E_j is a nondegenerated θ-image of E_{j-1}. If E_0 is of the form (a)–(c), then E_{j-1} is of the same form and E_j has the required tree by Lemma 5.6.2 or Lemma 5.5.6. If E_0 is of the form (d), then all of $E_0, \ldots, E_{j-1}$ are of type I. Let $0 \leq i < j$. If i is even, then E_i is of the form $xyz\cdots \rightrightarrows zy^2x\cdots$. If i is odd, then E_i is of the form $zy^2x\cdots \rightrightarrows xyz\cdots$. Assume first that j is even. Now E_j is of the form $yxzr\cdots \rightrightarrows zy^2x\cdots$, where $r \neq z$. This is the equation (b). Assume then that j is odd. Now E_j is of the form $y^2\cdots \rightrightarrows xy\cdots$ or $yzy\cdots \rightrightarrows xyz\cdots$. These are basic of the form B2 and B3.

The lemma has been proved for equations (a)–(d). The equation (e) is of the form (d) or (d'), so it has a basic tree. □

Lemma 5.6.4. *Supporting equations of the form* (5.21) *have basic trees.*

Proof. First, consider the equation $x^ay^bt\cdots \rightrightarrows zyx\cdots$, where $1 \leq a, b \leq 2$ and $t \neq y$. If $a = b = 1$, then this is basic of the form B4. If $a = 1$ and $b = 2$, then this is of type I and its images are of the form $zyx\cdots \rightrightarrows xy^2z\cdots$ or $yzxs\cdots \rightrightarrows xy^2z\cdots$, where $s \neq x$. These have basic trees by Lemma 5.6.3. Assume that $a = 2$. The equation is reduced by the substitutions $x \mapsto zx$, $x \mapsto zyx$ to the equations $xzxy\cdots \leftleftarrows yzxs\cdots$ and $xzy\cdots = zyx\cdots$, where $s \neq x$. The latter is basic of the form B5. If in the former $s = y$, then it is reduced by the substitution $y \mapsto xy$ to the equation of Lemma 5.5.5. If $s = z$, then the images of the equation are of the form $yzxz\cdots \leftleftarrows Dy$, where D is a conjugate of xzx. If $D = xzx$, then this image is basic of the form B4. If $D = zx^2$, then it is the equation (c) of Lemma 5.6.3. If $D = x^2z$, then it is reduced by the pair of substitutions $x \mapsto yx$ and $x \mapsto yzx$ to the

equations $zyxz\cdots \rightrightarrows xyxzy\cdots$ and $yzxz\cdots = xyzxzy\cdots$. The latter is basic of the form B5 and the former is reduced by the substitution $z \mapsto xz$ to the equation $zyx^2z\cdots = yx^2zy\cdots$, which has a basic tree by Lemma 5.5.5.

Second, consider the equation $x^ay^bt\cdots \rightrightarrows zxy\cdots$, where $1 \leq a, b \leq 2$ and $t \neq y$. If $a = 2$ or $a = b = 1$, then this is basic of the form B2 or B3. Assume that $a = 1$ and $b = 2$. If the fourth letter on the right is y, then the equation is reduced by the substitution $x \mapsto zx$ to the equation of Lemma 5.5.5. Otherwise, the θ-images of the equation are, by 2 of Lemma 5.6.1, basic equations or of the form $xy^2z\cdots \rightrightarrows zxyx\cdots$ or $yxzy\cdots \rightrightarrows zxzy\cdots$. The former has a basic tree by Lemma 5.6.3, the latter is reduced by the substitution $y \mapsto zy$ to the equation of Lemma 5.5.5. □

Lemma 5.6.5. *The equation $x^2yt\cdots \rightrightarrows zyzxy\cdots$, where $t \neq y$, has a basic tree.*

Proof. The images of this equation have basic trees by Lemma 5.6.4, except for the image under the morphism $x \mapsto zx$:

$$xzxyz\cdots \leftleftarrows yz^2xy\cdots .$$

This is reduced by the substitution $y \mapsto xy$ to the equation $zx^2yz\cdots = yz^2x^2\cdots$. Consider the corresponding one-sided equations.

The images of the equation

$$zx^2yz\cdots \rightrightarrows yz^2x^2\cdots \tag{5.23}$$

are of the form $zx^2yy^iz\cdots \leftleftarrows yzy^izx^2\cdots$, and the images of this under the morphisms $y \mapsto zy$, $y \mapsto zxy$, $y \mapsto zx^2y$ and under other morphisms are

$$x^2(zy)^{i+1}\cdots \rightrightarrows yz(zy)^izx^2\cdots , \tag{5.24}$$

$$xzx\cdots \rightrightarrows yz^2x\cdots , \tag{5.25}$$

$$zx^2yzx\cdots \rightrightarrows yz^2x^2y\cdots , \tag{5.26}$$

$$Dyzx\cdots \rightrightarrows yz^2x^2z\cdots , \tag{5.27}$$

where D is a conjugate of zx^2. The last two can be split into pairs of equations $zx^2yz \rightrightarrows yz^2x^2$, $x\cdots = y\cdots$ and $Dyz \rightrightarrows yz^2x^2$, $x\cdots = z\cdots$. These have only trivial solutions by Corollary 2.3.7. Consider the first two equations. They are nondegenerated images, so their images are θ-images of (5.23). These are equations of Lemma 5.6.4, except for the image of (5.24) under the morphism $x \mapsto yx$:

$$xyx(zy)^{i+1}\cdots \leftleftarrows z^2(yz)^i(yx)^2\cdots .$$

All images of this are again equations of Lemma 5.6.4, except for the image under the morphism $z \mapsto xz$:

$$yx(xzy)^{i+1}\cdots \rightrightarrows zxz(yxz)^i(yx)^2\cdots .$$

This is reduced to the equation

$$yx(xz^2y)^{i+1}\cdots = xz(zyxz)^i(zyx)^2\cdots ,$$

which can be split into the pair of equations $yx^2z^2 = xz^2yx$, $y\cdots = z\cdots$, which has only trivial solutions by Corollary 2.3.7. So (5.23) has a basic tree.

The images of the equation $zx^2yz\cdots \leftleftarrows yz^2x^2\cdots$ are of the following forms:

$$x^2zyz\cdots \rightrightarrows yz^2x^2\cdots , \tag{5.28}$$

$$xzxyz\cdots \rightrightarrows yz^2x^2\cdots , \tag{5.29}$$

$$zx^2yz\cdots \rightrightarrows yz^2x^2\cdots . \tag{5.30}$$

The images of (5.29) are equations of Lemma 5.6.4. The equation (5.30) is of the form (5.23). The images of (5.28) are equations of Lemma 5.6.4, except for the image under the morphism $x \mapsto yx$:

$$xyxzyz\cdots \leftleftarrows z^2(yx)^2\cdots .$$

All images of this are again equations of Lemma 5.6.4, except for the image under the morphism $z \mapsto xz$:

$$yx^2zyxz\cdots \rightrightarrows zxz(yx)^2\cdots .$$

This is reduced to the equation

$$yx^2z^2yxz\cdots = xz(zyx)^2\cdots ,$$

which can be split into the pair of equations $yx^2z^2 = xz^2yx$, $y\cdots = z\cdots$, which has only trivial solutions by Corollary 2.3.7. □

Lemmas 5.6.4 and 5.6.5 prove that if an equation has a supporting tree, then it has a basic tree.

Theorem 5.6.6. *Every supporting equation has a basic tree.*

Proof. By Lemma 5.6.4, it is enough to consider equations (5.22).

If $a = b = 1$, then the equation is $xyt\cdots \rightrightarrows z(yz)^dx\cdots$. Every image of this equation has a basic tree by Lemma 5.6.4.

If $a = 1$ and $b = 2$, then the equation is $xy^2t\cdots \rightrightarrows z(yz)^dx\cdots$. Its images are of the forms

$$xy^2z\cdots \leftleftarrows zyz\cdots, \tag{5.31}$$
$$xy^2z\cdots \leftleftarrows yzy\cdots, \tag{5.32}$$
$$xy^2z\cdots \leftleftarrows yz^2s\cdots, \tag{5.33}$$
$$xy^2z\cdots \leftleftarrows z^2yt\cdots, \tag{5.34}$$

where $s \neq z$, $t \neq y$. All images of (5.31) are of the form (5.21). All θ-images of (5.32) are, by 4 of Lemma 5.6.1, basic equations, supporting equations (5.21), or equations (b) of Lemma 5.6.3. All θ-images of (5.32) are, by 3 of Lemma 5.6.1, equations of Lemmas 5.6.3, 5.6.4 or 5.6.5. All images of (5.34) are supporting equations (5.21), except for the image under the morphism $z \mapsto xz$, which is the equation of Lemma 5.6.5.

If $a = 2$, then the equation is $x^2y^bt\cdots \rightrightarrows z(yz)^dx\cdots$. Its images are supporting equations (5.21), except for the image under the morphism $x \mapsto zx$:

$$xzxy\cdots \rightrightarrows (yz)^dzx\cdots.$$

If $d \geq 1$, then the images of this equation are supporting equations (5.21). If $d = 1$, then this is the equation (5.22) with $a = 1$ and $b = 2$. □

5.7 Main Theorem

Now we can start generalizing Theorem 5.6.6 and finally prove the main results of this chapter.

Lemma 5.7.1. *The equation $xy^azy^ps\cdots \rightrightarrows zy^bxy^qt\cdots$, where $a > 0$, $a + p = b + q$ and $s, t \neq y$, has a basic tree.*

Proof. If $a = 1$ and $b = 0$, then the equation is basic of the form B8. Consider other cases. The equation is reduced by the substitutions $x \mapsto zy^cx$ $(c = 0, \ldots, b)$ to the equations

$$xy^az\cdots \leftleftarrows y^{b-c}zy^cx\cdots \qquad (c = 0, \ldots, b-1), \tag{5.35}$$
$$xy^azy^psP = zy^bxy^qtQ. \tag{5.36}$$

When $b - c > 1$, (5.35) is basic of the form B2. When $b - c = 1$, its θ-images have a basic tree by 4 of Lemma 5.6.1 and by Lemmas 5.6.3 and 5.6.4.

If $a = b$, then (5.36) is basic of the form B6 or B7. Assume that $a < b$ (the case $a > b$ is similar). By using N5 and N4 we get the equation $y^dzy^ax\cdots = xy^bz\cdots$, where $d = b - a \geq 1$. Split this into one-sided

equations

$$y^d z y^a x \cdots \rightrightarrows x y^b z \cdots , \tag{5.37}$$

$$y^d z y^a x \cdots \leftleftarrows x y^b z \cdots . \tag{5.38}$$

If $d > 1$, then (5.37) is basic of the form B2. If $d = 1$, then its θ-images have a basic tree by 4 of Lemma 5.6.1 and by Lemmas 5.6.3 and 5.6.4. Equation (5.38) is reduced by the substitutions $x \mapsto y^c x$ $(c = 1, \ldots, d)$ to the equations

$$y^{d-c} z y^{a+c} x \cdots \rightrightarrows x y^b z \cdots \qquad \text{and} \qquad z y^{a+d} x \cdots = x y^b z \cdots .$$

Latter is basic of the form B6 or B7, former is of the form (5.37). □

Lemma 5.7.2. *The equation $E_0 : x y^a z \cdots \rightrightarrows z y^b x \cdots$, where $a > 0$, has a basic tree.*

Proof. The equation can be written in the form $x y^a z y^p u \cdots \rightrightarrows z y^b x y^q v \cdots$, where $u, v \neq y$. If $a + p = b + q$, then the claim follows from Lemma 5.7.1. Assume that $a + p \neq b + q$. Let $l \geq 4 + 2 \log_2(1 + |p - q|)$ be even. Like in Lemma 5.6.3, form a complete set of θ-images of E_0, a complete set of θ-images of these, and so on l times. These θ-images form chains $E_0, \ldots, E_l$. We show that each chain has an equation with a basic tree; this proves the claim.

First we consider chains of degenerated θ-images. There is a corresponding chain of ordinary images and we can use the rule N7. The equation E_l is replaced by the equation $x y^a z y^b x P \rightrightarrows z y^b x y^a z Q$, which has a basic tree by Lemma 5.7.1.

Then we consider nondegenerated chains. We assume that the part $E_0, \ldots, E_{j-1}$ of the chain is degenerated and that E_j is a nondegenerated θ-image of E_{j-1}. If $b = 0$, the equation E_0 is of the form $x y^a z \cdots \rightrightarrows z x \cdots$, and E_{j-1} is of the same form. Now by 1 of Lemma 5.6.1, E_j is a supporting equation and thus has a basic tree. If $b > 0$, then E_0 is of the form $x y^a z \cdots \rightrightarrows z y^b x \cdots$. Equation E_{j-1} is of the same form. Now E_j is of the form $y^c z y^d x \cdots \rightrightarrows x y^a z \cdots$, where $c + d = a$ and $c \geq 1$. If $c > 1$, then E_j is basic of the form B2. If $c = 1$, then E_j has a basic tree by 4 of Lemma 5.6.1 and by Lemmas 5.6.3 and 5.6.4. □

Lemma 5.7.3. *The equation $x y^a t \cdots \rightrightarrows z^c x B[x, z] y \cdots$, where $a, c \geq 1$ and $t \neq y$, has a basic tree.*

Proof. By 1 of Lemma 5.6.1, all θ-images of this equation are supporting equations or equations of Lemma 5.7.2. □

Lemma 5.7.4. *The equation $x^n y^m t \cdots \rightrightarrows z y A[y, z] x \cdots$, where $n, m \geq 1$ and $t \neq y$, has a basic tree.*

Proof. If $n = 1$, every image of the equation is of the form

$$xy^m z \cdots \leftleftarrows Dx \cdots , \tag{5.39}$$

where D is a conjugate of zyA. If $n > 1$, the image of the equation under the morphism $x \mapsto zx$ is

$$x(zx)^{n-1} y \cdots \leftleftarrows yAzx \cdots , \tag{5.40}$$

and all the other images are of the form

$$xzy \cdots \leftleftarrows Dx \cdots , \tag{5.41}$$

where D is a conjugate of zyA.

Consider (5.39). If $y^2 \leq D$, then this is basic of the form B2. If $yz \leq D$, then this is the equation of Lemma 5.7.3. If $z \leq D$, then this is of the form

$$x^a y^b s \cdots \rightrightarrows zy^d x \cdots , \tag{5.42}$$

where $a, b, d \geq 1$ and $s \neq y$. The case of (5.41) is similar. Equation (5.40) is of the form

$$x^a y^b s \cdots \rightrightarrows z(yz)^d x \cdots , \tag{5.43}$$

where $a, b, d \geq 1$ and $s \neq y$. It is enough to prove that (5.42) and (5.43) have basic trees.

Consider (5.42). Assume first that $a = 1$. Every image of this equation is of the form $xy^b z \cdots \leftleftarrows Dx$, where D is a conjugate of zy^d. If $z \leq D$, then this is the equation of Lemma 5.7.2. If $y^2 \leq D$, this is basic of the form B2. If $yz \leq D$, then this is the equation of Lemma 5.7.3. Assume then that $a > 1$. The image of the equation under the morphism $x \mapsto zx$ is $x(zx)^{a-1} y \cdots \leftleftarrows y^d zx \cdots$, and all other images are of the form $xzy \cdots \leftleftarrows Dx \cdots$, where D is a conjugate of zy^d. First of these is of type I and its images have basic trees by Theorem 5.6.6. The latter is the equation of Lemma 5.7.3 if $zy \leq D$; otherwise its images have basic trees by Theorem 5.6.6.

Consider (5.43). Assume first that $a = 1$. Every image of this equation is of the form $xy^b z \cdots \leftleftarrows Dx$, where D is a conjugate of $z(yz)^d$. If $yz \leq D$, this is the equation of Lemma 5.7.3. Otherwise $Dx = z^c ys$, where $1 \leq c \leq 2$ and $s \neq y$, and this image is of the form (5.42) and has a basic tree. Assume then that $a > 1$. The image of the equation under the morphism $x \mapsto zx$ is $x(zx)^{a-1} y \cdots \leftleftarrows (yz)^d zx \cdots$, and all other images are of the form $xzy \cdots \leftleftarrows Dx \cdots$, where D is a conjugate of $z(yz)^d$. First of these goes back to the case $a = 1$. The latter has a basic tree by Lemma 5.6.4. □

Theorem 5.7.5. *Every equation of length n with three unknowns has a basic tree of height $O(\log n)$.*

Proof. The trivial equation $U = U$ is a basic equation. All other equations can be reduced from the left and split into one-sided equations. By multiplication from the right, every one-sided equation can be turned into one of the equations

$$x^2 \cdots \rightrightarrows y^c x \cdots \tag{5.44}$$
$$xy \cdots \rightrightarrows y^c x \cdots \tag{5.45}$$
$$xz^a t \cdots \rightrightarrows y^c x B[x, y] z \cdots \tag{5.46}$$
$$x^a y^b s \cdots \rightrightarrows y^c z B[y, z] x \cdots \tag{5.47}$$
$$x^a z^b t \cdots \rightrightarrows y z B[y, z] x \cdots \tag{5.48}$$
$$x^a z^b t \cdots \rightrightarrows y^d z B[y, z] x \cdots , \tag{5.49}$$

where $a, b, c \geq 1$, $d > 1$, $t \neq z$ and $s \neq y$. We prove that all of these have basic trees.

Equation (5.44) is basic of the form B2. Equation (5.45) is reduced by the substitution $x \mapsto yx$ to the equation $xy \cdots = y^c x \cdots$, which is basic of the form B1. Equation (5.46) is the equation of Lemma 5.7.3. Equation (5.48) is the equation of Lemma 5.7.4.

Equation (5.47) is of type I and its images are of the form $xy \cdots \leftleftarrows Dx \cdots$, where D is a conjugate of $y^c z B$. If $y^2 \leq D$, then this is of the form (5.44), if $yz \leq D$, then of the form (5.46), and if $z \leq D$, then of the form (5.48). So every image of (5.47) and thus the equation itself has a basic tree.

Also (5.49) is of type I and its images are of the form $x(y \cdots)^{a-1} z^b y \cdots \leftleftarrows Dx \cdots$, where D is a conjugate of $y^d z B$. Again these are of the form (5.44), (5.46) or (5.48). So every image of (5.49) and thus the equation itself has a basic tree.

The constructions of trees in the lemmas produce trees of bounded height with two exceptions: Lemmas 5.6.3 and 5.7.2, where a tree with height of order $\log(1 + |p - q|)$ is constructed for the equation

$$xy^a zy^p \cdots \rightrightarrows zy^b xy^q \cdots . \tag{5.50}$$

We prove that the powers of y here cannot be more than n, which proves this theorem. In the definition of neighborhood, the rules N1, N2, N5 and N6 can produce higher powers than those in the initial equation. There is no need to use N6 to generate high powers and N5 is only used in Lemmas 5.5.5, 5.6.3 and 5.7.1, where it does not generate high powers. N2 is not used to generate higher powers than those that are already in the equation. Consider N1. Here an equation $xU(x, y, z) \rightrightarrows y^a xV(x, y, z)$ can be turned into $xU(y^i x, y, z) \leftleftarrows y^a xV(y^i x, y, z)$ for high values of i. But in order for y to be in the position of (5.50), the rules N1 or N2 must be used again. Then

y is replaced by xuy for some $u \in \{x, z\}^*$ and the powers of y disappear. The claim is proved. □

Theorem 5.7.6. *Every equation of length n with three unknowns has a parametric solution of length* $\exp(n^{O(1)})$.

Proof. By Theorem 5.7.5, every equation has a basic tree of height $O(\log n)$. By Theorem 5.3.2, the leaf equations have parametric solutions of bounded length. Now from Theorem 5.5.4 it follows that E has a parametric solution of length $O(n)^{52 \cdot 27^k}$, where $k = O(\log n)$, that is of length $\exp(n^{O(1)})$. □

It can be noted that two word parameters are sufficient, except for the trivial equation $U = U$: this is true for basic equations, and no new word parameters are added when the rules in the definition of a neighborhood are used. This could be expected because of the defect theorem. As for the numerical parameters, for basic equations their number is $O(1)$, and at each step when going up in the basic tree only $O(1)$ new parameters are added. Thus the number of different numerical parameters is $O(\log n)$. The same holds for the number of nested numerical parameters, that is the height of the parametric words.

Based on Theorem 5.7.6 we can prove that the shortest nontrivial solution is of exponential length. However, this is not trivial. For example, if we have a parametric word $(p^i q)^j$, then by giving the value 1 for the numerical parameters we get a short word, but the problem is that $i = j = 1$ does not necessarily satisfy the linear Diophantine relation. Thus we need to estimate the size of the minimal solution of the relation. We also need to make sure that the solution of the word equation is indeed nontrivial.

Theorem 5.7.7. *If an equation of length n with three unknowns has a nontrivial solution, it has a nontrivial solution of length* $\exp(n^{O(1)})$.

Proof. Consider an equation $E : x \cdots = y \cdots$ and its parametric solution

$$\{(h_j, R_j) \mid 1 \leq j \leq m\}$$

of length $\exp(n^{O(1)})$. If E has a nontrivial solution, it has a solution where x and y begin with the same letter but z begins with a different letter. Let $h \circ f \circ h_j$ be such a solution, where $h \circ f$ is a valuation. Now also $f \circ h_j$ is such a solution, and so is $g \circ h_j$ if $g \in R_j$ maps exactly the same numerical parameters to zero as f. Thus $g \circ h_j$ is a nontrivial solution. We must select g so that this solution is sufficiently short.

The lengths of the parametric words $h_j(t)$, where $t \in \{x, y, z\}$, are $\exp(n^{O(1)})$. By Theorems 5.5.4 and 5.7.5, every occurrence of a word parameter in $h_j(t)$ appears at most $g(i_1) \ldots g(i_k)$ times in $g(h_j(x))$, where $i_1, \ldots, i_k$ are the numerical parameters and $k = O(\log n)$. Thus the length of $g(h_j(t))$ is $g(i_1) \ldots g(i_k) \exp(n^{O(1)})$.

The conditions for g are that it must be in R_j and it must map exactly the same numerical parameters to zero as f. The latter condition can be handled by adding either the equation $i = 0$ (if $f(i) = 0$) or the inequality $i > 0$ (if $f(i) > 0$) to R_j for every $i \in \Lambda$. Inequalities $i > c$ can be replaced with $i = c + 1 + i'$, where i' is a new variable. In this way we get a linear Diophantine relation R'_j, which is a disjunction of linear systems of equations. Because $f \in R'_j$, at least one of these systems has a nonnegative integer solution.

According to [66], if a system of linear equations has a nonnegative integer solution, then it has one of size $O(lM)$, where l is the number of unknowns, M is an upper bound for the $r \times r$ subdeterminants of the augmented matrix of the system, and r is the rank of the system. Now r is at most $l = O(\log n)$. The system comes originally from the use of Theorem 5.3.2 on some equation in the basic tree of E. The lengths of the equations in this tree are exponential (by Theorem 5.5.4), and so are the coefficients in the system (by Theorem 5.3.2), so $M = \exp(n^{O(1)})$. Thus there is a nonnegative integer solution of size $\exp(n^{O(1)})$. This solution gives us a function g such that $g(i_1) \ldots g(i_k) \exp(n^{O(1)}) = \exp(n^{O(1)})$. This proves the theorem. □

Now we consider the satisfiability problem. Constant-free equations always have the solution where every unknown gets the value ε, and usually they have also other periodic solutions. The natural question is thus whether a constant-free equation has a nontrivial solution. This can be easily reduced to the satisfiability problem of equations with constants. In this way we get the result that the above-mentioned question is in NP for equations on three unknowns.

Theorem 5.7.8. *The existence of a nontrivial solution of a constant-free equation on three unknowns can be decided in nondeterministic polynomial time.*

Proof. The equation $xU = yV$, where $U, V \in \Xi^*$, has a nontrivial solution if and only if it has a solution $x = ax', y = ay', z = bz'$, where a and b are different letters and $x', y', z' \in \Sigma^*$. So we are interested in the existence of a solution for the equation obtained from $xU = yV$ by replacing x with ax', y with ay' and z with bz', where x', y', z' are now new unknowns. The length of this equation is twice the length of the original equation.

There is a nondeterministic algorithm (see [53]) that solves the existence of a solution for the last equation in time polynomial in $n \log N$, where n is the length of the equation and N is the length of the shortest solution. The claim now follows from Theorem 5.7.7. □

Chapter 6

Unique Decipherability in the Monoid of Languages

In this section we study unique decipherability in the monoids of unary and nonunary regular languages.

In the first three sections we are interested in problems related to unique decipherability in the monoid of unary languages. As stated in Section 2.6, this monoid is isomorphic to the additive monoid of sets of nonnegative integers. Thus we will formulate everything in terms of sets of numbers. Often we can allow the sets to contain also negative integers. We will mostly consider finite, or sometimes regular, sets.

We begin in Section 6.1 by giving the required definitions and by proving some results about powers of sets of integers. These results are related to the Frobenius problem, see e.g. [54] for a survey, [31] for a generalization for words and [19] for related algebraic results. The main result of this section is that if the elements of a set do not have a common divisor, then sufficiently large powers of the set contain almost all integers between their minimums and maximums. This result is very important in the later sections.

In Section 6.2 we consider the power equality problem. For example, we show that it is sufficient to consider powers that are of linear size with respect to the maximum of the sets. The results in this section form the basis for the solution of the unique decipherability problem, and are also of independent interest.

In Section 6.3 we give a characterization of codes in the additive monoid of finite sets of integers. In particular, we prove that a family of three sets is never a code, i.e. three sets always satisfy a nontrivial relation. We prove a similar result for certain infinite sets, including all infinite rational sets.

In Section 6.4 we prove that the unique decipherability problem is undecidable in the monoid of binary regular languages.

This chapter is based on the articles [57] and [36].

6.1 Additive Powers of a Set of Numbers

If $m, n \in \mathbb{Z}$, $k \in \mathbb{N}_0$ and $A, B \subseteq \mathbb{Z}$, then we use the following notation:

$$\begin{aligned} AB &= \{a + b \mid a \in A, b \in B\}, \\ A^k &= \{a_1 + \cdots + a_k \mid a_1, \ldots, a_k \in A\}, \\ A^* &= \bigcup_{k=0}^{\infty} A^k, \\ A + n &= \{a + n \mid a \in A\}, \\ A \cdot n &= \{an \mid a \in A\}, \\ A/n &= \{a/n \mid a \in A\}. \\ [m, n] &= \{a \in \mathbb{Z} \mid m \le a \le n\}, \\ [m, \infty) &= \{a \in \mathbb{Z} \mid a \ge m\}, \\ (-\infty, n] &= \{a \in \mathbb{Z} \mid a \le n\}, \end{aligned}$$

The first three come from the theory of formal languages, and that is why we use multiplicative notation even though we are dealing with sums.

We will often need to assume that the elements of a set do not have a common divisor, or that the minimum of a set is zero. Thus we let

$$S_n = \{A \subseteq [0, n] \mid 0, n \in A, \gcd A = 1\}.$$

If $A \in S_n$, then let $\widetilde{A} = \{n - a \mid a \in A\}$ be the "reverse" of A. Now $\widetilde{AB} = \widetilde{A}\widetilde{B}$.

Let $A = \{0, a_1, \ldots, a_r\} \subseteq \mathbb{N}_0$ and $\gcd A = 1$. It is well known that every sufficiently large integer can be represented in the form

$$a_1 x_1 + \cdots + a_r x_r, \tag{6.1}$$

where $x_1, \ldots, x_r \in \mathbb{N}_0$. The Frobenius problem asks what is the largest integer that cannot be represented in this way. This integer is called the Frobenius number of A and we denote it by $G(A)$. The numbers (6.1) form the set A^*, so $G(A)$ is the largest integer not in A^*.

We define $F_m(A)$ to be the smallest integer such that

$$A^* \cap [0, m] \subseteq A^{F_m(A)}.$$

We assume that $0 \in A$, so $A \subseteq A^2 \subseteq A^3 \subseteq \ldots$ and $F_m(A)$ exists for every m. The number $F_m(A)$ tells how large the coefficients $x_1, \ldots, x_r$ need to be: if $n \le m$ and n has a representation of the form (6.1), then n has such a representation where $x_1 + \cdots + x_r \le F_m(A)$.

There are many results concerning the size of the Frobenius number. We use the following result from [6].

Lemma 6.1.1. *If $A = \{a_0, \ldots, a_r\} \in S_n$, where $0 = a_0 < \cdots < a_r = n$, then $G(A) \le a_1 n - a_1 - n \le n^2 - 2n$.*

We also need an upper bound for $F_m(A)$.

Lemma 6.1.2. *If $A = \{a_0, \ldots, a_r\} \in S_n$, where $0 = a_0 < \cdots < a_r = n$, then $F_m(A) \le n - 1 + m/n$.*

Proof. Let $g = G(A)$ and $a \in A^* \cap [0, m]$. If $a \le g + n$, then $a \in A^k$, where $k = \lfloor (g+n)/a_1 \rfloor$. If $a > g + n$, then $a = g + i + n(a - g - i)/n$, where $i \in \{1, \ldots, n\}$ is such that $g + i \equiv a \mod n$. Now $g + i \in A^k$ and $n(a - g - i)/n \in A^l$, where again $k = \lfloor (g+n)/a_1 \rfloor$ and $l = (a - g - i)/n$, and thus $a \in A^{k+l}$. So with the help of Lemma 6.1.1 we get the result

$$F_m(A) \le k + l \le \frac{g+n}{a_1} + \frac{a-g-1}{n} \le n - 1 + \frac{m}{n}.$$

□

Next we examine the structure of A^k for large k. If $A \in S_n$, then $A^k \subseteq [0, kn]$. Because A^* contains almost every natural number, A^k contains almost every number from the interval $[0, kn]$. Only some numbers from the beginning and from the end are missing. These missing numbers will be essentially the same for all large values of k (of course the large missing numbers will be getting larger and larger as k grows). This is formalized by the following theorem.

Theorem 6.1.3. *Let $A \in S_n$ and $k \ge 2n - 2$. Let*

$$C = A^* \cap [0, n^2 - 2n] \qquad \text{and} \qquad \widetilde{D} = (\widetilde{A})^* \cap [0, n^2 - 2n].$$

Now

$$A^k = C \cup [n^2 - 2n + 1, kn - n^2 + 2n - 1] \cup (D + kn - n^2 + 2n).$$

Proof. Let $k \ge 2n - 2$. By Lemma 6.1.2,

$$F_{\lfloor kn/2 \rfloor}(A), F_{\lfloor kn/2 \rfloor}(\widetilde{A}) \le n - 1 + k/2 \le k.$$

Now we get

$$\begin{aligned} A^k \cap [0, \lfloor kn/2 \rfloor] &= A^* \cap [0, \lfloor kn/2 \rfloor] \\ &= (A^* \cap [0, n^2 - 2n]) \cup (A^* \cap [n^2 - 2n + 1, \lfloor kn/2 \rfloor]) \\ &= C \cup [n^2 - 2n + 1, \lfloor kn/2 \rfloor]. \end{aligned}$$

Here the first equality holds, because $k \ge F_{\lfloor kn/2 \rfloor}(A)$, and the last equality follows from Lemma 6.1.1. Similarly we get

$$\widetilde{A}^k \cap [0, \lfloor kn/2 \rfloor] = \widetilde{D} \cup [n^2 - 2n + 1, \lfloor kn/2 \rfloor].$$

The claim follows. □

6.2 Power Equality for Sets of Numbers

In this section we study the power equality problem, that is the problem of determining whether some powers of two finite sets $A, B \subseteq \mathbb{Z}$ are equal. The following lemma tells how this problem can be reduced to the case where $\min A = \min B = 0$.

Lemma 6.2.1. *Let* $\min A_i = m_i < \max A_i = n_i$, $A_i = B_i + m_i$, *and* $k, l > 0$. *Now* $A_1^k = A_2^l$ *if and only if* $B_1^k = B_2^l$ *and* $m_1 n_2 = m_2 n_1$.

Proof. The sets A_1^k and A_2^l are equal if and only if

$$B_1^k + km_1 = B_2^l + lm_2. \tag{6.2}$$

If the sets A_1^k and A_2^l are equal, then their minimums and maximums are equal, that is $km_1 = lm_2$ and $kn_1 = ln_2$. From this and (6.2) it follows that $B_1^k = B_2^l$ and $m_1 n_2 = m_2 n_1$.

On the other hand, if $B_1^k = B_2^l$, then

$$k(n_1 - m_1) = \max B_1^k = \max B_2^l = l(n_2 - m_2).$$

Multiplying this by $m_1 m_2$ gives

$$km_1(m_2 n_1 - m_1 m_2) = lm_2(m_1 n_2 - m_1 m_2).$$

If $m_1 n_2 = m_2 n_1$, then $km_1 = lm_2$ and (6.2) holds. □

It is clear that if $\min A = \min B = 0$, then some powers of A and B can be equal only if $\gcd A = \gcd B = d$, and if this is the case, then $A^k = B^l$ if and only if $(A/d)^k = (B/d)^l$. Thus we can assume that $d = 1$.

Example 6.2.2. Let $A^2 = B^2$. If the two smallest elements of A are 0 and a, then the two smallest elements of A^2 are also 0 and a. Thus 0 and a must also be the two smallest elements of B. Similarly the two largest elements of A and B must be the same.

The example of $A \neq B$, $A^2 = B^2$ where the largest element of A is as small as possible is $A = \{0, 1, 3, 4\}, B = \{0, 1, 2, 3, 4\}$ (or vice versa). In this case $A^2 = B^2 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$.

The sets A, B can also be selected to be maximal in the sense that they are not proper subsets of any set D such that $A^2 = D^2$. For example, if

$$A = \{0, 1, 3, 7, 8, 9\},\ B = \{0, 1, 3, 6, 8, 9\},\ C = \{0, 1, 2, 6, 8, 9\},$$

then $A^2 = B^2 = C^2 \neq D^2$ for all D such that $A \subsetneq D$, $B \subsetneq D$ or $C \subsetneq D$.

Theorem 6.2.3. *Let $m \leq n$, $A \in S_m$ and $B \in S_n$. There are $i, j > 0$ such that $A^i = B^j$ if and only if*

$$\begin{aligned} &A^k \cap [0, n^2 - 2n] = B^k \cap [0, n^2 - 2n] \quad \textit{and} \\ &\widetilde{A}^k \cap [0, n^2 - 2n] = \widetilde{B}^k \cap [0, n^2 - 2n], \end{aligned} \tag{6.3}$$

where $k = 2n - 2$.

Proof. If $A^i = B^j$, then $im = jn$ and $A^{ki} = B^{kj}$ for all k. Thus there are such i, j if and only if there are such $i, j \geq 2n - 2$. If $A^i = B^j$, where $i, j \geq 2n - 2$, then (6.3) holds by Lemma 6.1.2.

Next, assume that (6.3) holds. Consider two arbitrary integers $i, j \geq 2n - 2$ satisfying $im = jn$. Then $A^i = B^j$ by Theorem 6.1.3. □

Theorem 6.2.3 gives a condition for the existence of the required numbers i and j, and this leads to an algorithm for solving the power equality problem. The next theorem gives a similar condition, which is perhaps not as useful algorithmically, but may be easier in some other ways.

Theorem 6.2.4. *Let $A \in S_m$ and $B \in S_n$. There are $i, j > 0$ such that $A^i = B^j$ if and only if $A^* = B^*$ and $(\widetilde{A})^* = (\widetilde{B})^*$.*

Proof. If $m \leq n$, $A^i = B^j$ and $C \in \{A, B, \widetilde{A}, \widetilde{B}\}$, then

$$C^* = (C^{2n-2} \cap [0, n^2 - 2n]) \cup [n^2 - 2n + 1, \infty)$$

by Lemmas 6.1.1 and 6.1.2. Now $A^* = B^*$ and $(\widetilde{A})^* = (\widetilde{B})^*$ by Theorem 6.2.3. On the other hand, if $A^* = B^*$ and $(\widetilde{A})^* = (\widetilde{B})^*$, then (6.3) holds by Lemma 6.1.2. □

We can use Theorem 6.2.3 to prove that if $A^k = B^k$ holds for some k, then it holds for all sufficiently large k. We are not aware whether $A^k = B^k$ implies $A^{k+1} = B^{k+1}$.

Theorem 6.2.5. *If $A, B \in S_n$ and $A^k = B^k$ for some $k > 0$, then $A^k = B^k$ for all $k \geq 2n - 2$.*

Proof. If $A^k = B^k$ for some k, then by Theorem 6.2.3 equation (6.3) holds for $k = 2n - 2$, and by Lemma 6.1.2 it holds for all larger k as well. The claim now follows from Theorem 6.1.3. □

Theorem 6.2.5 raises the following question: if n is fixed, then what is the smallest number m such that if $A, B \in S_n$ and $A^k = B^k$ for some $k > 0$, then $A^k = B^k$ for all $k \geq m$? Theorem 6.2.5 proves that $m \leq 2n - 2$, and the following example proves that $m \geq n - 2$.

Example 6.2.6. Let $A = \{0, 1, n-1, n\}$. Now $A^{n-3} \neq [0, n]^{n-3}$, but $A^{n-2} = [0, n]^{n-2}$. The inequality holds, because $n - 2 \notin A^{n-3}$. The equality holds, because every element of $[0, n]^{n-2} = [0, (n-2)n]$ is of the form $an+b$, where $a \in [0, n-3]$ and $b \in [0, n]$, and if $a+b \leq n-2$, then $an+b \in \{0, 1, n\}^{n-2}$, and if $a+b > n-2$, then $an+b = (a+b-n+1)n+(n-b)(n-1) \in \{0, n-1, n\}^{n-2}$.

6.3 Unique Decipherability for Sets of Numbers

In this section we study the unique decipherability problem in the monoid of sets of integers. We defined in Section 2.6 that a subset of a semigroup is a code if the elements of the subset do not satisfy a nontrivial equation. The monoid of sets of integers is commutative, so every balanced equation is trivial. Now the definition of a code can be written as follows. A family of sets $\{A_1, \ldots, A_s\}$ is a code, or has the unique decipherability property, if there are no numbers $k_1, \ldots, k_s, l_1, \ldots, l_s$ such that $A_1^{k_1} \ldots A_s^{k_s} = A_1^{l_1} \ldots A_s^{l_s}$ and $k_i \neq l_i$ for some i.

Theorem 6.3.1. *Let $A_1, \ldots, A_s$ be finite sets of integers. Let $\min A_i = m_i$ and $\max A_i = n_i$. The sets A_i form a code if and only if $s = 1$ and $A_1 \neq \{0\}$ or $s = 2$ and $m_1 n_2 \neq m_2 n_1$.*

Proof. Let $A_1^{k_1} \ldots A_s^{k_s} = A_1^{l_1} \ldots A_s^{l_s}$. The minimums and maximums of these sets must be the same, that is

$$\begin{aligned} m_1(k_1 - l_1) + \cdots + m_s(k_s - l_s) &= 0 \quad \text{and} \\ n_1(k_1 - l_1) + \cdots + n_s(k_s - l_s) &= 0. \end{aligned}$$

This can be viewed as a pair of equations with s unknowns $k_i - l_i$ and coefficients m_i, n_i. This pair of equations has nontrivial solutions if and only if the rank of the matrix

$$\begin{pmatrix} m_1 & \ldots & m_s \\ n_1 & \ldots & n_s \end{pmatrix}$$

is smaller than s. The rank is s if and only if $s = 1$ and $A_1 \neq \{0\}$ or $s = 2$ and $m_1 n_2 \neq m_2 n_1$.

If the rank is s, then necessarily $k_i = l_i$ for all i. This means that the sets A_i form a code.

If the rank is smaller than s, then we can select the numbers k_i and l_i to be positive integers so that $k_j \neq l_j$ for some j. Let $A_i = B_i + m_i$. Let $d = \gcd(B_1 \cup \cdots \cup B_s)$ and $C_i = B_i/d$. If $D = C_1^{k_1} \ldots C_s^{k_s}$ and $E = C_1^{l_1} \ldots C_s^{l_s}$, then $D^* = (C_1 \cup \cdots \cup C_s)^* = E^*$ and $(\widetilde{D})^* = (\widetilde{C}_1 \cup \cdots \cup \widetilde{C}_s)^* = (\widetilde{E})^*$. Now from Theorem 6.2.4 it follows that $D^k = E^l$ for some k, l. Because

$$\begin{aligned} d \max D &= k_1(n_1 - m_1) + \cdots + k_s(n_s - m_s) \\ &= l_1(n_1 - m_1) + \cdots + l_s(n_s - m_s) = d \max E, \end{aligned}$$

it must be $k = l$. This means that

$$\begin{aligned}(A_1^{k_1} \dots A_s^{k_s})^k =& (B_1^{k_1} \dots B_s^{k_s})^k + k(k_1 m_1 + \cdots + k_s m_s) \\ =& D^k \cdot d + k(k_1 m_1 + \cdots + k_s m_s) \\ =& E^k \cdot d + k(l_1 m_1 + \cdots + l_s m_s) \\ =& (B_1^{l_1} \dots B_s^{l_s})^k + k(l_1 m_1 + \cdots + l_s m_s) = (A_1^{l_1} \dots A_s^{l_s})^k\end{aligned}$$

and the sets A_i do not form a code. □

A subset of a monoid is *rational* if it is obtained from finite sets by repeatedly using the operations of union, product and star. In other words, all finite sets are rational, and if A and B are rational, so are $A \cup B$, AB and A^*. In the case of the additive monoid $\mathbb{N}_0$, a set $A \subseteq \mathbb{N}_0$ is rational if and only if it is ultimately periodic, that is if there are finite sets B, C and a number n such that $A = B \cup C\{n\}^*$.

We have given a characterization of all codes in the additive monoid of finite sets of integers. Next it would be natural to study the unique decipherability problem for rational sets. We can indeed generalize Theorem 6.3.1, and the condition we need is actually weaker than rationality: some power of some set must contain an infinite rational set. The next lemma gives some equivalent conditions.

Lemma 6.3.2. *Let $A \subseteq \mathbb{Z}$ be infinite and $\min A > -\infty$. The following are equivalent:*

(i) A^k contains an infinite rational set for some k,

(ii) A^k contains an infinite arithmetic progression for some k,

If these conditions hold and $0 \in A$, then $\{\gcd A\}^* \setminus A^k$ is finite for some k. If also $\min A = 0$, then $A^* = A^l$ for some l (this is called the finite power property*).*

Proof. Every arithmetic progression is a rational set, and every infinite rational set contains an infinite arithmetic progression, so (i) and (ii) are equivalent.

Let $0 \in A$ and let $a, b \in \mathbb{Z}$ be such that $a + bn \in A^k$ for every $n \geq 0$. Because A^* contains all sufficiently large multiples of $\gcd A$, there are numbers c, l such that every multiple of $\gcd A$ that is in the interval $[c + 1, c + b]$ is also in A^l. Now A^{l+k} contains every number that is greater than $a + c$ and divisible by $\gcd A$.

Let $\min A = 0$ and let $\{\gcd A\}^* \setminus A^k$ be finite. Now $A^k \subseteq A^{k+1} \subseteq A^{k+2} \subseteq \cdots \subseteq \{\gcd A\}^*$ and only finitely many of these inclusions can be proper, so $A^l = A^{l+1} = A^{l+2} = \cdots = A^*$ for some l. □

It is not necessary for any of the conditions in Lemma 6.3.2 to hold for $k = 1$. For example, if A is the set of all squares, then $A^4 = \mathbb{N}_0 = A^*$ by Lagrange's four-square theorem.

Theorem 6.3.3. *Let $A_1, \ldots, A_s$ be sets of integers. Let $\min A_i = m_i > -\infty$ for all i. Let A_1^k contain an infinite arithmetic progression for some k. The sets A_i form a code if and only if $s = 1$ and $m_1 \neq 0$.*

Proof. If $s = 1$ and $m_1 = 0$, then $A_1^l = A_1^* = A_1^{l+1}$ for some l by Lemma 6.3.2. If $s = 1$ and $m_1 \neq 0$, then $A_1^k \neq A_1^l$ for all $k \neq l$, because $\min A_1^k = km \neq lm = \min A_1^l$ for all $k \neq l$.

Let $s \geq 2$. There are $k_1, k_2, l_1, l_2 > 0$ such that $k_1 m_1 + k_2 m_2 = l_1 m_1 + l_2 m_2$, but $k_1 \neq l_1$ or $k_2 \neq l_2$. Let $A_i = B_i + m_i$. Now B_1^k contains an infinite arithmetic progression, and the same is true for the sets $(B_1^{k_1} B_2^{k_2})^k$ and $(B_1^{l_1} B_2^{l_2})^k$. By Lemma 6.3.2, there is a number l such that $(B_1^{k_1} B_2^{k_2})^* = (B_1^{k_1} B_2^{k_2})^l$ and $(B_1^{l_1} B_2^{l_2})^* = (B_1^{l_1} B_2^{l_2})^l$. Also $(B_1^{k_1} B_2^{k_2})^* = (B_1 \cup B_2)^* = (B_1^{l_1} B_2^{l_2})^*$. Now

$$\begin{aligned}(A_1^{k_1} A_2^{k_2})^l =& (B_1^{k_1} B_2^{k_2})^l + l(k_1 m_1 + k_2 m_2) \\ =& (B_1^{k_1} B_2^{k_2})^* + l(k_1 m_1 + k_2 m_2) \\ =& (B_1^{l_1} B_2^{l_2})^* + l(l_1 m_1 + l_2 m_2) \\ =& (B_1^{l_1} B_2^{l_2})^l + l(l_1 m_1 + l_2 m_2) = (A_1^{l_1} A_2^{l_2})^l\end{aligned}$$

and the sets A_i do not form a code. □

In Theorem 6.3.3 we assumed that every infinite set is one-way infinite, i.e., has a finite minimum. The case where every set has a finite maximum is of course symmetric. We can also consider the two-way infinite case, i.e. the case when at least one set has arbitrarily large elements, and at least one (possibly the same) has arbitrarily small elements. This is done in the following theorem.

Theorem 6.3.4. *Let A and B be (not necessarily distinct) infinite sets of integers. Let the sets $(A \cap [0, \infty))^k$ and $(B \cap (-\infty, 0])^k$ contain infinite arithmetic progressions for some k. The sets A, B do not form a code.*

Proof. Now $(AB)^k$ contains increasing and decreasing infinite arithmetic progressions. Let $m \in (AB)^k$ and $C + m = (AB)^k$. Let a, b be such that

$$\gcd C = \gcd(C \cap [a, \infty)) = \gcd(C \cap (-\infty, b]).$$

By Lemma 6.3.2, there is a number l_1 such that $(C \cap [a, \infty))^{l_1}$ contains all but finitely many of the positive numbers divisible by $\gcd C$. Similarly, there is a number l_2 such that $(C \cap (-\infty, b])^{l_2}$ contains all but finitely many of

the negative numbers divisible by $\gcd C$. If $l > l_1, l_2$, then $C^l = \{\pm \gcd C\}^*$. Now

$$\begin{aligned}((AB)^k)^{l+\gcd C} &= C^{l+\gcd C} + ml + m \gcd C \\ &= \{\pm \gcd C\}^* + ml + m \gcd C \\ &= \{\pm \gcd C\}^* + ml \\ &= C^l + ml = ((AB)^k)^l\end{aligned}$$

and the sets A, B do not form a code. □

We have shown that three finite sets of integers do not form a code, and we have generalized this for certain infinite sets. However, no similar result holds for all infinite sets. The next example shows that there are arbitrarily large codes in the additive monoid of sets of integers.

Example 6.3.5. Let $A_i = \{1\} \cup \{(i + js)! \mid j \in \mathbb{N}_0\}$ for $i = 1, \ldots, s$. Let $B = A_1^{k_1} \ldots A_s^{k_s}$. We prove that the sets A_i form a code by showing that the set B uniquely determines the exponents k_i. Let j be such that $js > \min B = k_1 + \cdots + k_s$. Now $k(i + js)! + \min B - k \in B$ for $k \leq k_i$, but not for $k = k_i + 1$. Thus every k_i is determined by B.

6.4 Unique Decipherability for Languages

In this section we study the unique decipherability problem in the monoid of nonunary regular languages.

We fix two disjoint binary alphabets $\Sigma_1 = \{a, b\}$ and $\Sigma_2 = \{c, d\}$. We will use the following lemma that was proved in [11].

Lemma 6.4.1. *The unique decipherability problem is undecidable in the trace monoid $\Sigma_1^* \times \Sigma_2^*$.*

We will show that the monoid $\Sigma_1^* \times \Sigma_2^*$ can be effectively embedded in the monoid of regular languages over a non-unary alphabet.

For a word $w = a_1 \ldots a_n$, let $\mathrm{sw}(w)$ be the set of all (scattered) subwords of w, that is

$$\mathrm{sw}(w) = \{a_{i_1} \ldots a_{i_k} \mid 1 \leq i_1 < \cdots < i_k \leq n\}.$$

Let

$$X = (\Sigma_1 \cup \Sigma_2)^+ \setminus \Sigma_1^+ \setminus \Sigma_2^+$$

be the set of those words that contain letters from both Σ_1 and Σ_2. For all pairs of words $(u, t) \in \Sigma_1^* \times \Sigma_2^*$ we define a regular language

$$L(u, t) = \mathrm{sw}(u) \cup \mathrm{sw}(t) \cup X \tag{6.4}$$

over $\Sigma_1 \cup \Sigma_2$.

Lemma 6.4.2. *The mapping L defined by* (6.4) *is an injective morphism.*

Proof. First we show that if $u, u' \in \Sigma_1^*$ and $t, t' \in \Sigma_2^*$, then

$$L(u,t)L(u',t') = L(uu',tt').$$

Because $X \subseteq L(u,t)$ and $\varepsilon \in L(u',t')$, the set X is a subset of $L(u,t)L(u',t')$. Of the words in Σ_1^*, the language $L(u,t)L(u',t')$ contains exactly those that can be written as xy, where $x \in \mathrm{sw}(u)$ and $y \in \mathrm{sw}(u')$. These words form the set $\mathrm{sw}(uu')$. Similarly, $L(u,t)L(u',t') \cap \Sigma_2^* = \mathrm{sw}(tt')$. Thus

$$L(u,t)L(u',t') = \mathrm{sw}(uu') \cup \mathrm{sw}(tt') \cup X = L(uu',tt'),$$

and L is a morphism.

Next we show that if $u, u' \in \Sigma_1^*$ and $t, t' \in \Sigma_2^*$ and $L(u,t) = L(u',t')$, then $u = u'$ and $t = t'$. The longest words of Σ_1^* and Σ_2^* in $L(u,t)$ are u and t, and the longest words of Σ_1^* and Σ_2^* in $L(u',t')$ are u' and t'. Thus if $L(u,t) = L(u',t')$, then $u = u'$ and $t = t'$ and L is injective. □

Theorem 6.4.3. *The unique decipherability problem is undecidable in the monoid of regular languages over a non-unary alphabet.*

Proof. For the alphabet $\Sigma_1 \cup \Sigma_2$, this follows from Lemma 6.4.2. This alphabet has four letters, but $(\Sigma_1 \cup \Sigma_2)^*$ can be embedded in $\{a,b\}^*$, so the undecidability holds already for a binary alphabet. □

The question of the decidability of the unique decipherability problem in the monoid of finite languages could also be considered. It is noteworthy that in [11] everything is finite: the input is a finite collection of elements of the monoid $\Sigma_1^* \times \Sigma_2^*$. On the other hand, we obtain our result only for regular subsets of $(\Sigma_1 \cup \Sigma_2)^*$: the languages $L(u,t)$ contain an infinite regular part X, which is essential in the proof of Lemma 6.4.2. Actually, we do not know whether our result extends to finite collections of finite languages, so it remains an open question whether the problem is undecidable already in the monoid of finite languages.

The power equality problem for finite or regular nonunary languages is also interesting: given two languages A and B, does there exist two numbers $k, l \geq 1$ such that $A^k = A^l$? The decidability of this problem is not known. The easier problem of determining whether a language A has the finite power property, or whether there is a number k such that $A^k = A^*$, is known to be decidable, but the proofs are not trivial, see [60], [25] and [40].

Chapter 7

Conclusion

In Chapter 3 we, among other things, surveyed the following two open questions:

- How large is $\mathrm{IS}(n)$, the maximal size of an independent system of word equations on n unknowns?
- How large is $\mathrm{DC}(n)$, the maximal size of a decreasing chain of word equations on n unknowns?

For the first question we know that $3 \leq \mathrm{IS}(3) \leq \mathrm{UB}$ and $\Theta(n^4) \leq \mathrm{IS}(n) \leq \mathrm{UB}$. For the second question we know that $7 \leq \mathrm{DC}(3) \leq \mathrm{UB}$ and $\Theta(n^4) \leq \mathrm{IS}(n) \leq \mathrm{UB}$. Here the bound $\mathrm{DC}(3) \geq 7$ is new. Improving these estimates is one of the big open problems in combinatorics on words. In particular, the existence of a finite upper bound is a challenging problem.

In Chapter 4 we used polynomials and linear algebra to study word equations. One of the results we obtained in this way was that independent systems of word equation on three unknowns can have at most $|E|^2$ equations, where E is the shortest equation of the system. This bound works for n unknowns if the system remains independent when considering only solutions of rank $n-1$. These results suggest the following, possibly easier, variations of the above open question:

- Can we give a better bound than $|E|^2$ in the case of three unknowns?
- Can we give a similar bound in the case of n unknowns?

In Chapter 5 we reproved Hmelevskii's theorem and gave a bound for the size of parametric solutions and for the size of the shortest nontrivial solution. We concluded that solving the existence of a nontrivial solution for a constant-free equation on three unknowns is in NP.

In Chapter 6 we first analyzed the unique decipherability problem and other related problems in the monoid of unary languages, or equivalently, in

the monoid of sets of natural numbers. Then we proved that, in contrast to the unary case, the problem is undecidable in the monoid of binary regular languages. Decidability of this problem remains open for finite languages.

Bibliography

[1] M. H. Albert and J. Lawrence. The descending chain condition on solution sets for systems of equations in groups. *Proc. Edinburgh Math. Soc.*, 29(2):69–73, 1985.

[2] M. H. Albert and J. Lawrence. A proof of Ehrenfeucht's conjecture. *Theoret. Comput. Sci.*, 41(1):121–123, 1985.

[3] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences. Theory, applications, generalizations*. Cambridge University Press, 2003.

[4] Jean Berstel and Juhani Karhumäki. Combinatorics on words – a tutorial. In Gheorghe Paun, Grzegorz Rozenberg, and Arto Salomaa, editors, *Current Trends in Theoretical Computer Science. The Challenge of the New Century*. World Scientific, 2004.

[5] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2010.

[6] Alfred Brauer. On a problem of partitions. *Amer. J. Math.*, 64:299–312, 1942.

[7] L. G. Budkina and Al. A. Markov. F-semigroups with three generators. *Mat. Zametki*, 14:267–277, 1973.

[8] Christian Choffrut, Tero Harju, and Juhani Karhumäki. A note on decidability questions on presentations of word semigroups. *Theoret. Comput. Sci.*, 183(1):83–92, 1997.

[9] Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, 1997.

[10] Christian Choffrut and Juhani Karhumäki. Unique decipherability in the monoid of languages: an application of rational relations. *Theory Comput. Syst.*, 49(2):355–364, 2011.

[11] Marek Chrobak and Wojciech Rytter. Unique decipherability for partially commutative alphabet. *Fund. Inform.*, 10(3):323–336, 1986.

[12] Karel Culik, II and Juhani Karhumäki. Systems of equations over a free monoid and Ehrenfeucht's conjecture. *Discrete Math.*, 43(2–3):139–153, 1983.

[13] Elena Czeizler. Multiple constraints on three and four words. *Theoret. Comput. Sci.*, 391(1–2):14–19, 2008.

[14] Elena Czeizler and Juhani Karhumäki. On non-periodic solutions of independent systems of word equations over three unknowns. *Internat. J. Found. Comput. Sci.*, 18(4):873–897, 2007.

[15] Elena Czeizler and Wojciech Plandowski. On systems of word equations over three unknowns with at most six occurrences of one of the unknowns. *Theoret. Comput. Sci.*, 410(30–32):2889–2909, 2009.

[16] Volker Diekert. Makanin's algorithm. In M. Lothaire, editor, *Algebraic Combinatorics on Words*, pages 387–442. Cambridge University Press, 2002.

[17] Samuel Eilenberg and Marcel-Paul Schützenberger. Rational sets in commutative monoids. *J. Algebra*, 13:173–191, 1969.

[18] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16:109–114, 1965.

[19] Robert Gilmer. *Commutative Semigroup Rings*. University of Chicago Press, 1984.

[20] V. S. Guba. Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems. *Mat. Zametki*, 40(3):321–324, 1986.

[21] Tero Harju and Juhani Karhumäki. Morphisms. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 439–510. Springer-Verlag, 1997.

[22] Tero Harju and Juhani Karhumäki. Many aspects of defect theorems. *Theoret. Comput. Sci.*, 324(1):35–54, 2004.

[23] Tero Harju, Juhani Karhumäki, and Wojciech Plandowski. Independent systems of equations. In M. Lothaire, editor, *Algebraic Combinatorics on Words*, pages 443–472. Cambridge University Press, 2002.

[24] Tero Harju and Dirk Nowotka. On the independence of equations in three variables. *Theoret. Comput. Sci.*, 307(1):139–172, 2003.

[25] Kosaburo Hashiguchi. A decision procedure for the order of regular events. *Theoret. Comput. Sci.*, 8(1):69–72, 1979.

[26] Ju. I. Hmelevskiĭ. *Equations in free semigroups.* American Mathematical Society, 1976. Translated by G. A. Kandall from the Russian original: Trudy Mat. Inst. Steklov. 107 (1971).

[27] Štěpán Holub. In search of a word with special combinatorial properties. In *Computational and geometric aspects of modern algebra*, volume 275 of *London Math. Soc. Lecture Note Ser.*, pages 120–127. Cambridge Univ. Press, 2000.

[28] Štěpán Holub. Local and global cyclicity in free semigroups. *Theoret. Comput. Sci.*, 262(1–2):25–36, 2001.

[29] Štěpán Holub and Juha Kortelainen. On systems of word equations with simple loop sets. *Theoret. Comput. Sci.*, 380(3):363–372, 2007.

[30] Štěpán Holub and Juha Kortelainen. On partitions separating two words. In *Proceedings of the 7th International Conference on Words*, 2009.

[31] Jui-Yi Kao, Jeffrey Shallit, and Zhi Xu. The Frobenius problem in a free monoid. In *Proceedings of the 25th International Symposium on Theoretical Aspects of Computer Science*, pages 421–432, 2008.

[32] Juhani Karhumäki and Leonid P. Lisovik. The equivalence problem of finite substitutions on ab^*c with applications. *Internat. J. Found. Comput. Sci.*, 14(4):699–710, 2003.

[33] Juhani Karhumäki and Wojciech Plandowski. On the defect effect of many identities in free semigroups. In Gheorghe Paun, editor, *Mathematical aspects of natural and formal languages*, pages 225–232. World Scientific, 1994.

[34] Juhani Karhumäki and Aleksi Saarela. An analysis and a reproof of Hmelevskii's theorem. In *Proceedings of the 12th International Conference on Developments in Language Theory*, pages 467–478, 2008.

[35] Juhani Karhumäki and Aleksi Saarela. On maximal chains of systems of word equations. *Proc. Steklov Inst. Math.*, 274:116–123, 2011.

[36] Juhani Karhumäki and Aleksi Saarela. The unique decipherability in the monoid of regular languages is undecidable. *Fund. Inform.*, 110(1–4):197–200, 2011.

[37] David A. Klarner, Jean-Camille Birget, and Wade Satterfield. On the undecidability of the freeness of integer matrix semigroups. *Internat. J. Algebra Comput.*, 1(2):223–226, 1991.

[38] Juha Kortelainen. On the system of word equations $x_0u_1^ix_1u_2^ix_2\cdots u_m^ix_m = y_0v_1^iy_1v_2^iy_2\cdots v_n^iy_n$ $(i = 0, 1, 2, \cdots)$ in a free monoid. *J. Autom. Lang. Comb.*, 3(1):43–57, 1998.

[39] Werner Kuich. Semirings and formal power series. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 609–677. Springer-Verlag, 1997.

[40] Michal Kunc. Algebraic characterization of the finite power property. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, pages 120–131, 2006.

[41] Michal Kunc. The power of commuting with finite sets of words. *Theory Comput. Syst.*, 40(4):521–551, 2007.

[42] Markku Laine and Wojciech Plandowski. Word equations with one unknown. *Internat. J. Found. Comput. Sci.*, 22(2):345–375, 2011.

[43] M. Lothaire. *Combinatorics on Words*. Addison-Wesley, 1983.

[44] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[45] M. Lothaire. *Applied Combinatorics on Words*. Cambridge University Press, 2005.

[46] R. C. Lyndon and M. P. Schützenberger. The equation $a^M = b^Nc^P$ in a free group. *Michigan Math. J.*, 9:289–298, 1962.

[47] G. S. Makanin. The problem of the solvability of equations in a free semigroup. *Mat. Sb. (N.S.)*, 103(2):147–236, 1977. English translation in Math. USSR Sb. 32:129–198, 1977.

[48] Y. Matijasevic. Enumerable sets are diophantine (Russian). *Dokl. Akad. Nauk SSSR*, 191:279–282, 1970. Translation in Soviet Math Doklady, Vol 11, 1970.

[49] Filippo Mignosi, Jeffrey Shallit, and Ming-wei Wang. Variations on a theorem of Fine & Wilf. In *Proceedings of the 26th International Symposium on Mathematical Foundations of Computer Science*, pages 512–523, 2001.

[50] Dominique Perrin. Codes conjugués. *Information and Control*, 20:222–231, 1972.

[51] Wojciech Plandowski. Test sets for large families of languages. In *Developments in Language Theory*, pages 75–94, 2003.

[52] Wojciech Plandowski. Satisfiability of word equations with constants is in PSPACE. *J. ACM*, 51(3):483–496, 2004.

[53] Wojciech Plandowski and Wojciech Rytter. Application of Lempel-Ziv encodings to the solution of word equations. In *Proceedings of 25th International Colloquium on Automata, Languages, and Programming*, pages 731–742, 1998.

[54] Jorge L. Ramírez Alfonsín. *The Diophantine Frobenius Problem*. Oxford University Press, 2005.

[55] Aleksi Saarela. On the complexity of Hmelevskii's theorem and satisfiability of three unknown equations. In *Proceedings of the 13th International Conference on Developments in Language Theory*, pages 443–453, 2009.

[56] Aleksi Saarela. Systems of word equations and polynomials: a new approach. In *Proceedings of the 8th International Conference WORDS*, pages 215–225, 2011.

[57] Aleksi Saarela. Unique decipherability in the additive monoid of sets of numbers. *RAIRO Inform. Theor. Appl.*, 45(2):225–234, 2011.

[58] Arto Salomaa. The Ehrenfeucht conjecture: a proof for language theorists. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, 27:71–82, 1985.

[59] August Albert Sardinas and George W. Patterson. A necessary and sufficient condition for unique decomposition of coded messages. In *IRE Intern. Conv. Rec. 8 (1953)*, pages 104–108. Chapman and Hall, 1953.

[60] Imre Simon. Limited subsets of a free monoid. In *Proceedings of the 19th Annual Symposium on Foundations of Computer Science*, pages 143–150, 1978.

[61] Jean-Claude Spehner. *Quelques problémes d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre*. PhD thesis, Univ. Paris, 1976.

[62] Jean-Claude Spehner. Les systemes entiers d'équations sur un alphabet de 3 variables. In *Semigroups*, pages 342–357, 1986.

[63] Axel Thue. Über unendliche zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl.*, 7:1–22, 1906.

[64] Axel Thue. Über die gegenseitige lage gleicher teile gewisser zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl.*, 1:1–67, 1912.

[65] Paavo Turakainen. The equivalence of deterministic gsm replications on Q-rational languages is decidable. *Math. Systems Theory*, 20(4):273–282, 1987.

[66] Joachim von zur Gathen and Malte Sieveking. A bound on solutions of linear integer equalities and inequalities. *Proc. Amer. Math. Soc.*, 72(1):155–158, 1978.

Turku Centre for Computer Science
TUCS Dissertations

112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Communication and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
120. **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
121. **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
122. **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
123. **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
124. **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
125. **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
126. **Tuomo Saarni**, Segmental Durations of Speech
127. **Johannes Eriksson**, Tool-Supported Invariant-Based Programming
128. **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
129. **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
130. **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
131. **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
132. **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
133. **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
134. **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
135. **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign
136. **Matteo Brunelli**, Some Advances in Mathematical Models for Preference Relations
137. **Ville Junnila**, On Identifying and Locating-Dominating Codes
138. **Andrzej Mizera**, Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.
139. **Csaba Ráduly-Baka**, Algorithmic Solutions for Combinatorial Problems in Resource Management of Manufacturing Environments
140. **Jari Kyngäs**, Solving Challenging Real-World Scheduling Problems
141. **Arho Suominen**, Notes on Emerging Technologies
142. **József Mezei**, A Quantitative View on Fuzzy Numbers
143. **Marta Olszewska**, On the Impact of Rigorous Approaches on the Quality of Development
144. **Antti Airola**, Kernel-Based Ranking: Methods for Learning and Performace Estimation
145. **Aleksi Saarela**, Word Equations and Related Topics: Independence, Decidability and Characterizations

Turku Centre *for* Computer Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www. tucs.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics
- Institute of Information Systems Science

Åbo Akademi University
Division for Natural Sciences and Technology
- Department of Information Technologies

ISBN 978-952-12-2737-0
ISSN 1239-1883

Aleksi Saarela

Word Equations and Related Topics: Independence, Decidability and Characterizations

Aleksi Saarela

Word Equations and Related Topics: Independence, Decidability and Characterizations

Aleksi Saarela

Word Equations and Related Topics: Independence, Decidability and Characterizations