UNIVERSITY OF WARSAW
FACULTY OF MATHEMATICS, INFORMATICS AND MECHANICS

Szymon Toruńczyk

Languages of profinite words and the limitedness problem

PhD Thesis

Supervisor
dr hab. Mikołaj Bojańczyk

Institute of Informatics
University of Warsaw

March 2011

Authors declaration Aware of legal responsibility I hereby declare that I have written this dissertation myself and all the contents of the dissertation have been obtained by legal means.

March 31, 2011

Supervisors declaration The dissertation is ready to be reviewed.

March 31, 2011

Abstract

This thesis is about the limitedness problem. In the first part of the thesis, we give a self-contained proof of the decidability of the limitedness problem for B-automata – a model introduced by Bojańczyk and Colcombet [BC06]. These are automata with many counters, which can be incremented or reset, but not decremented. B-automata generalize distance automata of Hashiguchi [Has82] and the nested distance desert automata of Kirsten [Kir05]. Our proof brings to light some connections with the theory of profinite semigroups.

In the second part of the thesis, we develop a theory which further investigates the connections with the realm of profinite words. We view a B-automaton as defining a language of profinite words. Our theory gives an alternative description of the class of languages defined by B-automata – in terms of regular expressions, of logic, of recognition by homomorphisms and of congruences of finite index – where each of these notions is given a suitable interpretation in the context of profinite words.

Keywords: Limitedness problem, distance automata, nested distance desert automata, B-automata, star height problem, profinite words, MSO+$\mathbb{B}$ logic

ACM classification: D.2.4, F.4.3, F.1.1,

Streszczenie

Tematem niniejszej rozprawy jest problem ograniczoności. W pierwszej części rozprawy, prezentujemy niezależny dowód rozstrzygalności problemu ograniczoności dla B-automatów – modelu wprowadzonego przez Bojańczyka i Colcombeta [BC06]. Są to automaty wyposażone w liczniki które mogą być zwiększane o jeden, lub resetowane. B-automaty uogólniają automaty z odległością Hashiguchiego [Has82] oraz zagnieżdżone automaty Kirstena [Kir05]. Przedstawiony dowód ujawnia pewne związki z teorią półgrup proskończonych.

W drugiej części rozprawy rozwijamy teorię, która dalej bada związki pomiędzy problemem ograniczoności a słowami proskończonymi. W ramach tej teorii, B-automat definiuje język słów proskończonych. Przedstawiamy równoważne opisy klasy języków definiowanych przez B-automaty – w terminach wyrażeń regularnych, logiki, rozpoznawalności przez homomorfizmy oraz poprzez kongruencje skończonego indeksu – przy czym pojęciom tym przypisujemy odpowiednie znaczenie w kontekście słów proskończonych.

Acknowledgements

I would like to thank my brother-in-law, Mikołaj Bojańczyk, who has introduced me to the fascinating field of automata theory. But above all, I would like to thank my advisor, Mikołaj Bojańczyk, who has been guiding me through this field, supporting me with his inestimable generosity and knowledge, and has been an endless source of brilliant ideas, of which only a few I am beginning to grasp.

I thank Thomas Colcombet for many stimulating discussions. I thank Luc Segoufin for our collaboration in the year 2010, and for hosting my stay in Paris in 2011, when I finished the writing of this thesis.

Contents

CHAPTER 1

Introduction

The theme of this thesis is the limitedness problem. The underlying aspiration is to understand the problem and its solution better, and to find a theory which allows further generalizations. Below, we give a cursory historical introduction to this topic, and describe how the work of this dissertation fits into that background.

The star height and limitedness problems

The star height problem was introduced by L. C. Eggan in 1963 [Egg63]. It is the problem of computing the *star height* of a given regular language L, i.e. the smallest possible number of nestings of a Kleene star in a regular expression defining L. The regular expressions are allowed to use concatenation, union and the Kleene star, but not complementation (as opposed to the *generalized star height problem*). For example, the regular language L described by the expression $(a^*b)^*a^*$ can be also described by the expression $(a+b)^*$, so the star height of L is equal to 1. The star height problem remained open for 25 years, after which it was shown to be computable by K. Hashiguchi [Has88]. His proof is profound and insightful, but difficult to read, as many have commented.

Since then, various researchers have worked on understanding and simplifying the proof of the decidability of the star height problem and other, related language-theoretic problems. One of the related problems is the *finite power property* of a regular language L, introduced by J. Brzozowski in 1966: for a given regular language L, does there exist a finite number k such that $L^+ = L^1 \cup L^2 \cup \ldots \cup L^k$? During many years of research, it became apparent that the core combinatorial problem which underlies these sorts of problems is the *limitedness problem*.

The limitedness problem is a decision problem, which was introduced by K. Hashiguchi on his way to solving the star height problem. In its basic form, it can be formulated as follows. Given a nondeterministic automaton $\mathcal{A}$ whose transitions are additionally labeled by nonnegative, integer weights, does there exist a bound n such that every accepted word has some accepting run whose sum of weights is bounded by n?

Such automata, in which transitions have weights, are called *distance automata*. An example of a distance automaton over the alphabet $\{a, b\}$ is depicted in Figure 1.1. A distance automaton $\mathcal{A}$ can be seen as function assigning to an input word the minimal sum of the weights of

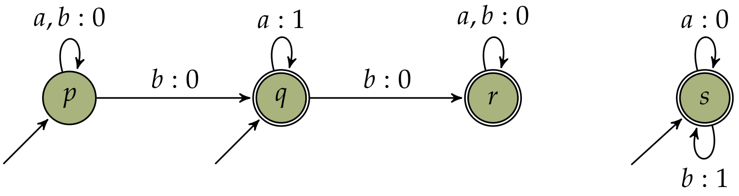


FIGURE 1.1: *A distance automaton. Its input alphabet is $\{a, b\}$, initial states are p, q, s and accepting states are q, r, s*

an accepting run over w, and $\perp$ if the word is not accepted[1] by $\mathcal{A}$. For instance, the automaton in the figure assigns to a word $a^{n_1}ba^{n_2}b \cdots ba^{n_k}$ the value $\min(n_1, n_2, \ldots, n_k, k)$. The limitedness problem then asks whether this function has a finite range. The automaton in the example is not limited, since the values assigned to the words $aba, aabaabaa, \ldots, (a^k b)^k a^k, \ldots$ grow unboundedly.

Let us illustrate the applicability of the limitedness problem in language-theoretic problems, by sketching a reduction from the finite power property to the limitedness problem for distance automata. Suppose that we want to check whether a language L accepted by a finite automaton $\mathcal{A}$ has the finite power property. The idea is simply to equip $\mathcal{A}$ with additional ε-transitions, which allow $\mathcal{A}$ to pass from any accepting state to any initial state (see Figure 1.2). These new transitions, however, are costly – we associate with them the weight 1 – as opposed to the original transitions of $\mathcal{A}$, which have weight 0. Now, we ask if the resulting distance automaton $\tilde{\mathcal{A}}$ is limited. If it is limited, this means that there exists a bound n such that any word accepted by $\tilde{\mathcal{A}}$ – and those are precisely the words in L^+ – can be accepted by a run which uses not more than n weighted transitions. In particular, such words belong to the language L^k, for some $k \leq n$, so this proves that L has the finite power property. The implication in the other direction is similar.

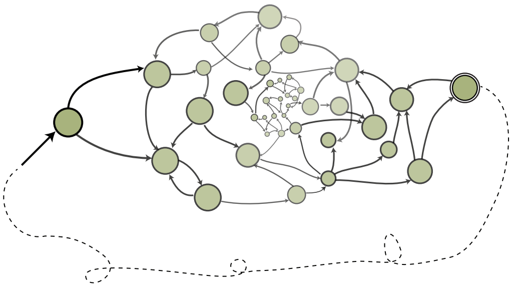


FIGURE 1.2: *Testing the finite power property: the automaton is extended by a "long" (in terms of distance) ε-transition from the accepting state to the initial state (depicted as a dashed edge)*

The tropical semiring. The limitedness problem of distance automata is closely related to problems concerning the *tropical semiring*. It was introduced by I. Simon [Sim78] (for a survey,

[1] Following K. Hashiguchi, the symbol used for $\perp$ is usually ∞. We, however, break with this tradition.

see [Pin98]), in his solution to the decidability proof for the finite power property. The tropical semiring consists of the natural numbers with the usual minimum and addition playing the role of semiring addition and multiplication. Moreover, the tropical semiring contains the extra element $\perp$, such that $\perp + n = n + \perp = \perp$, and $\min(\perp, n) = \min(n, \perp) = n$ for all numbers n.

The connection between automata and the tropical semiring is simple and comes from the following observation. In a distance automaton with n states, the weight of a run is the sum of the weights of its transitions, and the value of a word is the minimum of the weights of all possible runs. This is related to multiplication of $n \times n$ matrices over the tropical semiring, since if $s_1, s_2, \ldots, s_k$ are such matrices, in their product, the entry at position $[p, q]$ is the minimum over all sequences of indices $p_1, p_2, \ldots, p_{k-1} \in \{1, \ldots, n\}$ of the values

$$s_1[p, p_1] + s_2[p_1, p_2] + \cdots + s_k[p_{k-1}, q].$$

The above sum can be interpreted as the sum of weights of the transitions in a run from the state p to the state q. Basing on this observation, it is straightforward to show the equivalence of the limitedness problem for distance automata with the *finite section problem* for the tropical semiring, defined as follows. Given a set S of $n \times n$ matrices over the tropical semiring and a pair of indices $1 \leq p, q \leq n$, the $[p, q]$*-section* of S is the set of elements which occur on the $[p, q]$-coordinate of some matrix in S. The finite section problem asks about finiteness of the $[p, q]$-section of the set A^+ of matrices generated by multiplication from a given finite set of matrices A.

In his memorable paper, K. Hashiguchi [Has82] proved decidability of the limitedness problem for distance automata. As noted by I. Simon [Sim88], "the solution is very complicated and difficult to visualize and this led to further research to find other proofs of this result". Moreover, the proposed algorithm has bad complexity. In turn, K. Hashiguchi's proof of decidability of the star height problem goes via a complicated reduction to the limitedness problem.

A topological approach to limitedness. An alternative proof of the decidability of the limitedness problem was given by H. Leung [Leu88]. He considers an extension $\mathcal{T}$ of the tropical semiring by an element ω, ordered by

$$0 < 1 < 2 < \ldots < \omega < \perp.$$

Again, the semiring operations are minimum and +, where we let $x + \omega = \max(x, \omega)$ for $x \in \mathcal{T}$. The crucial property of the new element ω is topological – it is the limit point of any unbounded sequence of finite numbers. More precisely, we consider a metric over $\mathcal{T}$, in which the distance between a finite number n and ω is $1/(n+1)$, while the distance to $\perp$ is 1 independently from n. The topology of the semiring $\mathcal{T}$ is depicted in Figure 1.3. The basic idea behind the introduction of the new element ω is as follows. Let A be a finite set of matrices over the semiring $\mathcal{T}$, not using the value ω. The $[p, q]$-section of set of matrices A^+ is infinite if and only if its topological closure $\overline{A^+}$ contains a matrix whose $[p, q]$-coordinate is equal to ω. H. Leung considers another, finite, semiring $\mathcal{R}$, which is an abstracted version of the semiring $\mathcal{T}$ – its elements are

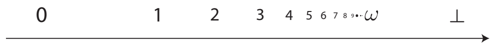


FIGURE 1.3: *The topology and order of the semiring $\mathcal{T}$*

$0, 1, \omega, \perp$, and 1 represents all finite, positive values. There is a natural *abstraction* mapping which converts a matrix over $\mathcal{T}$ to a matrix over $\mathcal{R}$, by replacing all finite, positive values by 1; this mapping is a homomorphism of semirings. As abstraction distinguishes ω from the finite values, the limitedness problem can be reduced to computing the abstraction of the set $\overline{A^+}$, for a given set of matrices A. For matrices over the semiring $\mathcal{R}$, H. Leung introduced a *stabilization* operation, which is a sort of acceleration – given an idempotent matrix e, its stabilization $e^\#$ is the abstraction of the limit of the sequence of matrices $e, e^2, e^3, \ldots$, treated as matrices over the semiring $\mathcal{T}$. For instance, if e is a 1×1 matrix with one entry equal to 1, then it is an idempotent in the semiring $\mathcal{R}$. However, treated as a matrix in the semiring $\mathcal{T}$, its successive powers are $1, 2, 3, 4, \ldots$, and their limit is ω. Therefore, the stabilization of the matrix e is the matrix with one entry equal to ω. The stabilization operation can be easily described in purely algebraic terms.

H. Leung proposed the following characterization of the abstractions of the elements of $\overline{A^+}$.

> *If A is a finite set of $n \times n$ matrices over $\mathcal{T}$, then the set of abstractions of the matrices in $\overline{A^+}$ is equal to the smallest set of matrices over $\mathcal{R}$ which is closed under multiplication and stabilization, and contains the abstractions of the matrices in A.*

Since the set of all $n \times n$ matrices over the semiring $\mathcal{R}$ is finite, this gives an efficient procedure for deciding the limitedness problem. H. Leung's original proof of correctness of the above characterization is complicated, and uses advanced techniques from semigroup theory – in particular, Green's relations and Brown's Theorem – in a careful analysis of the structure of the semigroup of matrices over $\mathcal{R}$.

I. Simon [Sim94] gave an independent proof of correctness of the characterization proposed by H. Leung. Instead of using Brown's Theorem, the main technical tool he used where his *factorization forests* – an ingenious, simple, yet powerful device to handle finite semigroups. Factorization forests have found many applications in problems which are not related to the star height problem (for an overview, see [Boj09a]).

Nested automata. D. Kirsten [Kir05] proposed a new model of automata – called *nested distance desert automata*. A usual distance automaton can be seen as a nondeterministic automaton with one counter, which allows one operation: increment. Roughly speaking, nested distance desert automata allow several counters, which are arranged into a strict hierarchy – an increment of a counter with higher priority entails resets of all counters with lower priorities. The value of a run is the highest value attained by any counter during this run. The value of a word w is the minimum of the values of all accepting runs over w. Hence, similarly to distance automata, a nested distance desert automaton also associates with an accepted input word a finite value, and with remaining words – the value $\perp$. Therefore, it makes sense to consider the *limitedness problem* for nested distance desert automata. D. Kirsten gave a proof of the decidability of this decision problem, elaborating the algebraic techniques used by K. Hashiguchi, I. Simon and H. Leung. Moreover, D. Kirsten presented an elegant and rather straightforward reduction of the star height problem to the limitedness problem for nested distance desert automata, thus giving a complete, alternative proof of the star height problem.

The first part of this dissertation contains yet another proof of the limitedness problem for an even more general model of automata, called *B-automata*, which allow resets and multiple counters, without the hierarchical constraint. We will come back to these automata further in this introduction. Our proof is inspired by the topological viewpoint of H. Leung, and uses the technique of I. Simon based on factorization forests. Following the topological viewpoint, we see the semigroup of matrices over the semiring $\mathcal{T}$ as a compact topological semigroup, and more precisely – as a profinite semigroup. Any such semigroup is naturally equipped with an operation called the *ω-power*, which, for a given element s, gives as a result the limit of the sequence $s, s^{2!}, s^{3!}, \ldots$. It turns out that the ω-power in the semiring of matrices over $\mathcal{T}$ corresponds precisely to the stabilization operation of H. Leung. We discover a crucial algebraic-topological property of the semigroup of matrices over $\mathcal{T}$:

> *If A is a finite set of $n \times n$ matrices over $\mathcal{T}$, then $\overline{A^+}$ is equal to the smallest set which is closed under multiplication and the ω-power, and contains A.*

This property generalizes the characterization proposed by H. Leung. Actually, we prove the above property for an extension of the semiring $\mathcal{T}$, suited for dealing with resets and multiple counters.

Our proof relies on an extension of the notion of a factorization forest to *stabilization semigroups*, and in this aspect resembles the proof presented by I. Simon [Sim94]. We believe that our proof is simpler than the other proofs, but most importantly, it gives an insight into the connection between the limitedness problem and profinite topology, which we further exploit in the second part.

Note that, together with D. Kirsten's straightforward reduction of the star height problem to the limitedness problem, our result gives yet another proof of decidability of the star height problem.

MSO+$\mathbb{B}$ and B/S-automata

The logic MSO+$\mathbb{B}$. In his dissertation, M. Bojańczyk (see also [Boj04]) has investigated the decidability of the following decision problem:

> *Is a given formula of the modal μ-calculus with backward modalities satisfiable in some finite structure?*

While dealing with this problem, M. Bojańczyk was lead to investigate the logic MSO over infinite trees, extended by a new quantifier $\mathbb{B}$. The definition of $\mathbb{B}$ is such that the formula

$\mathbb{B}X.\varphi(X)$ holds if and only if all the sets of positions X satisfying the formula φ have a commonly bounded size. Some fragments of this logic where shown decidable – enough to prove the decidability of the problem concerning modal μ-calculus; however, the question of decidability of the full logic MSO+$\mathbb{B}$ remains open. In fact, this logic is not even known to be decidable over infinite words.

The logic MSO+$\mathbb{B}$ has become a part of a research program (see [Boj10] for a survey), which tries to extend the notion of an ω-regular language, while preserving its robust properties. For instance, a language defined by MSO+$\mathbb{B}$ has a finite-index equivalence relation with finitely many equivalence classes, even for the strongest Arnold equivalence relation, commonly used for languages of infinite words. In [Boj09b, BT09], several examples of robust extensions of MSO have been considered. These where defined as extensions of the Weak MSO logic (which coincides with MSO over infinite words) by allowing the quantifier $\mathbb{B}$, or other, similar quantifiers. It appears that extensions of Weak MSO often admit equivalent automata models, thus allowing the use of automata techniques for solving the satisfiability problem. Unfortunately, this approach does not seem to work for the full logic MSO+$\mathbb{B}$.

Connection with limitedness. There is a connection between the logic MSO+$\mathbb{B}$ over infinite words and the limitedness problem for automata with counters (say, distance or nested distance desert automata). This is because limitedness of an automaton with counters $\mathcal{A}$, such as a nested distance desert automaton, can be expressed in MSO+$\mathbb{B}$ as follows:

> *For every infinite word of the form $w_1\$w_2\$w_3\$,\ldots$, if all words $w_1, w_2, \ldots$ are accepted by $\mathcal{A}$, then there exists:*
>
> – *an infinite word $\rho_1\$\rho_2\$\rho_3\$\ldots$ such that for every $i = 1, 2, \ldots$, the word ρ_i encodes an accepting run of $\mathcal{A}$ over w_i,*
>
> – *a common bound on the size of sets of positions in the words $\rho_1, \rho_2, \ldots$ which mark a sequence of increments of any counter which is not interrupted by a reset*

M. Bojańczyk and T. Colcombet [BC06], unaware of the existing work on the limitedness problem, have embarked on solving the problem of decidability of MSO+$\mathbb{B}$ over infinite words. The success was only partial – they managed to show decidability of only a fragment of the logic, but this fragment was big enough to capture the above formula describing limitedness. Therefore, this paper gives another, independent proof of decidability of the limitedness problem. Still, in the paper, the limitedness problem is not even mentioned!

B-automata. As a tool for analyzing this fragment of the logic MSO+$\mathbb{B}$, the authors introduced *B-automata*. They are automata over finite words, similar to nested distance desert automata, but the counters are not hierarchical, and allow two operations: increment and reset. The valuation of a word w under a given B-automaton is then the minimal value of an accepting run over w, and the value of a run is the maximal value attained by some counter. Also another model of automata was introduced, called *S-automata*; they are completely dual to B-automata

– we simply swap "minimal" with "maximal" in the definition of the valuation. The advantage of S-automata is that testing limitedness is much easier than for B-automata. The central duality result of the paper [BC06] relates B-automata with S-automata. It relies heavily on the factorization theorem of I. Simon.

For dealing with fragments of the logic MSO+$\mathbb{B}$ over infinite words, *ωB-automata* are considered instead of B-automata. These automata are defined like B-automata, but instead of evaluating an input word to a number, they give a Boolean answer, basing on whether the values of the counters remain bounded during some Büchi-accepting run. The duality result concerning B-automata implies decidability of the emptiness problem for Boolean combinations of ωB-automata. They, in turn correspond to a fragment of the logic MSO+$\mathbb{B}$ with restricted negation.

The model of ωB-automata appears to be a rather robust model of computation. A seemingly unrelated model of automata, considered in [ST11], in which counters are allowed to be increased and decreased (without having control on the differences in the increases and decreases) and tested for equality, turned out to be equi-expressive with ωB-automata over infinite words (the acceptance condition is Büchi). Note that allowing decrementation in ωB-automata quickly leads to undecidability, since a reduction from a two counter Minsky machine can be then easily provided. Such a reduction can be even given in the case corresponding to distance automata – of only one counter and no resets (see [DDG$^+$10], where such automata with incrementation and decrementation and bounds where viewed as energy games). In fact, it is difficult to come up with a reasonable extension of ωB-automata, for which emptiness and universality are both decidable.

Regular cost functions. Recently, T. Colcombet [Col09, Col10b] has created an entire framework which, among others, implies the central result of the paper [BC06] coauthored with M. Bojańczyk, and makes the connection to the limitedness problem more explicit, and gives it another proof. His framework is based on the paper with M. Bojańczyk, but also shares many similarities with the proofs by H. Leung and D. Kirsten. A key notion introduced by T. Colcombet is that of a *cost function*. This appears to be a good notion for comparing number-valued functions, when the limitedness problem is concerned. This is because a *cost function* is an abstraction which does not care about the precise values, but only cares about bounds. A formal definition is as follows. Let $f, g\colon A^+ \to \mathbb{N} \cup \{\omega\}$ be two functions over the set of words over an alphabet A. We say that *f dominates g* if for any set of words $K \subseteq A^+$, if f is bounded over K, then also g is bounded over K. The domination relation defines a partial preorder over the set of functions. The equivalence classes of this relation are called *cost functions*. For example, if $A = \{a, b\}$, the function computing the length of an input word is equivalent to the function computing the size of the largest block of a's plus the number of b's in an input word.

For a B-automaton $\mathcal{A}$, let $f_\mathcal{A}$ denote the function which maps any input word w accepted by $\mathcal{A}$ to its valuation under $\mathcal{A}$. The limitedness problem is then equivalent to deciding whether, given a B-automaton $\mathcal{A}$, the function $f_\mathcal{A}$ is dominated over its domain by the function constantly equal to 0. Note that this does not depend on the function $f_{\mathcal{A}}$ itself, but only on its equivalence class with respect to the domination relation.

The following, deep result can be derived from the central result of M. Bojańczyk and T. Colcombet [BC06], but the below statement is due to T. Colcombet [Col09], who proved it in an algebraic framework.

Theorem 1.1 *The class of cost functions defined by B-automata coincides with the class of cost functions defined by S-automata. This correspondence is effective. As a consequence, the domination relation is decidable for this class of cost functions.*

It is important that we only distinguish the equivalence class of the function computed by an automaton, and not the exact function itself. It follows from a result of D. Krob [Kro92] that it cannot be decided whether two given distance automata define the same function.

Following T. Colcombet, we call the class of cost functions defined by B- or S-automata *regular cost functions*. The framework of T. Colcombet gives a description of regular cost functions not only in terms of automata, but also gives an equivalent description in terms of algebra, and in terms of a counting extension of the MSO logic over finite words, and in terms of regular expressions.

The full MSO+$\mathbb{B}$ logic It is not clear if the framework of regular cost functions, which is based on counter automata, is the right approach to satisfiability of MSO+$\mathbb{B}$ over infinite words, and not just for the fragment with restricted quantifier use. A noticeable problem with extending the approach of cost functions to larger fragments of the logic MSO+$\mathbb{B}$ is that there is no reasonable way to define the complement of a function, or of a cost function. More substantive doubts are based on the paper [HST10], which proves that MSO+$\mathbb{B}$ can define non-Borel languages of infinite words. This implies that there can be no nondeterministic automaton model for MSO+$\mathbb{B}$ that has a Borel acceptance condition, which excludes all known nondeterministic automata models that use counters. One has to keep in mind that the non-Borel result still leaves room for automata; a distant analogy is that parity automata on infinite trees recognize non-Borel sets.

A topological approach to the limitedness problem

The raison d'être of this work can be seen as a desire of approaching the decidability problem of the logic MSO+$\mathbb{B}$. We do not even solve any new fragment of this logic, but only wish to establish a framework in which at least the limitedness problem can be presented in an elegant and intuitive way.

The cleanest and most general available proofs of the decidability of the limitedness problem (for distance automata, or nested distance desert automata, or for B-automata) are the most recent ones – by D. Kirsten and T. Colcombet. However, H. Leung's insight of considering the problem from a topological perspective does not play a big role in these proofs, as they rely mostly on algebra, making their solutions "difficult to visualize", using the phrasing of I. Simon. We believe that there is a connection with profinite topology, which should be exploited in order to simplify the proof, and perhaps further extend the result. Basing on this idea, we establish a theory of B- and S-automata which is comparable with the classical theory of finite automata, and which seems to be the natural environment for problems like the limitedness or domination problem.

In the classical theory, as well as in the theory of Büchi automata (see [Büc62, Tho90]), apart from the automata themselves, we have notions of regular expressions, recognition by homomorphisms, syntactic congruence, and MSO logic, which all turn out to be equivalent, with respect to the languages they define.

The second part of this dissertation presents a theory analogous to the classical theory, which is suitable for B- and S-automata. One difficulty which arises is that it is not clear what type of object such an automaton $\mathcal{A}$ recognizes. Considering simply the set of words for which $\mathcal{A}$ has an accepting run cannot bear interesting results, as it ignores the counter operations. The solution of T. Colcombet is to consider the cost function $f_{\mathcal{A}}$ computed by $\mathcal{A}$. The drawback of this approach is that a cost function is something of a completely different nature than a language of words. It makes no sense to consider the complement of a cost function, nor to talk about membership of a word in a cost function, which are both needed to define the syntactic congruence, for instance. Neither does it make sense to perform standard set-theoretic constructions, such as the quotient by the syntactic congruence. As a consequence, the theory of T. Colcombet lacks of a construction à la the Myhill-Nerode quotient by the syntactic congruence.

Our approach will overcome these difficulties, as we will again see automata as recognizing languages – not of words, but of profinite words. To illustrate the connection, consider the distance automaton from Figure 1.1. There is a single *profinite* word which witnesses the fact that the automaton is not limited – this word can be written as $(a^{\omega}b)^{\omega}a^{\omega}$, and can be defined as the limit of the sequence of words $((a^{n!}b)^{n!}a^{n!})_{n=1}^{\infty}$. We say that this profinite word does not *belong to the language* of the automaton.

We will discover the corresponding notions of regular expressions, recognition by homomorphisms and MSO logic, and syntactic congruence – each capable of defining a genuine language of profinite words. Remarkably, in this viewpoint, the problems of limitedness and domination simply appear as the language universality and inclusion problems. (The distance automaton from the example is not universal, because the profinite word $(a^{\omega}b)^{\omega}a^{\omega}$ does not belong to its language.) Since, as usual, inclusion can be reduced to emptiness by using intersection and complementation, the central algorithmic problem which emerges is computing complementation. One of the consequences of the theory we present is that it can be effectively computed, which implies the decidability results regarding limitedness or domination.

Outline. The first chapter of this thesis contains preliminary notions, which are well established in the literature – in particular, of topology, semigroups, and of profinite semigroups or profinite words – so the reader can use it as a reference for some definitions and results which are used in the other parts. The rest is divided into two parts. The first part deals with the limitedness problem for distance automata, and more generally, of B-automata. In the second part, the profinite theory for B- and S-automata is developed. We end the second part with a discussion which compares our theory with the theory of T. Colcombet and to the theory of ω-regular languages. We give a rough idea of how we believe the theory can be generalized, in order to approach the problem of decidability of the logic MSO+$\mathbb{B}$.

Preliminaries

In these preliminaries, we give definitions and state some basic properties of the objects with which we will be working with. In Section 2.1 we introduce the basic notions from general topology. For a reference, any book on general topology can be used, e.g. [Wil70, Eng89]. Most of these notions are only used in Chapter 4 of these preliminaries, but some of them are also used in Part II of this thesis. In Section 2.2 we describe a connection between finite topological spaces and finite partially ordered sets. This connection will be used only in Part II of this thesis.

In Chapter 3 we define the basic notions from semigroup theory, recall the relation between finite automata and finite semigroups, and introduce some elementary tools of the structure theory of finite semigroups. These tools will be used for proving in Part I a Factorization Theorem for stabilization semigroups.

In Chapter 4 we introduce metrizable profinite semigroups using projective limits of projective sequences, which is a restricted version of the more general definition using projective systems (see e.g. Almeida [Alm05]). We characterize them as totally disconnected compact topological semigroups, which are metrizable. We further define the free profinite semigroup in a way very similar to the standard definition [Alm05, Pin09] using the profinite completion. Algebraic properties of the ω-power in profinite semigroups will be used in both Part I and Part II of this thesis. However, the free profinite semigroup and profinite words will only appear in Part II.

CHAPTER 2

Topology

2.1 Metric and topological spaces

A *metric space* (X, d) is a set X equipped with a *metric*, i.e. a function $d\colon X \to \mathbb{R}$, called the *distance*, such that for all $x, y, z \in X$:

- $d(x,z) \leq d(x,y) + d(y,z)$ (*triangle inequality*)
- $d(x,y) = d(y,x)$
- $d(x,y) = 0$ if and only if $x = y$

An example of a metric space is $\mathbb{R}$ itself, equipped with the metric $d(x,y) = |x - y|$. Another example is a *discrete metric space*, i.e. any set X equipped with a metric d such that $d(x,y) = 1$ if $x \neq y$.

A *topological space* is a set X of *points* together with a distinguished family of *open sets*, called the *topology* of X. This family must satisfy the following three axioms: 1) the empty set and the set X are open sets, 2) a union of an arbitrary family of open sets is again an open set, 3) an intersection of a *finite* family of open sets is again an open set.

An open set which contains a point x may be called a *neighborhood* of x. A topology on X is often specified by providing its *base*, i.e. a family of open sets, called *basic open sets*, such that any open set is a union of basic open sets. Equivalently, a base is any family of "arbitrarily small" open neighborhoods, i.e. such that any neighborhood of a point x contains some neighborhood of x which is a basic open set. For a metric space (X,d) we define the topology by specifying the base consisting of all *open balls*, i.e. sets of the form $B_\varepsilon(x) = \{y \in X : d(x,y) < \varepsilon\}$. Therefore, the open sets in a metric space are arbitrary unions of open balls.

Complements of open sets are called *closed sets*. Therefore, the family of closed sets is preserved by arbitrary intersections and finite unions. If $M \subseteq X$ is any subset, then by $\overline{M}$ we denote the *closure* of M, i.e. the intersection of all closed subsets containing M. We say that M is *dense* in X if $\overline{M} = X$. A *clopen set* set is a set that is both closed and open in X. Clopen sets are closed under arbitrary finite Boolean operations, but not necessarily under arbitrary unions or intersections.

A mapping $f\colon X \to Y$ of two topological spaces is *continuous* if the inverse image of an open set in Y is an open set in X. Equivalently, the inverse image of a closed set in Y is a closed set in X. The mapping f is called a *homeomorphism* if it is a bijection and its inverse is also continuous.

If a set X is endowed with two topologies, $\mathcal{U}$ and $\mathcal{U}'$, then we say that the topology $\mathcal{U}$ is *stronger* than the topology $\mathcal{U}'$, if $\mathcal{U}' \subseteq \mathcal{U}$. If $\mathcal{U}$ is stronger than $\mathcal{U}'$, then we also say that $\mathcal{U}'$ is *weaker* than $\mathcal{U}$.

A subset Y of a topological space X is equipped with the *subspace topology*, in which open sets are the restrictions of open sets in X to Y. If (X, d) is a metric space, and Y is its subset, then the subspace topology on Y coincides with the topology of the metric space (Y, d'), where $d' = d|_{Y \times Y}$ is the metric d restricted to Y.

If X and Y are two topological spaces, then we can consider the *product topology* on $X \times Y$, whose basic open sets are the sets of the form $U \times V$, where U is an open set in X and V is an open set in Y. If (X, d) and (Y, d') are metric spaces, then the product topology on $X \times Y$ coincides with the topology induced by the *product metric* over $X \times Y$, which is defined as the coordinatewise maximum of the two metrics d, d'.

Separation axioms A topological space X is called:

- a *discrete space* if the points are open, i.e. for any $x \in X$, the set $\{x\}$ is open
- a *totally disconnected space* if any two distinct points $x, y \in X$ have some disjoint neighborhoods which are clopen
- a T_2 *space*, or a *Hausdorff space*, if any two distinct points $x, y \in X$ have some disjoint neighborhoods
- a T_1 *space* if the points are closed, i.e. for any $x \in X$, the set $\{x\}$ is closed
- a T_0 *space* if any pair of points can be separated by an open set, i.e. if for any two points $x, y \in X$ there exists an open set which contains precisely one of the points x, y.

The classes of topological spaces defined by the above axioms form a hierarchy ordered by inclusion, where the class of discrete topological spaces is the smallest and the class of T_0 spaces is the largest. Each of these inclusions is strict. We also note that any metric space induces a topology which is Hausdorff, which follows immediately from the triangle inequality.

Example 2.1. Consider the set $\overline{\mathbb{N}} = \mathbb{N} \cup \{\omega\}$ with the following metric:

$$d(m,n) = \left|\frac{1}{n+1} - \frac{1}{m+1}\right|,$$

for all $m, n \in \overline{\mathbb{N}}$, where we assume $\frac{1}{\omega+1} = 0$. This metric is depicted in Figure 2.1. A subset F of $\overline{\mathbb{N}}$ is closed if and only if it is finite or contains ω. This metric space is totally disconnected. It is homeomorphic to the subspace $\{1/n : n \in \mathbb{N}\} \cup \{0\}$ of $\mathbb{R}$.

Example 2.2. The *Sierpiński space* is the set $\{1, \omega\}$, with a topology consisting of $\emptyset$, $\{1\}$ and $\{1, \omega\}$. It is not a T_1 space, since $\{1\}$ is not a closed set. It is, however a T_0 space, since the points $1, \omega$ can be separated by the open set $\{1\}$.

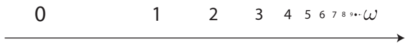


FIGURE 2.1: *The topology of* $\overline{\mathbb{N}}$

The mapping $f\colon \overline{\mathbb{N}} \to \{1, \omega\}$ which maps every finite number to 1 and ω to ω is a continuous mapping, since the inverse images of the open sets are, respectively: $\emptyset, \mathbb{N}$ and $\overline{\mathbb{N}}$, which are all open in $\overline{\mathbb{N}}$.

Convergence and limits A sequence $x_1, x_2, \ldots$ of elements of a topological space X is *convergent* to $x \in X$ if any neighborhood of x contains almost all the elements of the sequence $x_1, x_2, \ldots$ In this definition, the word *neighborhood* could be replaced by *basic neighborhood*.

It should be stressed that, as far as the notion of *convergence* makes sense in any topological space, the notion of *the limit* of a convergent sequence only makes sense in *Hausdorff* topological spaces. Indeed, if $x_1, x_2, \ldots$ is a sequence which is convergent to some element x and also to some other element y, then, provided that X is Hausdorff, x must be equal to y. We then call x the *limit* of the sequence $x_1, x_2, \ldots$ However, in the Sierpiński space from Example 2.2, the sequence $1, 1, 1 \ldots$ is convergent to 1, but it is also convergent to ω.

In a metric space, x is the limit of a sequence of points $x_1, x_2, \ldots$ if and only if their distances to x converge to 0. The closure $\overline{M}$ of a subset M of a metric space is equal to the set of all points which are limits of sequences from M.

Continuity in metric spaces Let (X, d) and (Y, d') be two metric spaces and let $f\colon X \to Y$ be a mapping. Then:

- f is continuous if and only if for every $x \in X$ and for every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $y \in X$ satisfies $d(x, y) < \delta$, then $d'(f(x), f(y)) < \varepsilon$. Equivalently, for any sequence $x_1, x_2, \ldots$ with limit x, the sequence $f(x_1), f(x_2), \ldots$ has limit $f(x)$.
- f is called *uniformly continuous*, if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $x, y \in X$ satisfy $d(x, y) < \delta$, then $d'(f(x), f(y)) < \varepsilon$.
- f is called *non-expansive*, if for every $x, y \in X$

$$d'(f(x), f(y)) \leq d(x, y).$$

Any non-expansive mapping is uniformly continuous and any uniformly continuous mapping is continuous. In general, the converse statements do not hold. Continuous mappings, uniformly continuous mappings and non-expanding mappings are preserved by composition of mappings.

Compactness A topological X space is *compact* if it is Hausdorff and for any family of open sets which covers X, there exists a finite sub-family which covers X. More precisely, if $\{U_i\}_{i \in I}$

is a family of open sets such that

$$\bigcup_{i \in I} U_i = X,$$

then there exists a finite set of indices $\{i_1, \ldots, i_n\} \subseteq I$ such that

$$U_{i_1} \cup \ldots \cup U_{i_n} = X.$$

By complementation, a space X is compact iff it is Hausdorff and for any family $\mathcal{F}$ of closed subsets of X such that every finite subfamily of $\mathcal{F}$ has a nonempty intersection, $\mathcal{F}$ itself has a nonempty intersection.

A metric space is compact if and only if it is *sequentially compact*, i.e. any infinite sequence $x_1, x_2, \ldots$ of points has a subsequence which is convergent.

Compact spaces are important because of their several good properties: a closed subset of a compact space is a compact space, any Cartesian product of any family of compact spaces is compact, and an image of a compact space under a continuous mapping to a Hausdorff space is compact. As a consequence of these facts, a continuous mapping between compact spaces maps closed sets to closed sets. Moreover, if the compact spaces are metric spaces, then the mapping is uniformly continuous.

We say that a subset Y of X is *compact* if Y is compact when the subspace topology is considered. Thus, as mentioned above, any closed subset of a compact set is a compact set. Compact subsets of $\mathbb{R}$ are precisely those which are closed and contained in some interval of bounded length.

Connectedness We say that a subset K of a topological space X is *connected* if K is not contained in a union of two disjoint, open subsets of X, each of which has a nonempty intersection with K. For instance, the subset $\{0, 2\}$ of $\mathbb{R}$ is not connected, since the intervals $(-1, 1)$ and $(1, 2)$ are open, disjoint subsets of $\mathbb{R}$, each intersecting $\{0, 2\}$. It is not difficult to prove that the set of reals $\mathbb{R}$ is connected. In particular, connected sets are not preserved by taking subspaces. However, the image of a connected set under any continuous mapping is again a connected set. In particular, the image of a nonempty connected set under a continuous mapping to a discrete space is a singleton. Observe that a set K is connected if and only if there exists a clopen set U which intersects K, but does not contain K.

Recall that a topological space X is totally disconnected if and only if any pair of its points can be separated by two disjoint clopen sets. This is equivalent to saying that every connected subset of X has at most one point. An example of a totally disconnected space is the set $\mathbb{Q}$ of rational numbers (the topology on $\mathbb{Q}$ is the topology induced by the standard metric over $\mathbb{Q}$). In this thesis, we will mostly deal with compact and totally disconnected metric spaces.

Example 2.3. A canonical example of a compact, totally disconnected metric space is the *Cantor space* C. As a set, C is the set of all infinite binary sequences, $\{0, 1\}^{\mathbb{N}}$. The metric on C is defined by considering the distance between two distinct sequences $a = (a_n)_{n \in \mathbb{N}}$ and $b = (b_n)_{n \in \mathbb{N}}$

defined by

$$d(a,b) \overset{def}{=} 2^{-r} \qquad \textit{where} \quad r = \min\{n : a_n \neq b_n\}.$$

Therefore, two sequences are close to one another if they share a long prefix. If a and b are two sequences which differ at position r then the open balls around a and b of radius 2^{-r} separate a from b. Each of these balls is also a closed set, since it is equal to the closed ball with a corresponding center and radius 2^{-r-1}. Therefore, a and b can be separated by clopen sets, so C is a totally disconnected space. It is also not difficult to prove that it is compact.

2.2 Finite topologies

A *finite topological space* is a topological space with finitely many points, i.e. a finite set X equipped with a topology. Throughout Section 2.2 we assume that X is finite. In this section, a topology on X is specified by the family of closed subsets of X (rather than the family of open subsets of X). Such a family defines the structure of a topological space on X if it is closed under union, intersection and contains $\emptyset$ and X.

Note that if the space X is a T_1 space, i.e. a space in which points are closed, then any subset of X is a closed set as a finite union of closed sets, and so the topology of X is the topology of the discrete space. Therefore, in this section we are mostly interested in finite topological spaces which are not T_1. An example of such a space is the Sierpiński space from Example 2.2.

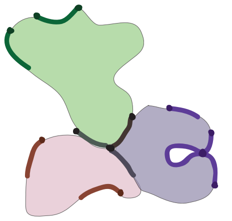


FIGURE 2.2: *A finite topological space. Points are represented by two, one and zero-dimensional areas, and one area is in the closure of another area if it is contained in its boundary. Closed points are precisely the zero-dimensional points.*

Given a finite topological space X, we define its *specialization preorder* by

$$x \leq y \quad \text{iff} \quad x \in \overline{\{y\}}.$$

Equivalently, we may write

$$x \leq y \quad \text{iff} \quad \overline{\{x\}} \subseteq \overline{\{y\}}.$$

It is trivial to verify that the relation $\leq$ is transitive and reflexive, i.e. is a *preorder*. The condition that $\leq$ is antisymmetric is precisely equivalent to the condition that X is a T_0 topological space.

Example 2.4. In the Sierpiński space, the induced preorder results in $\omega < 1$.

We say that a subset $K \subseteq X$ of a partially preordered set is *downward-closed*, if whenever $x \in K$ and $y \in X$ is such that $y \leq x$, then also $y \in K$. It is trivial to check that a subset of X is closed iff it is downward-closed with respect to $\leq$. Conversely, if $(X, \leq)$ is a finite partially preordered set, then we may define a topology on X for which $\leq$ is the specialization preorder, by defining the closed sets as precisely the sets which are downward-closed with respect to the preorder on X. Clearly, downward-closed sets are preserved by unions and intersections, and contain $\emptyset$ and X. Therefore, this yields a valid topology over X.

We say that a mapping $f\colon X \to Y$ of two preordered set is *order-preserving* if whenever $x \leq x'$ in X, then $f(x) \leq f(x')$. Equivalently, the inverse image under f of a downward-closed set in Y is a downward-closed set in X.

It is therefore clear that continuous mappings between finite topological spaces is nothing else than order-preserving mappings between finite preordered sets. The product topology over a Cartesian product of two topological spaces corresponds to the coordinatewise product preorder over the Cartesian product of two preordered sets.

We denote the smallest downward-closed set containing a given set $Y \subseteq X$ by $\downarrow Y$, or by $\downarrow y$ in the case when $Y = \{y\}$.

Corollary 2.1 *There is an isomorphism between the category of finite topological spaces with continuous mappings, and the category of finite partially preordered sets with monotone mappings. Via this isomorphism, T_0 topological spaces correspond to partially ordered sets.*

Because of the above correspondence, we can specify the topology on a finite set by defining a partial preorder on its elements.

CHAPTER 3

Semigroups

A *semigroup* is a set S equipped with a mapping from $S \times S$ to S, denoted $\cdot$ and called the *product* or *multiplication* in S, which is *associative*, i.e. for all $s, t, u \in S$ satisfies

$$(s \cdot t) \cdot u = s \cdot (t \cdot u).$$

The semigroup is *commutative* if moreover

$$s \cdot t = t \cdot s.$$

As an example, consider the set $\mathbb{N}$ equipped with addition as semigroup multiplication, or the set of integers with addition modulo n some finite modulus. These are commutative semigroups. Other examples of semigroups include groups, vector spaces, rings, and many others. A *homomorphism* is a mapping $f\colon S \to T$ between semigroups which preserves multiplication, i.e. for all $s, s' \in S$,

$$f(s \cdot s') = f(s) \cdot f(s').$$

An important example is the *free semigroup* generated by a set A. It is the set A^+ of all nonempty words over the alphabet A, with word concatenation as the semigroup operation. It has the property that any mapping $\alpha\colon A \to S$ from the alphabet A to any semigroup S extends in a unique way to a homomorphism $\tilde{\alpha}\colon A^+ \to S$, by setting

$$\tilde{\alpha}(a_1 a_2 \ldots a_k) = \alpha(a_1) \cdot \alpha(a_2) \cdots \alpha(a_k) \qquad \textit{for } a_1, a_2, \ldots a_k \in A.$$

A *neutral element* in a semigroup S is an element, sometimes denoted 1, which satisfies

$$s \cdot 1 = 1 \cdot s = s \qquad \textit{for } s \in S,$$

and a *zero element* is an element, sometimes denoted 0, which satisfies

$$s \cdot 0 = 0 \cdot s = 0 \qquad \textit{for } s \in S.$$

A semigroup which has a neutral element is called a *monoid*. Not every semigroup has a neutral (or a zero element), but if it doesn't, it is easy to extend it by adding such an element. We denote by S^1 the monoid S with an neutral element added, if necessary. The *free monoid*A^* generated by a set A, is the set A^+ extended by the empty word, denoted ε. It is isomorphic to the monoid $(A^+)^1$.

Ordered semigroups A *preordered semigroup* is a semigroup S equipped with a *preorder* $\leq$ (i.e. a transitive and reflexive relation), which is consistent with multiplication, i.e. for every $s, s', t, t' \in S$,

$$s \leq s' \wedge t \leq t' \implies s \cdot t \leq s' \cdot t'.$$

A *homomorphism* of preordered semigroups is a homomorphism f of semigroups which also preserves the preorder, i.e. $s \leq t \implies f(s) \leq f(t)$. If, in the above definition, "preorder" is replaced by "partial order" or "linear order", we obtain definitions of *partially ordered semigroups* or *linearly ordered semigroups* and of homomorphisms of partially ordered semigroups or of linearly ordered semigroups .

Semirings A semiring is like a ring, except that addition is not required to be invertible. More precisely, a *semiring* is a set equipped with two monoid structures, called *addition* and *multiplication*. Addition is required to be commutative and multiplication is required to distribute over addition, from both sides. Moreover, the neutral element of addition, denoted 0, is required to act as a zero element with respect to multiplication. An example of a semiring is the set $\mathbb{N}$ with usual addition and multiplication. A *homomorphism* of semirings is required to preserve addition and multiplication.

If $(S, \cdot, \leq)$ is a linearly ordered semigroup with a zero element, which is the largest element with respect to $\leq$, then S can be transformed into a semiring $(S, \min, \cdot\,)$, called the *min-semiring* induced by $(S, \cdot, \leq)$, in which $\cdot$ acts as semiring multiplication and min with respect to the linear order acts as semiring addition. Note that a homomorphism of linearly ordered semigroups with a largest zero element is automatically a homomorphism of the induced min-semirings.

For example, by considering the usual linear order over $\overline{\mathbb{N}}$, we can view $(\overline{\mathbb{N}}, \min, +)$ as a min-semiring. In this semiring, addition plays the role of semiring multiplication, and min plays the role of semiring addition.

Dually, assuming that the zero element of S is the smallest element with respect to $\leq$, we could define the induced *max-semiring* $(S, \max, \cdot\,)$.

Semiring of matrices Let S be a semiring with multiplication $\cdot$ and addition $+$. If Q is a finite set, then a $Q \times Q$ *matrix over* S is a matrix indexed by pairs from $Q \times Q$, whose entries are elements of S. We denote by $m[p, q]$ the entry in m at position (p, q), and we call it the entry in row p and column q of m. We denote by $\mathbb{M}_Q S$ the set of $Q \times Q$ matrices over S. It has a semiring structure, defined by:

$$\begin{aligned}(m+n)[p,q] &\overset{def}{=} m[p,q]+n[p,q],\\ (m\ \cdot\ n)[p,q] &\overset{def}{=} \sum_{r\in Q} m[p,r]\cdot n[r,q].\end{aligned}$$

3.1 Semigroups and automata

Transformation semigroups of finite automata Let B denote the *Boolean semiring*, i.e. the set $\{0,1\}$ equipped with $\vee$ as semiring addition and $\wedge$ as semiring multiplication. We describe the well-known description of automata in terms of the semigroup of binary relations over the set of states, seen as matrices over the semiring B.

Let $\mathcal{A}$ be a finite nondeterministic automaton over the alphabet A, with states Q and with transition relation $\delta \subseteq Q\times A\times Q$. Consider the set $\mathbb{M}_Q B$ of $Q\times Q$ matrices over the min-semiring B. Each letter a induces a matrix $\delta_{\mathcal{A}}(a)\in\mathbb{M}_Q B$ defined by

$$\delta_{\mathcal{A}}(a)[p,q] \overset{def}{=} \begin{cases}1 & \text{if } (p,a,q)\in\delta\\ 0 & \text{if } (p,a,q)\notin\delta\end{cases}.$$

The mapping $\delta_{\mathcal{A}}\colon A\to\mathbb{M}_Q B$ extends to a unique homomorphism from the free monoid:

$$\tilde{\delta}_{\mathcal{A}} \quad : \quad A^* \quad \longrightarrow \quad \mathbb{M}_Q B\,.$$

Then, for $w\in A^*$ and $p,q\in Q$, $\tilde{\delta}_{\mathcal{A}}(w)[p,q]=1$ if and only if the automaton $\mathcal{A}$ has a run over the word w from state p to state q. The *transformation semigroup* of the automaton $\mathcal{A}$ is then the image of the mapping $\tilde{\delta}_{\mathcal{A}}$, treated as a semigroup equipped with matrix multiplication. Any finite semigroup is (isomorphic to) a transformation semigroup of some finite automaton.

Syntactic congruence Let S be a semigroup. We will consider *terms* using multiplication and arbitrary elements of S as constants (leafs of the term). The considered terms will have one variable, which appears only once in the term. We denote terms by τ,τ'. Note that a term τ with one free variable can be seen as a mapping $\tau\colon S\to S$.

Let $L\subseteq S$ be any set. We define the equivalence relations $\simeq_1$ and $\simeq_2$ over S as follows.

$$\begin{array}{llcll} x\simeq_1 y & \textit{iff} & u\cdot x\cdot v\in L \iff u\cdot y\cdot v\in L & \textit{for every } u,v\in S^1,\\ x\simeq_2 y & \textit{iff} & \tau(x)\in L \iff \tau(y)\in L & \textit{for every term } \tau \textit{ with one free variable,}\\ & & & \textit{which appears once in } \tau. \end{array}$$

Both equivalence relations $\simeq_1,\simeq_2$ are *congruences* with respect to $\cdot$ i.e. satisfy

$$x\simeq y \quad \textit{and} \quad x'\simeq y' \quad \Longrightarrow \quad x\cdot x'\simeq y\cdot y'.$$

Remark 3.1. The definition of the congruence $\simeq_2$ generalizes naturally to algebras over other signatures – it suffices to consider terms allowing all operations from the signature. However, it is important that in semigroups, this congruence has a simple form given by $\simeq_1$, as stated in the following easy lemma.

Lemma 3.1. *Let $L \subseteq S$. Then the induced congruences $\simeq_1$ and $\simeq_2$ over S, defined above, coincide.*

We call either of the relations $\simeq_1, \simeq_2$ the *syntactic congruence* of the set L, and denote it $\simeq_L$. Clearly, the syntactic congruence *saturates* L, i.e. L is a union of equivalence classes of $\simeq_L$. The *syntactic semigroup* of L is the quotient of $S/\simeq_L$. As always in the case of congruences, the quotient inherits a structure of a semigroup from S.

Example 3.1. Let A be a finite alphabet and let L be a subset of the semigroup A^*. Then the syntactic congruence of L is the classical *Myhill-Nerode equivalence* for languages.

For instance, let $A = \{a, b\}$ and $L \subseteq A^*$ be the set of all words with an even number of a's. Then the syntactic congruence of the set $L \subseteq S$ partitions S into two equivalence classes: L and $S - L$. Its syntactic semigroup is isomorphic to the group of order 2.

The following theorem is classical. A language satisfying any of the equivalent conditions below is called a *regular language.*

Theorem 3.2. *Let $L \subseteq A^*$ be a language. The following conditions are equivalent.*

1. *L is recognized by a finite automaton*
2. *There exists a homomorphism $\alpha \colon A^* \to S$ and a subset $F \subseteq S$, such that $\alpha^{-1}(F) = L$.*
3. *The Myhill-Nerode equivalence of L has finite index*

Proof. *1 ⇒ 2.* If L is recognized by a finite automaton $\mathcal{A}$, then let $\delta_{\mathcal{A}} \colon A^* \to S$ be the homomorphism induced by $\mathcal{A}$, and let $F \subseteq S$ be the set of all matrices m such that $m[p, q] = 1$ for some initial state p and some accepting state q of $\mathcal{A}$. Then, w is mapped to F by α if and only if w is accepted by $\mathcal{A}$. Therefore, $L = \alpha^{-1}(F)$.

2 ⇒ 3. We show that the syntactic congruence of L is coarser than the congruence induced by α. Indeed, assume that $\alpha(x) = \alpha(y)$ for some two words $x, y \in A^*$. Then, for any $u, v \in A^*$,

$$u \cdot x \cdot v \in L \iff \alpha(u)\alpha(x)\alpha(v) \in L \iff \alpha(u)\alpha(y)\alpha(v) \in L \iff u \cdot y \cdot v \in L.$$

Hence, $x \simeq_L y$. Since the congruence induced by α has a finite index and $\simeq_L$ is coarser, it follows that $\simeq_L$ also has a finite index.

3 ⇒ 1. The states of the automaton are the equivalence classes of $\simeq_L$, its initial state is the equivalence class $[\varepsilon]$ of the empty word, its accepting states are the equivalence classes which intersect L, and its transition relation is deterministic, and allows a transition labeled by the letter a from the equivalence class $[w]$ to the equivalence class $[wa]$. □

3.2 Structure of finite semigroups

Idempotents

An element s of a semigroup S is called an *idempotent* if $s \cdot s = s$. We denote idempotents by symbols e, f. Observe that if e is idempotent, then $e = e^2 = e^3 = \ldots$

If $s \in S$, then an *idempotent power* of s is any idempotent element of the form s^k, for some $k \geq 1$. In general, not every element has an idempotent power – consider for instance the element 1 in the semigroup $(\mathbb{N}, +)$ of natural numbers with addition. However, in *finite* semigroups, every element has an idempotent power.

Lemma 3.3. *Let S be a finite semigroup. Then, for any $s \in S$ there exists an idempotent power of s, and it is unique, i.e. if s^k and s^l are both idempotent powers of s then $s^k = s^l$.*

Moreover, if $n = |S|$, then for all $s \in S$, $s^{n!}$ is the unique idempotent power of s.

Proof. For uniqueness, note that if $k, l \geq 1$ are such that s^k and s^l are idempotents, then

$$s^k = (s^k)^l = s^{kl} = (s^l)^k = s^l.$$

Now we will show by induction on $n = |S|$ that for every $s \in S$, $s^{n!}$ is the idempotent power of s. If $|S| = 1$ and $s \in S$, then s must already be an idempotent. Otherwise, consider the set of powers of s, starting from the second power:

$$s^{\geq 2} = \{s^2, s^3, s^4, \ldots\}.$$

If s appears in the set $s^{\geq 2}$, then we have that $s = s^k$ for some $k > 1$. Note that we may assume that $k \leq n+1$, since by the pigeonhole principle, the elements $s^{n+2}, s^{n+3}, \ldots$ must appear earlier in the sequence $s^2, s^3, \ldots$ If $k = 2$, then already s is idempotent, and if $k > 2$, then

$$s^{k-1} \cdot s^{k-1} = s^{2k-2} = s^k \cdot s^{k-2} = s \cdot s^{k-2} = s^{k-1},$$

so s^{k-1} is idempotent. By our assumption $k - 1 \leq n$, so $k - 1$ divides $n!$. It follows that $s^{n!}$ is also idempotent.

The other case is that s does not appear in the set $s^{\geq 2}$. Then $s^{\geq 2}$ is a strict subsemigroup of S, so we can apply the inductive assumption and conclude that $s^{n!}$ is an idempotent power of s^n, so it is also an idempotent power of s. □

We call any number N such that s^N is idempotent for all $s \in S$ an *idempotent exponent* of S. The above lemma says that for a finite semigroup S, the number $|S|!$ is an idempotent exponent of S. We denote by s^ω the unique idempotent power of s. Since homomorphisms map idempotents to idempotents, we deduce the following.

Corollary 3.4. *Assume that $f\colon S \to T$ is a homomorphism of finite semigroups. Then for all $s \in S$,*

$$f(s^\omega) = f(s)^\omega.$$

Green's Relations

If U, V are two subsets of a semigroup, then we write UV for the set of all possible products of the form $u \cdot v$, where $u \in U$ and $v \in V$. The set U^1V then denotes the union $UV \cup V$, and similarly, UV^1 denotes the union $UV \cup U$.

A *two-sided ideal* (respectively, a *left-sided ideal*, a *right-sided ideal*) in a semigroup S is a nonempty subset $T \subseteq S$ such that $S^1TS^1 \subseteq T$ (respectively, $S^1T \subseteq T$, $TS^1 \subseteq T$). A two sided ideal T is *trivial*, if it contains only one element, or if it is equal to T.

Two-sided ideals are important, since if I is such an ideal, then the finest equivalence relation which identifies all the elements of I is a semigroup congruence. We denote by S/I the *quotient* by this congruence. We will often omit the words "two-sided" and refer to I as simply an *ideal* in S.

The smallest two-sided ideal containing an element $s \in S$ is S^1sS^1. The *Green relation* $\mathcal{J}$ is defined as follows: $s\mathcal{J}t$ if and only if $S^1sS^1 = S^1tS^1$, i.e. s and t generate the same two-sided ideal. A $\mathcal{J}$-class is an equivalence class of the relation $\mathcal{J}$. We also consider the *$\mathcal{J}$-preorder* $\leq_{\mathcal{J}}$, which is defined so that $s \leq_{\mathcal{J}} t$ iff $S^1sS^1 \subseteq S^1tS^1$. We define the Green relation $\mathcal{L}$ by writing $s\mathcal{L}t$ if $S^1s = S^1t$, and we define the Green relation $\mathcal{R}$ by writing $s\mathcal{R}t$ if $sS^1 = tS^1$. It follows from the definitions that if $s\mathcal{R}t$ or $s\mathcal{L}t$, then $s\mathcal{J}t$. The *Green relation* $\mathcal{H}$ is the intersection of the relations $\mathcal{L}$ and $\mathcal{R}$, i.e. $s\mathcal{H}t$ if and only if $s\mathcal{L}t$ and $s\mathcal{R}t$.

Green's relations are useful for proving properties of the semigroup by induction on their structure. A simple structural induction might consider the following cases:

S has some nontrivial ideal I Then we prove our property for the quotient S/I, and show how the property can be lifted via the quotient mapping to S.

S has only trivial ideals This means either that the only ideal in S is S itself – we then call S a *simple* semigroup – or that the only ideals in S are S itself and an ideal with only one element. Note that there can be only one ideal with one element, and that its only element is the zero element of the semigroup. If the only ideals in S are S and $\{0\}$, then we can further distinguish two cases. One case is a degenerate case, when the product of any two elements in S is equal to 0 (actually, in this case, S can contain at most one non-zero element, since otherwise there would be a nontrivial ideal in S). Otherwise, we call S a 0-*simple* semigroup.

The inductive analysis therefore boils down to the cases of simple and 0-simple semigroups. Note that in both cases all non-zero elements are in a single $\mathcal{J}$-class, and that $S \cdot S = S$, since $S \cdot S$ is a nonzero ideal in S. There is a theorem, due to Rees [Ree40], which gives a precise characterization of finite simple and 0-simple semigroups. We will not need the full characterization, but we will need some intermediate observations. Although they follow from the results of Rees, and trace their roots to the work of Suschkewitsch [Sus28], we use the language introduced by Green [Gre51].

We turn to an analysis of simple and 0-simple semigroups. Two important facts will play a role for us, proved in Lemma 3.7 and Lemma 3.8, respectively. The first lemma implies that in a simple or 0-simple semigroup, the $\mathcal{H}$-class of a product $s_1 s_2 s_3 \dots s_n$ is determined by s_1 and s_n. The second lemma says that if H is an $\mathcal{H}$-class, then either $(H \cdot H) \cap H = \emptyset$, or H is a group.

We prove the following lemmas, which hold in any finite semigroup. However, we will be mostly interested in the case of simple and 0-simple semigroups.

The lemma below implies that if $u <_{\mathcal{R}} s$, then $u <_{\mathcal{J}} s$.

Lemma 3.5. *For any $s, t \in S$,*

$$\text{if} \quad s \mathcal{J} (s \cdot t) \quad \text{then} \quad s \mathcal{R} (s \cdot t).$$

Proof. Assume that $s\mathcal{J}(st)$, i.e. there exist $u, v \in S^1$ such that $ustv = s$. Therefore,

$$\begin{aligned} s &= u^1 \cdot s \cdot (tv)^1 = u \cdot (ustv) \cdot tv \\ &= u^2 \cdot s \cdot (tv)^2 = u^2 \cdot (ustv) \cdot (tv)^2 \\ &= \dots = \\ &= u^\omega \cdot s \cdot (tv)^\omega = (u^\omega \cdot s \cdot (tv)^\omega)(tv)^\omega = s \cdot (tv)^\omega \\ &= st \cdot ((vt)^{\omega - 1} v). \end{aligned}$$

Therefore, $s \in stS^1$, and obviously, $st \in sS^1$. It follows that $sS^1 = stS^1$, so $s\mathcal{R}st$. □

Corollary 3.6. *If e, f are two idempotents such that $ef = fe = f$ and $e\mathcal{J}f$, then $e = f$. If e, f are two idempotents such that $e\mathcal{H}f$, then $e = f$.*

Proof. For the first part, observe that since $e\mathcal{J}ef$, by the previous lemma, there exists an element u such that $e = (ef)u$. Since $ef = f = ff$, we have that $e = (ff)u = f(fu) = fe = f$.

For the second part, if $e\mathcal{H}f$, then $f = es$ for some $s \in S^1$, and so $f = es = ees = ef$. By symmetry, $f = fe$. Then, by we conclude that $e = f$ by the first part. □

Lemma 3.7. *Assume that $s = s_1 \cdot s_2 \cdots s_n$ and s, s_1, s_n are all $\mathcal{J}$-equivalent. Then, the $\mathcal{H}$-class of s is the intersection of the $\mathcal{R}$-class of s_1 with the $\mathcal{L}$-class of s_n.*

Proof. Since $s\mathcal{J}s_1$, by the previous lemma, $s\mathcal{R}s_1$. By symmetry, $s\mathcal{L}s_n$. By definition, the $\mathcal{H}$-class of s is the intersection of its $\mathcal{L}$-class with its $\mathcal{R}$-class. Therefore, the $\mathcal{H}$-class of s is also the intersection of the $\mathcal{L}$-class of s_n with the $\mathcal{R}$-class of s_1. □

Lemma 3.8. *If H is an $\mathcal{H}$-class such that $H \cdot H$ intersects H, then H is a group.*

Proof. We first show that if $H \cdot H$ intersects H, then H is closed under multiplication, i.e. $H \cdot H \subseteq H$.

Let $p, q \in H$ be such that $pq \in H$. Let $s, t \in H$ be arbitrary. The relation $s\mathcal{L}p$ implies $st\mathcal{L}pt$. Similarly, $t\mathcal{R}q$ implies $pt\mathcal{R}pq$, so $st\mathcal{L}pt\mathcal{R}pq$. This implies that $st\mathcal{J}pq$. Therefore, $st\mathcal{J}s$, so by Lemma 3.5, $st\mathcal{R}s$. By symmetry, $st\mathcal{L}t$. It follows that the $\mathcal{H}$-class of st is H. Therefore, H is closed under multiplication.

Now, we show that multiplication in H is invertible. Let $s, t \in H$. Since $st \in H$, there exists an element $u \in S^1$ such that $stu = s$. Then, by iteratively substituting s by stu, we obtain that

$$s(tu)^\omega = s.$$

We show that $(tu)^\omega \in H$. Clearly, $(tu)^\omega \geq_{\mathcal{J}} s$ and $(tu)^\omega \leq_{\mathcal{J}} t$. Since $s\mathcal{J}t$ it follows that $(tu)^\omega \mathcal{J} t$. Then, by Lemma 3.5, we deduce that $(tu)^\omega \mathcal{R} t$. Then, $(tu)^\omega \mathcal{J} t \mathcal{J} s = s(tu)^\omega$, so from a symmetric version of Lemma 3.5, we deduce that $(tu)^\omega \mathcal{L} s$. Altogether, it follows that $(tu)^\omega \in H$.

Let $r = (ut)^{\omega-1}u$. Then, it is easy to see that $(tu)^\omega \mathcal{H} r$, so $r \in H$. Moreover, $str = s(tu)^\omega = s$ and tr and rt are both idempotents in H. By Corollary 3.6, tr and rt are both the unique idempotent of H.

We have therefore shown that for every $s, t \in H$ there exists an $r \in H$ such $tr = rt = e$ is the unique idempotent in H and that $str = s$. Then, $utr = u$ for every $u \in H$, since u can be written as vs, and $utr = vstr = vs = u$. We denote the element r by t^{-1}.

From what we have shown, $te = t$ and $t\,t^{-1} = t^{-1}t = e$ for every $t \in H$. It follows that for every $t \in H$, $et = t\,t^{-1}t = te = t$. Therefore, $(H, \cdot\,, {}^{-1}, e)$ is a group. □

3.3 Topological semigroups

Topological algebras We use terminology from universal algebra, which includes the notions of a *signature* κ and an *algebra* over the signature κ. For instance, any semigroup is an algebra over the signature $\kappa = \langle \cdot \rangle$, and any group is an algebra over the signature $\kappa = \langle \cdot\,, {}^{-1}, 1\rangle$, where $\cdot$ is a binary function, ${}^{-1}$ is a unary function and 1 is a function of arity 0. (Note that the converse statements are not true, since semigroups and groups additionally require some axioms.)

We say that $\mathcal{S}$ is a *topological algebra* over the signature κ if it has a structure of a topological space and of an algebra over the signature κ, and the instantiations of the operations in κ are continuous mappings over $\mathcal{S}$. A *homomorphism* of two topological algebras over the signature κ is a mapping which is both continuous and a homomorphism of algebras. If a homomorphism of topological algebras is a homeomorphism, then we say that it is an *isomorphism*.

Topological semigroups A *topological semigroup* is a topological algebra over the signature $\langle \cdot \rangle$, which is also a semigroup, i.e. the mapping $\cdot$ is associative. Similarly we define topological monoids, topological semirings, topological groups, etc. A mapping f of two topological semigroups is a *homomorphism* if it is a homomorphism of topological algebras, and f is an *isomorphism* if it is an isomorphism of topological algebras.

Example 3.2. The semigroup $(\overline{\mathbb{N}}, +)$ is a topological semigroup (or monoid), if we consider the topology over $\overline{\mathbb{N}}$ defined in Example 2.1 – addition is a continuous mapping from $\overline{\mathbb{N}} \times \overline{\mathbb{N}}$ to $\overline{\mathbb{N}}$. The reals with addition $(\mathbb{R}, +)$ and the reals with multiplication $(\mathbb{R}, \cdot)$ are both topological monoids, and $(\mathbb{R}, +, \cdot\,, 0, 1)$ is a topological field.

Finite topological semigroups

A *finite topological semigroup* is a topological semigroup S, whose set of elements is finite. As a consequence of Corollary 2.1, finite topological semigroups correspond precisely to finite preordered semigroups. We describe this correspondence in more detail.

Let $(S, \ \cdot\ )$ be a finite topological semigroup and let $\leq$ be its specialization preorder. We claim that $(S, \ \cdot\ , \leq)$ is a preordered semigroup. By assumption, the mapping $\cdot \colon S \times S \to S$ is a continuous mapping. Therefore, it is an order-preserving mapping, i.e. if $(s, t) \leq (s', t')$ in $S \times S$, then $s \cdot t \leq s' \cdot t'$. Hence, $(S, \cdot, \leq)$ is a preordered semigroup. Conversely, in the same way we can show that if $(S, \cdot, \leq)$ is a preordered semigroup, then multiplication is continuous with respect to the topology on S induced by the preorder. Note that again, T_0 topological semigroups correspond to partially ordered semigroups, and T_1 topological semigroups correspond to semigroups ordered by the identity relation.

Example 3.3. Consider the Sierpiński space $\{1, \omega\}$ from Example 2.2. We impose on it a semigroup structure, in which 1 is the neutral element and 0 is the zero element. Then, $\{1, \omega\}$ is a finite topological semigroup. Note that the mapping from $(\overline{\mathbb{N}}, +)$ to $\{1, \omega\}$, which maps all finite numbers to 1 and ω to ω, is a homomorphism of topological semigroups.

CHAPTER 4

Profinite semigroups

4.1 Metrizable profinite semigroups

Projective sequences We fix a set of indices $I = \{1, 2, 3, \ldots\}$, equipped with the natural linear order. We call a *projective sequence* an infinite family $\{S_i\}_{i \in I}$ of semigroups equipped with a family of *connecting homomorphisms* $\{\alpha_i\}_{i \in I}$, as depicted below.

$$S_1 \xleftarrow{\alpha_1} S_2 \xleftarrow{\alpha_2} S_3 \xleftarrow{\alpha_3} S_4 \leftarrow \ldots. \tag{1}$$

We will say that a sequence $s = (s_1, s_2, s_3, \ldots)$ such that $s_i \in S_i$, is *consistent* if $\alpha_i(s_{i+1}) = s_i$ for $i \in I$. We call s_i the *component* of s in S_i. Let S denote the set of consistent sequences. We call S the *projective limit* of the projective sequence (1). For $i \in I$, there is a canonical projection which maps $s \in S$ to its component in S_i, and is a homomorphism of semigroups.

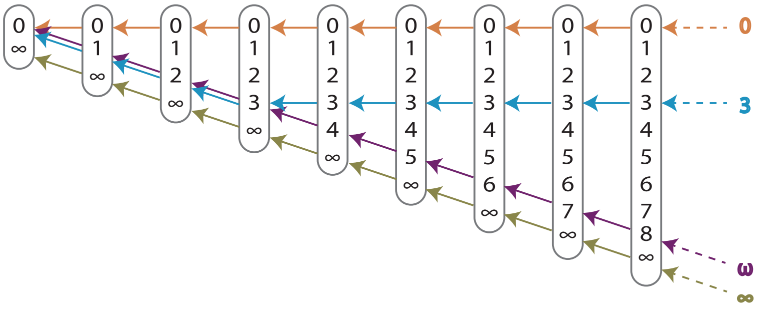


FIGURE 4.1: *A projective sequence and some consistent sequences*

The projective limit S carries a semigroup structure, defined by coordinatewise multiplication:

$$(s_1, s_2, s_3, \ldots) \cdot (t_1, t_2, t_3, \ldots) = (s_1 \cdot t_1, s_2 \cdot t_2, s_3 \cdot t_3, \ldots).$$

The product of two consistent sequences is again a consistent sequence, by assumption that the connecting mappings are homomorphisms.

Remark 4.1. In a more general definition (see e.g. [Alm05]), one considers *projective systems*, in which the set of indices is assumed to be any directed set I. The notion of a projective limit generalizes to such systems. However, for our needs, it is sufficient to consider limits of projective sequences.

Example 4.1. For $i \in I$, let $\mathcal{T}_{/i=\omega}$ denote the set $\{0, 1, \ldots, i-2, i-1, \omega, \perp\}$. We equip $\mathcal{T}_{/i=\omega}$ with an associative and commutative operation $+$, which, for $k, l \in \{0, 1, 2, \ldots, i-1, \omega\}$ acts as addition up to threshold $i-1$, i.e. for example $(i-1)+1 = \omega = \omega + 1$. Moreover,

$$k + \perp = \perp + k = \perp$$

for all $k \in \mathcal{T}_i$. Thus, in $\mathcal{T}_{/i=\omega}$, the element ω represents finite numbers greater than $i-1$, while the element $\perp$ denotes an undefined value.

For $i \in I$, let $\alpha_{(i=\omega)} \colon \mathcal{T}_{/i+1=\omega} \to \mathcal{T}_{/i=\omega}$ be the mapping which maps the element i to ω, and leaves the other elements unchanged. Consider the projective sequence

$$\mathcal{T}_{/1=\omega} \xleftarrow{\alpha_{(1=\omega)}} \mathcal{T}_{/2=\omega} \xleftarrow{\alpha_{(2=\omega)}} \mathcal{T}_{/3=\omega} \leftarrow \ldots .$$

Some consistent sequences of this projective sequence are depicted in Figure 4.1. We denote the consistent sequences $0, 1, 2, \ldots, \omega, \perp$, with a natural meaning. The projective limit of the above projective sequence is therefore

$$\mathcal{T} = \{0, 1, 2, \ldots, \omega, \perp\}.$$

We denote the semigroup operation in $\mathcal{T}$ by $+$, since it acts as addition over the elements $0, 1, 2, \ldots \in \mathcal{T}$. If s or t is an element of the set $\{\omega, \perp\}$, then $s + t$ is equal to the maximum of the two elements s, t with respect to the ordering $0 < 1 < \ldots < \omega < \perp$.

The canonical projection from $\mathcal{T}$ to $\mathcal{T}_{/i=\omega}$ maps all elements $i, i+1, i+2, \ldots, \omega$ to ω, and leaves the others unchanged.

Metric structure We equip the projective limit S of the sequence (1) with a distance. The idea is that two consistent sequences s, t are similar, if they share a long prefix. Therefore, we define the distance by

$$d(s,t) \overset{def}{=} 2^{-r} \qquad \textit{where} \quad r = \min\{i : s_i \neq t_i\}. \tag{2}$$

for $s \neq t$, and obviously we put $d(s,t) = 0$ if $s = t$.

Proposition 4.1. *The distance d turns S into a compact metric space. Multiplication is continuous with respect to this metric, so S is also a compact topological semigroup. The canonical projections from S to S_i are also continuous homomorphisms, where S_i is considered with the discrete topology.*

Proof. Symmetry of the distance is obvious. The definition makes it is impossible that $d(s,t) = 0$ for $s \neq t$. It therefore suffices to verify the triangle inequality. We prove that d is even a *ultrametric*, i.e. it satisfies

$$d(s,u) \leq \max\{d(s,t), d(t,u)\}$$

for any s, u, t which are consistent families. This follows immediately from the fact that for any i, if $s_i \neq u_i$ then $s_i \neq t_i$ or $t_i \neq u_i$.

We check that multiplication is continuous in S. This is true because if $d(s, s') \leq 2^{-n}$ and $d(t', t') \leq 2^{-n}$, then for any $i < n$, $s_i = s'_i$ and $t_i = t'_i$, so obviously $s_i \cdot t_i = s'_i \cdot t'_i$. This implies that $d(s \cdot t, s' \cdot t') \leq 2^{-n}$ as well. Hence multiplication is even uniformly continuous.

Also, for each $i \in I$, the canonical projection from S to S_i is uniformly continuous, since if $d(s, s') < 2^{-i}$, then by definition, $s_i = s'_i$.

It remains to prove that the metric over S is compact. Let $s^1, s^2, \ldots$ be an infinite sequence of elements of S. We must show that this sequence has a subsequence which is convergent to some $s \in S$. Let s^n be the consistent sequence $\{s^n_i\}_{i \in I}$.

For $i \in I$, we will say that a sequence $s^1, s^2, \ldots$ of elements of S is *ultimately constant* with respect to the coordinate i, if $s^k_i = s^{k+1}_i = s^{k+2}_i = \ldots$ for some $k \in \mathbb{N}$.

Let $i \in I$. Since S_i is finite, by the pigeonhole principle, we can choose an infinite subsequence of $s^1, s^2, \ldots$ which is ultimately constant with respect to S_i. Since the set I is countable, we can use the diagonal construction, and repeat this procedure indefinitely, obtaining an infinite subsequence $s^{n_1}, s^{n_2}, \ldots$ which is ultimately constant with respect to every coordinate $i \in I$.

Without loss of generality, we assume that the original sequence $s^1, s^2, \ldots$ has already the property that it is ultimately constant with respect to every coordinate $i \in I$, i.e. for every coordinate $i \in I$ there exists an element s_i such that $s^k_i = s_i$ for almost all k. We define the "suspected" limit consistent sequence in the obvious way, as $s = \{s_i\}_{i \in I}$. Then, clearly, for any given n, if k is sufficiently large, then s^k agrees with s on all the coordinates $1, 2, 3, \ldots, n$, implying that $d(s^k, s) < 2^{-n}$. Hence, the sequence $s^1, s^2, \ldots$ converges to s. This finishes the proof of compactness of S. □

Example 4.2. One can get a sense of the topology of the projective limit $\mathcal{T}$ from Example 4.1 by looking at Figure 4.2: the element $\omega \in \mathcal{T}$ is closer to $3 \in \mathcal{T}$ than to $0 \in \mathcal{T}$, since the corresponding sequences share a longer prefix. However, all elements are equally far away from $\perp \in \mathcal{T}$.

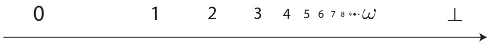


FIGURE 4.2: *The topology and order of the profinite semigroup $\mathcal{T}$*

Remark 4.2. In the definition of the metric, we used the function $r \mapsto 2^{-r}$, which, as we saw, didn't play a big role in the proof. If we considered instead any other monotonically decreasing function, we would still obtain an ultrametric, yielding an isomorphic topological semigroup.

Definition 1. We will call a topological semigroup which is isomorphic to a projective limit of a projective sequence of semigroups a *metrizable profinite semigroup*.

Remark 4.3. As already mentioned, in general, projective systems can be indexed by arbitrary directed sets. The corresponding projective limits are then general *profinite semigroups*. We will later on see a justification for our use of the attribute "metrizable" – a general profinite semigroup S is a metrizable profinite semigroup if and only if there exists a metric which induces the topology of S. Note that not every profinite semigroup is metrizable.

Remark 4.4. Note that the assumption that the projective sequence consists of semigroups and their homomorphisms was used only to conclude that its projective limit also carries a semigroup structure. Therefore, the notion of a *metrizable profinite semigroup* extends without any difficulty to other algebraic objects – instead of considering projective limits of semigroups, we might as well consider projective limits of projective sequences of sets, or of groups, or of semirings, etc., where the mappings of the projective sequences (or systems in general) are required to be homomorphisms of the respective algebraic structures. We can then analogously as above define (metrizable) *profinite sets*, *profinite groups*, *profinite semirings*, etc., which are then compact topological algebras in the respective categories. For instance, the Cantor space is a metrizable profinite set.

Example 4.3. The semigroup $\mathcal{T}$ is a metrizable profinite semigroup. Actually, it is even a metrizable profinite linearly ordered monoid and min-semiring. Indeed, each of the semigroups $\mathcal{T}_{/i=\omega}$ which was considered in the previous example possesses a natural linear order, in which $0 < 1 < \ldots < i-1 < \omega < \perp$. The semigroup operation is compatible with this order, giving rise to a linearly ordered semigroup. Moreover, each mapping from $\mathcal{T}_{/i+1=\omega}$ to $\mathcal{T}_{/i=\omega}$, for $i \in I$ clearly preserves the order. Therefore, the projective sequence is actually a projective sequence of linearly ordered monoids, and also of the induced min-semirings. Hence, its projective limit $\mathcal{T}$ is a linearly ordered metrizable profinite monoid, in which $0 < 1 < \ldots < \omega < \perp$, and also a metrizable profinite semiring, where addition plays the role of semiring multiplication and min plays the role of semiring addition.

Example 4.4. For $i \in I$, let Z_i be the cyclic group of order 2^i. The elements of Z_i are thus integers with addition, modulo 2^i. For each $i \in I$ there is a natural homomorphism from Z_{i+1} to Z_i, which maps an integer k modulo 2^{i+1} to the same number k, modulo 2^i. Let Z denote the projective limit of this projective sequence. Then Z is a metrizable profinite group.

If $b_1, b_2, b_3, \ldots$ is a sequence of 0's and 1's, then by taking $k_1 = 1$ and for each $n \geq 1$, $k_{n+1} = k_n + b_n \cdot 2^n$, we obtain a sequence $k_1, k_2, \ldots$ which is a family consistent with the projective sequence. This gives a precise correspondence between the elements of Z and infinite binary sequences. This correspondence is actually a homeomorphism between Z and the Cantor space, considered in Example 2.3. The elements of Z are called *2-adic integers*. In particular, there are uncountably many 2-adic integers.

Note that the considered projective sequence is not only a projective sequence of semigroups, but also of rings, where Z_n is equipped with multiplication modulo 2^n. Therefore, the projective limit Z is actually a metrizable profinite ring, and, in particular, a metrizable profinite semiring.

The ω-power

We now see that the notion of an idempotent power considered for finite semigroups can be extended to metrizable profinite semigroups – any element $s \in S$ will induce a unique idempotent s^{ω}, which is not strictly an idempotent power of s, but can be arbitrarily closely approximated by powers of s. We proceed to the formal construction.

Let S be a metrizable profinite semigroup, specified as a projective limit of a projective sequence $S_1 \xleftarrow{\alpha_1} S_2 \xleftarrow{\alpha_2} \ldots$ of finite semigroups. Recall that in each finite semigroups S_i, for every element s_i there is a unique idempotent power s_i^{ω} of s_i, and that $\alpha_i(s_{i+1}^{\omega}) = \alpha_i(s_{i+1})^{\omega}$ by Corollary 3.4. It follows that if $s \in S$, i.e. $s = (s_1, s_2, \ldots)$ is a consistent sequence of elements, then

$$s^{\omega} \quad \overset{def}{=} \quad (s_1^{\omega}, s_2^{\omega}, \ldots)$$

is also a consistent sequence of elements. We call the operation $s \mapsto s^{\omega}$ over a metrizable profinite semigroup the *ω-power*. It follows from the definitions that if S is a finite semigroup, treated as a discrete profinite semigroup, then the ω-power of an element s is the same as its unique idempotent power.

There is a related operation to the ω-power in profinite semigroups, denoted $s \mapsto s^{\omega-1}$. We first define this operation for finite semigroups. Let S be a finite semigroup, and let N be its idempotent power, so that $s^{\omega} = s^N$ for every $s \in S$. Then, for any given $s \in S$, the element $u = s^{N-1} \cdot s^N$ clearly satisfies:

$$u \cdot s = s^{\omega} \tag{3}$$

$$u \cdot s^{\omega} = u. \tag{4}$$

Moreover, an element $u \in S$ satisfying the above two equations is unique, since

$$u \overset{(4)}{=} u \cdot s^{\omega} = u \cdot s^N = u \cdot s \cdot s^{N-1} \overset{(3)}{=} s^{\omega} \cdot s^{N-1} = s^N \cdot s^{N-1}.$$

For $s \in S$, we denote by $s^{\omega-1}$ the unique element which satisfies the equations (3) and (4).

It follows from uniqueness that if $\alpha \colon S \to T$ is a homomorphism of finite semigroups, then $\alpha(s^{\omega-1}) = \alpha(s)^{\omega-1}$. Hence, if S is a metrizable profinite semigroup, specified as a projective limit of a projective sequence $S_1 \xleftarrow{\alpha_1} S_2 \xleftarrow{\alpha_2} \ldots$ of finite semigroups, and $s = (s_1, s_2, \ldots)$ is a consistent sequence of elements, then

$$s^{\omega-1} \quad \overset{def}{=} \quad (s_1^{\omega-1}, s_2^{\omega-1}, \ldots)$$

is also a consistent sequence of elements. Moreover, it follows that $s^{\omega-1}$ can be also specified as the unique element $u \in S$ which satisfies the equations (3) and (4).

Remark 4.5. Similarly, one defines the ω-power and the operation $s \mapsto s^{\omega-1}$ in general profinite semigroups. All the properties of the ω-power listed further on hold also for general profinite semigroups.

Example 4.5. Let $\mathcal{T}$ be the profinite semigroup from the previous examples. Clearly, for any i, in the semigroup $\mathcal{T}_{/i=\omega}$, the only idempotent elements are $0, \omega$ and $\perp$, and for any element $k \in \mathcal{T}_{/i=\omega}$ where $1 \leq k \leq \omega$, the idempotent power of k is ω. It follows that in the profinite semigroup $\mathcal{T}$, the ω-power maps 0 to 0 and $\perp$ to $\perp$, and any other element is mapped to ω.

Proposition 4.2. *Let S be a metrizable profinite semigroup. The ω-power is continuous mapping from S to S and satisfies the following properties.*

$$s^\omega \cdot s^\omega = s^\omega \tag{5}$$

$$s \cdot (t \cdot s)^\omega = (s \cdot t)^\omega \cdot s \tag{6}$$

$$(s^n)^\omega = s^\omega \qquad \textit{for } n = 1, 2, 3 \ldots \tag{7}$$

$$(s^\omega)^\omega = s^\omega \tag{8}$$

Moreover,

$$\lim_{n \to \infty} s^{n!} = s^\omega \tag{9}$$

Proof. Assume that S is defined as a projective limit of a projective sequence $S_1 \leftarrow S_2 \leftarrow \ldots$

Continuity follows just as in the case of multiplication, or even simpler – if $s, t \in S$ are such that $d(s,t) \leq 2^{-n}$, then for every $i < n$, $s_i = t_i$, so also $s_i^\omega = t_i^\omega$ and hence $d(s^\omega, t^\omega) \leq 2^{-n}$, proving that the ω-power is non-expanding, and in particular, continuous.

To prove the five listed properties of the ω-power, we first show that they all hold in finite semigroups. Assume that S is a finite semigroup. Then the equation (5) is immediate, since the ω-power of an element s of a finite semigroup is its idempotent power. To verify the equation (6), we just note that it obviously holds by associativity if we replace ω by the idempotent exponent of S. The last equations (7) and (8) also follow easily.

The equation (9) follows from the fact that if n is larger than the size of S, then $n!$ is an idempotent exponent of S. Therefore, $s^{n!} = s^\omega$ for sufficiently large n.

Now, from the fact that the equations (5)-(9) hold for all finite semigroups, it follows that they also hold in the profinite semigroup S. Indeed, let $s = (s_1, s_2, \ldots)$ be a consistent sequence, and consider for instance the equation (5). By what we have proved for finite semigroups, for all $i \in I$,

$$s_i^\omega \cdot s_i^\omega = s_i^\omega.$$

But this means that the element $s^\omega \cdot s^\omega$ of S is the same as the element s^ω, proving the equation (5). The other equalities (6)-(8) follow in the same way.

To prove the equation (9), take any $k \geq 0$. We will show that for sufficiently large n, the sequences $s^{n!}$ and s^ω share a prefix of length at least k. Indeed, it suffices to consider a number n larger than the size of the semigroups $S_1, S_2, \ldots, S_k$. Then, for each of these semigroups, $n!$ is its idempotent exponent. It follows that the sequences s^ω and $s^{n!}$ share a prefix of length k, so $d(s^{n!}, s^\omega) < 2^{-k}$ for sufficiently large n. This proves equation (9). □

4.2 General profinite semigroups

In this section, we give an alternative definition of a general profinite semigroup, without using the notions of projective limits. The equivalence of the two definitions is due to Numakura [Num57].

Residually free profinite semigroups Let S be a topological semigroup. We say that two elements $s, t \in S$ can be *distinguished* by a finite semigroup T, if there is a continuous homomorphism $\varphi\colon S \to T$ such that $\varphi(s) \neq \varphi(t)$. Here, we assume that the topology of T is that of a discrete topological space. We say that S is *residually finite* if any pair of distinct elements of S can be distinguished by a finite semigroup.

Note that any metrizable profinite semigroup is residually finite, thanks to Proposition 4.1.

Example 4.6. As an example of a topological semigroup which is *not* residually finite, consider the semigroup $(\mathbb{R}, +)$, with the usual topology. Indeed, since $\mathbb{R}$ is connected, any mapping from $\mathbb{R}$ to a discrete space is a constant mapping, which cannot distinguish any two points in $\mathbb{R}$.

The following lemma implies that any compact totally disconnected topological semigroup is residually finite. Note that, as far as the syntactic congruence can be defined in any algebraic structure, the following lemma is specific to semigroups, in which the syntactic congruence has a simple form as described in Lemma 3.1.

Lemma 4.3 (Hunter's Lemma). *Let S be a compact topological semigroup and let L be a clopen subset of S. Then the equivalence classes of $\simeq_L$ are clopen. In particular, $\simeq_L$ has finite index.*

Proof. We will show that the $\simeq_L$-equivalence class of any element x is closed and open. The first property holds in abstract algebras, if we consider $\simeq_L$ as the congruence $\simeq_2$. However, the property of being open relies on the description of $\simeq_L$ in terms of the relation $\simeq_1$, and is specific to semigroups. Below, we prove directly that the class of x is clopen.

An element y is *not* $\simeq_L$-equivalent to x if there exist $u, v \in S$ such that precisely one of the elements $u \cdot x \cdot v$, $u \cdot y \cdot v$ belongs to L. This leads to considering the following subsets of $S \times S \times S$.

$$\begin{aligned} M_1 &\overset{def}{=} \{(u,y,v) : \quad (u \cdot y \cdot v \in L) \wedge (u \cdot x \cdot v \notin L)\}, \\ M_2 &\overset{def}{=} \{(u,y,v) : \quad (u \cdot y \cdot v \notin L) \wedge (u \cdot x \cdot v \in L)\}. \end{aligned}$$

By continuity of multiplication, and because both L and $S - L$ are clopen, it follows that the sets M_1, M_2 are clopen subsets of $S \times S \times S$. Hence, so is their union, denoted N.

Let π_2 be the projection onto the second coordinate of $S \times S \times S$. Then, $\pi_2(N)$ is precisely the complement of the $\simeq_L$-equivalence class of x. We shall conclude the lemma, by proving that $\pi_2(N)$ is clopen.

The space $S \times S \times S$ is compact, as a Cartesian product of such spaces. Therefore the set N is compact, as a closed set in a compact space. As any projection, π_2 maps open sets to open sets, and is continuous, so it maps compact sets to compact sets. Therefore, the set $\pi_2(N)$ is both a compact and open subset of S, so it is clopen.

Hence, we have proved that the $\simeq_L$-equivalence class of x, i.e. the set $S - \pi_2(N)$ is clopen. Because $\simeq_L$-equivalence classes form an partition of S, and they are all open, it follows from compactness that $\simeq_L$ has finite index. □

General profinite semigroups We will provide an alternative definition of a general profinite semigroup, which explains the chosen terminology for metrizable profinite semigroups. First, we prove the following, standard theorem (see [Alm05, Theorem 3.1] for instance).

Theorem 4.4. *Let S be a compact topological semigroup. Then the following conditions are equivalent.*

1. *S is residually finite*

2. *S is totally disconnected*

Proof. 1 ⇒ 2. Let U be a connected subset of S. Assume that s, t are two points in U. We will show that then $s = t$. Let $\varphi\colon S \to T$ be any continuous homomorphism to a finite semigroup, with the discrete metric. Since φ is continuous, the image of the connected set U is connected in T. But the only connected nonempty subsets of T singletons, so φ does not distinguish s from t. Since S is residually finite, this must mean that $s = t$.

2 ⇒ 1. Let s, t be two distinct elements in S. Since S is totally disconnected, there exists a clopen set $U \subseteq S$ which contains s and does not contain t. Then, s is not related to t under the syntactic congruence $\simeq_U$ of the set U. By Lemma 4.3, $\simeq_U$ has finitely many equivalence classes and each of them is clopen. Therefore, the quotient mapping from S to $S/{\simeq_U}$ is a homomorphism to a finite semigroup, which is continuous and distinguishes s from t. □

Definition 2. A *profinite semigroup* is a compact topological semigroup which satisfies either of the two equivalent conditions of Theorem 4.4.

From Proposition 4.1 it follows immediately that any metrizable profinite semigroup (according to Definition 1) is a profinite semigroup. We prove that conversely, any profinite semigroup which can be equipped with a metric which is compatible with its topology is indeed a metrizable profinite semigroup, justifying the chosen terminology.

Proposition 4.5. *If S is a profinite semigroup and S can be equipped with a metric compatible with its topology, then S is a metrizable profinite semigroup.*

Proof. We will actually use an a priori weaker assumption, that the topology of S has a countable basis consisting of clopen sets. It is a consequence of a general and simple fact from topology (see e.g. [Wil70, Theorem 29.7]), that any compact totally disconnected metric space has a countable basis which consists of clopen sets.

Let $U_1, U_2, \ldots$ be all the elements of the base, and we assume that they are all clopen. We will construct a projective sequence, whose limit is isomorphic to the topological semigroup S.

For $n = 1, 2, \ldots$, let $\simeq_n$ be the intersection of the syntactic congruences of the first n sets of the sequence $U_1, U_2, \ldots$ By Hunter's lemma, the congruences induced by the sets $U_1, U_2, \ldots$

have finitely many clopen equivalence classes, so $\simeq_n$ also has finitely many clopen equivalence classes, as a finite intersection of such congruences. We define S_n as the quotient of S by the congruence $\simeq_n$. Since for each n, $\simeq_{n+1}$ is finer than $\simeq_n$, it follows that there is a natural projection from S_{n+1} to S_n, which is a semigroup homomorphism. Our projective sequence is then

$$S_1 \leftarrow S_2 \leftarrow S_3 \leftarrow \ldots,$$

equipped with the natural projections.

It remains to prove that the projective limit S' of this sequence is isomorphic to S. There is a natural mapping φ from S to S', which maps an element S to the sequence of its equivalence classes with respect to $\simeq_1, \simeq_2, \ldots$ The mapping φ is clearly a homomorphism, and is continuous. Indeed, if n is any number, then any two elements which are $\simeq_n$-equivalent are mapped to the same element in S_n, so continuity follows from the fact that $\simeq_n$ has finitely many clopen equivalence classes.

The mapping φ is moreover bijective. To see that, let $s_1, s_2, s_3, \ldots$ be a consistent sequence in S'. For each $n \in \mathbb{N}$, s_n is an equivalence class of the congruence $\simeq_n$, and s_{n+1} is contained in s_n. Therefore, $s_1, s_2, s_3, \ldots$ is a descending sequence of nonempty closed sets in a compact space, so their intersection is nonempty. Let $\psi(s_1, s_2, s_3, \ldots)$ denote any element in this intersection. Then, it is easy to see that ψ and φ are mutual inverses.

Therefore, φ is a continuous bijection of compact metric spaces. It follows that φ is a homeomorphism of topological spaces. Moreover, φ is a homomorphism of semigroups. Therefore, φ is an isomorphism of topological semigroups. □

Corollary 4.6. *The closure $\overline{T}$ of a subsemigroup T of a metrizable profinite semigroup S is a metrizable profinite semigroup.*

Proof. Note that the closure of T is a semigroup – from continuity of multiplication in S it follows that if $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ are two convergent sequence in T, then the sequence $x_1 \cdot y_2, x_2 \cdot y_2, \ldots$ is also a convergent sequence in T. Therefore, $\overline{T}$ is a closed subsemigroup of S. It is therefore compact, metrizable and residually-finite. □

4.3 Profinite words

Fix a finite alphabet A. We consider regular languages over A, which sometimes will be seen as subsets of A^*, and sometimes as subsets of A^+, obtained by simply removing the empty word if necessary.

Consider a sequence of finite words over the alphabet A,

$$w_1,\ w_2,\ w_3, \ldots$$

We will say that the sequence of words is *convergent* if for every given regular language $L \subseteq A^+$, there exists a position $n \in \mathbb{N}$ such that all the words $w_n, w_{n+1}, \ldots$ either all belong to L, or all belong to the complement of L.

It makes sense to say whether a convergent sequence *ultimately belongs* to a regular language L – we just check whether this property holds for sufficiently large elements of the sequence. We say that two convergent sequences are *equivalent* if for every regular language L, either both sequences ultimately belong to L, or none of them ultimately belongs to L.

Example 4.7. Any sequence of words which is constant, or ultimately constant, is convergent. The sequence of words

$$a,\ a^2,\ a^3,\ a^4,\ \ldots$$

is *not* convergent. Indeed, let L be the regular language of words of even length. Then, every second word in the above sequence belongs to L. The sequence of words

$$a,\ a^{2!},\ a^{3!},\ a^{4!},\ \ldots$$

is convergent. To see that, note that if L is a regular language accepted by an automaton $\mathcal{A}$, and $m, n \geq |\mathcal{A}|$, then $\mathcal{A}$ cannot distinguish between $a^{n!}$ and $a^{m!}$, so both words either belong to L, or do not belong to L.

Similarly, the sequence of words

$$(aa),\ (aa)^{2!},\ (aa)^{3!},\ (aa)^{4!},\ \ldots$$

is also convergent, and equivalent to the previous one.

Definition 3. The *free profinite semigroup*, denoted $\widehat{A^+}$, is the set of equivalence classes of convergent sequences. We call the elements of $\widehat{A^+}$ *profinite words* over the alphabet A, and denote them using symbols $x, y, z, \ldots$ Thus, a profinite word is an equivalence class of convergent sequences. The *free profinite monoid*, denoted $\widehat{A^*}$, is the set $\widehat{A^+} \cup \{\varepsilon\}$.

It is not difficult to verify that if $w_1, w_2, \ldots$ and $v_1, v_2, \ldots$ are two convergent sequences, then

$$w_1 \cdot v_1, w_2 \cdot v_2, \ldots$$

is also a convergent sequence. This defines a multiplication operation on convergent sequences. It is also easy to see that this multiplication operation is a congruence with respect to equivalence of convergent sequences. Therefore, the free profinite semigroup $\widehat{A^+}$ indeed has a structure of a semigroup.

We now specify a topology over $\widehat{A^+}$. We introduce a notion of distance between profinite words. Let us fix a notion of *size* of a regular language L. We will consider this to be the size of the smallest semigroup recognizing L, i.e. the size of the syntactic semigroup of L, or equivalently, the index of the Myhill-Nerode equivalence of L. However, one could equally well define the size of L to be the size of the smallest automaton recognizing L, or the size of the smallest regular expression describing L, – all these notions lead to the same topology over $\widehat{A^+}$.

We say that a regular language L *distinguishes* two profinite words x and y if precisely one of the two profinite words ultimately belongs to L, in the sense described before. We define the *distance* between x and y as

$$d(x,y) = 2^{-r} \qquad \textit{where } r = \min\{\textit{size}(L) : \textit{ L distinguishes x from y}\}.$$

This value evaluates to 0 if and only if $x = y$. It is straightforward to check that this distance defines a metric over $\widehat{A^+}$.

There is a natural embedding of A^+ which maps a word w to the equivalence class of the sequence which is constantly equal to w. We may thus view A^+ as a subset of $\widehat{A^+}$.

The following properties of the free profinite semigroup are crucial. The last property is called the *universal property of the free profinite semigroup*.

Proposition 4.7. *1. $\widehat{A^+}$ is a metrizable profinite semigroup*

2. The set of finite words A^+ is dense in $\widehat{A^+}$

3. If $\alpha\colon A^+ \to S$ is a homomorphism from A^+ to a metrizable profinite semigroup S, then there exists a unique extension of α to a continuous mapping $\hat{\alpha}\colon \widehat{A^+} \to S$. Moreover,

- *The mapping $\hat{\alpha}$ is a homomorphism*
- *The image of $\hat{\alpha}$ is the closure of the image of α in S*
- *For any clopen set $F \subseteq S$, $\hat{\alpha}^{-1}(F) = \overline{\alpha^{-1}(F)}$.*

Sketch of proof. It is quite clear that multiplication is continuous with respect to the metric on $\widehat{A^+}$. Moreover, the metric is compact. To prove this, we observe that for each n there are only finitely many distinct regular languages of size at most n, and then use the pigeonhole argument similar as in Proposition 4.1. Moreover, the metric is totally disconnected – this follows just as in the case of the Cantor set, considered in Example 2.3. Therefore, $\widehat{A^+}$ is a metrizable profinite semigroup, by Proposition 4.5.

To prove the density of A^+ in $\widehat{A^+}$, we need to show that any profinite word $x \in \widehat{A^+}$ is in the closure of A^+. Let x be the equivalence class of a convergent sequence $w_1, w_2, \ldots$. We will show that in fact the sequence of finite words $w_1, w_2, \ldots$ is convergent to x. Let N be an arbitrary number. Let $L_1, L_2, \ldots, L_k$ be all the regular languages of size at most N. Since the sequence $w_1, w_2, \ldots$ is convergent, it follows that there is a position m such that for all $1 \leq i \leq k$, all the words $w_m, w_{m+1}, w_{m+2}, \ldots$ either belong to the language L_i, or belong to its complement. Then the distance between the word w_m and the profinite word x is smaller than 2^{-N}, since no regular language of size at most N can distinguish them.

We move on to proving the universal property of the free profinite semigroup, i.e. the third condition of the proposition.

We consider the metric over A^+ which is the restriction of the metric over $\widehat{A^+}$. First, we show that any mapping from A^+ to a metrizable profinite semigroup S is uniformly continuous. Assume that S is defined as the projective limit of a projective sequence

$$S_1 \leftarrow S_2 \leftarrow S_3 \leftarrow \ldots$$

Let $\varepsilon > 0$ be given, and let $n \in \mathbb{N}$ be such that $2^{-n} < \varepsilon$. Let K be larger than the sizes of the semigroups $S_1, S_2, \ldots, S_n$. Then, by definition of the metric over A^+, it follows that if w, v are two finite words with $d(w,v) < 2^{-K}$ for $i = 1, 2, \ldots, n$, then the semigroup S_i does not distinguish w from v, so the first n components of $\varphi(w)$ and $\varphi(v)$ coincide. In particular, if $d(w,v) < 2^{-K}$, then $d(\varphi(w), \varphi(v)) < 2^{-n} < \varepsilon$. Hence, φ is uniformly continuous.

The rest of the proof uses only arguments from general topology. By uniform continuity of φ, it is easy to see that any sequence of elements of A^+ which is convergent to some element of $\widehat{A^+}$ is mapped to a convergent sequence in S. Therefore, by density of A^+ in $\widehat{A^+}$, the mapping α extends to a unique continuous mapping $\hat{\alpha}\colon \widehat{A^+} \to S$, which is moreover a semigroup homomorphism (from continuity of multiplication in S, it follows that if w_n and v_n are two convergent sequences, then $\varphi(w_n \cdot v_n)$ is convergent to the same limit as $\varphi(w_n) \cdot \varphi(v_n)$). It is also clear from the definition, that the image of $\hat{\alpha}$ is contained in the closure of the image of α. Since $\hat{\alpha}$ is continuous and defined over the compact set $\widehat{A^+}$, its image is also compact, so it must be equal to the closure of the image of α.

It remains to show that if F is clopen in S, then

$$\hat{\alpha}^{-1}(F) = \overline{\alpha^{-1}(F)}.$$

Since $\hat{\alpha}^{-1}(F)$ is closed, the right-to-left inclusion is immediate. We prove the left-to-right inclusion. Assume that $\hat{\alpha}(x) \in F$. Since $A^+ = \alpha^{-1}(F) \cup \alpha^{-1}(S - F)$, and A^+ is dense in $\widehat{A^+}$, it follows that either $x \in \overline{\alpha^{-1}(F)}$ or that $x \in \overline{\alpha^{-1}(S - F)}$. The second possibility is impossible, since it would imply that $\hat{\alpha}(x) \in S - F$, by continuity of $\hat{\alpha}$ and the fact that $S - F$ is closed. Therefore, $x \in \overline{\alpha^{-1}(F)}$. □

Corollary 4.8. *For any finite set A with at least one element, the free profinite semigroup $\widehat{A^+}$ is uncountable.*

Proof. Let Z denote the set of 2-adic integers. It is a metrizable profinite semigroup, which is uncountable. Moreover, the semigroup generated by 1 is dense in Z. Let $\hat{\alpha}$ be the continuous extension to $\widehat{A^+}$ of the homomorphism $\alpha\colon A^+ \to Z$ which maps every element in A to $1 \in Z$. Then, the image of $\hat{\alpha}$ is equal to Z, so it is uncountable. Hence, $\widehat{A^+}$ must be uncountable. □

Regular languages and clopen sets

We say that a set $L \subseteq \widehat{A^+}$ is *recognizable* if there exists a *continuous* homomorphism $\alpha\colon \widehat{A^+} \to S$ to a finite semigroup S (with the discrete topology) and a subset F of S, such that $L = \alpha^{-1}(F)$.

Theorem 4.9. *Let $L \subseteq \widehat{A^+}$. Then the following conditions are equivalent.*

1. *$L \cap A^+$ is a regular language and $L = \overline{L \cap A^+}$*
2. *L is recognizable*
3. *L clopen*
4. *the syntactic congruence $\simeq_L$ of L has finite index and $L = \overline{L \cap A^+}$*

Proof. *1 ⇒ 2.* Since $L \cap A^+$ is regular, there exists a homomorphism $\alpha\colon A^+ \to S$ to a finite semigroup (the transformation semigroup of an automaton, for instance), and a set $F \subseteq S$, such that $L \cap A^+ = \alpha^{-1}(F)$. By Proposition 4.7, there exists a unique continuous homomorphic extension of α to a mapping $\hat{\alpha}\colon \widehat{A^+} \to S$, and moreover

$$\hat{\alpha}^{-1}(F) = \overline{\alpha^{-1}(F)} = \overline{L \cap A^+} = L,$$

where the last equality follows from the assumption. Therefore, L is recognizable.

2 ⇒ 3. This is because any subset of a discrete space is clopen, and the inverse image of a clopen subset under a continuous mapping is clopen.

3 ⇒ 4. Assume that L is clopen. From Lemma 4.3 it follows that $\simeq_L$ has a finite index.

We show that $L = \overline{L \cap A^+}$. The right-to-left inclusion follows from the assumption that L is closed. Therefore, we only need to show that if $x \in L$, then arbitrarily close to x there is a word $w \in A^+ \cap L$. Since L is open, it follows that there exists an $\varepsilon > 0$ such that the ε-ball around x is contained in L. Since A^+ is dense in $\widehat{A^+}$, it follows that arbitrarily closely to x we can find a word $w \in A^+$. If $d(w, x) < \varepsilon$ then additionally $w \in A^+ \cap L$. This proves that $L \subseteq \overline{L \cap A^+}$.

4 ⇒ 1. Recall that the syntactic congruence is a congruence over $\widehat{A^+}$. Let $\alpha\colon \widehat{A^+} \to S$ be the quotient mapping. In particular, α is a homomorphism of semigroups.

Let $F \subseteq S$ be such that $L = \alpha^{-1}(F)$. Let $\beta\colon A^* \to S^1$ be the restriction of α to A^+, extended to ε by putting $\beta(\varepsilon) = 1 \in S^1$. Then, β is a homomorphism from A^* to a finite semigroup, and

$$\beta^{-1}(F) = \alpha^{-1}(F) \cap A^* = L \cap A^+.$$

Therefore, $L \cap A^+$ is a regular language, by Theorem 3.2. □

PART I

Limitedness and stabilization semigroups

CHAPTER 5

Overview

In this chapter, we give an overview of the results we will be proving in Part I of this thesis. We will use distance automata and the tropical semiring for illustrating the basic concepts and proof techniques.

Consider the distance automaton $\mathcal{A}$ over the alphabet $\{a, b\}$, which is depicted in Figure 5.1.

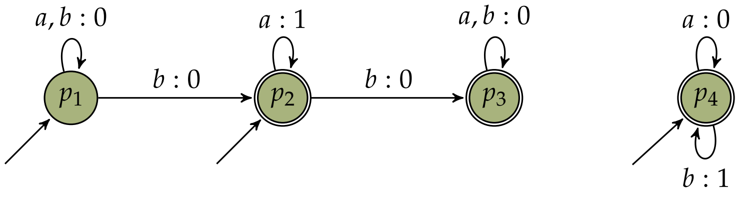


FIGURE 5.1: *A distance automaton $\mathcal{A}$*

Each input word w induces a 4×4 *transition matrix* $\delta_{\mathcal{A}}(w)$. The entry at position $[i, j]$ of $\delta_{\mathcal{A}}(w)$ is the minimal sum of weights in a run over w, which starts in the state p_i and ends in the state p_j. If there is no such run, the entry is equal to $\perp$. For the automaton from the diagram, the word *abaaba* induces the following matrix.

$$\delta_{\mathcal{A}}(abaaba) = \begin{bmatrix} 0 & 1 & 2 & \perp \\ \perp & \perp & 1 & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & 2 \end{bmatrix}$$

For instance, the entry at position $[1, 2]$ is at most 1 because of the run

$$p_1 \xrightarrow{a:0} p_1 \xrightarrow{b:0} p_1 \xrightarrow{a:0} p_1 \xrightarrow{a:0} p_1 \xrightarrow{b:0} p_2 \xrightarrow{a:1} p_2,$$

and it is not equal to 0, because every run from p_1 to p_2 over *abaaba* has at least one increment.

Now, let p_1, p_2, p_4 be the initial states and p_2, p_3, p_4 be the accepting states of $\mathcal{A}$. We will call a position $[i, j]$ in a matrix an *accepting position* if p_i is an initial state and p_j is an accepting state of $\mathcal{A}$. (In the matrices depicted below on this page, accepting positions lie on the intersections of the highlighted columns and rows.) The minimal value of an accepting run over a word w, denoted $f_{\mathcal{A}}(w)$, is equal to the minimal value at an accepting position of $\delta_{\mathcal{A}}(w)$.

Assume that $\mathcal{A}$ is limited. This means that there is some bound N such that for every accepted word w, there is an accepting run of value smaller than N. This can be checked effectively for a fixed bound N, since it can be expressed as a problem of finite automata, counting up to N. Therefore, to prove limitedness of $\mathcal{A}$, it suffices to check if it is limited by 1, then check if it is limited by 2, etc. However, in the example given above, this algorithm will never terminate, since $\mathcal{A}$ is not limited. So we need a way of verifying effectively that $\mathcal{A}$ is not limited. How can we find a witness for non-limitedness of $\mathcal{A}$?

Non-limitedness of $\mathcal{A}$ is equivalent to the existence of a sequence of words $w_1, w_2, \ldots$ such that the values $f_{\mathcal{A}}(w_1), f_{\mathcal{A}}(w_2), \ldots$ are finite and converge to ω. For example, if we take w_n to be $(a^n b)^n a^n$, then the corresponding matrix is

$$\delta_{\mathcal{A}}(w_n) = \begin{bmatrix} 0 & n & n & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & n \end{bmatrix}$$

and so $f_{\mathcal{A}}(w_n) = n$. In particular, $\mathcal{A}$ is not limited, since $\lim_{n\to\infty} f_{\mathcal{A}}(w_n) = \omega$. However, we would like a *finite* witness for non-limitedness of $\mathcal{A}$. Note that the sequence of matrices $\delta_{\mathcal{A}}(w_1), \delta_{\mathcal{A}}(w_2), \ldots$ converges to the matrix

$$x = \lim_{n\to\perp} \delta_{\mathcal{A}}(w_n) = \begin{bmatrix} 0 & \omega & \omega & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & \omega \end{bmatrix}$$

To simplify notation, let us denote $u = \delta_{\mathcal{A}}(a)$ and $v = \delta_{\mathcal{A}}(b)$, and let $A = \{u, v\}$. Note that the mapping $\delta_{\mathcal{A}}$ is a homomorphism from the semigroup $\{a, b\}^+$ to the semigroup of matrices over the semiring $\mathcal{T}$. Let A^+ denote the set of matrices generated from A by matrix multiplication.

The matrix x above has the following two properties: its minimal entry at an accepting position is equal to ω, and x is in the closure $\overline{A^+}$ of A^+. The first property is trivial to check. But how can we effectively determine whether a given matrix x belongs to the set $\overline{A^+}$? In the case of our particular matrix x, it follows from the definition of the sequence $w_1, w_2, \ldots$ that

$$x = \lim_{n\to\infty} (u^n v)^n v^n,$$

so x indeed lies in the closure of the set A^+. But not every convergent sequence of matrices has such a simple description. In fact, there are usually uncountably many convergent sequences of matrices from A^+.

In general, not always a sequence of matrices of the form $y, y^2, y^3, \ldots$ is convergent; however, a sequence of matrices of the form $y, y^{2!}, y^{3!}, \ldots$ is always convergent, and its limit is y^ω, the ω-power of y. This is a useful property of profinite semigroups, and the semigroup of matrices over the semiring $\mathcal{T}$ is a profinite semigroup. Moreover, it is easy to compute the ω-power of a given matrix y. Another useful property of profinite semigroups is that multiplication is continuous. It follows that

$$x = (u^\omega \cdot v)^\omega \cdot v^\omega.$$

The above formula is a finite witness of the fact that $x \in \overline{A^+}$. This witness is a term which uses multiplication, the ω-power and the matrices u, v, and it evaluates to x. It is not a big surprise that x has a witness which is such a term, since we defined it as a limit of matrices of the form $(u^n v)^n u^n$. What is surprising, however, is that *every* limit matrix $y \in \overline{A^+}$ has some finite witness of this form. This is proved in the following theorem, which follows from our results, and generalizes the characterization given by H. Leung.

Theorem 5.1. *Let Q be a finite set and let A be a finite set of $Q \times Q$ matrices over the semiring $\mathcal{T}$. Then*

$$\overline{A^+} = A^{\langle \,\cdot\, , \omega \rangle},$$

where $A^{\langle \,\cdot\, , \omega \rangle}$ denotes the set of matrices which can be obtained by multiplication and the ω-power from matrices in A.

What is really useful, and is the hard part of the proof, is the left-to-right inclusion.

To demonstrate the power of the above theorem, notice that it yields a straightforward, though inefficient procedure for deciding limitedness of a given distance automaton $\mathcal{A}$. The procedure runs two algorithms in parallel. The first algorithm checks, for every finite number $N \in \mathbb{N}$, whether $\mathcal{A}$ is limited by N. If $\mathcal{A}$ is limited by some number N, then the algorithm halts with success. The second algorithm enumerates all the elements x of $A^{\langle \,\cdot\, , \omega \rangle}$, where A is the set of transition matrices of the form $\delta_{\mathcal{A}}(a)$, induced by all letters a in the input alphabet. If it enumerates an element x such that the minimal entry at an accepting position in x is ω, then the algorithm halts with failure.

A more efficient algorithm also easily follows from the above theorem, and will be described later.

B-automata Now, suppose that we are given not a distance automaton, but a B-automaton. These differ from distance automata by allowing resets and several counters. Both these new features lead to difficulties which we now describe. In a distance automaton, a single run can be represented by a single natural number, corresponding to the sum of the weights. In B-automata, this is no longer the case, and both resets and multiple counters contribute to this problem. Assuming that we have only one counter, but allowing resets, the relevant information about a single run is either a natural number describing the number of increments in the run, or – in case some resets where performed – a triple: the longest block of increments not interrupted by a reset, the number of increments after the last reset, and finally the number of increments before the first reset.

Next, if we allow multiple counters, we need to store such a triple for every counter separately. Therefore, the relevant information about a single run is a vector of natural numbers. A problem which arises is when two runs merge. In distance automata, we simply "remember" only the smaller of the two values corresponding to the runs, relying on the fact that natural numbers are linearly ordered. This approach, however, does not work when dealing with vectors. How can we tell if it is better to increment a counter 100 times and the reset it, or to reset it and then increment 100 times? Or how can we tell if a run which resets counter c and increments 10 times counter d is better then a run which does the opposite? This lack of a linear order leads to technical difficulties. However, by using more complicated data structures and semigroups, these difficulties can be resolved.

The main result The main theorem of Part I of this thesis is a theorem which generalizes Theorem 5.1 stated above, by replacing the semiring $\mathcal{T}$ by a more complex semiring, suited for B-automata. The semiring, and the theorem, are described in Section 7.1. The proof of the main theorem is in Section 7.2. From this result, decidability of the limitedness problem for B-automata easily follows. This implies decidability of the emptiness problem for nested distance desert automata of D. Kirsten, which are a special case of B-automata. Limitedness of D. Kirsten's automata is the heart of his proof of the decidability of the star height problem. Therefore, together with D. Kirsten's elegant and simple reduction of the star height problem to the limitedness problem of nested distance desert automata, our result implies decidability of the star height problem.

Our proof technique generalizes the techniques of H. Leung, I. Simon and D. Kirsten. In Section 6.1, we define a notion of a *stabilization semigroup*, designed to reflect the properties of the ω-power of a profinite semigroup in a finite algebraic object. Then, in Section 6.3, we generalize the Factorization Theorem of I. Simon to abstract stabilization semigroups. This is the key technical tool used in the proof of the main theorem, proved in Chapter 7.

In the following section, we introduce some notation, basic concepts and their properties, exemplified by the classical setting of distance automata and the tropical semiring.

5.1 The basic semigroups

We recall the semiring $\mathcal{T}$, which was introduced by H. Leung, extending the tropical semiring of I. Simon. First, we view $\mathcal{T}$ as a linearly ordered semigroup, whose underlying ordered set is:

$$0 < 1 < 2 < \ldots < \omega < \bot.$$

The semigroup operation in $\mathcal{T}$, which we denote by $+$, works as follows:

$$s + t = \begin{cases} s + t & \text{if } s, t \in \mathbb{N} \\ \max(s, t) & \text{if } s \text{ or } t \text{ belongs to } \{\omega, \perp\} \end{cases} .$$

Since $\mathcal{T}$ is a linearly ordered semigroup, with the zero element $\perp$ being the largest element, it also has a structure of a min-semiring, where the role of semiring addition is played by the operation min and role of semiring multiplication is the operation $+$.

The approximative semirings We introduce some semirings related to the semiring $\mathcal{T}$. The following notation is borrowed from [Pin98]. Let k, l be two elements of the semiring $\mathcal{T}$. By $\mathcal{T}_{/k=l}$ we denote the quotient of the semiring $\mathcal{T}$ by the finest congruence of ordered semigroups, which identifies k with l. We denote by $\alpha_{(k=l)}$ the quotient mapping. In particular, $\alpha_{(k=l)}$ is a homomorphism of ordered semigroups, and, consequently, of semirings, where min is semiring addition. The elements in $\mathcal{T}_{/k=l}$ are denoted by their smallest representatives in $\mathcal{T}$, except for the class of ω, which is denoted ω.

For every $N = 1, 2, \ldots$, there are two finite semirings of interest: $\mathcal{T}_{/N=\omega}$ and $\mathcal{T}_{/N=N+1}$. The elements of $\mathcal{T}_{/N=\omega}$ are

$$0 < 1 < 2 < \ldots < N - 1 < \omega < \perp,$$

and $(N-1) + 1 = \omega$. The elements of $\mathcal{T}_{/N=N+1}$ are

$$0 < 1 < 2 < \ldots < N < \omega < \perp,$$

and $N + 1 = N$. These semirings and the quotient mappings are depicted in Figure 5.2, in the case $N = 3$. From this figure it is visible that the mapping $\alpha_{(3=\omega)}$ is continuous, unlike the mapping $\alpha_{(3=4)}$, at least if the discrete topology is considered on $\mathcal{T}_{/3=4}$. This is because the sequence $3, 4, 5, 6, \ldots$ is convergent to ω, and is mapped by $\alpha_{(3=4)}$ to the sequence $3, 3, 3, 3, \ldots$ which is not convergent to the image of ω.

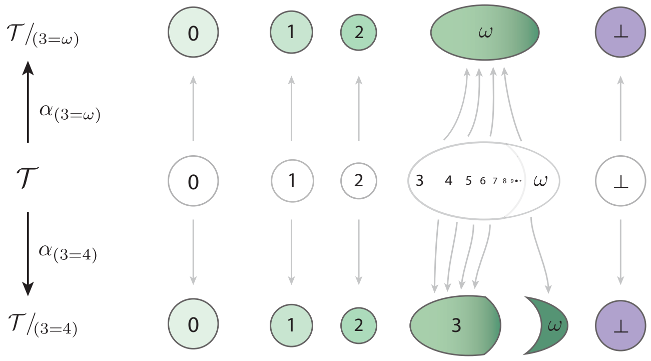


FIGURE 5.2: *The mappings $\alpha_{(3=\omega)}$ and $\alpha_{(3=4)}$.*

Since the congruence in $\mathcal{T}$ identifying N with $N+1$ is finer than the congruence identifying N with ω, it follows that the mapping $\alpha_{(N=\omega)}$ factorizes through the mapping $\alpha_{(N=N+1)}$, i.e. there exists a semigroup homomorphism which makes the diagram in Figure 5.3 commute. We denote the resulting mapping again by $\alpha_{(N=\omega)}$, since it further identifies N with ω in $\mathcal{T}_{/N=N+1}$. Similarly, the homomorphism $\alpha_{(N=\omega)}\colon \mathcal{T} \to \mathcal{T}_{/N=\omega}$ factorizes through the homomorphism $\alpha_{(N+1=\omega)}\colon \mathcal{T} \to \mathcal{T}_{/N+1=\omega}$.

$$\begin{tikzcd}
& \mathcal{T}_{/N=N+1} \arrow[rd, dashed, "\alpha_{(N=\omega)}"] & \\
\mathcal{T} \arrow[ru, "\alpha_{(N=N+1)}"] \arrow[rr, "\alpha_{(N=\omega)}"] & & \mathcal{T}_{/N=\omega}
\end{tikzcd}$$

FIGURE 5.3: *The factorization of $\alpha_{(N=\omega)}$ via $\alpha_{(N=N+1)}$.*

The approximative semigroups of matrices Since the mappings $\alpha_{(N=\omega)}$ and $\alpha_{(N=N+1)}$ are homomorphisms of semirings, it follows that they lift to homomorphisms of the semigroups of $Q \times Q$ matrices:

$$\begin{aligned}
\alpha_{(N=\omega)} &: \mathbb{M}_Q\mathcal{T} \longrightarrow \mathbb{M}_Q\mathcal{T}_{/N=\omega} \\
\alpha_{(N=N+1)} &: \mathbb{M}_Q\mathcal{T} \longrightarrow \mathbb{M}_Q\mathcal{T}_{/N=N+1}.
\end{aligned}$$

We use the same symbols for the induced homomorphisms. Note that on a computational level, the mapping $\alpha_{(N=\omega)}$ simply replaces all finite entries greater than $N-1$ by ω, and the mapping $\alpha_{(N=N+1)}$ replaces all finite entries greater than N by N. Note also that the commutative diagram from Figure 5.3 lifts to a commutative diagram on the level of matrices.

The profinite structure Recall that $\mathcal{T}$ has a structure of a metrizable profinite semigroup, which was considered in Example 4.1 in the Preliminaries. There, $\mathcal{T}$ was defined as the projective limit of the semigroups $\mathcal{T}_{/N=\omega}$. The semigroup $\mathbb{M}_Q\mathcal{T}$ of matrices over the semiring $\mathcal{T}$ also has a structure of a profinite semigroup, which we now describe. We may view the set of $Q \times Q$ matrices as a Cartesian product of $|Q|^2$ copies of $\mathcal{T}$, and consider the product metric over this set. The resulting metric is compact, as a Cartesian product of compact metrics. Moreover, a product of totally disconnected topological spaces is again a totally disconnected topological space (this is because a product of clopen sets is a clopen set, so in the product space two points again can be separated by clopen sets). Finally, matrix multiplication is continuous with respect to this metric, as it is described by a formula using the two continuous semiring operations in $\mathcal{T}$. Therefore, $\mathbb{M}_Q\mathcal{T}$ is a totally disconnected, compact metrizable topological semigroup, so it is a metrizable profinite semigroup by Proposition 4.5.

5.2 Limitedness of distance automata

The crucial idea in H. Leung's solution to the limitedness problem for distance automata is to use the approximative homomorphism $\alpha_{(1=2)}$. We sketch roughly the reasoning and recall the key result.

Let $\mathcal{A}$ be a distance automaton and let $A \subseteq \mathbb{M}_Q\mathcal{T}$ be the set of matrices determined via $\delta_{\mathcal{A}}$ by each letter of the input alphabet. Recall that in order to decide non-limitedness of $\mathcal{A}$, it suffices to decide whether there exists a matrix $x \in \overline{A^+}$ which has entries equal to ω at some positions. This can be determined basing on the image $\alpha_{(1=2)}(\overline{A^+})$ of A^+.

To compute the set $\alpha_{(1=2)}(\overline{A^+})$, H. Leung proposed considering a *stabilization* operation, denoted #, and which he described as follows. Assume that $s \in \mathbb{M}_Q\mathcal{T}$ is a matrix, such that $\alpha_{(1=2)}(s)$ is an idempotent in $\mathbb{M}_Q\mathcal{T}_{/1=2}$. H. Leung proved that in this case, the sequence of matrices $s, s^2, s^3, \ldots$ converges to a matrix s^* in $\mathbb{M}_Q\mathcal{T}$. Its image $\alpha_{(1=2)}(s^*)$ is then defined to be the stabilization $\alpha_{(1=2)}(s)^\#$ of $\alpha_{(1=2)}(s)$. It depends only on $\alpha_{(1=2)}(s)$, and not on s.

Example 5.1. Let s be equal to the matrix $\delta_{\mathcal{A}}(a)$, for $\mathcal{A}$ considered in the overview. Then, for $n = 1, 2, \ldots$

$$s^n = \delta_{\mathcal{A}}(a^n) = \begin{bmatrix} 0 & \perp & \perp & \perp \\ \perp & n & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & 0 \end{bmatrix}$$

For each n, the image of s^n under $\alpha_{(1=2)}$ is a matrix e which looks just like the matrix s^1. In particular, e is idempotent, since

$$e^2 = \alpha_{(1=2)}(s)^2 = \alpha_{(1=2)}(s^2) = e.$$

However, the stabilization of e is a different matrix:

$$e^\# = \alpha_{(1=2)}(s^*) = \begin{bmatrix} 0 & \perp & \perp & \perp \\ \perp & \omega & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & 0 \end{bmatrix}$$

The characterization given by H. Leung [Leu98] is as follows.

Theorem 5.2. *Let Q be a finite set and let $A \subseteq \mathbb{M}_Q\mathcal{T}$ be a finite set of $Q \times Q$ matrices over the semiring $\mathcal{T}$. Then,*

$$\alpha_{(1=2)}(\overline{A^+}) = \alpha_{(1=2)}(A)^{\langle \cdot, \# \rangle}.$$

As a consequence, limitedness is decidable for distance automata.

Above, $\alpha_{(1=2)}(A)^{\langle \cdot, \# \rangle}$ denotes the closure of $\alpha_{(1=2)}(A)$ under stabilization and multiplication in the semigroup $\mathbb{M}_Q\mathcal{T}_{/1=2}$.

The above result has been proved in various forms by I. Simon, H. Leung, D. Kirsten and others. It also follows immediately from Theorem 5.1. This is because for $s \in \mathbb{M}_Q\mathcal{T}$ such that

$\alpha_{(1=2)}(s)$ is idempotent, s^* is precisely equal to the ω-power of s, so

$$\alpha_{(1=2)}(s)^{\#} = \alpha_{(1=2)}(s^{\omega}).$$

From this observation, and from Theorem 5.1, it follows that

$$\alpha_{(1=2)}(\overline{A^+}) = \alpha_{(1=2)}(A^{\langle\,\cdot\,,\omega\rangle}) = \alpha_{(1=2)}(A)^{\langle\,\cdot\,,\#\rangle}.$$

Algorithm for the limitedness problem Theorem 5.2 implies decidability of the limitedness problem for distance automata: to test if $\mathcal{A}$ is not limited, using a fix-point algorithm, compute the set $\alpha_{(1=2)}(A)^{\langle\,\cdot\,,\#\rangle}$, and check whether it contains a matrix s such that $\min_{p,q} s[p,q] = \omega$ where p ranges through the initial states and q ranges through the accepting states of $\mathcal{A}$. The presented algorithm works in exponential space, since there are only $4^{|Q|^2}$ matrices in $\mathbb{M}_Q\mathcal{T}_{/1=2}$.

However, as proved by D. Kirsten [Kir05], the fix-point computation stabilizes after at most $|Q|-1$ steps, where each step consists of closing under stabilization, and then under multiplication. From this, it is not difficult to deduce that the computation can be carried out by a nondeterministic algorithm working in polynomial space. By Savitch's theorem, this implies the existence of a PSPACE-algorithm for solving the limitedness problem. This complexity bound is also optimal. For details, see [Kir05].

CHAPTER 6

Stabilization semigroups and the Factorization Theorem

In this chapter, we introduce an abstract notion of a *stabilization semigroup*, in which multiplication and stabilization are required to satisfy certain axioms, resembling the properties of the ω-power in profinite semigroups. The semigroup $\mathbb{M}_Q\mathcal{T}_{/1=2}$ with its stabilization operation is an example of a stabilization semigroup. In Section 6.3 we develop a technical tool – a factorization theorem for stabilization semigroups. It is our key technical tool for proving in Chapter 7 the main theorem, which generalizes Theorem 5.1 and implies the decidability of the limitedness problem of B-automata.

6.1 Stabilization semigroups

Stabilization semigroups are finite algebraic objects, designed to capture the essence of the ω-power of profinite semigroups. There is already one definition of stabilization semigroups, provided in [Col09], but ours differs slightly.

To avoid confusion with the ω-power in profinite semigroups, stabilization in an abstract stabilization semigroup will be denoted by the symbol #. Recall from Proposition 4.2 that in any profinite semigroup, the ω-power satisfies the following relations (below, we write # instead of ω):

$$
\begin{aligned}
(s^n)^\# &= s^\# && \textit{for } n = 1, 2, \ldots && \text{(A1)}\\
(s\,t)^\# s &= s(t\,s)^\# && && \text{(A2)}\\
s^\# s^\# &= s^\# && && \text{(A3)}\\
\left(s^\#\right)^\# &= s^\# && && \text{(A4)}\\
e\,e^\# &= e^\# && \textit{if } e \textit{ is idempotent} && \text{(A5)}
\end{aligned}
$$

A *stabilization semigroup* is a semigroup S equipped with a unary operation $s \mapsto s^{\#}$ called *stabilization*, which satisfies the above axioms. Intuitively, stabilization represents the limit of a sequence of a growing numbers of iterations. Note that (A2) and (A5) imply a stronger version of (A5):

$$e\,e^{\#} = e^{\#}e = e^{\#} \qquad\qquad \textit{if } e \textit{ is idempotent.} \tag{A5'}$$

A *homomorphism* $\varphi\colon S \to T$ of stabilization semigroups is a mapping which is a homomorphism of semigroups, such that $\varphi(s^{\#}) = \varphi(s)^{\#}$ for all $s \in S$.

Example 6.1. As suggested before, all profinite semigroups are examples of stabilization semigroups, when equipped with stabilization which is equal to the ω-power. In particular, the semigroup $\mathcal{T}$ is naturally equipped with a stabilization operation. So is the set $\mathbb{M}_Q\mathcal{T}$ of $Q \times Q$ matrices over the semigroup $\mathcal{T}$. Note that this is an advantage of the profinite approach, since we do not need to define stabilization in $\mathbb{M}_Q\mathcal{T}$ and check its algebraic properties by hand, as was done by H. Leung and I. Simon.

A continuous homomorphism φ of profinite semigroups is a homomorphism of stabilization semigroups, since we have:

$$\varphi(x^{\omega}) = \varphi(\lim_{n\to\infty} x^{n!}) = \lim_{n\to\infty} \varphi(x)^{n!} = \varphi(x)^{\omega}.$$

Example 6.2. Consider the finite semigroup $S = (\{0,1\}, \max)$. There are two different stabilization operations over S. One is the trivial stabilization, which is the identity mapping. The other one is the stabilization which maps 0 to 1 and 1 to 1. It is easy to check, using axiom (A5') that these are the only two possible ways of equipping S with a stabilization operations.

Example 6.3. More generally, any finite semigroup S has one distinguished *trivial stabilization* operation, which is the identity over idempotents, and maps s to the idempotent power of s. This is a special of Example 6.1, since any finite semigroup is also a profinite semigroup.

Example 6.4. Wilke algebras [Wil93] can be seen as precisely the stabilization semigroups which satisfy the identity $s^{\#} \cdot t = s^{\#}$. The elements of the form $t \cdot s^{\#}$ represent ultimately periodic words.

The approximative stabilization semigroups For each positive integer N, the mapping

$$\alpha_{(N=\omega)} \quad : \quad \mathbb{M}_Q\mathcal{T} \quad \longrightarrow \quad \mathbb{M}_Q\mathcal{T}_{/N=\omega}$$

is a continuous homomorphism of profinite semigroups. In particular, it is a homomorphism of stabilization semigroups. How about the following mapping?

$$\alpha_{(N=N+1)} \quad : \quad \mathbb{M}_Q\mathcal{T} \quad \longrightarrow \quad \mathbb{M}_Q\mathcal{T}_{/N=N+1}$$

Let us first consider the case with no states, i.e. of the mapping $\alpha_{(N=N+1)} \colon \mathcal{T} \to \mathcal{T}_{/N=N+1}$. Recall that it is not continuous for the discrete topology over the codomain. The ω-power in $\mathcal{T}$ maps every nonzero, finite number to ω, and leaves $0, \omega$ and ∞ unchanged. Therefore, if we define the mapping # in $\mathcal{T}_{/N=N+1}$, so that it also maps every nonzero, finite number to ω, and leaves $0, \omega$ and ∞ unchanged, then we have that for every $s \in \mathcal{T}$,

$$\alpha_{(N=N+1)}(s^{\omega}) = \alpha_{(N=N+1)}(s)^{\#}.$$

Note that # differs from the idempotent power in $\mathcal{T}_{/N=N+1}$, since N is idempotent, but $N^{\#} = \omega$.

The following lemma saves us from proving by hand that # satisfies the axioms of a stabilization semigroup.

Lemma 6.1. *Let $\alpha \colon S \to T$ be a surjective homomorphism of semigroups from a profinite semigroup S to a semigroup T which is equipped with an operation # such that*

$$\alpha(s^{\omega}) = \alpha(s)^{\#} \qquad \textit{for every } s \in S.$$

Then S equipped with # is a stabilization semigroup, and α is a homomorphism of stabilization semigroups.

Proof. Since S equipped with $\cdot$ and ω satisfies all the equations (A1)-(A4), it immediately follows (since α is surjective) that these equations also hold in T, if the operations $\cdot$ and # are considered instead. To verify the axiom (A5), choose any idempotent $e \in T$. Let $s \in S$ be such that $\alpha(s) = e$, and let $u = s^{\omega-1}$. By definition of u, we have that

$$s \cdot u = s^{\omega}, \tag{1}$$

$$s^2 \cdot u = s \cdot s^{\omega}. \tag{2}$$

Since $\alpha(s) = e = e^2 = \alpha(s^2)$, it follows that

$$e^{\#} = \alpha(s^{\omega}) = \alpha(s \cdot u) = \alpha(s) \cdot \alpha(u) = \alpha(s^2) \cdot \alpha(u) = \alpha(s^2 \cdot u) = \alpha(s) \cdot \alpha(s^{\omega}) = e \cdot e^{\#}.$$

This proves that T satisfies all the axioms of a stabilization semigroup, and that α is a homomorphism of stabilization semigroups. □

As mentioned, the semigroup $\mathbb{M}_Q\mathcal{T}$ is naturally equipped with the ω-power, since it is profinite. Moreover, as the following proposition states, the semigroup homomorphism

$$\alpha_{(N=N+1)} \quad : \quad \mathbb{M}_Q\mathcal{T} \quad \longrightarrow \quad \mathbb{M}_Q\mathcal{T}_{/N=N+1}$$

induces a stabilization operation over $\mathbb{M}_Q\mathcal{T}_{/N=N+1}$, stemming from the ω-power in $\mathbb{M}_Q\mathcal{T}$.

Proposition 6.2. *Let N be a finite number and let Q be a finite set. There is a unique operation # such that the following diagram commutes:*

$$\begin{tikzcd}
\mathbb{M}_Q\mathcal{T} \arrow[r, "\alpha_{(N=N+1)}"] \arrow[d, "\omega"'] & \mathbb{M}_Q\mathcal{T}_{/N=N+1} \arrow[d, dashed, "\#"] \\
\mathbb{M}_Q\mathcal{T} \arrow[r, "\alpha_{(N=N+1)}"] & \mathbb{M}_Q\mathcal{T}_{/N=N+1}
\end{tikzcd}$$

The mapping # has moreover the following description for idempotent $e \in \mathbb{M}_Q\mathcal{T}_{/N=N+1}$:

$$e^{\#}[p,q] = \min_{r \in Q} \big(e[p,r] + (e[r,r])^{\#} + e[r,q] \big).$$

Consequently, endowed with the operation #, the semigroup $\mathbb{M}_Q\mathcal{T}_{/N=N+1}$ becomes a stabilization semigroup, and the mapping $\alpha_{(N=N+1)}$ becomes a homomorphism of stabilization semigroups.

We skip the proof of this proposition, as it will follow from the more general Proposition 7.3. Note only that the last sentence of the proposition follows easily from the rest of the proposition. This is because the axioms of a stabilization semigroup hold in the profinite semigroup $\mathbb{M}_Q\mathcal{T}$, and as a consequence also hold in $\mathbb{M}_Q\mathcal{T}_{/N=N+1}$ (except for the axiom (A5), which follows easily from the above formula for $e^{\#}$).

6.2 $\langle\ \cdot\ , \omega\rangle$-locally closed profinite semigroups

In this section, we introduce *$\langle\ \cdot\ ,\omega\rangle$-locally closed* profinite semigroups. These are profinite semigroupswhich satisfy the equality $A^{\langle\ \cdot\ ,\omega\rangle} = \overline{A^+}$ for every finite set of elements A. (Recall that if A is a subset of a profinite semigroup, then by $A^{\langle\ \cdot\ ,\omega\rangle}$ we denote the set of elements generated from A by using multiplication and the ω-power.) If α is a homomorphism from such a profinite semigroup to a finite stabilization semigroup, then computing the image of $\alpha(\overline{A^+})$ is equivalent to computing the set $\alpha(A)^{\langle\ \cdot\ ,\omega\rangle}$ of elements generated from $\alpha(A)$ by stabilization and multiplication.

Note that the closure of A^+ contains $A^{\langle\ \cdot\ ,\omega\rangle}$, since for any $s \in S$, the element s^ω is the limit of the sequence $(s^{n!})_{n=1}^{\infty}$. Therefore,

$$\overline{A^{\langle\ \cdot\ ,\omega\rangle}} = \overline{A^+},$$

so the requirement that $A^{\langle\ \cdot\ ,\omega\rangle} = \overline{A^+}$ is equivalent to the requirement that the set $A^{\langle\ \cdot\ ,\omega\rangle}$ is closed. This characterization explains why we use the term *$\langle\ \cdot\ ,\omega\rangle$-locally closed.*

Example 6.5. The semigroup $\mathcal{T}$ is a $\langle\ \cdot\ ,\omega\rangle$-locally closed profinite semigroup. This is true since in $\mathcal{T}$, any set containing ω is closed.

Example 6.6. As a semigroup that is not $\langle\ \cdot\ ,\omega\rangle$-locally closed, consider the the free profinite semigroup $\widehat{A^+}$ generated by A. In this semigroup, the closure of A^+ is uncountable, while $A^{\langle\ \cdot\ ,\omega\rangle}$ is countable.

The following proposition follows immediately from the definitions, and from the fact that a closed subsemigroup of a profinite semigroup is closed under stabilization (and multiplication).

Proposition 6.3. *A closed subsemigroup of a $\langle\ \cdot\ ,\omega\rangle$-locally closed profinite semigroup is again $\langle\ \cdot\ ,\omega\rangle$-locally closed.*

Theorem 5.1 stated in the overview says that the profinite semigroup $\mathbb{M}_Q\mathcal{T}$ of matrices over the semiring $\mathcal{T}$ is $\langle\ \cdot\ ,\omega\rangle$-locally closed. Later on, we will prove a more general theorem, analogous to Theorem 5.1, but concerning a more general semiring related to B-automata. The semigroup $\mathbb{M}_Q\mathcal{T}$ will then be a closed subsemigroup of a more general $\langle\ \cdot\ ,\omega\rangle$-locally closed profinite semigroup.

6.3 Factorization trees for stabilization semigroups

In this chapter we prove a version of Simon's Factorization Forest Theorem for stabilization semigroups. It has been formulated in the special case of the semigroup $\mathbb{M}_Q\mathcal{T}_{/1=2}$ in the work [Sim94] of I. Simon, but we believe our proof is cleaner and less involved.

Let A be a finite alphabet, S a finite stabilization semigroup, and $\alpha\colon A \to S$ a mapping. A *factorization tree* f is tied to the following objects: the mapping α, an *input* word $w \in A^+$, and an *output* $s \in S$. The definition is inductive.

Base rule. Each letter $a \in A$ is a factorization tree, with input a and output $\alpha(a)$.

Binary rule. Suppose that f, g are factorization trees with inputs $v, w \in A^+$ and outputs $s, t \in S$, respectively. Then (f, g) is a factorization tree with input vw and output st.

Stabilization rule. Suppose that factorization trees $f_1, \ldots, f_n$ have inputs $v_1, \ldots, v_n \in A^+$, but the same idempotent $e \in S$ on output. Then $\#(f_1, \ldots, f_n)$ is a factorization tree with input $v_1 \cdots v_n$ and output $e^\#$.

A factorization tree can be seen as a tree, where the base rule corresponds to leaves, the binary rule corresponds to nodes of outdegree two, and the stabilization rule corresponds to nodes of unbounded outdegree. When talking about the *height* of a factorization tree, we refer to the tree representation (we assume the leaves to have height 1). The objective of this theory is to find, for a given input word, a factorization tree of small height. Note that, unlike in classical factorization trees of Simon, here, two factorization trees with the same input might yield different outputs. This is because out of two trees with output e we can construct, using the binary rule, a tree with output e, or, using the stabilization rule, a tree with output $e^\#$, and in general, $e \neq e^\#$.

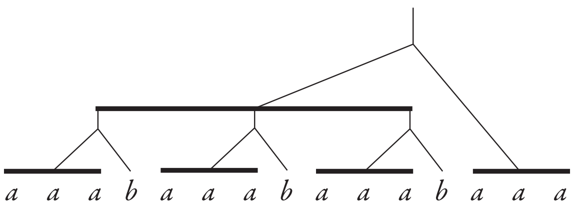


FIGURE 6.1: *A factorization tree of height 4, input $(a^3b)^3a^3$ and output $(e^\# f)^\# e^\#$*

Example 6.7. Let S be the stabilization semigroup $\mathbb{M}_4\mathcal{T}_{/1=2}$ of 4×4 matrices over the semiring $\mathcal{T}_{/1=2}$ and let us once again consider the distance automaton $\mathcal{A}$ from Figure 5.1 in the overview, whose input alphabet is $\{a, b\}$, and which induces the mapping $\delta_{\mathcal{A}}\colon \{a, b\} \to S$. We will be constructing factorization trees with respect to the mapping $\delta_{\mathcal{A}}$. Let $e = \alpha_{(1=2)}(\delta_{\mathcal{A}}(a))$ and $f = \alpha_{(1=2)}(\delta_{\mathcal{A}}(b))$. Those are both idempotent elements of S. For $n \geq 1$, let

$$w_n = (a^n b)^n a^n$$

be an input word.

We can construct a factorization tree over the input word w_n, which does not use the stabilization rule. Its height is approximately $\log n$ and its output is the matrix

$$efe = \begin{bmatrix} 0 & 1 & 1 & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & 1 \end{bmatrix}$$

But, thanks to the stabilization rule, we can also construct a factorization tree of height 4 with input w_n, independently of the number n, as illustrated in Figure 6.1 in the case $n = 3$. The output of this tree is the matrix

$$(e^\# f)^\# e^\# = \begin{bmatrix} 0 & \omega & \omega & \perp \\ \perp & \perp & \perp & \perp \\ \perp & \perp & 0 & \perp \\ \perp & \perp & \perp & \omega \end{bmatrix}$$

In fact, it is easy to see that any word over the alphabet $\{a, b\}$ has a factorization tree of height at most 5.

For a stabilization semigroup S, we write $\|S\|$ for the smallest number h such that for all mappings $\alpha\colon A \to S$, each input word in A^+ has a factorization tree (we place no restrictions on its output) of height at most h. The key result is that this number is finite. This is stated in the following “factorization theorem”.

Theorem 6.4. *For any finite stabilization semigroup S, $\|S\|$ is finite.*

Historical note The factorization trees described above are an extension of the notion of factorization trees introduced by I. Simon. There, the stabilization rule takes a simpler form of an *idempotent rule*, which outputs e instead of $e^\#$. I. Simon [Sim90] proved the factorization theorem for trees with the idempotent rule. Using the stabilization rule instead of the idempotent rule gives a more general result, since any finite semigroup can be treated as a stabilization semigroup, with $s^\#$ defined as s^ω. In another paper [Sim94] I. Simon introduced his *tropical trees* – the definition is virtually identical to the above notion of factorization trees, with the only difference that S was specifically assumed to be the semigroup $\mathbb{M}_Q\mathcal{T}_{/1=2}$ of matrices over the semiring $\mathcal{T}_{/1=2}$.

6.3.1 Proof of the factorization theorem

The rest of this section is devoted to showing the factorization theorem, whose proof is basically identical to the one for the standard theorem on factorization trees of I. Simon (see e.g. [Boj09a]). The proof proceeds by induction on the size of the stabilization semigroup S. The base case, when S has one element, is obvious, and $\|S\|$ is two (each factorization tree is a stabilization rule applied to a sequence of base rules, and base rules have height 1).

In the proof we will extend α from single letters in A to words in A^+, using the semigroup structure of S. We will use the term *type of* $w \in A^+$ for the value $\alpha(w)$. Note that a factorization tree for w might not output the type of w, as it might happen in the case of a stabilization when $e^\# \neq e$.

We consider two cases, depending on whether S has a nontrivial ideal (i.e. an ideal T different than S and than $\{0\}$), or not.

S has some nontrivial ideal Let T be an ideal in S such that T has more than one element and that $T \neq S$. By axiom (A5'), two-sided ideals are closed under stabilization, so the quotient S/T is a stabilization semigroup with a zero, where zero corresponds to all elements in T. Let $\beta\colon A \to S/T$ be the quotient morphism.

A factorization tree has *normal form* if the arguments of each stabilization rule are nonzero idempotents, with the possible exception of the topmost rule. Note that the notion of normal form is appropriate only when the stabilization semigroup has a zero, as is the case for S/T.

Lemma 6.5. *A factorization tree can be transformed into normal form, without increasing its height.*

Proof. If some stabilization node x uses 0's as its arguments and x is not the root of the tree, then the tree can be "rotated", pushing the children of x toward the root. This is illustrated in Figure 6.2. In the figure, the node x is assumed to be a right child of a binary node. The case when x is a left child of a binary node is symmetric. Finally, if x is a child of a stabilization node y, then x can be removed, by plugging the children of x directly into the node y. □

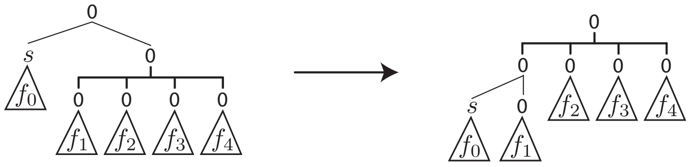


FIGURE 6.2: *A rotation*

We use the inductive assumption on the size of the stabilization semigroup to get a β-factorization tree f for w. By using the above lemma, we may assume that f has normal form. If f applies the stabilization rule only for nonzero idempotents, then it is a legal factorization tree also for α. Otherwise, the topmost rule of f is the stabilization rule applied to trees which evaluate to 0 in S/T. By applying the following fact we conclude that w has an α-factorization tree of height at most $\|S/T\| + \|T\| - 1$, which is finite by the induction assumption.

Fact 1. *Let $T \subseteq S$ be a proper subsemigroup closed under stabilization. Suppose input words $w_1, \ldots, w_n$ have factorization trees of height at most N with outputs in T. Then $w_1 \cdots w_n$ has a factorization tree of height at most $N + \|T\| - 1$.*

Proof. By substituting factorization trees for $w_1, \ldots, w_n$ into the a factorization tree inside T. □

S has only trivial ideals In this case, only ideals in S are S itself, and perhaps a single ideal containing one element – the zero of S. If S has a zero, we denote it by 0.

No matter if S has or doesn't have a zero, we have the following.

Fact 2. *For any idempotent e, either $e^{\#} = e$ or $e^{\#} = 0$.*

Proof. By axiom (A5′), $e^{\#}e = e\, e^{\#} = e^{\#}$. Therefore, $e^{\#} \leq_{\mathcal{J}} e$. If conversely, $e \leq_{\mathcal{J}} e^{\#}$, then by Corollary 3.6 of the Preliminaries, $e = e^{\#}$. Otherwise, if $e >_{\mathcal{J}} e^{\#}$, then the ideal generated by $e^{\#}$ is strictly smaller than the ideal generated by e. Therefore, $e^{\#} \subseteq \{0\}$. □

Let $w = a_1 \cdots a_n$ be a word. In general, the type of w might be equal to 0 (if S has a zero at all). We first reduce this case to the case when the type of w is not equal to 0. If the type of w is zero, then w can be decomposed as $w = w_1 b_1 w_2 b_2 \cdots w_k b_k w_{k+1}$ where the types of $w_1, \ldots, w_{k+1}$ are not zero, and the types of $w_1 b_1, \ldots, w_k b_k$ are zero. It is not difficult to construct a factorization tree for w, whose height is the maximal height of factorization trees for $w_1, \ldots, w_{k+1}$, plus three (a binary rule to combine w_i with b_i, a stabilization rule to combine all the pairs $w_1 b_1, \ldots, w_k b_k$, and a binary rule to append w_{k+1}).

We are left with the case when the type of w is not zero. We will provide an α-factorization for w, by induction on the number of distinct two letter infixes ab that appear in w. The induction base, where there is no such infix, corresponds to the case of a single letter, when an α-factorization tree can easily be provided.

Consider now the induction step, and let ab be an infix that appears in w. Consider a factorization

$$w = w_0 ab w_1 ab w_2 \cdots ab w_k ab w_{k+1}$$

so that ab does not appear in any of the words $w_0, \ldots, w_{k+1}$. We can use the inductive assumption to produce, for each w_i, an α-factorization tree, which we then extend to a factorization tree f_i for $bw_i a$. Note that f_i outputs either zero, or the type of $bw_i a$, which is not zero. For the same reason as previously, the interesting case is when all the factorization trees $f_1, \ldots, f_k$ for the words $bw_1 a, bw_2 a, \ldots, bw_k a$ have nonzero outputs, say $s_1, \ldots, s_k$.

The key observation is that since all infixes of w have types in the same $\mathcal{J}$-class, then the $\mathcal{H}$-class of the type of each infix is uniquely determined by its first and last letter. This follows from Lemma 3.7 of the Preliminaries. Therefore, all of the elements $s_1, \ldots, s_k$ are in the same $\mathcal{H}$-class (since they all start with b and end with a), call it T. If $k \geq 2$, then $T \cdot T$ intersects T, so T is a group by Lemma 3.8. Let e be the unique idempotent of T. Then, by Fact 2, $e^\# = e$ or $e^\# = 0$. In either case, $T \cup \{e^\#\}$ is a stabilization subsemigroup of S, such that T is a group. We can therefore apply Fact 1 to get the desired result, as long as we can prove the theorem for $T \cup \{e^\#\}$, where T is any subgroup of S and e is its neutral element.

Group case Assume that G is a subgroup of S and that 1 is its neutral element, and that $1^\# = 1$ or $1^\# = 0 \notin G$.

Let $w = a_1 \cdots a_n$ be a word, such that all letters a_i have types in G. We will provide an α-factorization for $a_1 \cdots a_n$, by induction on the size of

$$P_w = \{\alpha(a_1) \cdots \alpha(a_i) \in G : \quad i \in \{1, \dots, n-1\}\} ,$$

which is the set of types of nontrivial prefixes of w. The induction base, when the set above is empty, corresponds to words of length 1, when clearly an α-factorization can be found. For the induction step, consider some $g \in P_w$, and let $w = w_1 \cdots w_k$ be a factorization such that the only nontrivial prefixes of w with type g are $w_1, w_1w_2, \dots, w_1 \cdots w_{k-1}$. Since G is a group, we have

$$P_{w_i} = g^{-1} g P_{w_i} \subseteq g^{-1}(P_w - \{g\})$$

and therefore each P_{w_i} has smaller size than P_w. By applying the induction assumption, we get an α-factorization for each w_i. Some of these factorizations can output 0, the others output 1, since 1 is the type of each $w_2, \dots, w_{k-1}$.

We have thus finished the inductive step in the proof of Theorem 6.4. The produced bound on $\|S\|$ is polynomial in the size of S, but is not optimal. It is not difficult to enhance the above proof (exactly as in the case of usual semigroups) to obtain an upper bound of $3|S|$ for $\|S\|$.

CHAPTER 7

Limitedness of B-automata

Theorem 5.1 phrased in the overview states a property of the semiring $\mathcal{T}$ related to distance automata. In this chapter, we will formulate and prove an analogous result, but in the more general case of B-automata, which extend distance automata by allowing resets and multiple counters. The main problem which we encounter is the lack of a linear order over the set of counter operations, when resets or multiple counters are allowed. We resolve this problem by considering more general algebraic structures.

7.1 B-automata

In this section, we formally define B-automata and the appropriate semigroups for analyzing their limitedness.

Let A be a fixed finite alphabet. A *B-automaton* $\mathcal{A}$ is described by the following components, each being a finite set:

- A set of *states* Q
- A set of *initial states*, $I \subseteq Q$
- A set of *accepting states*, $F \subseteq Q$
- A set of *counters* C
- A *transition relation*, $\delta \subseteq Q \times A \times (\{inc, reset\}^*)^C \times Q$.

A B-automaton has an underlying finite automaton, obtained by removing the counters and ignoring the third component of the transition relation.

A *run* of the automaton $\mathcal{A}$ over a word w is a sequence of transitions, which corresponds to a run of the underlying finite automaton. Similarly, we define *accepting runs* and *accepted words*, by lifting the respective notions from finite automata.

Let ρ be a run and $c \in C$ a counter. Then, ρ induces a sequence $\rho[c] \in \{inc, reset\}^*$ of operations performed over the counter c. The *maximal value* of c in the run ρ, denoted $maxval_c(\rho)$ is the length of the longest block of letters *inc* in the word $\rho[c]$.

The *value* of the run ρ is defined as

$$val_{\mathcal{A}}(\rho) = \max_{c} maxval_c(\rho) \in \mathbb{N},$$

and the *valuation* of an input word w under $\mathcal{A}$ is defined as

$$f_{\mathcal{A}}(w) = \min_{\rho} val_{\mathcal{A}}(\rho) = \min_{\rho} \max_{c} maxval_c(\rho) \in \overline{\mathbb{N}},$$

where ρ ranges through all *accepting* runs of $\mathcal{A}$ over w. The value $f_{\mathcal{A}}(w)$ is defined as $\perp$ if $\mathcal{A}$ does not accept w. We say that a B-automaton $\mathcal{A}$ is *limited*, if there exists a bound $N \in \mathbb{N}$ such that $f_{\mathcal{A}}(w) \leq N$ for every w accepted by $\mathcal{A}$.

Distance automata are a special case of B-automata, in which the reset operation is never used, and there is only one counter.

Example 7.1. Consider the B-automaton $\mathcal{A}$ with one counter over the alphabet $\{a, b, \$\}$, depicted in Figure 7.1. Then,

$$f_{\mathcal{A}}(a^{n_{11}}ba^{n_{12}}b\ldots\$a^{n_{21}}ba^{n_{22}}b\ldots\$\ldots\$a^{n_{k1}}ba^{n_{k2}}b\ldots) \quad = \quad \min_{i} \max_{j} n_{ij}.$$

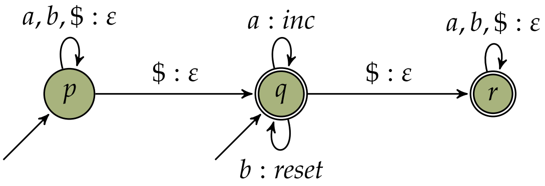


FIGURE 7.1: *A B-automaton with one counter*

7.1.1 The semigroup for B-automata

The semigroup suited for describing transitions of a distance automaton with states Q is the semigroup $\mathbb{M}_Q\mathcal{T}$. It is defined in several stages – the basic semigroup is the tropical semiring, which is further extended by the element ω, and then matrices are formed over this semiring. In the case of B-automata, the appropriate semigroup is more complex, and we also define it in several stages.

The basic object is the semigroup $\mathbb{B}$, suited for describing the behavior of a single counter in a single run of a B-automaton. As a set,

$$\mathbb{B} \subseteq \mathbb{N} \cup \mathbb{N}^3.$$

We denote the elements of the first summand by (n) and the elements of the second summand by (k, l, m), where $n, k, l, m \in \mathbb{N}$. The set $\mathbb{B}$ consists of all elements of the form (n), and all elements of the form (k, l, m), where k and m are not larger than l. Let us consider a "syntactic mapping" σ from the set $\{inc, reset\}^*$ to $\mathbb{B}$, which is defined according to the following rules.

$$\begin{array}{rcl} inc^n & \mapsto & (n) \\ inc^{n_1} reset\ inc^{n_2} \ldots inc^{n_{k-1}} reset\ inc^{n_k} & \mapsto & (n_1, \max(n_1, n_2, \ldots, n_k), n_k) \\ inc^{n_1} reset\ inc^{n_2} & \mapsto & (n_1, \max(n_1, n_2), n_2) \end{array}$$

Therefore, the element $(n) \in \mathbb{B}$ represents a sequence of n increments, while the element $(k, l, m) \in \mathbb{B}$ represents any sequence of counter operations, in which the number of increments before the first reset is equal to k, and the number of increments after the last reset is equal to m, and the longest block of increments has length l.

It is clear that the above mapping σ from $\{inc, reset\}^*$ to $\mathbb{B}$ preserves concatenation, i.e. it induces a congruence with respect to concatenation. Therefore, the set $\mathbb{B}$ inherits a structure of a semigroup. Multiplication in this semigroup is given in Figure 7.2.

$$\begin{aligned} (m) \cdot (n) &= (m + n) \\ (m) \cdot (n_1, n_2, n_3) &= (m + n_1, \max(n_2, m + n_1), n_3) \\ (n_1, n_2, n_3) \cdot (m) &= (n_1, \max(n_2, n_3 + m), n_3 + m) \\ (n_1, n_2, n_3) \cdot (m_1, m_2, m_3) &= (n_1, \max(n_2, m_2, n_3 + m_1), m_3) \end{aligned}$$

FIGURE 7.2: *The multiplication table in* $\mathbb{B}$.

The above construction of $\mathbb{B}$ is "natural" with respect to the definition of the B-automaton, since it yields a semigroup isomorphic to the the quotient of $\{inc, reset\}^*$ by the coarsest congruence which respects the mapping $maxval\colon \{inc, reset\}^* \to \mathbb{N}$.

We define a partial ordering over $\mathbb{B}$, in which $(m) \leq (n)$ if $m \leq n$ and $(m_1, m_2, m_3) \leq (n_1, n_2, n_3)$ if $m_i \leq n_i$ for $i = 1, 2, 3$. This way, $\mathbb{B}$ becomes a partially ordered semigroup. Note however, that unlike $\mathcal{T}$, it is not linearly ordered. This is because it does not make sense to compare, for instance, the element $(100, 100, 0)$ with the element $(0, 100, 100)$ in $\overline{\mathbb{B}}$.

The semigroup $\overline{\mathbb{B}}$ We extend the monoid $\mathbb{B}$ by an element ω, which represents the limit of a sequence of counter operations, in which the counter attains arbitrarily high values.

More formally, we consider the one point compactification of $\mathbb{B}$, denoted $\overline{\mathbb{B}}$. This is the set $\mathbb{B} \cup \{\omega\}$, where ω is the limit point of any infinite set of elements of $\mathbb{B}$. For this, we define a metric over $\overline{\mathbb{B}}$. We first define $\|s\|$ for $s \in \mathbb{B}$ as the largest value appearing in s (i.e. $\|(n_1, n_2, n_3)\| = n_2$ and $\|(n)\| = n$). The distance between any element $s \in \mathbb{B}$ and ω is equal to $1/(\|s\| + 1)$. The distance between any two elements $s, t \in \mathbb{B}$ is defined as the sum of the distances from s and t to ω. This metric is compact.

We assume that ω is larger than all the other elements in $\mathbb{B}$. The semigroup structure extends to $\overline{\mathbb{B}}$, by treating ω as the zero element, i.e. $\omega \cdot s = s \cdot \omega = \omega$ for all $s \in \overline{\mathbb{B}}$. This way, $\overline{\mathbb{B}}$ becomes a (partially ordered) profinite semigroup. There are two ways of seeing this. One way is by observing that $\overline{\mathbb{B}}$ is a totally disconnected, compact topological semigroup, which is moreover metrizable. Another way is to see that $\overline{\mathbb{B}}$ is isomorphic to the projective limit of a sequence of semigroups, where the N-th semigroup, denoted $\overline{\mathbb{B}}_{/N=\omega}$, is the quotient of the semigroup $\overline{\mathbb{B}}$ by the congruence which identifies with ω all elements s with $\|s\| \geq N$. The behavior of the ω-power in $\overline{\mathbb{B}}$ is described in Figure 7.3.

$$\begin{aligned} (0)^{\omega} &= (0) \\ (m)^{\omega} &= \omega \qquad \text{for } m \geq 1 \\ (n_1, n_2, n_3)^{\omega} &= (n_1, n_2, n_3)^2 = (n_1, \max(n_2, n_1 + n_3), n_3) \end{aligned}$$

FIGURE 7.3: *The ω-power in* $\overline{\mathbb{B}}$.

The semigroup for multiple counters If C is a finite set, then $\overline{\mathbb{B}}^C$ is given the structure of the product semigroup, which is partially ordered, by the coordinate-wise ordering. This is the semigroup which is suited for representing a single transition in a B-automaton with counters C and one state. Note that considering many counters C further amplifies the problem of defining a natural linear order over $\overline{\mathbb{B}}^C$.

The partial semigroup of transitions A *partial semigroup* is a set S equipped with a multiplication operation, denoted $\cdot$, which is partially defined over $S \times S$ and is associative in the sense that for all $s, t, u \in S$, if one of the sides of the equation below is defined, then so is the other, and the equality holds:

$$(s \cdot t) \cdot u = s \cdot (t \cdot u).$$

(A partial semigroup can be seen as a *small category*, i.e. a category, whose objects and morphisms form a set, or it can be seen as a semigroup with zero, with the zero element removed.)

Let Q be a fixed set of states. For a semigroup S, the set $Q \times S \times Q$ has a natural structure of a partial semigroup $Q \times S \times Q$, in which multiplication is partially defined by

$$(p, s, q) \cdot (q, t, r) = (p, s \cdot t, r).$$

If S is a partially ordered semigroup, then we consider the induced coordinate-wise partial order over $Q \times S \times Q$, stemming from the identity ordering over Q.

Now, if S is an ordered partial semigroup (i.e. a partial semigroup equipped with a partial order, such that multiplication is consistent with the order), then $P_{\uparrow}S$ denotes the set whose elements are upward-closed subsets of S, equipped with the associative operation

$$I \cdot J \overset{def}{=} \uparrow\{s \cdot t : \ s \in I, \ t \in J, \text{ and } s \cdot t \text{ is defined}\}.$$

This way, $P_{\uparrow}S$ becomes a partially ordered semigroup, where the partial order is set inclusion.

Let C be a fixed set of counters. Then $Q \times \overline{\mathbb{B}}^C \times Q$ represents the set of possible transitions of a B-automaton with counters C and states Q, and their limits. We denote its elements by

τ, σ, etc. Let $\tau = (p, o, q)$. We call p the *source state* and q the *target state* of the transition τ. For $c \in C$ by $\tau[c] \in \overline{\mathbb{B}}$ we denote the component of o corresponding to counter c. We consider the coordinatewise partial ordering on the set of transitions, i.e. $\tau \leq \sigma$ if they have the same source states, and they have the same target states, and for each $c \in C$, $\tau[c] \leq \sigma[c]$.

For a transition $\tau \in Q \times \overline{\mathbb{B}}^C \times Q$, by $\|\tau\|$ we denote the largest *finite* entry appearing in τ (or 0 if there are none):

$$\|\tau\| = \max\{\|\tau[c]\| : \quad c \in C,\ \tau[c] \neq \omega\}.$$

The semigroup of sets of transitions We will only consider upward-closed sets of transitions of B-automata (a justification for this will be given later), i.e. elements of the set

$$P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q).$$

The semigroup $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ has a structure of a profinite semigroup. There are at least three ways of seeing this, each resulting in an equivalent profinite structure. We will use only one of the constructions, but we also describe the others.

One way is to consider the Hausdorff metric over $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$. The Hausdorff metric is a metric over the family of compact subsets of a compact metric space. The distance between two sets M, N is the infimum of all numbers ε such that any point in M is at most ε-distant from a point in N, and any point in N is at most ε-distant from a point in M. The Hausdorff metric is a compact metric, and if the underlying metric space is totally disconnected, then the Hausdorff metric is also totally disconnected. Observe that in our case, any upward-closed subset of $Q \times \overline{\mathbb{B}}^C \times Q$ is compact, so the Hausdorff metric can be considered. Also, it is easy to see that the product is continuous with respect to this metric. This way, $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ becomes a metrizable profinite semigroup – it is compact, totally disconnected, and metrizable.

Another way of equipping $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ with a profinite structure is as follows. For each $N \in \mathbb{N}$, we consider the finite semigroup $\overline{\mathbb{B}}_{/N=\omega}$ defined earlier. We then view $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ as the projective limit of the sequence of semigroups $P_{\uparrow}(Q \times (\overline{\mathbb{B}}_{/N=\omega})^C \times Q)$. This is the profinite structure that we will be using. We denote by $\alpha_{(N=\omega)}$ the canonical mapping of the two semigroups:

$$\alpha_{(N=\omega)} : \quad P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q) \quad \longrightarrow \quad P_{\uparrow}(Q \times (\overline{\mathbb{B}}_{/N=\omega})^C \times Q).$$

The mapping $\alpha_{(N=\omega)}$ is therefore a continuous homomorphism from a profinite semigroup to a finite semigroup.

The last way of equipping $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ with a profinite structure is to see that all the used constructions are functors from respective categories of algebraic structures. More precisely: the Cartesian product (used to define the semigroup $\overline{\mathbb{B}}^C$) is an endofunctor in the category of finite ordered semigroups; the "semidirect product" (used to define the partial semigroup $Q \times \overline{\mathbb{B}}^C \times Q$) is a functor from the category of finite ordered semigroups to the category of finite ordered partial semigroups; finally the operator $P_{\uparrow}$ is a functor from finite ordered partial semigroups to finite semirings. Those are in fact “continuous functors”, so they they map profinite objects to profinite objects.

The main theorem We are now ready to formulate the main theorem of the first part of this dissertation, which is a generalization of the Theorem 5.1 stated in the overview.

Theorem 7.1. *Let C, Q be finite sets. Then the profinite semigroup*

$$P_{\uparrow}(Q \times \overline{\mathbb{B}}^{C} \times Q)$$

is $\langle \,\cdot\, , \omega \rangle$-locally closed.

The case of distance automata Theorem 5.1 stated in the overview is the special case of Theorem 7.1 when C has one element, and there are no resets. Formally, we derive it as follows. Since $\mathcal{T}$ is linearly ordered, it is isomorphic to $P_{\uparrow}\overline{\mathbb{N}}$ ($\perp \in \mathcal{T}$ corresponds to $\emptyset \subseteq \overline{\mathbb{N}}$). In turn $P_{\uparrow}\overline{\mathbb{N}}$ is a closed subsemiring of $P_{\uparrow}\overline{\mathbb{B}}$ (the semigroup $P_{\uparrow}\overline{\mathbb{B}}$ has a natural semiring structure, in which addition is set union), and so $\mathbb{M}_{Q}\mathcal{T}$ is isomorphic to a closed subsemigroup of $\mathbb{M}_{Q}(P_{\uparrow}(\overline{\mathbb{B}})) \simeq P_{\uparrow}(Q \times \overline{\mathbb{B}} \times Q)$. Therefore, Theorem 5.1 follows, since a closed subsemigroup of a $\langle \,\cdot\, , \omega \rangle$-locally closed semigroup is again $\langle \,\cdot\, , \omega \rangle$-locally closed. Also, a proof of Theorem 5.1 can be obtained by adopting and simplifying the proof of Theorem 7.1.

Deciding limitedness of B-automata Completely analogously as in the case of distance automata, Theorem 7.1 implies decidability of the limitedness problem for B-automata. We only sketch the characterization of non-limitedness for B-automata. Let $\mathcal{A}$ be a B-automaton with states Q and counters C. Then, $\mathcal{A}$ induces a transition function $\delta_{\mathcal{A}}$, which is a homomorphism defined over the set of all finite input words, by specifying that for each input letter a,

$$\delta_{\mathcal{A}}(a) \quad \overset{def}{=} \quad \uparrow\{(p, o, q) : \quad (p, a, \tilde{o}, q) \in \delta, \quad o = \sigma^{C}(\tilde{o})\} \qquad \in P_{\uparrow}(Q \times \overline{\mathbb{B}}^{C} \times Q),$$

where

$$\sigma^{C} \quad : \quad (\{inc, reset\}^{*})^{C} \quad \longrightarrow \quad \overline{\mathbb{B}}^{C}$$

is the coordinatewise extension of the syntactic mapping σ. It is easy to see that the valuation function $f_{\mathcal{A}}$ induced by $\mathcal{A}$ factorizes as the composition of $\delta_{\mathcal{A}}$ and a continuous mapping

$$\tilde{f}_{\mathcal{A}} \quad : \quad P_{\uparrow}(Q \times \overline{\mathbb{B}}^{C} \times Q) \quad \longrightarrow \quad \overline{\mathbb{N}} \cup \{\perp\}$$

defined by

$$\tilde{f}_{\mathcal{A}}(x) = \min_{\tau \in x} \max_{c \in C} \left(\textit{maximal value in } \tau[c]\right),$$

where τ ranges through all transitions in x whose source state is an initial state and target state is an accepting state of $\mathcal{A}$, and $\min \emptyset = \perp$. Note how the min above is compatible with our restriction to upward-closed sets of transitions.

Let $A \subseteq P_{\uparrow}(Q \times \overline{\mathbb{B}} \times Q)$ be the finite set of all elements $\delta_{\mathcal{A}}(a)$ induced by the input letters. Then, clearly a B-automaton $\mathcal{A}$ is *not* limited, if and only if there exists an element $x \in \overline{A^+}$ such that $\tilde{f}_{\mathcal{A}}(x) = \omega$. Equivalently, x has the following property:

> *There exists a transition τ in x from an initial state to an accepting state, but for every such transition τ, $\tau[c] = \omega$ for some counter $c \in C$.*

As in the case of distance automata, the above characterization yields an effective procedure for determining non-limitedness of $\mathcal{A}$, by enumerating all elements of $\overline{A^+} = A^{\langle \cdot ,\omega\rangle}$.

This time, however, it is less obvious how to enumerate all the elements of $A^{\langle \cdot ,\omega\rangle}$. We will only give a hint on how this can be done, but the decidability result will actually also follow from other results.

First, let us observe that each element of $P_{\uparrow}(Q \times \overline{\mathbb{B}} \times Q)$ has a finite description – this is one of the advantages of considering upward-closed subsets of $Q \times \overline{\mathbb{B}}^C \times Q$, since they are determined by a *finite* number of elements, as stated in the following lemma.

Lemma 7.2. *Let $s \in P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$. Then s, as a subset of $Q \times \overline{\mathbb{B}}^C \times Q$, has only finitely many minimal elements.*

Proof. Recall that a partially ordered set X is called a *well quasi-order* if every infinite sequence $x_1, x_2, x_3, \ldots$ of elements of X has an increasing pair, i.e. a pair $x_i \leq x_j$ with $i < j$. In particular, every upward-closed set can only have a finite set of minimal elements.

The set of natural numbers $\mathbb{N}$, ordered in the usual way, and every finite set Q, ordered by the identity relation, are both well quasi-ordered sets. Standard results from the theory of well quasi-orders (see e.g. [Kru72]) imply that the operations used to define $(Q \times \overline{\mathbb{B}}^C \times Q)$ – most importantly, the Cartesian product – preserve well quasi-orders. □

We denote by $\|x\|$ the largest of the values $\|\tau\|$, where $\|\tau\|$ ranges over the minimal elements of x. Therefore, by Lemma 7.2, $\|x\|$ is finite for $x \in P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ (we assume $\|x\| = 0$ if $x = \emptyset$). We interpret $\|x\|$ as the largest number needed to describe the element x.

It is quite clear that given two elements x, y, $\|x \cdot y\| \leq \|x\| + \|y\|$. It is also not difficult to prove that $\|x^{\omega}\| \leq D \cdot \|x\|$ for some constant D depending on Q and C, but not on x. From this, one can easily derive an algorithm for computing x^{ω} for any given element x, giving rise to an efficient procedure which enumerates all elements of $A^{\langle \cdot ,\omega\rangle}$, thus proving decidability of the limitedness problem for B-automata. However, we can also derive the decidability result in another, more efficient way – just as in Leung's solution for distance automata – via a mapping $\alpha_{(1=2)}$ to a finite stabilization semigroup, which is a homomorphism of stabilization semigroups, and distinguishes the value ω from all finite values. We will define such a mapping in the following section, thus implying decidability of the limitedness problem for B-automata.

The rest of this chapter is devoted to the proof of Theorem 7.1.

7.2 The main theorem

In this section, we prove Theorem 7.1. We fix a finite set of counters C and a finite set of states Q. By $\mathbb{P}$ we denote the profinite semigroup $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$.

We begin with defining the appropriate finite stabilization semigroups.

7.2.1 The approximative semigroups for B-automata

We define the *approximative semigroups* for B-automata. Let N be a fixed positive integer. By $\overline{\mathbb{B}}_{/N=N+1}$ denote the set

$$\{0,1,\ldots,N\} \cup \{0,1\ldots,N\}^3 \cup \{\omega\}.$$

Consider a semigroup structure over $\overline{\mathbb{B}}_{/N=N+1}$ defined as in Figure 7.2, but in which addition is only up to threshold N, i.e. if k, l are finite numbers, then $k+l$ evaluates to the usual sum, provided that it does not exceed N; otherwise, $k+l$ evaluates to N. Note that $\overline{\mathbb{B}}_{/N=N+1}$ has also a structure of a stabilization semigroup, defined as in Figure 7.3, where again addition is considered up to threshold N.

There is a natural mapping $\alpha_{(N=N+1)} \colon \overline{\mathbb{B}} \to \overline{\mathbb{B}}_{/N=N+1}$, which replaces finite values larger than N by N. It is clear that this is a homomorphism of partially ordered semigroups, and also of stabilization semigroups.

By $\mathbb{P}_{/N=N+1}$ we denote the approximative semigroup $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C_{/N=N+1} \times Q)$. Then, the homomorphism $\alpha_{(N=N+1)}$ lifts to a homomorphism of semigroups

$$\alpha_{(N=N+1)} \quad : \quad \mathbb{P} \quad \longrightarrow \quad \mathbb{P}_{/N=N+1}\,.$$

This follows from the mentioned fact that all the used constructions (Cartesian product, the "semi-direct" product, the functor $P_{\uparrow}$) are functors in respective categories, so they preserve homomorphisms.

We will prove that via the homomorphism $\alpha_{(N=N+1)}$, the ω-power in $\mathbb{P}$ induces a stabilization operation in $\mathbb{P}_{/N=N+1}$. Recall that $\overline{\mathbb{B}}_{/N=N+1}$ has a structure of a stabilization semigroup and that the mapping $\alpha_{(N=N+1)} \colon \overline{\mathbb{B}} \to \overline{\mathbb{B}}_{/N=N+1}$ which replaces numbers larger than N by N is a homomorphism of stabilization semigroups.

Remark 7.1. T. Colcombet [Col10b] considers a general construction of a stabilization semigroup $P_{\uparrow}S$ of upward-closed subsets of a *finite* stabilization semigroup S, equipped with a partial order, satisfying certain axioms (described in more detail in Section 11.2.1 of Part II of this thesis). However, that construction does not relate stabilization in $\mathbb{P}_{/N=N+1}$ with the ω-power in $\mathbb{P}$.

For $\sigma \in Q \times \overline{\mathbb{B}}^C_{/N=N+1} \times Q$, if $\sigma = (q, o, q)$ for some $q \in Q$ and $o \in \overline{\mathbb{B}}^C_{/N=N+1}$, then we define

$$\sigma^{\#} \quad \overset{def}{=} \quad (q,\ o^{\#}, q).$$

Proposition 7.3. *There is a unique operation # such that the following diagram commutes:*

$$\begin{tikzcd}
\mathbb{P} \arrow[r, "\alpha_{(N=N+1)}"] \arrow[d, "\omega"'] & \mathbb{P}_{/N=N+1} \arrow[d, dashed, "\#"] \\
\mathbb{P} \arrow[r, "\alpha_{(N=N+1)}"] & \mathbb{P}_{/N=N+1}
\end{tikzcd}$$

Equipped with this operation, $\mathbb{P}_{/N=N+1}$ becomes a stabilization semigroup. The mapping # has moreover the following description for idempotent $e \in \mathbb{P}_{/N=N+1}$:

$$e^{\#} \ = \uparrow\{\sigma_1 \cdot \sigma_e^{\#} \cdot \sigma_2 : \qquad \sigma_1, \sigma_e, \sigma_2 \in e \quad \textit{and} \quad \sigma_e = \sigma_e^2\}.$$

First, we prove a simple combinatorial property of $\overline{\mathbb{B}}^C$.

Lemma 7.4. *Let $\tau \in \overline{\mathbb{B}}^C$. Then there exists a number $k_0 \in \mathbb{N}$ with the following property. For any $k \geq k_0$ and sequence $\tau_1, \tau_2, \ldots, \tau_k$ of operations in $\overline{\mathbb{B}}^C$, if*

$$\tau_1 \cdot \tau_2 \cdots \tau_{k-1} \cdot \tau_k = \tau, \tag{1}$$

then there exist indices $i < j$ such that for $n = 1, 2, \ldots$

$$(\tau_1 \cdot \tau_2 \cdots \tau_{i-1}) \cdot (\tau_i \cdot \tau_{i+1} \cdots \tau_{j-1})^n \cdot (\tau_j \cdot \tau_{j+1} \cdots \tau_k) = \tau. \tag{2}$$

Proof. Let N be larger than $\|\tau\|$. Let k_0 be a number with the following property.

> *For any sequence $\sigma_1, \sigma_2, \ldots, \sigma_{k-1}, \sigma_k$ of elements of $\overline{\mathbb{B}}^C_{/N=\omega}$ with $k \geq k_0$, there exist indices $i < j$ such that $\sigma_i \cdot \sigma_{i+1} \cdots \sigma_{j-1}$ is idempotent.*

The existence of such a number k_0 is an elementary property of finite semigroups, which follows easily from the pigeonhole principle and the existence of idempotent powers in finite semigroups.

Let $\tau_1, \tau_2, \ldots, \tau_k$ be a sequence which satisfies (1), with $k \geq k_0$, and let $\sigma_1, \sigma_2, \ldots, \sigma_k$ be the respective images under $\alpha_{(N=\omega)}$. Apply above property of k_0 and conclude the existence of indices $i < j$ such that $\sigma_i \sigma_{i+1} \cdots \sigma_{j-1}$ is idempotent. We claim that equation (2) holds. To verify this, consider any counter $c \in C$, and let $\tau' = \tau_i \cdots \tau_{j-1}$. If $\tau[c] < \omega$, then $\|\tau[c]\| < N$, and it follows that $\|\tau'[c]\| < N$, so the image of $\tau'[c]$ in $\overline{\mathbb{B}}_{/N=\omega}$ looks just like $\tau'[c]$. Since this image is idempotent, it follows that $\tau'[c]$ itself is idempotent (this follows from multiplication in $\overline{\mathbb{B}}_{/N=\omega}$).

Therefore, we have shown that if $\tau[c] < \omega$, then $\tau'[c] = (\tau'[c])^n$ for $n \geq 1$, so the equation (2) holds when restricted to such coordinates c. For the remaining coordinates c, where $\tau[c] = \omega$, the equation (2) also obviously holds. □

Proof of Proposition 7.3. First we will verify that if $s \in \mathbb{P}$ is such that $e = \alpha_{(N=N+1)}(s)$ is idempotent, then

$$\alpha_{(N=N+1)}(s^{\omega}) \ = \uparrow\{\sigma_1 \cdot \sigma_e^{\#} \cdot \sigma_2 : \qquad \sigma_1, \sigma_e, \sigma_2 \in e \quad \text{and} \quad \sigma_e = \sigma_e^2\}. \tag{3}$$

($\supseteq$). To show the inclusion $\supseteq$ in the equality (3), consider any $\sigma_1, \sigma_e, \sigma_2 \in e$, with σ_e idempotent. In particular, the source and target states of σ_e coincide. Since $e = \alpha_{(N=N+1)}(s)$, there are corresponding transitions $\tau_1, \tau_e, \tau_2 \in s$, such that

$$\sigma_1 = \alpha_{(N=N+1)}(\tau_1),$$
$$\sigma_e = \alpha_{(N=N+1)}(\tau_e),$$
$$\sigma_2 = \alpha_{(N=N+1)}(\tau_2).$$

Then, for any $n \geq 3$, we may consider the sequence

$$\tau_1, \quad \underbrace{\tau_e, \quad \tau_e, \ldots, \tau_e,}_{n!-2} \quad \tau_2.$$

Since for $\tau \in \overline{\mathbb{B}}^C$, $\tau^{n!-2} \leq \tau^\omega$, we conclude that for every $n \geq 3$ there is a transition $\sigma_n \in s^{n!}$, with

$$\sigma_n \leq \tau_1 \cdot \tau_e^\omega \cdot \tau_2.$$

Let τ be a limit of some convergent subsequence of $\sigma_1, \sigma_2, \ldots$. It follows that

$$\tau \leq \tau_1 \cdot \tau_e^\omega \cdot \tau_2,$$

and since the sequence $(s^{n!})_{n=1}^\infty$ converges to s^ω, we deduce that $\tau \in s^\omega$.

Then,

$$\alpha_{(N=N+1)}(\tau) \leq \sigma_1 \cdot \sigma_e^\# \cdot \sigma_2,$$

because $\alpha_{(N=N+1)} : \overline{\mathbb{B}}^C \to \overline{\mathbb{B}}^C_{/N=N+1}$ is an order preserving homomorphism of stabilization semigroups. Since $\alpha_{(N=N+1)}(s^\omega)$ is an upward-closed set, it follows that it contains $\sigma_1 \cdot \sigma_e^\# \cdot \sigma_2$. This ends the proof of the inclusion $\supseteq$.

($\subseteq$). Now, we show the inclusion $\subseteq$ in the equation (3). Assume that

$$\tau \in s^\omega.$$

Let N be larger than $\|\tau\|$. Let us take n such that the distance between $s^{n!}$ and s^ω is smaller than 2^{-N}, i.e. s^ω and $s^{n!}$ have the same image in $\overline{\mathbb{B}}^C{}_{/N=\omega}$. This implies that $s^{n!}$ contains a transition τ' which agrees with τ on all entries smaller than N. In particular, $\tau' \leq \tau$, since $\|\tau\| < N$.

Since $\tau' \in s^{n!}$, there is a sequence of transitions in s

$$\tau'_1, \quad \tau'_2, \quad \tau'_3, \ldots \tau'_{n!-1}, \quad \tau'_{n!}$$

whose product is at most equal to τ'.

Using Lemma 7.4 and the pigeonhole principle (to make sure the state component recurs), it is straightforward to show that if $n!$ is sufficiently large, then there exists a pair of indices

$1 \leq i < j \leq n!$ such that:

$$\tau_1 \cdot \tau_e^\omega \cdot \tau_2 \leq \tau',$$

where

$$\begin{array}{rcl}\tau_1 & \overset{def}{=} & \tau'_1 \cdot \tau'_2 \cdots \tau'_{i-1}, \\ \tau_e & \overset{def}{=} & \tau'_i \cdot \tau'_{i+1} \cdots \tau'_{j-1}, \\ \tau_2 & \overset{def}{=} & \tau'_j \cdot \tau'_{j+1} \cdots \tau'_{n!}.\end{array}$$

It follows that

$$\tau_1 \cdot \tau_e^\omega \cdot \tau_2 \leq \tau.$$

Therefore, since $\alpha_{(N=N+1)} : \overline{\mathbb{B}}^C \to \overline{\mathbb{B}}^C_{/N=N+1}$ is an order preserving homomorphism of stabilization semigroups,

$$\alpha_{(N=N+1)}(\tau_1) \cdot \alpha_{(N=N+1)}(\tau_e)^\# \cdot \alpha_{(N=N+1)}(\tau_2) \leq \alpha_{(N=N+1)}(\tau).$$

Moreover,

$$\alpha_{(N=N+1)}(\tau_1), \alpha_{(N=N+1)}(\tau_e), \alpha_{(N=N+1)}(\tau_2) \in \alpha_{(N=N+1)}(s),$$

and $\alpha_{(N=N+1)}(\tau_e)$ is idempotent. This proves that $\alpha_{(N=N+1)}(\tau)$ belongs to the right-hand side of the equation (3), finishing the proof of the left-to-right inclusion.

We now extend the operation # to all elements of $\mathbb{P}_{/N=N+1}$, by defining

$$u^\# \overset{def}{=} e^\#,$$

where e is the idempotent power of u. It is clear that for any $s \in \mathbb{P}$,

$$\alpha_{(N=N+1)}(s^\omega) = \alpha_{(N=N+1)}(s)^\#. \tag{4}$$

It follows from Lemma 6.1 that $\mathbb{P}_{/N=N+1}$ is a stabilization semigroup and that $\alpha_{(N=N+1)}$ is a homomorphism of stabilization semigroups. □

7.2.2 Proof of Theorem 7.1

We only need to prove that for any finite set $A \subseteq \mathbb{P}$,

$$\overline{A^+} \subseteq A^{\langle\,\cdot\,,\omega\rangle}, \tag{1}$$

because the other inclusion holds in every profinite semigroup.

We give an informal sketch of the proof in case of $\mathbb{M}_Q\mathcal{T}$ instead of $\mathbb{P}$. In order to prove the inclusion (1), we need to show (roughly) that if there are matrices in A^+ with arbitrarily large entries at some set of positions, then in $A^{\langle\,\cdot\,,\omega\rangle}$ there is a matrix y having the ω entry at this set of positions. The way we prove it is as follows. Take a matrix v which is a product of a sequence of matrices in A, and which has (sufficiently) large entries at some set of positions. Then, consider a factorization tree f of this sequences of matrices in A, with respect to a mapping $\alpha_{(N=N+1)}$. By the factorization theorem, the tree can be assumed to have a height bounded by a number independent of the choice of v. The factorization tree f necessarily uses some stabilization rules (since it has many leaves, and a small height). The crucial step is to show that the output s of the tree f must have entries equal to ω where v has large entries. Then, the factorization tree f provides a recipe for producing (using multiplication and the ω-power) a matrix $y \in A^{\langle\, \cdot\, ,\omega\rangle}$, which has values equal to ω at the positions where v has large entries.

We now proceed to a formal proof of the Theorem 7.1. First, we prove a couple of lemmas.

Lemma 7.5. *Let $N \in \mathbb{N}$ and let $x, y \in \mathbb{P}$ be such that $\alpha_{(N=N+1)}(x) = \alpha_{(N=N+1)}(y)$ and $\|x\| < N$. Then $x = y$.*

Proof. It is easy to see that the property stated in the lemma holds for $\overline{\mathbb{B}}$ instead of $\mathbb{P}$. We now deduce that it holds for $\mathbb{P}$.

Assume that $\tau \in x$ is a minimal element. Then $\|\tau\| \leq \|x\| < N$ and there exists a $\sigma \in y$ such that $\alpha_{(N=N+1)}(x) = \alpha_{(N=N+1)}(y)$. Using the property of $\overline{\mathbb{B}}$, we deduce that $\tau = \sigma$. Hence, $\tau \in y$.

Now assume that $\sigma \in y$ is a minimal element. If $\|\sigma\| < N$ then we proceed as above and conclude that $\sigma \in x$. If $\|\sigma\| \geq N$, then $\alpha_{(N=N+1)}(\sigma)$ contains an entry equal to N. But $\alpha_{(N=N+1)}(\sigma) \in x$, so there is some minimal element $\tau \in x$ such that $\alpha_{(N=N+1)}(\tau) \leq \alpha_{(N=N+1)}(\sigma)$. Since τ is minimal, $\tau \in y$ by the reasoning above, and moreover, $\|\tau\| < N$. It follows that $\tau \leq \sigma$, so $\sigma = \tau$ by minimality of σ. This is a contradiction, since $\|\tau\| < N \leq \|\sigma\|$. □

Lemma 7.6. *Let f be an $\alpha_{(N=N+1)}$-factorization forest with output s. Then there exists an element $y \in A^{\langle\, \cdot\, ,\omega\rangle}$ such that $\alpha_{(N=N+1)}(y) = s$.*

Proof. The lemma is proved by induction on the structure of f. The most interesting case is when the topmost rule is the stabilization rule; the base and binary rules are trivial. Assume that $f = \#(f_1, f_2, \ldots, f_l)$, and the outputs of the trees $f_1, f_2, \ldots, f_l$ are all equal to e. By inductive assumption, there exists an element $y_1 \in A^{\langle\, \cdot\, ,\omega\rangle}$ such that $\alpha_{(N=N+1)}(y_1)$ is equal to the output of f_1. We take y to be equal to $y_1^{\#}$. Then, since $\alpha_{(N=N+1)}$ is a homomorphism of stabilization semigroups, it follows that $\alpha_{(N=N+1)}(y) = e^{\#}$, which is the output of the tree f. This proves the lemma. □

Let $\mathbb{P}_{/N=\omega}$ denote the finite semigroup $P_{\uparrow}(Q \times \overline{\mathbb{B}}^{C}_{/N=\omega} \times Q)$, and let

$$\alpha_{(N=\omega)} \quad : \quad \mathbb{P} \quad \longrightarrow \quad \mathbb{P}_{/N=\omega}$$

denote the quotient mapping, induced from the mapping $\alpha_{(N=\omega)} \colon \overline{\mathbb{B}} \to \overline{\mathbb{B}}_{/N=\omega}$.

The following easy lemma states that from the output of an $\alpha_{(N=N+1)}$-factorization tree, we can determine the image of its input under $\alpha_{(N=\omega)}$.

Lemma 7.7. *Let N be a finite number and A be a finite alphabet. Let f be an $\alpha_{(N=N+1)}$-factorization tree with input $v \in A^+$ and output $s \in \mathbb{P}_{/N=N+1}$. Then*

$$\alpha_{(N=\omega)}(v) = \alpha_{(N=\omega)}(s).$$

Proof. In the statement of the lemma, we consider the image $\alpha_{(N=\omega)}(v)$ of v under the mapping from $\mathbb{P}$ to $\mathbb{P}_{/N=\omega}$, and the image $\alpha_{(N=\omega)}(s)$ under the mapping from $\mathbb{P}_{/N=N+1}$ to $\mathbb{P}_{/N=\omega}$. Both these mappings are defined so that they replace any element τ with $\|\tau\| \geq N$ by ω. Basically, the reason for why $\alpha_{(N=\omega)}(v) = \alpha_{(N=\omega)}(s)$ is that in $\mathbb{P}_{/N=\omega}$ we have trivial stabilization, so the output of a factorization tree is equal to the image of its input. A formal proof requires some abstract nonsense arguments, which we now present.

Clearly, the following diagram commutes.

$$\begin{tikzcd}
& \mathbb{P}_{/N=N+1} \arrow[dr, "\alpha_{(N=\omega)}"] & \\
\mathbb{P} \arrow[ur, "\alpha_{(N=N+1)}"] \arrow[rr, "\alpha_{(N=\omega)}"] & & \mathbb{P}_{/N=\omega}
\end{tikzcd}$$

The mappings in the diagram are homomorphisms of semigroups, and, as we shall show, also of stabilization semigroups where stabilization in $\mathbb{P}_{/N=\omega}$ is equal to the ω-power. We only need to check this for the upper-right mapping in the diagram, as for the other two it was already argued before. Let $s \in \mathbb{P}_{/N=N+1}$. Since $\alpha_{(N=N+1)}$ is surjective, $s = \alpha_{(N=N+1)}(\tilde{s})$ for some $\tilde{s} \in \mathbb{P}$. We then have that:

$$\begin{aligned}\alpha_{(N=\omega)}(s^{\#}) = \alpha_{(N=\omega)}(\alpha_{(N=N+1)}(\tilde{s})^{\#}) &= \alpha_{(N=\omega)}(\alpha_{(N=N+1)}(\tilde{s}^{\omega})) \\ &= \alpha_{(N=\omega)}(\tilde{s}^{\omega}) = \alpha_{(N=\omega)}(\tilde{s})^{\omega} = \alpha_{(N=\omega)}(s)^{\omega},\end{aligned}$$

proving that the mapping $\alpha_{(N=\omega)}$ from $\mathbb{P}_{/N=N+1}$ to $\mathbb{P}_{/N=\omega}$ is a homomorphism of stabilization semigroups.

Let f be an $\alpha_{(N=N+1)}$-factorization tree as in the formulation of the lemma. Then, we may consider an induced factorization tree $\alpha_{(N=\omega)}(f)$, which is obtained from the factorization tree f by applying the mapping $\alpha_{(N=\omega)}$ to the outputs at each node of f. Then, $\alpha_{(N=\omega)}(f)$ is a factorization tree with respect to the mapping $\alpha_{(N=\omega)}\colon \mathbb{P} \to \mathbb{P}_{/N=\omega}$. Its input is v and its output is $\alpha_{(N=\omega)}(s)$. On the other hand, since stabilization is trivial in $\mathbb{P}_{/N=\omega}$, the output of an factorization tree is the image of its input, so

$$\alpha_{(N=\omega)}(s) = \alpha_{(N=\omega)}(v),$$

proving the lemma. □

The following lemma is crucial. It says that if the input of a tree has small values, then its output must have small values. In the sketch of the proof we presented earlier, the crucial step was the contrapositive of this implication.

We introduce some notation. For any operation $o \in \overline{\mathbb{B}}$ and number $L \in \mathbb{N}$, we define the operation $o + L \in \overline{\mathbb{B}}$ in the obvious way: it is equal to $(n + L)$ if $o = (n)$, $(n_1 + L, n_2 + L, n_3 + L)$ if $o = (n_1, n_2, n_3)$ and ω if $o = \omega$. It is trivial to check that for any $\sigma, \tau \in \overline{\mathbb{B}}$ and $L \in \mathbb{N}$,

$$(\tau + L) \cdot \sigma \leq (\tau \cdot \sigma) + L \quad \text{and} \quad \tau \cdot (\sigma + L) \leq (\tau \cdot \sigma) + L.$$

This addition with elements of $\mathbb{N}$ lifts from $\overline{\mathbb{B}}$ to $\overline{\mathbb{B}}^C$ and further to $Q \times \overline{\mathbb{B}}^C \times Q$, still preserving the above inequalities.

If $\sigma \in \overline{\mathbb{B}}^C_{/N=N+1}$, then by $\sigma + L$ we denote the element $\tilde{\sigma} + L \in \overline{\mathbb{B}}^C$, where $\tilde{\sigma}$ is the unique operation in $\overline{\mathbb{B}}^C$ which has exactly the same entries as σ.

Lemma 7.8. *Let N be a finite number and A be a finite set of elements of $\mathbb{P}$ with $\|u\| < N$ for $u \in A$. Let f be an $\alpha_{(N=N+1)}$-factorization tree of height h, with input $v \in A^+$ and output $s \in \mathbb{P}_{/N=N+1}$. Then, for every $\sigma \in s$ there exists $\tau \in v$ such that*

$$\alpha_{(N=\omega)}(\tau) = \alpha_{(N=\omega)}(\sigma), \tag{2}$$
$$\tau \leq \sigma + L, \tag{3}$$

where $L = L(h, N)$ depends only on h and N.

Proof. Let $L(h, N) = N \cdot 6^h$. Note that since v is upward-closed, it suffices to consider only the case of a minimal element $\sigma \in s$. We prove the property by induction on the structure of the tree f. The only interesting case is when the topmost rule is the stabilization rule, but we consider all three cases.

Base rule If f is a factorization tree consisting only of the base rule, then $h = 0$, $v \in A$ and $s = \alpha_{(N=N+1)}(v)$. We assume that $\sigma \in s$ is minimal. Then, $\|\sigma\| < N$, since $\|v\| < N$ and σ is a minimal element in $\alpha_{(N=N+1)}(v)$. Let $\tau \in v$ be a minimal element such that

$$\alpha_{(N=N+1)}(\tau) = \alpha_{(N=N+1)}(\sigma).$$

Then, $\|\tau\| < N$, and so τ and σ have exactly the same entries. Therefore, $\tau \leq \sigma + 0$, and also $\alpha_{(N=\omega)}(\tau) = \alpha_{(N=\omega)}(\sigma)$.

Stabilization rule The interesting case is when the topmost rule is the stabilization rule, that is when $f = \#(f_1, \ldots, f_l)$. Let $v_1, \ldots, v_l \in A^+$ be the inputs of $f_1, \ldots, f_l$ and let e be the output of $f_1, \ldots, f_l$. The output of f is $s = e^\#$. Suppose all the factorization trees $f_1, \ldots, f_l$ have height h.

Let $L = L(h, N)$. Choose any minimal $\sigma \in e^\#$. We show that there exists a $\tau \in v$ such that $\tau \leq \sigma + 6L$. Since σ is minimal in $e^\#$, there exist $\sigma_1, \sigma_e, \sigma_l \in e$ with σ_e idempotent and such that

$$\sigma = \sigma_1 \cdot \sigma_e^\# \cdot \sigma_l.$$

By the inductive assumption, there exist:

- $\tau_1 \in v_1$ such that $\alpha_{(N=\omega)}(\tau_1) = \alpha_{(N=\omega)}(\sigma_1)$ and $\tau_1 \leq \sigma_1 + L$,
- $\tau_k \in v_k$ such that $\alpha_{(N=\omega)}(\tau_k) = \alpha_{(N=\omega)}(\sigma_e)$ and $\tau_k \leq \sigma_e + L$, for $k = 2, 3, \dots l-1$,
- $\tau_l \in v_l$ such that $\alpha_{(N=\omega)}(\tau_l) = \alpha_{(N=\omega)}(\sigma_l)$ and $\tau_l \leq \sigma_l + L$.

Let $\tau_e = \tau_2 \cdot \tau_3 \cdots \tau_{l-1}$. We check that for each coordinate $c \in C$, the inequality $\tau_e[c] \leq \sigma_e^{\#}[c] + 4L$ holds. We do this by considering all possible cases for $\sigma_e[c]$.

1. $\sigma_e[c] = 0$. Then $\tau_k[c] = 0$ for $2 \leq k \leq l-1$, so $\tau_e[c] = 0 = \sigma_e^{\#}[c]$.
2. $\sigma_e[c] \in \{1, \dots, \omega\}$. Then $\sigma_e^{\#}[c] = \omega$ and $\tau_e[c] \leq \omega$.
3. $\sigma_e[c] = (n_1, n_2, n_3)$. Then $\sigma_e^{\#}[c] = (n_1, n_2, n_3)$ since σ_e is idempotent and
$$\tau_e[c] \leq (n_1 + L, \max(n_2 + L, n_1 + n_3 + 2L), n_3 + L) \leq (n_1, n_2, n_3) + 4L,$$
since $n_1, n_3 \leq N \leq L$.

It follows that $\tau_e \leq \sigma_e^{\#} + 4L$. Taking $\tau = \tau_1 \cdot \tau_e \cdot \tau_l$, we have $\tau \in v$ and
$$\tau \leq (\sigma_1 + L) \cdot (\sigma_e^{\#} + 4L) \cdot (\sigma_2 + L) \leq (\sigma_1 \cdot \sigma_e^{\#} \cdot \sigma_2) + 6L = \sigma + L(h+1, N).$$

Moreover, it follows from idempotency of $\alpha_{(N=\omega)}(\sigma_e) = \alpha_{(N=\omega)}(\tau_e)$ that $\alpha_{(N=\omega)}(\tau) = \alpha_{(N=\omega)}(\sigma)$. This finishes the inductive step in the case of a stabilization rule.

Binary rule The case of the binary rule is only simpler than the previous case. In fact, the calculations are identical as in the degenerate stabilization case, when $l = 2$ and σ_e does not exist (or is the neutral element). □

Proof of Theorem 7.1 We proceed with the proof of Theorem 7.1, and more precisely, of the inclusion $\overline{A^+} \subseteq A^{\langle \cdot, \omega \rangle}$. Let $x \in \overline{A^+}$. Let N be any number larger than $\|x\|$ and $\|u\|$ for every $u \in A$. Let $h = \|\mathbb{P}_{/N=N+1}\|$ and let $L = L(h, N)$ be as in the lemma.

Since $x \in \overline{A^+}$, there exists an element $v \in A^+$ which is very close to x – we may assume that x and v have the same image under $\alpha_{(N+L+1=\omega)}$. For v treated as a word over A, there is a factorization tree f of height at most h. Let $s \in \mathbb{P}_{/N=N+1}$ be its output.

We will prove that
$$\alpha_{(N=N+1)}(x) = s. \tag{4}$$

From the above equality, the theorem follows – by Lemma 7.6, we may find an element $y \in A^{\langle \cdot, \omega \rangle}$ such that $\alpha_{(N=N+1)}(y) = s = \alpha_{(N=N+1)}(x)$. Since $\|x\| < N$, Lemma 7.5 implies that $x = y$ and so $x \in A^{\langle \cdot, \omega \rangle}$, concluding the theorem.

We now prove equation (4), by showing inclusions in both ways.

For the left-to-right inclusion, it suffices to show that the minimal elements of x are mapped by $\alpha_{(N=N+1)}$ to s. Let τ be a minimal element in x. Then, every entry of τ is either equal to ω, or is strictly smaller than N. Since $\alpha_{(N=\omega)}(x) = \alpha_{(N=\omega)}(v) = \alpha_{(N=\omega)}(s)$, it follows that there exists a $\sigma \in s$ such that $\alpha_{(N=\omega)}(\tau) = \alpha_{(N=\omega)}(\sigma)$. Then $\sigma \leq \alpha_{(N=N+1)}(\tau)$, since τ has all finite entries smaller than N. Therefore, $\alpha_{(N=N+1)}(\tau) \in s$ since s is upward-closed.

For the other inclusion, we show that if $\sigma \in s$ then $\sigma \in \alpha_{(N=N+1)}(x)$. Apply Lemma 7.8 to A and N and the factorization tree f. Let τ satisfy the property stated in the lemma. Since x is close to v it follows that there is some $\tau' \in x$ such that τ and τ' agree with each other on all entries not larger than $N + L$.

Assume that the transition σ has an entry equal to N. Then, by the inequality (3) in Lemma 7.8, the corresponding entry in τ is at most equal to $N + L$. Hence, it matches the entry in τ'. It follows that $\alpha_{(N=N+1)}(\tau') \leq \sigma$. Because $\tau' \in x$ and $\alpha_{(N=N+1)}(x)$ is upward-closed, it follows that $\sigma \in \alpha_{(N=N+1)}(x)$.

This ends the proof of the equality (4), finishing the proof of Theorem 7.1.

PART II

B/S-automata and languages of profinite words

CHAPTER 8

Overview

In Part II of this thesis, we create a profinite theory for B-automata. In this chapter, we give a broad overview of this theory. Apart from a construction in Section 11.3, this part is independent from Part I.

B- and S-automata First, let us recall the definition of B- and S-automata. The definition of B-automata which we now present differs slightly from the one used in Part I, as we consider only one unachievable value ω, instead of two distinct values ω and $\perp$. Also, for a homogeneous definition of B- and S-automata, only the values attained by a counter prior to a reset are taken into account. These slight differences do not restrict the expressivity of the models.

B-automata and *S-automata* are nondeterministic automata over finite words, which are moreover equipped with a finite number of counters. There are two counter operations available for each counter: *inc* and *reset* – *inc* increases the current value of the counter by 1 and *reset* sets the value to 0. A transition of an automaton may trigger any sequence of operations on its counters. If the operation *reset* is performed in a run ρ on a counter which currently stores a value n, then we say that *n is a reset value* in the considered run ρ.

First we define the *valuation* of a word under a given B-automaton $\mathcal{A}$. Recall that $\mathcal{A}$ is nondeterministic, so there might be many runs over a single word. For a particular run ρ of $\mathcal{A}$, we define the *value* of ρ as its maximal reset value, i.e. the number

$$\max\{n : \textit{ in the run } \rho\textit{, the value } n \textit{ is a reset value}\}.$$

We will always assume that $\max(\emptyset)$ is equal to 0. Next, the valuation $f_{\mathcal{A}}(w)$ of an input word w under the automaton $\mathcal{A}$ is the minimum of the values of all accepting runs ρ over w:

$$f_{\mathcal{A}}(w) = \min_{\rho} \max\{n : \textit{ in the run } \rho\textit{, the value } n \textit{ is a reset value}\}. \tag{1}$$

Note that min ranges only over the accepting runs ρ of $\mathcal{A}$. In Part II we use the convention that $\min(\emptyset) = \omega$ (this differs from the Part I, where $\min(\emptyset) = \perp$).

If $\mathcal{A}$ is an S-automaton, the definition of a valuation $f_{\mathcal{A}}(w)$ of an input word w is completely dual. We simply swap min with max in the formula (1). Therefore, the value of the run is the minimal reset value (and ω if there are none), and the value of the word is the maximal value of all accepting runs.

Example 8.1 (The running example). Consider the following deterministic B-automaton $\mathcal{A}$ with one state and one counter, over the alphabet $A = \{a, b\}$. The counter is incremented whenever a letter a is encountered. When reading a letter b, the automaton resets its counter. Then, for an input word $w = a^{n_1}ba^{n_2}\ldots ba^{n_k}$,

$$f_{\mathcal{A}}(w) = \max\{n_1, n_2, \ldots, n_k\}.$$

Formally, due to the definition chosen in Part II, there is a slight complication, since the counter is not reset after the last block of a's. We may, however, declare that the automaton resets some specified counters after reaching the end of the input word. This does not change the expressivity of the model, since the automaton may guess the end of the word using nondeterminism. This shows that the model of B-automata considered in Part II can simulate the model from Part I. Conversely, in the model from Part I, to avoid taking into account the values of a counter c after the last reset, the automaton can guess that this was the last reset and skip all further increments of counter c.

The function $f_{\mathcal{A}}$ described above can be also defined by a nondeterministic S-automaton $\mathcal{B}$, depicted in Figure 8.1. The automaton $\mathcal{B}$ has three states p, q, r, the first two being initial and the last accepting. There is one counter. By using nondeterminism and the initial states p and q, $\mathcal{B}$ may decide to find itself in state q at the beginning of a chosen block of a's of the input word, without having touched the counter previously. In the state q, the automaton counts the length of the current block of a's, and at any moment up to the first encountered b, it may choose to reset the counter and move to the idle, accepting state r.

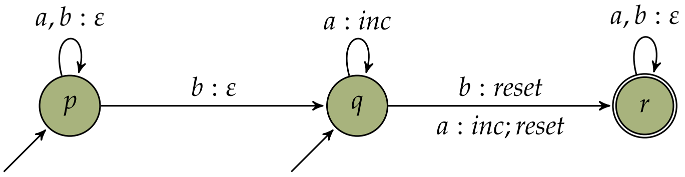


FIGURE 8.1: *An S-automaton with one counter*

Therefore, an accepting run of $\mathcal{B}$ may count the length of a single consecutive block of a's of the input word, and this block of $a's$ may be chosen arbitrarily using nondeterminism. It follows that $f_{\mathcal{B}}$ computes the largest block of a's in the input word, so $f_{\mathcal{B}}$ is precisely the same function as $f_{\mathcal{A}}$.

Example 8.2. Let $\mathcal{A}$ be a finite nondeterministic automaton. We may view $\mathcal{A}$ as a B-automaton with no counters. Then, it defines a function $f_{\mathcal{A}}$ which gives 0 for any word accepted by $\mathcal{A}$ and ω for any rejected word. Dually, if we treat $\mathcal{A}$ as an S-automaton, it defines a function $f_{\mathcal{A}}$

which gives ω for any accepted word, and 0 for any rejected word. Therefore, when defining the semantic of $f_{\mathcal{A}}$, it is important to know whether $\mathcal{A}$ is a B- or S-automaton.

Limitedness The central *limitedness problem* for B- or S-automata is the following decision problem:

> *For a given B-automaton $\mathcal{A}$, decide whether the function $f_{\mathcal{A}}$ has a finite range.*

The B-automaton in the example is not limited, since $f_{\mathcal{A}}(a^n) = n$ for any $n \in \mathbb{N}$.

Remark 8.1. Note that the range of $f_{\mathcal{A}}$ may contain the element ω and still be finite. This is the case for instance for finite automata, as described in Example 8.2, where the function $f_{\mathcal{A}}$ attains only the values 0 and ω. However, when concerning the limitedness problem, it is harmless to assume that the range of $f_{\mathcal{A}}$ does not contain the element ω, i.e. $f_{\mathcal{A}}(w) < \omega$ for every $w \in A^+$. This corresponds to the assumption that the underlying finite automaton accepts all finite words. (This is why we do not need the value $\perp$ which was considered in Part I of this thesis.) Indeed, consider for instance a B-automaton $\mathcal{A}$ such that the finite automaton underlying $\mathcal{A}$ accepts the regular language $L \subseteq A^+$. Let $\mathcal{A}'$ be a finite automaton accepting the complement of L, and we will treat $\mathcal{A}'$ as a B-automaton, as in Example 8.2. We construct a new B-automaton $\mathcal{B}$ as a disjoint union of $\mathcal{A}$ with $\mathcal{A}'$. Then we have that $f_{\mathcal{B}}(w) < \omega$ for every $w \in A^+$. Moreover, $\mathcal{A}$ is limited if and only if $\mathcal{B}$ is limited, since the range of $f_{\mathcal{A}}$ differs from the range of $f_{\mathcal{B}}$ at most by the single element ω.

Outline The correspondence between B- and S-automata illustrated in the running example is part of a bigger theory. In the rest of this chapter, we give an overview of the profinite theory of B- and S-automata, developed in Part II. In the profinite theory, a B- or S-automaton defines a language of profinite words, i.e. a subset of $\widehat{A^+}$. The limitedness problem for B-automata then appears as the universality problem. One of the consequences of this theory is another proof of decidability of the limitedness problem for B-automata, which is mostly independent from the proof presented in Part I.

The main theorem, analogously to the classical theorems from the theory of regular languages, talks about equivalent characterizations of some classes of languages of profinite words – via B- and S-automata, via B- and S-regular expressions, via recognizability by semigroups, via MSO+inf logic, and finally, via a finite index property. Each of these notions will be analyzed in detail in further sections.

Let us fix once and for all a finite alphabet A. Many notions will implicitly assume this alphabet: finite words are assumed to be elements of A^+ and profinite words – elements of $\widehat{A^+}$. In the examples, if not announced otherwise, we will more concretely assume the alphabet $A = \{a, b\}$.

Languages recognized by B- and S-automata

The essential idea underlying our theory is to consider not only finite words, but also profinite words. It turns out that for a given B- or S-automaton $\mathcal{A}$, we can easily evaluate the function $f_{\mathcal{A}}$ over any *profinite* word, as we now describe.

Let $\mathcal{A}$ be a B- or S-automaton. The following, simple observation is vital. For any $n \in \mathbb{N}$, the set

$$f_{\mathcal{A}}^{-1}([0,n]) = \{w \in A^+ : \ f_{\mathcal{A}}(w) \le n\}$$

is a regular language, so it makes sense to say that a given profinite word x satisfies $f_{\mathcal{A}}(x) \le n$ if x lies in the closure of this regular set. Formally, for $x \in \widehat{A^+}$, we define

$$\widehat{f_{\mathcal{A}}}(x) = \min\{n \in \mathbb{N} : \ x \in \overline{f_{\mathcal{A}}^{-1}([0,n])}\}.$$

This value may happen to be ω. It is straightforward to show that $\widehat{f_{\mathcal{A}}}$ is a continuous function from $\widehat{A^+}$ to $\overline{\mathbb{N}}$. By density of A^+ in $\widehat{A^+}$, the extension of $f_{\mathcal{A}}$ to a continuous function defined over $\widehat{A^+}$ is unique, so we will further identify $f_{\mathcal{A}}$ with the continuous mapping $\widehat{f_{\mathcal{A}}}$ defined over the entire set $\widehat{A^+}$.

Similarly to the idea underlying cost functions, we are not interested in the exact values of the function $f_{\mathcal{A}}$. We are only interested in sequences of words over which $f_{\mathcal{A}}$ grows indefinitely. By continuity of $f_{\mathcal{A}} \colon \widehat{A^+} \to \overline{\mathbb{N}}$ and compactness of $\widehat{A^+}$, this relevant information is encoded in the closed set

$$\{x \in \widehat{A^+} : \ f_{\mathcal{A}}(x) = \omega\}.$$

Following this idea, we define the languages of profinite words recognized by B- and S-automata. An S-automaton $\mathcal{A}$ defines the closed set $L(\mathcal{A})$, consisting of all profinite words x such that $f_{\mathcal{A}}(x) = \omega$. For a B-automaton $\mathcal{A}$, it is useful to define $L(\mathcal{A})$ dually, as the open set of profinite words x for which $f_{\mathcal{A}}(x) < \omega$. In either case, we call $L(\mathcal{A})$ the *language recognized by* $\mathcal{A}$. The motivation for the distinct definitions of $L(\mathcal{A})$ is that S-automata try to maximize, while B-automata try to minimize the value of a run.

Example 8.3. Let $\mathcal{A}$ be the B-automaton from Example 8.1, computing the largest block of a's. Then $L(\mathcal{A})$ is the language of all profinite words for which every block of a's has uniformly bounded length:

$$\begin{aligned} L(\mathcal{A}) &= \{x \in \widehat{A^+} : \ f_{\mathcal{A}}(x) < \omega\} = \\ &= \bigcup_{n \in \mathbb{N}} \{x \in \widehat{A^+} : \ \text{every block of } a\text{'s in } x \text{ has length bounded by } n\} \end{aligned}$$

It is not difficult to show that a profinite word has arbitrarily long blocks of a's if and only if it contains a^ω as an infix. Therefore, if $\mathcal{B}$ is the S-automaton from Example 8.1 (recall that

$f_{\mathcal{A}} = f_{\mathcal{B}}$), we deduce that

$$\begin{aligned} L(\mathcal{B}) &= \{x \in \widehat{A^+} : \; f_{\mathcal{B}}(x) = \omega\} = \widehat{A^+} - L(\mathcal{A}) = \\ &= \{x_1 \cdot a^{\omega} \cdot x_2 : \; x_1, x_2 \in \widehat{A^+}\}. \end{aligned}$$

Limitedness Assume that we want to test for limitedness of a B-automaton $\mathcal{A}$. As mentioned earlier, we may safely assume that the underlying finite automaton accepts all finite words. Then, $\mathcal{A}$ is limited if and only if $L(\mathcal{A}) = \widehat{A^+}$. Indeed, assume that $\mathcal{A}$ is limited. Then, there exists a bound N such that

$$A^+ \subseteq f_{\mathcal{A}}^{-1}([0, N]),$$

so $f_{\mathcal{A}}(x) \leq N$ for every $x \in \widehat{A^+}$ and therefore $L(\mathcal{A}) = \widehat{A^+}$. Conversely, assume that $L(\mathcal{A}) = \widehat{A^+}$, i.e. $f_{\mathcal{A}}(x) < \omega$ for every $x \in \widehat{A^+}$. By compactness and continuity, $f_{\mathcal{A}}$ attains its supremum in $\widehat{A^+}$, i.e. there exists a point $x_0 \in \widehat{A^+}$ such that $f_{\mathcal{A}}(x) \leq f_{\mathcal{A}}(x_0)$ for all $x \in \widehat{A^+}$. Since $f_{\mathcal{A}}(x_0) < \omega$, this proves that the range of $f_{\mathcal{A}}$ is finite.

B- and S-regular expressions

We consider regular expressions, extended by two new operations. Except for the Kleene star, which represents arbitrary, unrestricted iteration, a *B-regular expression* allows the use of the *bounded iteration*, denoted $L^{<\infty}$. Dually, an *S-regular expression* allows the use of the *infinite iteration* denoted L^{∞}.

In the classical theory, a regular expression E defines a language $L_E \subseteq A^*$. However, in our setting, the languages corresponding to regular expressions will be subsets of $\widehat{A^*}$. In particular, even a regular expression which does not use the operations $L^{<\infty}$ nor L^{∞} will be interpreted as the closure $\overline{L_E}$ of the regular language L_E. For instance, if $A = \{a, b\}$, then the regular expression $(a+b)^*$ describes the set of all profinite words over A (including the empty word). More generally, B-regular expressions define open subsets of $\widehat{A^*}$, which we will call *B-regular languages*, while S-regular expressions define closed subsets of $\widehat{A^*}$, which we will call *S-regular languages*. (They do not cover all the closed and open subsets, since there are uncountably many of those.) For instance, the S-regular expression $(a+b)^{\infty}$ describes the set of all *infinite* profinite words over $A = \{a, b\}$, i.e. elements of $\widehat{A^*} - A^*$, while the B-regular expression $(a+b)^{<\infty}$ describes precisely its complement, i.e. the elements of A^*. The precise semantics of these expressions will be given in Section 9.3. Now we will only give two examples which should be enough to pass the intuitions.

Example 8.4. The B-regular expression

$$(a^{<\infty}\, b)^*\, a^{<\infty}$$

describes precisely the language accepted by the B-automaton $\mathcal{A}$ from the previous examples, i.e. the language of profinite words for which every block of a's has a finite length (as noted in the previous example, this is equivalent to the existence of a common bound on the length of all a-blocks).

The S-regular expression

$$(a+b)^* \, a^\infty \, (a+b)^*$$

describes precisely the complement of the expression above. This is the language accepted by the S-automaton $\mathcal{B}$ from the previous examples. We will often view B- or S-regular expressions as defining subsets of $\widehat{A^+}$ instead of $\widehat{A^*}$, by simply ignoring the empty word ε.

Remark 8.2. Note that the expression a^∞ defines a language which is uncountable – it contains all profinite words of infinite length, which use only the letter a. It is a different language than $\{a^\omega\}$ which contains only a single word. In fact, any S-regular expression defines either a finite language of finite words or an uncountable language of profinite words. This is because already one use of a Kleene star or of the infinite iteration produces a language which is uncountable.

Connection with cost functions There is a close connection between the framework of cost functions and *closed* languages in $\widehat{A^+}$. For any closed language $L \subseteq \widehat{A^+}$, we may consider the function $f_L \colon A^+ \to \overline{\mathbb{N}}$ defined by

$$f_L(w) = \frac{1}{d(w,L)},$$

where $d(w,L)$ is the distance in $\widehat{A^+}$ between w and L. We view the cost function $[f_L]$, i.e. the $\approx$-equivalence class of f_L, as the cost function induced by the language L.

The mapping $L \mapsto [f_L]$ from closed languages to cost functions is an order-preserving isomorphism onto its image. This image does not contain all cost functions, but it contains all the ones which are relevant for the theory. For instance, if $L = L(\mathcal{A})$ for some S-automaton $\mathcal{A}$, then f_L is $\approx$-equivalent to the function $f_{\mathcal{A}}$ computed by $\mathcal{A}$. Similarly, if $L = \widehat{A^+} - L(\mathcal{B})$ for some B-automaton $\mathcal{B}$, then f_L is $\approx$-equivalent to the function $f_{\mathcal{B}}$ computed by $\mathcal{B}$.

Syntactic congruence

Just as multiplication is intimately related with regular languages, multiplication together with the ω-power over $\widehat{A^+}$ appear to be of central importance for B- and S-regular languages. For notational reasons, we will view $(\widehat{A^+}, \cdot, \omega)$ as an algebra over the signature $\langle \cdot, \# \rangle$, where the ω-power of $\widehat{A^+}$ plays the role of the operation # of the signature. Hence, in any reference to the operation # in the profinite semigroup $\widehat{A^+}$, we implicitly refer to the ω-power. The signature $\langle \cdot, \# \rangle$ will appear in many introduced notions, such as the following.

Let $L \subseteq \widehat{A^+}$. Its *$\langle \cdot, \# \rangle$-syntactic congruence* $\simeq_L$ is the coarsest equivalence relation over $\widehat{A^+}$ which preserves multiplication, the ω-power, and membership in L. This equivalence can be defined explicitly, similarly to the definition of the Myhill-Nerode equivalence. We will give such a definition in Section 11.1. It is slightly more complicated than the usual Myhill-Nerode equivalence due to the fact that we are considering both multiplication and the ω-power.

Example 8.5. Let $L = (a^{<\infty}\ b)^*\ a^{<\infty}$ be the language of the B-automaton which computes the minimal length of a block of a's. It is easy to check that the equivalence classes of $\simeq_L$ are:

$$a^{<\infty}, \qquad (a^{<\infty}\ b)^+\ a^{<\infty}, \qquad (a+b)^*\ a^{\infty}\ (a+b)^*.$$

As usual, the syntactic congruence for the language L coincides with the syntactic congruence for the complement language, $\widehat{A^+} - L$.

Stabilization semigroups

We will consider languages $L \subseteq \widehat{A^+}$ whose $\langle\ \cdot\ , \#\rangle$-syntactic congruence has a finite index. Such a set yields a finite $\langle\ \cdot\ , \#\rangle$*-syntactic algebra*, i.e. the quotient $S_L = \widehat{A^+}/{\simeq_L}$. Since $\simeq_L$ is a congruence, the syntactic algebra is equipped with two operations – the usual multiplication, and *stabilization*, denoted #, which stems from the ω-power in the profinite semigroup. The syntactic algebra also naturally inherits a topology from $\widehat{A^+}$, which is usually non-Hausdorff, i.e. there might be singleton sets which are not closed. (In fact, the topology of S_L is Hausdorff if and only if L is clopen). Multiplication and stabilization in S_L are continuous with respect to the topology, and also satisfy several properties which are easily derived from the properties of multiplication and ω-power over $\widehat{A^+}$, for instance that stabilization yields idempotents. It turns out that syntactic algebras are nothing else than the stabilization semigroups introduced by Colcombet in [Col09].

Example 8.6. Let S_L denote the quotient set induced by the language L from the previous examples. As observed in Example 8.5, it consists of three equivalence classes, which we will denote by $[a], [b]$ and $[a^\infty]$, respectively. Multiplication, stabilization and topology over S_L flow from the properties of the three equivalence classes: multiplication is commutative and each element is idempotent, $[a^\infty]$ is the zero element and $[a]$ is the neutral element; stabilization maps $[a]$ to $[a^\infty]$ and is the identity over the remaining elements; $[a^\infty]$ is contained in the closure of $[a]$ and in the closure of $[b]$. In terms of the specialization preorder on S_L, this means that $[a^\infty] < [a]$ and $[a^\infty] < [b]$; the elements $[a]$ and $[b]$ are incomparable.

Recognizability by finite stabilization semigroups

We consider an analogue of of the notion of recognizability by semigroups in the classical theory. Recall that a subset $L \subseteq \widehat{A^+}$ is *recognizable* if there is a mapping $\alpha\colon A \to S$ to a finite semigroup such that for the induced homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ we have $L = \hat{\alpha}^{-1}(F)$ for some $F \subseteq S$. The *induced homomorphism* is the unique continuous homomorphism from $\widehat{A^+}$ to S (with the discrete topology) which extends α. The existence and uniqueness of this extension follows easily from the definition of $\widehat{A^+}$.

Instead of semigroups, we will be dealing with finite topological stabilization semigroups. These objects are assumed to satisfy several axioms which bind stabilization # and multiplication – for instance, that $(st)^\# s = s(ts)^\#$ for all $s, t \in S$. Moreover, there are some assumptions on compatibility of stabilization with the topology of S, most importantly, that for any $s \in S$, the sequence $s, s^{2!}, s^{3!}, \dots$ is convergent to $s^{\#}$. Note that one has to be careful with this notion of convergence, since, usually, the topology is not Hausdorff. Otherwise, recall that finite Hausdorff topological spaces have a discrete topology, so the last axiom boils down to saying that $s^{\#}$ is the idempotent power of S, which means that S is a classical semigroup equipped with the idempotent power as stabilization. However, we will be only considering stabilization semigroups which satisfy the T_0 separability axiom (corresponding to the fact that the specialization preorder is a partial order).

The following result plays a pivotal role in the theory, and its proof is difficult comparing to the classical case.

Theorem 8.1. *Let $\alpha\colon A \to S$ be any mapping from a finite alphabet A to a finite T_0 stabilization semigroup S. Then there exists a unique canonical continuous homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ extending α.*

Above, and in the future (unless stated otherwise), a *homomorphism* from $\widehat{A^+}$ to a stabilization semigroup S is required to preserve multiplication and map the ω-power in $\widehat{A^+}$ to stabilization in S. The *canonical* homomorphism $\hat{\alpha}$ mentioned in the theorem can be characterized, among others, by either of the following conditions.

– *For any closed set $F \subseteq S$, the language $\hat{\alpha}^{-1}(F)$ is S-regular*

– *For any open set $F \subseteq S$, the language $\hat{\alpha}^{-1}(F)$ is B-regular.*

Note that, unlike in the classical situation, the canonical continuous homomorphic extension is not necessarily the *unique* continuous homomorphic extension of α. We will call the canonical extension $\hat{\alpha}$ the *homomorphism induced* by α. We say that a language $L \subseteq \widehat{A^+}$ is *recognized* by the canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ if $L = \alpha^{-1}(F)$ for some $F \subseteq S$.

Example 8.7. Let S be the topological stabilization semigroup $\widehat{A^+}/{\simeq_L}$ from the previous example, whose elements are $[a], [b], [a^\infty]$. Let $\alpha\colon A \to S$ map a to $[a]$ and b to $[b]$. In this case, we will see that the quotient mapping $\alpha_L\colon \widehat{A^+} \to S$ is the homomorphism induced α.

The algebraic and topological structure of S_L are designed in such a way that α_L is indeed a continuous homomorphism. There are a couple of ways of seeing that α_L is a canonical extension of α. One way is to verify that the inverse images of the (nontrivial) closed subsets of S under α_L are $[a^\infty]$, $[a^\infty] \cup [a]$ and $[a^\infty] \cup [b]$, and each of these sets happens to have a representation as an S-regular language:

$$\begin{aligned}
[a^\infty] &= (a+b)^* a^\infty (a+b)^*,\\
[a^\infty] \cup [a] &= (a+b)^* a^\infty (a+b)^* \cup a^*\\
[a^\infty] \cup [b] &= (a+b)^* b (a+b)^* \cup a^\infty
\end{aligned}$$

This shows that α_L is the homomorphism induced by α. As we will see in Example 8.11, in fact α_L is *not* a unique continuous homomorphism extending α, i.e. there exists a continuous homomorphism $\beta\colon \widehat{A^+} \to S$ which extends α, but such that $\beta \neq \alpha_L$. Therefore, β is not canonical.

Logic

We define a variant of the MSO logic over profinite words. A formula of this logic defines a set of profinite words. Usually, in the case of finite words (see e.g. [Tho97]), one sees such a word as a model whose elements are positions of the word, and so a formula of MSO speaks about the positions of the word. The same can be done for infinite words indexed by natural numbers. However, in profinite words, "positions" are not well-defined. (One can define the first, second, last position, etc. but what is the set of positions of $(ab)^{\omega}$?) Still, one can talk about properties of a profinite word, such as e.g.:

– The first letter is an a
– There are exactly 5 letters a before the first b
– There are at least n letters a
– For every $n \in \mathbb{N}$, there are at least n letters a

The first three properties listed above are regular properties, i.e. they can be seen as defining regular languages of finite words. Because of that, for any profinite word x, we can say that x satisfies a regular property P iff for any sequence of finite words which converges to x, almost all words in the sequence satisfy the property P. Therefore, regular properties of finite words lift to well-defined properties of profinite words. The last property listed above is a conjunction of regular properties, so it is also a well defined property of profinite words. We will often contract this property to: "there are infinitely many a's".

Therefore, we see that we do not really need to quantify over sets of positions to describe properties of profinite words. To define the logic MSO over profinite words, we view the constructs of MSO as operations on languages of profinite words – for instance, the existential quantifier $\exists$ corresponds to language projection. Technically, a word over the alphabet $A \times \{0,1\}$ can be split into a profinite word over the alphabet A, and a *marking* – a profinite word over the alphabet $\{0,1\}$. For example, the formula $a(x)$ holds in those profinite words over $A \times \{0,1\}$, which contain precisely one symbol $(a,1) \in A \times \{0,1\}$ and no other symbols with a 1 on the last coordinate. A formula beneath a quantifier $\exists$ defines a language over the extended alphabet $A \times \{0,1\}$, and the projection forgets about the second coordinate. In a similar fashion we can interpret the the usual constructs of MSO: the second- and first-order quantifiers $\exists, \forall$, the Boolean connectives $\wedge, \vee, \neg$, the binary predicates $<, \in$ and the unary predicates $a(x)$, per each letter $a \in A$.

So far, the described logic allows the construction of precisely the class of clopen sets. To go beyond that, we add a new predicate $\inf(X)$, which, intuitively, checks if X marks infinitely many positions. A bit more formally, $\inf(X)$ holds if the marking of the profinite word contains infinitely many 1's. This is a closed property of profinite words over the alphabet $A \times \{0,1\}$. We will also use the negation of this predicate, $\mathrm{fin}(X)$, which tests for finiteness rather than infiniteness.

Example 8.8. The language $(a+b)^*a^{\infty}(a+b)^*$ which was considered before is precisely the set of profinite words which satisfy the following formula of MSO+inf:

$$\exists X.\, \mathrm{inf}(X) \ \wedge\ \forall x, y, z. \big(x \in X \ \wedge\ z \in X \ \wedge\ (x < y < z) \implies a(y)\big)$$

We denote the logic MSO extended by the quantifier inf by MSO+inf, and distinguish the syntactic fragment MSO+inf$^+$ (respectively, MSO+fin$^+$) where the predicate inf (respectively, fin) appears only under a positive number of negations. The formula in the example above is a formula of MSO+inf$^+$.

It follows easily from the definitions that the logic MSO+inf defines the smallest class of languages of profinite words which is closed under projection, direct images under letter-to-letter homomorphisms, Boolean operations, contains all clopen sets and the language $b^*(a \cdot b^*)^\infty$ of profinite words with infinitely many a's.

The main theorem

We are now ready to state the main result of the theory, which binds all the notions introduced above. By $A^{\langle\,\cdot\,,\omega\rangle}$ we denote the set of elements of $\widehat{A^+}$ which can be generated from A by applying multiplication and the ω-power. Note that $A^{\langle\,\cdot\,,\omega\rangle}$ is different than $\widehat{A^+}$, since the first set is countable, while the latter is not (an explicit profinite word in the difference is e.g. $a^{\omega-1}$).

Theorem 8.2. *Let $L \subseteq \widehat{A^+}$. The following conditions are equivalent:*

1. *L is defined by an S-regular expression,*
2. *$L = L(\mathcal{A})$ for some S-automaton $\mathcal{A}$,*
3. *L is definable in* MSO+inf$^+$*,*
4. *L is closed and is recognized by some canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ to some finite T_0 stabilization semigroup,*
5. *$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}$ and its $\langle\,\cdot\,,\#\rangle$-syntactic congruence has finite index.*

Dually, the following conditions are equivalent:

1′. *L is defined by a B-regular expression,*

2′. *$L = L(\mathcal{A})$ for some B-automaton $\mathcal{A}$,*

3′. *L is definable in* MSO+fin$^+$*,*

4′. *L is open and is recognized by some canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ to some finite T_0 stabilization semigroup,*

5′. *The complement $\widehat{A^+} - L$ satisfies either of the conditions 4-3.*

Moreover, L is recognized by a canonical mapping to a finite T_0 stabilization semigroup if and only if it is a Boolean combination of languages satisfying either of the above conditions.

Theorem 8.2 immediately yields that B-regular languages are precisely the complements of S-regular languages. It also implies the theorem of Colcombet [Col09] mentioned in the introduction, that S-automata and B-automata define the same classes of cost functions. Similar results where already proven by Bojańczyk and Colcombet [BC06] (regular expressions considered in this paper have a different semantic, and cost functions are not mentioned). Central to our framework are its algebraic foundations, where stabilization semigroups play a dominant role. This is very similar to what happens in the framework of Colcombet [Col09]. The novelty here is the description in terms of languages of profinite words, rather than cost functions, which allows defining the syntactic congruence, and leads to the discovery of the link between stabilization semigroups and the two operations, multiplication and ω-power, in the profinite semigroup.

The translations between the equivalent models listed in Theorem 8.2 – i.e. between regular expressions, automata, and logic formulas – are effective. It also follows easily from the developed theory that one can effectively test for emptiness of Boolean combinations of languages defined by these models. This decidability result extends the decidability results of K. Hashiguchi and D. Kirsten. It also extends the decidability result of T. Colcombet, since in our theory it makes sense to ask (and effectively test) for emptiness of a Boolean combination of MSO+inf$^+$ formulas, which is meaningless in the theory of T. Colcombet. (The theory of Colcombet considers a logic, called *cost* MSO , whose syntax is similar to the syntax of MSO+inf. However, there is no reasonable semantic of formulas in which the predicate corresponding to inf appears both positively and negatively.)

A rather easy application of the theorem is to use is it for determining whether a language of profinite words (or, equivalently, a cost function) is S- or B-regular, as demonstrated in the following example.

Example 8.9. Let $A = \{a\}$ and let $L = \{a^\omega\} \subseteq \widehat{A^+}$. It is not difficult to establish that the mapping from A^+ to $\overline{\mathbb{N}}$:

$$a^k \mapsto \min\{n \in \mathbb{N} : \ n \text{ does not divide } k\}$$

is a representative of the cost function $[f_L]$ corresponding to L.

We will show that L is not an S-regular language. One way of seeing this is by using Remark 8.2, which says that an S-regular language L is either finite and consists of finite words, or is uncountable.

Another way of proving that L is not S-regular is by checking that it has infinite index. For $k \in \mathbb{Z}$, let

$$x_k = a^{\omega+k}.$$

If $k, l \in \mathbb{Z}$, and $k \neq l$, then $x_k \not\simeq_L x_l$. This is because $x_k \cdot a^{\omega-k} \in L$ while $x_l \cdot a^{\omega-k} \notin L$ and $\simeq_L$ is required to be preserved multiplication by arbitrary profinite words and to respect membership in L. Therefore, each of the profinite words x_k, where $k \in \mathbb{Z}$, is in a distinct equivalence class with respect to $\simeq_L$. This implies that the syntactic congruence of L has an infinite index, so L is not an S-regular language, and $[f_L]$ is not a regular cost function.

Necessity of the assumptions

In Theorem 8.2, one of the equivalent conditions for L being S-regular is that

$$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}} \textit{ and its } \langle\,\cdot\,,\#\rangle\textit{-syntactic congruence has finite index.} \qquad (*)$$

The first part of this condition is an elementary property of S-regular languages and other classes of languages of closed subsets of $\widehat{A^+}$, as shown in the following lemma (we will see later on that the class of S-regular languages satisfies the assumptions of the lemma).

Lemma 8.3. *Let $\mathcal{L}$ be a class of languages of profinite words over A with the following properties.*

1. *$\mathcal{L}$ consists of closed languages, i.e. for every $L \in \mathcal{L}$, L is closed*
2. *Every nonempty language $L \in \mathcal{L}$ contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$*
3. *$\mathcal{L}$ is closed under intersections with clopen sets, i.e. if $L \in \mathcal{L}$ and $U \subseteq \widehat{A^+}$ is clopen, then $L \cap U \in \mathcal{L}$.*

Then, for every $L \in \mathcal{L}$,

$$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}. \tag{2}$$

Proof. Let $L \in \mathcal{L}$. The right-to-left inclusion in equation (2) is obvious, since L is closed and contains $L \cap A^{\langle\,\cdot\,,\omega\rangle}$.

For the left-to-right inclusion, let $x \in L$. Let U be a clopen neighborhood of x. Then, $L \cap U$ is nonempty, since it contains x, and $L \cap U \in \mathcal{L}$ by the third assumption on $\mathcal{L}$. By the second assumption on $\mathcal{L}$, $L \cap U$ contains an element $y \in A^{\langle\,\cdot\,,\omega\rangle}$. Therefore, $y \in A^{\langle\,\cdot\,,\omega\rangle} \cap L \cap U$, so y is an element in the neighborhood of U from the set $L \cap A^{\langle\,\cdot\,,\omega\rangle}$. Since U is an arbitrary clopen set, and clopen sets form a base of the topology of $\widehat{A^+}$, this implies that $x \in \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}$, finishing the left-to-right inclusion of equation (2). □

It is natural to ask, whether the condition $(*)$ for being S-regular does not follow from the weaker requirement:

L is closed and its syntactic congruence has finite index.

We will see a negative answer to this question. Another suspicious matter is the formulation of Theorem 8.1, which only assures of the uniqueness of the *canonical* continuous and homomorphic extension of α, instead of uniqueness among all continuous homomorphic extensions of α. In fact, both these points at issue are related, and we will see that the formulations are as strong as necessary.

The following example illustrates that the requirement $L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}$ in the theorem cannot be relaxed to $L = \overline{L}$.

Example 8.10. Let $A = \{a, b\}$. Let F denote the set of profinite words over A which contain all finite words over A as infixes. More precisely, for $w \in A^+$, let F_w be the set of profinite words which contain w as an infix, i.e.

$$F_w = \{x \cdot w \cdot y : \; x, y \in \widehat{A^*}\},$$

and define

$$F = \bigcap_{w \in A^+} F_w.$$

Then, F is a closed ideal in $\widehat{A^+}$, as an intersection of closed ideals. From compactness of $\widehat{A^+}$ we deduce that F is nonempty, since for any finite set of words $I \subseteq A^+$, the intersection $\bigcap_{w \in I} F_w$ is clearly nonempty. Moreover, F is a prime ideal, i.e.

$$x \cdot y \in F \quad \Longleftrightarrow \quad x \in F \vee y \in F. \tag{3}$$

This is a consequence of the simple fact that

$$x \notin F_v,\ y \notin F_w \quad \Longrightarrow \quad x \cdot y \notin F_{v \cdot w}.$$

Similarly, we can prove that for all $x \in \widehat{A^+}$,

$$x \in F \quad \Longleftrightarrow \quad x^\omega \in F, \tag{4}$$

which follows from the fact that for any infinite profinite word x (i.e. element of $\widehat{A^+} - A^+$),

$$x \notin F_v \quad \Longrightarrow \quad x^\omega \notin F_{v \cdot v},$$

and for all finite words x, $x \notin F$ and also $x^\omega \notin F$.

The equivalences (3) and (4) together state that F is a prime ideal with respect to both multiplication and the ω-power. As a consequence, the $\langle \cdot, \# \rangle$-syntactic congruence $\simeq_F$ has two equivalence classes: F and $\widehat{A^+} - F$. Hence, the syntactic congruence of F has a finite index and F is closed. However, F contains no element of the set $A^{\langle \cdot, \omega \rangle}$ (otherwise, F would contain an element of A, by the fact that F is a prime ideal with respect to the two operations). Therefore,

$$\emptyset = \overline{F \cap A^{\langle \cdot, \omega \rangle}} \neq F,$$

so F is not S-regular.

As a side remark, we deduce that the function

$$f_F(w) \approx \min\{|v| : v \in A^+ \text{ is not a factor of } w\}$$

is not a regular cost function.

From the above example we also obtain a negative answer to the question about the uniqueness of the homomorphic extensions, which arises naturally in view of Theorem 8.1. Recall that in Example 8.7 we have seen that the mapping α_L is the *canonical* continuous homomorphic extension of α. Now we will see that it is not unique, however.

Example 8.11. Let $F \subseteq \widehat{A^+}$ be the set from the previous example, and let $\alpha\colon A \to S$ be the mapping from Example 8.7, where $S = \{[a], [b], [a^\infty]\}$. Let $\beta\colon \widehat{A^+} \to S$ map elements of $\widehat{A^+}$ according to the following rules:

$$\begin{array}{rcl} a^{<\infty} & \mapsto & [a], \\ \big((a^{<\infty}\, b)^{+}\, a^{<\infty}\big)\ - F & \mapsto & [b], \\ \big((a+b)^{*}\, a^{\infty}\ (a+b)^{*}\big)\ \cup F & \mapsto & [a^{\infty}]. \end{array}$$

Using the observations made in the previous example, we verify that β is a continuous homomorphism, which clearly extends the mapping $\alpha\colon A \to S$ considered in Example 8.7. Indeed, over $\widehat{A^{+}} - F$, the mapping β agrees with the canonical homomorphism α_L computed in Example 8.7. Therefore, if $x, y \notin F$, then also $x \cdot y \notin F$, and so

$$\beta(x \cdot y) = \alpha_L(x \cdot y) = \alpha_L(x) \cdot \alpha_L(y) = \beta(x) \cdot \beta(y).$$

On the other hand, if $x \in F$ or $y \in F$, then $x \cdot y \in F$, so

$$\beta(x \cdot y) = [a^{\infty}] = \beta(x) \cdot \beta(y).$$

Therefore, β is a homomorphism with respect to multiplication. Similarly we prove that β preserves stabilization. Moreover, β is continuous, since we can easily verify that the inverse images of all the closed subsets of S are closed.

Since β extends the mapping α considered in Example 8.7, we see that the mapping α_L induced by α is not the unique continuous homomorphism extending α.

This completes the overview. In the following chapters, we formally define the notions and prove the results stated in this overview.

CHAPTER 9

Languages of profinite words

In this chapter, we will discuss several ways of describing a *language of profinite words*, i.e. a set of profinite words. Recall that a language of profinite words which is both closed and open corresponds to a regular language in A^+. An archetypical example extending this class of languages is the language of all *infinite profinite words*, i.e. elements of $\widehat{A^+} - A^+$. This language is a closed set, but is not an open set in $\widehat{A^+}$. Therefore, it does not correspond to a regular language in A^+. We will be mostly interested in languages which are either open or closed subsets of $\widehat{A^+}$, such as the set of all infinite profinite words, or its complement, $A^+ \subseteq \widehat{A^+}$. A convenient way to describe such languages is with the use of uniformly continuous functions defined over A^+, as we will describe in Section 9.1. Via this description, we will easily define the languages of profinite words recognized by B-automata, which are open languages, and the languages recognized by S-automata, which are closed languages. In Section 9.3 we will describe languages of profinite words using B- and S-regular expressions. They provide an alternative way of defining languages recognized by B- or S-automata. Finally, in Section 9.4, languages will be defined using an extension of the MSO logic to profinite words. This logic gives rise to a very large class, which includes languages described by B- and S-regular expressions, but also contains languages with a much more complicated topological structure. We will, however, distinguish a syntactical fragment of the logic, which corresponds precisely to the languages defined by B- and S-automata.

9.1 Uniformly continuous functions

We consider the profinite metric d over the space $\widehat{A^+}$, and also its restriction to A^+. Recall that if v, w are finite words, then $d(v, w) = 2^{-n}$ where n is the size of the smallest semigroup which distinguishes v from w. With this metric, A^+ is a discrete topological space, i.e. every word in A^+ is isolated. In particular, any function $f\colon A^+ \to \overline{\mathbb{N}}$ is continuous.

The situation is different in $\widehat{A^+}$, which has a more complicated topology ($\widehat{A^+}$ can be defined as the completion of A^+ with respect to the metric d). Not every function $f\colon \widehat{A^+} \to \overline{\mathbb{N}}$ is continuous. Consider for instance the function which to every finite word assigns 0 and to every infinite profinite word assigns 1. It is not continuous, since an infinite profinite word is a limit of a sequence of finite words, over which the value is 0; however the value of the limit is 1.

Recall that a function $f\colon A^+ \to \overline{\mathbb{N}}$ is *uniformly continuous* if

$$\forall_{\delta>0}\exists_{\varepsilon>0}\forall_{v,w\in A^+}\quad d(v,w)<\varepsilon \implies d'(f(v),f(w))<\delta.$$

Above, d' is a fixed compact metric over $\overline{\mathbb{N}}$ in which ω is the unique limit point, such as the metric considered in Example 2.1 of the Preliminaries. Since $\widehat{A^+}$ is compact, every continuous function over $\widehat{A^+}$ is automatically uniformly continuous. It follows from a result from general topology that uniformly continuous functions over A^+ are precisely the restrictions of (uniformly) continuous functions over $\widehat{A^+}$. Below we state a result which is more specific to our setting.

Proposition 9.1. *Let $f\colon A^+ \to \overline{\mathbb{N}}$ be any function. Then, the following conditions are equivalent:*

1. *f is uniformly continuous*
2. *f extends to a continuous function $\hat{f}\colon \widehat{A^+} \to \overline{\mathbb{N}}$*
3. *For any number n, the set $f^{-1}([0,n])$ is a regular language*
4. *For any number n, the set $f^{-1}(\{n\})$ is a regular language.*

Remark 9.1. It is a simple and known fact that a continuous mapping of metric spaces is uniquely determined by its restriction to a dense subset of the domain. Since A^+ is dense in $\widehat{A^+}$, it follows that the continuous extension $\hat{f}\colon \widehat{A^+} \to \overline{\mathbb{N}}$ of a uniformly continuous function f is *unique*.

Proof. As mentioned earlier, the equivalence of the first two conditions follow from simple topological reasonings. The equivalence of the last two conditions is immediate, by closure of regular languages under Boolean combinations. We now show the equivalence between the second and third conditions of the proposition.

Assume that f extends to a continuous function $\hat{f}\colon \widehat{A^+} \to \overline{\mathbb{N}}$, and let $n \in \mathbb{N}$. Then, the interval $[0,n]$ is a clopen subset of $\mathbb{N}$, so its inverse image under $\hat{f}$ is also a clopen set. Therefore, its intersection of with A^+,

$$\hat{f}^{-1}([0,n]) \cap A^+ = f^{-1}([0,n])$$

is a regular language by Proposition 4.9 stated in the Preliminaries.

Conversely, assume that for every n, the set $f^{-1}([0,n])$ is a regular language. Then, for any profinite word x we define

$$\hat{f}(x) = \min\{n \in \overline{\mathbb{N}} :\ x \in \overline{f^{-1}([0,n])}\}.$$

It is trivial to check that $\hat{f}$ extends f. Moreover, the inverse image of any closed interval $[0,n]$ is the set

$$\hat{f}^{-1}([0,n]) = \overline{f^{-1}([0,n])},$$

which is clopen by Proposition 4.9. It follows easily that $\hat{f}$ is a continuous function. □

Example 9.1. To test for uniform continuity of a function $f\colon A^+ \to \overline{\mathbb{N}}$, it is usually most convenient to use the characterization given by the third condition of the proposition. All the functions which, for a given input word w, compute the values listed below, are uniformly continuous.

- The length of w,
- The exponential of the length of w,
- The sum of the number of a's on even positions of w with the number of b's in w,
- 0 if $w \in L$ and ω if $w \notin L$, where L is a fixed regular language,
- The smallest/largest number n such that w has an infix $ba^n b$,
- The largest number n such that w has n infixes of the form ba^n,
- The smallest prime number n which divides the length of w,
- The length of the shortest word which is/isn't an infix of w.

On the other hand, a function which computes any of the following values listed below is *not* uniformly continuous.

- 0 if $w \in L$ and ω if $w \notin L$, where L is a fixed non-regular language,
- The difference between the number of a's and the number of b's in w, if it is positive, and 0 otherwise,
- The largest prime number n which divides the length of w.

Remark 9.2. Let $f_1, f_2, \ldots, f_n \colon A^+ \to \overline{\mathbb{N}}$ be uniformly continuous functions, and let $g \colon \overline{\mathbb{N}}^n \to \overline{\mathbb{N}}$ be continuous. Then, the function $h \colon A^+ \to \overline{\mathbb{N}}$

$$h(w) = g(f_1(w), f_2(w), \ldots, f_n(w))$$

is uniformly continuous. This follows from the fact that a composition of continuous functions is continuous.

The following proposition follows immediately from the third condition Proposition 9.1.

Proposition 9.2. *Let $\mathcal{A}$ be a B- or S-automaton. Then $f_{\mathcal{A}}$ is a uniformly continuous function.*

In fact, the same result could be obtained for any reasonable model of automata with counters which cannot be decremented, for instance alternating automata with counters.

Notation. We will often identify a uniformly continuous function $f \colon A^+ \to \overline{\mathbb{N}}$ with its unique continuous extension $\hat{f} \colon \widehat{A^+} \to \overline{\mathbb{N}}$. In particular, if $\mathcal{A}$ is a B- or S-automaton, then for any $x \in \widehat{A^+}$, $f_{\mathcal{A}}(x)$ is a well-defined element of $\overline{\mathbb{N}}$.

Let f be a uniformly continuous function and let $\hat{f}$ be its unique extension to $\widehat{A^+}$. Then we define the *ω-set* of f to be the set $L_f = \hat{f}^{-1}(\omega) \subseteq \widehat{A^+}$. By continuity of $\hat{f}$, the ω-set is a closed language of profinite words.

Remark 9.3. Any closed language L of profinite words can be obtained as an ω-set of some uniformly continuous function $f\colon A^+ \to \overline{\mathbb{N}}$. Indeed, we may take

$$f_L(x) = \frac{1}{d(x,L)},$$

where

$$d(x,L) = \inf\{d(x,y) :\ y \in L\}$$

is the distance between x and L. It is an elementary fact from topology that $x \mapsto d(x,L)$ is a continuous function from $\widehat{A^+}$ to $\overline{\mathbb{R}}$. In our case, this function has its values in the set

$$\{2^{-n} :\ n \in \mathbb{N}\} \cup \{0\}.$$

It follows that f_L is well defined and continuous over $\widehat{A^+}$, and its ω-set is precisely L.

9.2 Languages defined by B- and S-automata

If $\mathcal{A}$ is an S-automaton, then we define $L(\mathcal{A})$ as the ω-set of the function $f_{\mathcal{A}}$, i.e.

$$L(\mathcal{A}) = \{x \in \widehat{A^+} :\ f_{\mathcal{A}}(x) = \omega\}.$$

In particular, $L(\mathcal{A})$ is a closed language of profinite words. Dually, if $\mathcal{A}$ is a B-automaton, then we define $L(\mathcal{A})$ as the complement of the ω-set of the function $f_{\mathcal{A}}$, i.e.

$$L(\mathcal{A}) = \{x \in \widehat{A^+} :\ f_{\mathcal{A}}(x) < \omega\},$$

and so $L(\mathcal{A})$ is an open language of profinite words.

Example 9.2. Let $\mathcal{A}$ be the deterministic S-automaton which computes in its only counter the length of the input word. Then $L(\mathcal{A})$ is the set of all infinite profinite words. Formally, this is not true, since $\mathcal{A}$ needs to reset its counter at the end of the word. To do that, $\mathcal{A}$ needs to use nondeterminism to guess the last position of the word, and reset its counter there.

Proposition 9.3. *Both the class of languages accepted by B-automata and the class of languages accepted by S-automata contains all clopen subsets of $\widehat{A^+}$. Both classes are closed under finite unions and intersections.*

Proof. Let $U \subseteq \widehat{A^+}$ be any clopen set. Then, there exists a regular language $L \subseteq A^+$ such that $U = \overline{L}$, and there exist finite nondeterministic automata $\mathcal{A}$ and $\mathcal{B}$, recognizing the language $L \subseteq A^+$ and its complement $A^+ - L$, respectively.

If we view $\mathcal{A}$ as a B-automaton with no counters, then the function $f_{\mathcal{A}}$ over A^+ is the function which assigns 0 to elements of L and ω to elements of $A^+ - L$. Since both $\overline{L}$ and $\widehat{A^+} - \overline{L}$ are clopen sets, it follows that the unique continuous extension of $f_{\mathcal{A}}$ to $\widehat{A^+}$ assigns 0 to $\overline{L}$ and ω to $\widehat{A^+} - \overline{L}$. Therefore, $L(\mathcal{A}) = \overline{L}$.

Completely dually, if we view $\mathcal{B}$ as an S-automaton with no counters, then $f_{\mathcal{B}}$ maps elements of L to ω and elements of $A^+ - L$ to 0. It follows that $L(\mathcal{B}) = \overline{L}$.

It remains to show that the classes of languages accepted by B- or S-automata are closed under binary unions and intersections. This follows from a Cartesian product construction for automata – one can construct a product of two B-automata, or a product of two S-automata in the expected way. By defining the accepting states appropriately, we can assure that the resulting automaton either recognizes the union or the intersection of the languages accepted by the original automata. The details are standard. □

Proposition 9.4. *Let $\mathcal{A}$ be a B-automaton and let L be the regular language of finite words which are accepted by the finite automaton which underlies $\mathcal{A}$. Then $\mathcal{A}$ is limited if and only if*

$$\overline{L} = L(\mathcal{A}).$$

In particular, if the underlying automaton accepts all finite words, then limitedness of $\mathcal{A}$ is equivalent to the universality of $L(\mathcal{A})$.

Proof. Note that the inclusion

$$\overline{L} \supseteq L(\mathcal{A})$$

holds for any B-automaton $\mathcal{A}$, since a profinite word x belongs to $L(\mathcal{A})$ iff it is a limit of a sequence of finite words $w_1, w_2, \ldots$ such that $f_{\mathcal{A}}(w_n) = k$ for every n and for some fixed $k \in \mathbb{N}$. In particular, for each $w_1, w_2, \ldots \in L$ and hence $x \in \overline{L}$.

We now prove the left-to-right implication of the equivalence stated in the proposition. Assume that $\mathcal{A}$ is limited. Then there exists a bound $N \in \overline{\mathbb{N}}$ such that for every word $w \in L$, $f_{\mathcal{A}}(w) \leq N$. In particular, if x is a limit point of words from L, then $f_{\mathcal{A}}(x) \leq N < \omega$ by continuity of $f_{\mathcal{A}}$ over $\widehat{A^+}$, so $x \in L(\mathcal{A})$. This proves that $\overline{L} \subseteq L(\mathcal{A})$.

For the other implication, assume that $\overline{L} \subseteq L(\mathcal{A})$. Let $N \in \overline{\mathbb{N}}$ be the supremum of $f_{\mathcal{A}}$ over the set $\overline{L}$. Since $f_{\mathcal{A}}$ is continuous, it attains its supremum over the compact set $\overline{L}$, i.e. there exists a point $x \in \overline{L}$ such that $f_{\mathcal{A}}(x) = N$. Since $x \in \overline{L} \subseteq L(\mathcal{A})$, it follows that $f_{\mathcal{A}}(x) = N < \omega$. Therefore, $f_{\mathcal{A}}$ is bounded by $N < \omega$ over L. □

9.2.1 Connection with cost functions

As was mentioned in the introduction to this thesis, the limitedness problem has a simple description in the framework of cost functions, which is quite similar to the one described in Proposition 9.4. This similarity can be explained by the following connection between the two frameworks.

We denote by $\preccurlyeq$ the domination relation over A^+ functions, i.e. $f \preccurlyeq g$ if for every subset K of A^+, if g is bounded over K, then f is bounded over K. We write $f \approx g$ iff $f \preccurlyeq g$ and $g \preccurlyeq f$. For a function $f\colon A^+ \to \overline{\mathbb{N}}$, we denote by $[f]$ the *cost function* induced by f, i.e. its $\approx$-equivalence class. We say that a cost function is *continuous*, if it has *some* representative which is a uniformly continuous function from A^+ to $\overline{\mathbb{N}}$. Note that a cost function might have two representatives, of which only one is uniformly continuous (consider for instance the function constantly equal to 0 and the function equal to 1 over an irregular language and 0 elsewhere).

For a closed set L, let f_L denote the uniformly continuous function considered in Remark 9.3, defined by:

$$f_L(w) = \frac{1}{d(x, L)}.$$

Proposition 9.5. *The two mappings depicted in the diagram below are mutually inverse isomorphisms of lattices.*

$$\begin{tikzcd}
\boxed{\begin{array}{c}\textit{Closed subsets of } \widehat{A^+}, \\ \textit{ordered by } \subseteq\end{array}} \arrow[r, bend left, "L \mapsto {[f_L]}"] & \boxed{\begin{array}{c}\textit{Continuous cost functions,} \\ \textit{ordered by } \preccurlyeq\end{array}} \arrow[l, bend left, "L_f \mapsfrom {[f]}"]
\end{tikzcd}$$

Consequently, via the above mappings, the language $L(\mathcal{A})$ defined by an S-automaton $\mathcal{A}$ corresponds to the cost function $[f_{\mathcal{A}}]$, and the complement of the language $L(\mathcal{B})$ defined by a B-automaton $\mathcal{B}$ corresponds to the cost function $[f_{\mathcal{B}}]$.

Lemma 9.6. *Let $f, g \colon A^+ \to \overline{\mathbb{N}}$ be uniformly continuous functions and let L_f and L_g be the ω-sets of the continuous extensions to $\widehat{A^+}$ of f and g, respectively. Then,*

$$L_f \subseteq L_g \qquad \iff \qquad f \preccurlyeq g,$$

where $\preccurlyeq$ is the domination relation.

Proof. Assume that $L_f \subseteq L_g$. Suppose that f is unbounded over a set $K \subseteq A^+$. Then, by compactness of $\widehat{A^+}$, there is a sequence $w_1, w_2, \ldots$ of elements of K which is convergent to some profinite word x and such that $f(w_1), f(w_2), \ldots$ converges to ω. In particular, $f(x) = \omega$, so $x \in L_f \subseteq L_g$, hence $g(x) = \omega$. By continuity of g, this implies that g is unbounded over the sequence $w_1, w_2, \ldots$, and therefore over K as well. Since K was arbitrary, this proves that $f \preccurlyeq g$.

For the other implication, we proceed similarly. Assume that $f \preccurlyeq g$ and $x \in L_f$. Let $\varepsilon > 0$ and let $w_1, w_2, \ldots$ be any sequence of finite words such that $d(w_n, x) < \varepsilon/n$ for all n. Then, by continuity of f, the sequence of values $(f(w_n))_{n=1}^{\infty}$ converges to $f(x) = \omega$. In particular, f is unbounded over $K = \{w_n : n \in \mathbb{N}\}$. Since $f \preccurlyeq g$, g is also unbounded over the set K, which in turn is contained in the ε-neighborhood of x. Since $\varepsilon > 0$ is arbitrary, and g is continuous in x, this shows that $g(x) = \omega$. Hence $x \in L_g$. □

Proof of Proposition 9.5. From the lemma it follows that if f and g are $\approx$-equivalent cost functions, then $L_f = L_g$. Therefore, for a given cost function $[f]$, the set L_f is independent of the choice of the uniformly continuous representative f of $[f]$. It follows from Remark 9.3 that, if we start with a closed set L, then the ω-set of the function f_L is again the set L. Therefore, the mapping $L \mapsto [f_L]$ is the right-inverse of the mapping $[f] \mapsto L_f$. In particular, $[f] \mapsto L_f$ is surjective. From Lemma 9.6 it follows that this mapping is an order-preserving isomorphism onto its image. It follows that it is an isomorphism, and that the mapping $L \mapsto [f_L]$ is its inverse. □

9.3 B- and S-regular expressions

As in the case of languages of finite words, a convenient way of describing a language of profinite words is by using regular expressions. We define two dual types of regular expressions, called *B-regular expressions* and *S-regular expressions*. These names come from [BC06], where very similar expressions, defining sets of sequences of finite words, were considered. Here, B-regular expressions will define open sets and S-regular expressions will define closed sets in $\widehat{A^*}$. As we will see in Chapter 13, B-regular expressions correspond precisely to B-automata, while S-regular expressions correspond precisely to S-automata. However, until we prove this equivalence, by a *B-regular language* (resp. S-*regular language*) we mean mean a language defined by a B-regular (resp. S-regular) expression.

Recall that the basic syntax of regular expressions over an alphabet A allows the following operations: union, concatenation, Kleene star, the empty set $\emptyset$ and singleton sets corresponding to the letters in A and the empty word ε. If E is a regular expression, then it defines a regular language in A^*, but we are more interested in its closure, i.e. the corresponding clopen set, which we denote by $L(E) \subseteq \widehat{A^*}$. We can interpret the basic operations $\cup, \cdot, *$ directly as operations over clopen languages of profinite words. As one can expect, $K \cup L$ corresponds to the union of K and L, and $K \cdot L$ corresponds to the set of products of the form $x \cdot y$, where $x \in K$ and $y \in L$. However, note that for a clopen set L, the language L^* usually does *not* correspond to the union $\bigcup_{n=1}^{\infty} L^n$, but to its closure (recall that $L(E)$ should again be a clopen set). Without the closure, the resulting operation would correspond precisely to *bounded iteration*, which we are about to define.

Informally, the *bounded iteration* of a language L, denoted $L^{<\infty}$ is the set of profinite words which are products of a finite number of words from L. Dually, the *infinite iteration* of L, denoted L^{∞}, corresponds to the set of profinite words which can be obtained as a product of an infinite number of words from L. We proceed with a formal definition.

We use the following notation, where n is any positive number.

$$\begin{aligned} E^n &\overset{def}{=} \overbrace{E \cdot E \cdots E}^{n \text{ times}} \\ E^{\geq n} &\overset{def}{=} E^n \cdot E^* \\ E^{<n} &\overset{def}{=} \{\varepsilon\} \cup E^1 \cup E^2 \cup \ldots \cup E^{n-1}. \end{aligned}$$

S-regular expressions We extend the syntax of classical regular languages by adding a unary operation $L \mapsto L^{\infty}$, called *infinite iteration*. The extended expressions are called *S-regular expressions*. The semantics of the extended regular expressions is defined as follows. Let E be an S-regular expression. For each number $n \in \mathbb{N}$, define $E_{\infty:=n}$ to be the regular expression obtained from E by replacing the exponent ∞ of infinite iteration with the exponent $\geq n$. Then, the expression E evaluates to

$$L(E) = \bigcap_{n \in \mathbb{N}} L(E_{\infty:=n}),$$

where $L(E_{\infty:=n})$ is the clopen set defined by the regular expression $E_{\infty:=n}$. Notice that each of the languages $L(E_{\infty:=n})$ is a closed set, and therefore $L(E)$ is closed as an intersection of closed sets.

Example 9.3. Let $A = \{a, b\}$ and consider the S-regular expression E

$$(a^{\infty}\, b)^{\infty}\, a^{\infty}.$$

Then $L(E)$ is the set of all profinite words such that every block of a'a has infinite length, and there are infinitely many such blocks. It is equivalent to the cost function which computes the length of the minimum of two numbers for a given finite word w: the shortest block of a's in w and the total number of b's in w.

B-regular expressions Now we define a dual extension of classical regular expressions, where except for the usual constructs of regular expressions, we may use the operation $L \mapsto L^{<\infty}$, called *bounded iteration*. We call these extended expressions *B-regular expressions*. For such an expression E, let $E_{\infty:=n}$ denote the regular expression obtained from E by substituting each occurrence of the exponent $<\infty$ with the exponent $<n$. Then, the expression E evaluates to

$$L(E) = \bigcup_{n \in \mathbb{N}} L(E_{\infty:=n}).$$

Since for each $n \in \mathbb{N}$, $E_{\infty:=n}$ is a regular expression, each of the sets $L(E_{\infty:=n})$ is open. Therefore, $L(E)$ is an open set, as a union of open sets.

Example 9.4. Consider the S-regular language $L(E)$ from the previous example. Its complement is defined by the B-regular expression

$$(a^{*}\, b)^{*}\, a^{<\infty}\, (b\, a^{*})^{*} \quad \cup \quad (a^{*}\, b)^{<\infty}\, a^{*}.$$

Indeed, a profinite word is in the complement of L if and only if either it has some finite block of a's or it has finitely many b's.

Complementation The main result of the paper [BC06] is that languages defined by B-regular expressions and S-regular expressions are complements of each other. However, the expressions considered there defined languages of infinite words, rather than profinite words. We will discover an analogous result in the profinite setting, by proving the following theorem (which follows from our main theorem, Theorem 8.2).

Theorem. *Languages defined by S-regular expressions are precisely the complements of the languages defined by B-regular expressions.*

9.4 MSO+inf logic

Following the tradition initiated by Büchi, we present a logic over profinite words which captures languages recognized by B- or S-automata. The logic is an extension of the MSO logic over profinite words, which we define here. We will then distinguish two syntactic fragments which capture precisely the classes of B- and S-regular languages. These two fragments correspond to two extensions (called *cost* MSO) introduced by Colcombet [Col10b]; in these extensions, a formula defines a cost function.

Monadic Second Order Logic (MSO) over profinite words is analogous to MSO over finite or ω-words (for a reference, see [Tho97, PP04]). In fact, the syntax of the logic is exactly the same as for finite or ω-words – we allow first and second order quantification, Boolean operations, set inclusion tests, linear order tests and label tests. Remarkably, a formula φ of this logic describes precisely the clopen language of profinite words which corresponds to the regular language of finite words described by the very same formula φ.

The semantic of MSO over profinite words, however, is defined differently than in the case of finite words. For finite words, we treat a word as an algebraic structure whose underlying set is its set of positions, i.e. a set of natural numbers. For a profinite word, though, it is impossible to define reasonably its set of positions. Because of this, we need to give a different definition of the semantic of MSO over profinite words.

Our definition of the semantic interprets constructs of the logic as operations over languages, such as union, projection, etc. and the predicates as languages of profinite words (perhaps over an extended alphabet, for encoding the valuation of the free variables).

Formula	*Expression*	*Description*
$\forall x.a(x)$	a^*	*"only a's"*
$\exists x.\exists y.a(x) \wedge b(y) \wedge x < y$	$(a+b)^*a(a+b)^*b(a+b)^*$	*"some a before some b"*
$\exists X.a(X) \quad \wedge \quad \mathrm{inf}(X)$	$b^*(ab^*)^\infty$	*"infinitely many a's"*
$\forall X.b(X) \implies \mathrm{fin}(X)$	$a^*(ba^*)^{<\infty}$	*"finitely many b's"*

FIGURE 9.1: *Formulas of* MSO+inf *over* $\{a,b\}$ *and equivalent B/S-regular expressions*

To reach beyond the clopen sets, we furthermore extend the MSO logic by a second-order unary predicate $\mathrm{inf}(X)$ which – informally – tests whether X is infinite. A bit more precisely, the formula $\mathrm{inf}(X)$ with one free second-order variable X corresponds to the language of profinite words over the alphabet $A \times \{0,1\}$ (where the second coordinate corresponds to the "set of positions" X) which contain infinitely many letters of the form $(a,1)$. Using the predicate $\mathrm{inf}(X)$, it is straightforward to construct a formula defining the language $b^*(ab^*)^\infty$ of profinite words with infinitely many a's (for this formula, and other examples, see Figure 9.1). Since this language is closed, but not clopen, it follows that the predicate inf is not definable in terms of the remaining ones. A dual predicate, $\mathrm{fin}(X)$ is defined as the negation of inf. We will discover that S-regular languages correspond to the fragment MSO+inf$^+$ of the logic, in which inf is allowed only to appear positively; dually, B-regular languages correspond to the fragment MSO+fin$^+$. However, the full logic MSO+inf is far larger than the union of these two fragments.

Syntax To simplify the definitions, in the the formal syntax we will only allow second-order constructs. Using them, it is straightforward to further interpret the symbols which correspond to first-order logic.

Let A be a fixed finite alphabet. A *formula of* MSO over profinite words is a formula obtained from the following constructs:

- the quantifiers $\exists, \forall$ which bind second-order variables, denoted $X, Y, \ldots$,
- a unary predicate $a(X)$ for each $a \in A$,
- a binary predicate $X \subseteq Y$,
- a binary predicate $X < Y$,
- Boolean connectives $\wedge, \vee, \neg$.

A *formula of* MSO+inf additionally allows:

- a unary predicate $\inf(X)$.

Using the above constructs of MSO, we can further define additional useful predicates, in an obvious way:

$$\begin{aligned} \mathit{empty}(X) &\equiv \forall Y.(Y \subseteq X \implies X \subseteq Y), \\ \mathit{singleton}(X) &\equiv \neg \mathit{empty}(X) \wedge \forall Y.\Big(Y \subseteq X \implies \big(X \subseteq Y \vee \mathit{empty}(Y)\big)\Big). \end{aligned}$$

We may then consider first-order variables, denoted $x, y, \ldots$, which are implicitly assumed to be guarded by the formula $\mathit{singleton}(x)$.

Semantic Let $\mathscr{X}$ be a finite set of second-order variables, denoted $X, Y, \ldots$. The *$\mathscr{X}$-valuation alphabet* over A is the alphabet $A \times \{0,1\}^{\mathscr{X}}$. We will call the first component of this alphabet the *A-component*, and a component corresponding to $X \in \mathscr{X}$ will be called the *X-component*. We denote letters in the alphabet $A \times \{0,1\}^{\mathscr{X}}$ by λ, λ'. The A-component of the letter λ is denoted λ_A, and its X-component is denoted λ_X. Given a profinite word x over the alphabet A, an *$\mathscr{X}$-valuation* over x is a profinite word ν over the $\mathscr{X}$-valuation alphabet, such that its projection onto the A-component is x.

We interpret an $\mathscr{X}$-valuation ν over x as an assignment of a "marking" of x to each variable $X \in \mathscr{X}$. Although the set of positions of this marking cannot be formally defined, we can talk about certain properties of such markings. For instance, "the marking X is contained in the marking Y" if for every letter λ in ν, if λ has a 1 on the X-component, then it has a 1 on the Y-component. In this fashion, we will define the semantic for all the logical symbols.

For defining the semantic of the quantifiers, we consider an *erasing* mapping, for each variable $X \in \mathscr{X}$. This mapping is defined via a mapping

$$\mathit{erase}_X \quad : \quad A \times \{0,1\}^{\mathscr{X}} \quad \longrightarrow \quad A \times \{0,1\}^{\mathscr{X} - \{X\}},$$

which maps a letter λ of the $\mathscr{X}$-valuation alphabet to the letter obtained by omitting its X-coordinate with 0. The resulting letter is a letter of a valuation alphabet over the set of variables $\mathscr{X} - \{X\}$. The mapping $erase_X$ naturally extends to a homomorphism from profinite words to profinite words, which maps an $\mathscr{X}$-valuation ν over x to a $(\mathscr{X} - \{X\})$-valuation $erase_X(\nu)$ over x.

We define, for each formula φ of MSO+inf with free variables $\mathscr{X}$, and each valuation ν over x, the expression "x satisfies $\varphi[\nu]$", denoted $x \models \varphi[\nu]$, according to the table in Figure 9.2.

$(\vee)$	$x \models (\varphi \vee \psi)[\nu]$	iff	$x \models \varphi[\nu]$ or $x \models \psi[\nu]$
$(\wedge)$	$x \models (\varphi \wedge \psi)[\nu]$	iff	$x \models \varphi[\nu]$ and $x \models \psi[\nu]$
$(\neg)$	$x \models (\neg\varphi)[\nu]$	iff	not $x \models \varphi[\nu]$
(a)	$x \models (a(X))[\nu]$	iff	for every letter λ appearing in ν, if $\lambda_X = 1$ then $\lambda_A = a$
$(\subseteq)$	$x \models (X \subseteq Y)[\nu]$	iff	for every letter λ appearing in ν, if $\lambda_X = 1$ then $\lambda_Y = 1$
$(<)$	$x \models (X < Y)[\nu]$	iff	for every pair of letters λ, λ' such that ν factorizes as $\nu_1\lambda\nu_2\lambda'\nu_3$ with $\nu_1, \nu_2, \nu_3 \in \widehat{A^*}$, if $\lambda_Y = 1$ then $\lambda'_X = 0$
(inf)	$x \models (\mathrm{inf}(X))[\nu]$	iff	ν contains infinitely many letters λ such that $\lambda_X = 1$
$(\exists)$	$x \models (\exists X.\varphi)[\nu]$	iff	for some $(\mathscr{X} \cup \{X\})$-valuation ν' such that $erase_X(\nu') = \nu$, $x \models \varphi[\nu']$
$(\forall)$	$x \models (\forall X.\varphi)[\nu]$	iff	for every $(\mathscr{X} \cup \{X\})$-valuation ν' such that $erase_X(\nu') = \nu$, $x \models \varphi[\nu']$

FIGURE 9.2: *The semantic of the logic* MSO+inf

For a formula φ with no free variables, the appropriate valuation alphabet – the $\emptyset$-valuation alphabet – is simply A, and there is precisely one valuation over a given profinite word x – namely x itself. We then define the *language* of the formula φ as the set

$$L(\varphi) = \{x : \quad x \models \varphi[x]\}.$$

We say that a language $L \subseteq \widehat{A^+}$ is *definable in* MSO+inf , if $L = L(\varphi)$ for some formula φ.

Remark 9.4. Without the unary predicate $\mathrm{inf}(X)$, the logic would capture precisely clopen sets. Indeed, all the remaining predicates are clopen, a projection of a clopen set is again a clopen set, and also clopen sets are closed under Boolean combinations.

Example 9.5. MSO+inf over profinite words can define sets which are not closed nor open. For instance, the conjunction of the two last formulas in Figure 9.1 defines the language L of all profinite words which contain infinitely many a's and only finitely many b's, which is not closed neither open. By a further projection of this language (corresponding to applying $\exists Y.$ and further relativizing to Y), we obtain a language $\exists L$ which is not even a Boolean combination of open sets.

It is not difficult to see that the following Proposition holds. We leave it without a proof, as it is not used in this thesis.

Proposition 9.7. *The class of languages definable in* MSO+inf *over all finite alphabets is the smallest class of languages which is closed under Boolean combinations, projections, inverse images under letter-to-letter homomorphisms, and contains all clopen sets and the language $b^*(ab^*)^\infty$.*

The restricted fragments We distinguish two fragments of the logic MSO+inf. One fragment, denoted MSO+inf$^+$, is the syntactic fragment where the predicate inf appears only *positively*, i.e. under a positive number of negations. We define the predicate $\mathrm{fin}(X)$ as the negation of the predicate $\mathrm{inf}(X)$, i.e. $\mathrm{fin}(X) \equiv \neg\,\mathrm{inf}(X)$. The fragment dual to MSO+inf$^+$ is the fragment MSO+fin$^+$, where the predicate fin appears only positively (and does not use the predicate inf otherwise). Obviously, the negation of a formula of MSO+fin$^+$ is a formula of MSO+inf$^+$, and vice-versa. By a simple inductive reasoning, we observe that the fragment MSO+inf$^+$ defines only closed sets, and that the fragment MSO+fin$^+$ defines only open sets – indeed, each of the predicates corresponds to a closed set, and both closed sets and open sets are preserved by finite unions, intersections and under projection, and negation turns open sets to closed sets and vice-versa.

We will see later on that languages accepted by S-automata correspond precisely to the fragment MSO+inf$^+$, while languages accepted by B-automata correspond to the fragment MSO+fin$^+$. Converting an automaton to a formula is straightforward – it suffices to define the language of accepting runs of the automaton and project it onto the input alphabet, using the existential quantifier. We do this in Proposition 10.4. This conversion yields a normal form to the formulas of MSO+inf$^+$ and MSO+fin$^+$.

Converting a formula into an automaton is difficult – it requires showing that automata are closed under the constructs comprising the logic. B- and S-automata are closed under union and intersection by Proposition 9.3 and it is easy to see that they are closed under projection (as usually the case with nondeterministic automata). Moreover, B-automata capture the unary predicate fin and S-automata capture the unary predicate inf. The only missing part is closure under complementation. As was mentioned in Section 9.3, we will prove that languages accepted by B-automata are precisely the complements of languages accepted by S-automata. Using this, the conversion from the restricted fragments of logic to automata follows easily.

CHAPTER 10

Elementary constructions

In this chapter we prove some simple results about the classes of languages defined in the previous chapter. In Section 10.1, we show how to convert a regular expression into an automaton. In Section 10.2, we characterize the profinite words accepted by B- and S-automata using formulas of MSO+inf. In Section 10.3, we discuss the decision procedures for determining emptiness of a language recognized by a a given B- or S-automaton.

10.1 From regular expressions to automata

Just like in the case of the usual regular expressions, it is easy to translate a B-regular expression into an equivalent B-automaton, or an S-regular expression into an S-automaton. Basically, one needs to show that B- and S-automata are closed under all the operations which comprise the respective expressions. The usual operations are dealt with just as in the classical setting, and closure under bounded (resp. infinite) iteration is performed similarly as closure under usual iteration, but with introducing an additional counter which counts the number of performed iterations. We describe this construction in more detail.

Proposition 10.1. *An S-regular language is recognized by an S-automaton. A B-regular language is recognized by a B-automaton. Both translations are effective.*

Proof. The proof technique is completely standard. Consider S-regular languages for instance. We show closure of S-automata under all the operations comprising S-regular expressions. Closure under union, concatenation and Kleene star is done as usual. To prove closure under infinite iteration, we proceed as in the case of the Kleene star, and additionally introduce a new counter, which is incremented whenever a corresponding loop is completed; this counter is reset whenever a loop corresponding to a super-expression is completed.

This way, it is not difficult to inductively construct out of a given S-regular expression E an S-automaton $\mathcal{A}$ such that for all $w \in A^+$ and $n \in \mathbb{N}$

$$w \in L\left(E_{\infty:=n}\right) \qquad \Longleftrightarrow \qquad f_{\mathcal{A}}(w) \geq n. \tag{1}$$

From this it immediately follows that $L\left(E\right) = L\left(\mathcal{A}\right)$. In the case of B-automata, we proceed identically, with the only difference that in the above invariant, $\geq n$ should be replaced by $< n$.

As in the usual case, one can either introduce ε-transitions, or deal directly with nondeterministic automata. The formal construction is tedious, but we present it below for the sake of completeness.

As usual, we proceed by induction on the structure of the expression E. The base case is trivial. Therefore, it suffices to show that for S-automata, we can perform the operations which correspond to union, concatenation, the Kleene star, and of infinite iteration, while preserving the invariant (1). The most interesting case is the case of infinite iteration, which we consider now.

Let $\mathcal{A} = (Q, I, F, C, \delta)$ be an S-automaton which satisfies (1), where Q are its states, I are its initial states, F are its accepting states, C are its counters, and

$$\delta \subseteq Q \times A \times (\{\mathit{inc}, \mathit{reset}\}^*)^C \times Q$$

is the transition relation. Let $\mathit{reset}(C)$ denote the sequence of operations which resets each counter in C. We construct an automaton $\mathcal{A}^\infty$ corresponding to the expression $F = E^\infty$:

$$\mathcal{A}^\infty = (Q \cup \{q_0\},\ I,\ \{q_0\},\ C \cup \{c_0\},\ \delta'),$$

where δ' is from δ by extending by the following extra transitions:

- for all transitions $(p, a, op, q_f) \in \delta$ such that $q_f \in F$, and for all $q_i \in I$, we add to δ' the transition

$$\big(p, \quad a, \quad (op;\ \mathit{reset}(C);\ \mathit{inc}(c_0)), \quad q_i\big), \tag{A}$$

- for all transitions $(p, a, op, q_f) \in \delta$ such that $q_f \in F$, we add to δ' the transition

$$\big(p, \quad a, \quad (op;\ \mathit{reset}(C);\ \mathit{inc}(c_0);\ \mathit{reset}(c_0)), \quad q_0\big). \tag{B}$$

We verify that that $\mathcal{A}^\infty$ satisfies the following condition

$$w \in L\left(F_{\infty:=n}\right) \qquad \Longleftrightarrow \qquad f_{\mathcal{A}^\infty}(w) \geq n. \tag{2}$$

To prove the right-to-left implication, assume that $\mathcal{A}$ has an accepting run ρ over w of value at least n. We will show that then $w \in L\left(F_{\infty\mapsto n}\right)$. Since ρ is accepting, it must end in the state q_0, so by construction, ρ has at least one reset. By assumption, each reset value in the run ρ is at least n. In particular, c_0 is incremented at least n times during the run ρ. It follows that the run ρ can be decomposed as

$$\rho = \rho_1, \tau_1, \rho_2, \tau_2, \ldots, \rho_{k-1}, \tau_{k-1}, \rho_k, \tau_k,$$

where $k \geq n$, $\tau_1, \ldots, \tau_{k-1}$ are transitions of type (A), and τ_k is a transition of type (B), and the runs $\rho_1, \ldots, \rho_k$ only use the original transitions of δ.

We decompose the word w accordingly to the decomposition of ρ:

$$w = w_1 a_1 w_2 a_2 \dots w_{k-1} a_{k-1} w_k a_k.$$

Over each of the words $w_i a_i$, where $1 \leq i \leq k$, basing on the run $\rho_i \tau_i$, it is straightforward to construct an accepting run ρ_i' of $\mathcal{A}$, which moreover has value at least n.

By inductive assumption, it follows that for each $1 \leq i \leq k$,

$$w_i a_i \in L\left(E_{\infty \mapsto n}\right).$$

Therefore, w is a concatenation of $k \geq n$ copies of words from $L\left(E_{\infty \mapsto n}\right)$, so

$$w \in L\left(\left(E_{\infty \mapsto n}\right)^{\geq n}\right) = L\left(F_{\infty \mapsto n}\right).$$

This proves the right-to-left implication of (2).

The proof in the other direction is very similar. Assume that $w \in L\left(F_{\infty \mapsto n}\right)$. We will show that $f_{\mathcal{A}}(w) \geq n$. Since $w \in L\left(\left(E_{\infty \mapsto n}\right)^{\geq n}\right)$, we can write

$$w = w_1 w_2 \dots w_k,$$

where $k \geq n$ and $w_1, w_2, \dots, w_k \in L\left(E_{\infty \mapsto n}\right)$. By inductive assumption, for each $1 \leq i \leq k$, there exists an accepting run ρ_i of $\mathcal{A}$ over w_i, such that $val_{\mathcal{A}}(\rho_i) \geq n$. Basing on the runs $\rho_1, \dots, \rho_n$, in a straightforward way we construct a run ρ of $\mathcal{A}^\infty$, such that $val_{\mathcal{A}}(\rho) \geq n$, proving that $f_{\mathcal{A}}(w) \geq n$. This shows $\mathcal{A}^\infty$ satisfies the condition (2).

To complete the picture of the translation of S-regular expressions into S-automata, we also need to deal with union, concatenation and the Kleene star, and also the base case of a single-letter expression. However, this can be done in a totally standard way, so we omit these constructions here.

Now, let us comment on the translation of B-regular expressions into B-automata. The construction proceeds analogously to the case of S-regular expressions. In this setting, we construct by induction on the structure of a B-regular expression E a B-automaton $\mathcal{A}$ such that for all $w \in A^+$,

$$w \in L\left(E_{\infty := n}\right) \qquad \Longleftrightarrow \qquad f_{\mathcal{A}}(w) < n. \tag{3}$$

In fact, the entire construction for S-automata and its proof of correctness can be repeated word by word, with only replacing each exponent ∞ by $<\infty$ and each inequality $\geq n$ by $< n$. □

Note that the described translation produces hierarchical automata, i.e. automata in which the counters are linearly ordered into a hierarchy, and any counter operation on a counter c entails resets of all counters which are below c in the hierarchy.

Corollary 10.2. *B-regular languages are recognized by hierarchical B-automata, and S-regular languages are recognized by hierarchical S-automata.*

As we mentioned in the introduction, hierarchical B-automata correspond precisely to nested distance desert automata of Kirsten.

10.2 From automata to logic

In this section, we will describe the languages defined by B- and S-automata using formulas of MSO+fin$^+$ and MSO+inf$^+$, respectively.

Let $\mathcal{A}$ be a B- or S-automaton, and let

$$\delta \subseteq Q \times A \times (\{inc, reset\}^*)^C \times Q$$

be its transition relation. We call elements of δ *transitions*. For a transition $\tau = (p, a, op, q) \in \delta$, we call p its *source state* and q its *target state*, and a its *label*. We say that τ *resets* a counter c if the c-coordinate of op contains the operation *reset*; similarly we may refer to *increments* performed by τ. A *profinite run* of $\mathcal{A}$ is a profinite word ρ over the alphabet δ, such that each two consecutive transitions τ, τ' in ρ are composable, i.e. the target state τ is the same as the source state of τ'. We say that a profinite run ρ is *initial* if the source state of the first transition in δ is an initial state of $\mathcal{A}$. Dually, we say that ρ is *final* if the target state of the last transition in δ is a accepting state of $\mathcal{A}$. All the above properties of runs are in fact regular properties of words, so they can be expressed by formulas of MSO over profinite words.

If ρ decomposes as $\rho = \pi \cdot \rho' \cdot \sigma$, where $\pi, \rho', \sigma \in \widehat{\delta^*}$, then we say that ρ' is a *value trace* in ρ of a counter c, provided that:

- only the last transition of ρ' resets the counter c
- π is empty or its last transition resets the counter c.

Moreover, we say that a value trace ρ' in ρ of a counter c is *bounded by n* if it contains at most n transitions that increment c, and we say that it is *finite* if it is bounded by some number $n \in \mathbb{N}$; otherwise, we say that it is *infinite*.

Let $\pi \colon \widehat{\delta^*} \to \widehat{A^*}$ be the natural homomorphism induced by the mapping from δ to A which assigns to a transition its label. We say that a profinite run ρ is a run *over* the profinite word $\pi(\rho)$.

Lemma 10.3. *Let $\mathcal{A}$ be a B-automaton with transition relation δ and set of counters C. Let $x \in \widehat{A^+}$. Then, $f_{\mathcal{A}}(x) < \omega$ if and only if*

$$\text{there exists a profinite run } \rho \in \widehat{\delta^+} \text{ over } x \text{ which is initial and final, and such that every value trace in } \rho \text{ of any counter } c \text{ is finite.} \tag{4}$$

Dually, let $\mathcal{A}$ be an S-automaton with transition relation δ and set of counters C. Let $x \in \widehat{A^+}$. Then, $f_{\mathcal{A}}(x) = \omega$ if and only if

$$\text{there exists a profinite run } \rho \in \widehat{\delta^+} \text{ over } x \text{ which is initial and final, and such that every value trace in } \rho \text{ of any counter } c \text{ is infinite.} \tag{5}$$

Proof. First, consider the case when $\mathcal{A}$ is a B-automaton. Suppose that $f_{\mathcal{A}}(x) < \omega$. Then, there exists a number $k \in \mathbb{N}$ and sequence of finite words $w_1, w_2, \ldots$ which is convergent to x and such that $f_{\mathcal{A}}(w_n) \leq k$ for all $n = 1, 2, \ldots$. In particular, for each n, there is a run $\rho_n \in \delta^*$ of $\mathcal{A}$ over w_n, such that $val_{\mathcal{A}}(\rho) \leq k$. By compactness of $\widehat{\delta^*}$, we may assume without loss of generality that the sequence of runs $\rho_1, \rho_2, \ldots$ is convergent to a profinite run $\rho \in \widehat{\delta^*}$. By continuity of $\pi \colon \widehat{\delta^*} \to \widehat{A^*}$, ρ is a profinite run over the profinite word x.

Let L_k denote the set of profinite runs $\rho \in \widehat{\delta^*}$ which satisfy the following property:

ρ is an initial and final run and every value trace in ρ of a counter c is bounded by k.

Since L_k corresponds to a regular property, it is a clopen set in $\widehat{\delta^*}$. Moreover, the runs $\rho_1, \rho_2 \ldots$ are all in L_k. Therefore, also $\rho \in L_k$. We have thus shown that the condition (4) holds.

Conversely, assume that the condition (4) holds and let ρ be the profinite run which witnesses that. We will see that there exists a bound k such that $\rho \in L_k$. Indeed, otherwise, there is an infinite sequence $k_1 < k_2 < \ldots$ such that for each n, there exists a value trace ρ'_{k_n} in ρ of some counter c_{k_n} which is not bounded by k_n. By restricting to a subsequence if necessary, we may assume that $c_{k_1} = c_{k_2} = \ldots = c$ for some counter c. For each n, let $\rho = \pi_{k_n} \cdot \rho'_{k_n} \cdot \sigma_{k_n}$ be the decomposition of ρ, which comes from the definition of value traces of ρ. By compactness of $\widehat{\delta^*}$, we may assume that each of the three sequences of profinite words, $(\pi_{k_n})_n$, $(\rho'_{k_n})_n$ and $(\sigma_{k_n})_n$, is convergent. Let π, ρ' and σ denote their respective limits. We then have that $\rho = \pi \cdot \rho' \cdot \sigma$ and moreover, ρ' is a value trace in ρ of the counter c which is infinite. Therefore, ρ does not satisfy the condition (4), contrary to the assumption.

We have thus shown that $\rho \in L_k$ for some number $k \in \mathbb{N}$. Since L_k is an open set in $\widehat{\delta^*}$, it follows that any sufficiently close finite run ρ_0 also belongs to the set L_k. From this it follows that x is a limit of a sequence of finite words, each of which has an accepting run of value at most k. Therefore, $f_{\mathcal{A}}(x) \leq k$. This finishes the proof of the equivalence in the case when $\mathcal{A}$ is a B-automaton.

Now, let us consider the case when $\mathcal{A}$ is an S-automaton. The proof then is quite similar, so we only sketch it.

Assume that $f_{\mathcal{A}}(x) = \omega$. Then, there exists a sequence of finite words $w_1, w_2, \ldots$ which is convergent to x and such that $f_{\mathcal{A}}(w_n) > n$ for all $n = 1, 2, \ldots$. Let $\rho_n \in \delta^*$ be an accepting run over w_n with value larger than n. By compactness of $\widehat{\delta^*}$, we may assume without loss of generality that the sequence of runs $\rho_1, \rho_2, \ldots$ is convergent to a profinite run $\rho \in \widehat{\delta^*}$. Moreover, ρ is a profinite run over x.

For any number $k \in \mathbb{N}$, let M_k denote the set of profinite runs $\rho \in \widehat{\delta^*}$ which satisfy the regular property:

ρ is an initial and final run and every value trace in ρ of any counter c is not bounded by k.

Since for every number k, almost all runs $\rho_1, \rho_2, \ldots$ belong to the open set M_k, it follows that $\rho \in M_k$ for every $k \in \mathbb{N}$. Therefore, ρ witnesses the property (5).

Conversely, let ρ witness the property (5). Clearly, $\rho \in M_k$ for every number $k \in \mathbb{N}$. Since M_k is open, for any k there exists a finite run $\rho^k \in M_k$ over some finite word w_k, such that

$d(\rho, \rho^k) < 2^{-k}$. Then, $val_{\mathcal{A}}(\rho_k) \geq k$, so $f_{\mathcal{A}}(w_k) \geq k$. Moreover, the sequence $w_1, w_2, \ldots$ converges to x, proving that $f_{\mathcal{A}}(x) = \omega$. □

Proposition 10.4. *If $\mathcal{A}$ is a B-automaton, then $L(\mathcal{A})$ is definable in* MSO+fin$^+$. *If $\mathcal{A}$ is an S-automaton, then $L(\mathcal{A})$ is definable in* MSO+inf$^+$.

Proof. By a straightforward translation of the characterizations from the lemma into formulas of MSO+inf. Moreover, the translation yields formulas of a certain normal form. □

10.3 Emptiness of B- and S-automata

The emptiness problems for the two models differ significantly. Determining emptiness of a B-automaton $\mathcal{A}$ is straightforward. Indeed, from continuity of $f_{\mathcal{A}}$ and openness of $L(\mathcal{A})$ it follows easily that $L(\mathcal{A})$ is nonempty if and only if it contains a finite word. Therefore, it suffices to determine whether there exists a single finite word in A^+ over which $\mathcal{A}$ has an accepting run. This yields a reduction to the emptiness problem for finite automata.

Corollary 10.5. *The emptiness problem of B-automata is in* LOGSPACE, *and is complete for this class.*

On the other hand, determining emptiness of an S-automaton $\mathcal{A}$ is more difficult. Nonemptiness of $L(\mathcal{A})$ is equivalent to the existence of a *sequence* of finite words $w_1, w_2, \ldots$, such that the values $f_{\mathcal{A}}(w_1), f_{\mathcal{A}}(w_2), \ldots$ converge to ω. (This is equivalent to non-limitedness of the automaton $\mathcal{A}$.)

This resembles the situation in Part I of this thesis, where we dealt with limitedness of distance or B-automata. However, limitedness of S-automata is much easier than limitedness of distance automata. This is because of the quantifier alternations in the definitions of the respective problems. Limitedness of distance automata is expressed by

> *There exists a bound N, such that for every accepted word, there exists an accepting run of value not larger than N,*

while limitedness of S-automata is expressed by:

> *There exists a bound N, such that every accepting run has value not larger than N.*

Reassuming, the emptiness problem for B-automata is trivial, and the emptiness (or limitedness) problem for S-automata is moderate; finally, the universality (or limitedness) problem for B-automata is very difficult. (One could also consider the universality problem for S-automata – it is equivalent to universality of finite nondeterministic automata, so it fits in between “trivial” and “moderate” in the above scale.)

We now turn to an analysis of the emptiness problem of S-automata. As usually when dealing with emptiness, we might as well assume that the automaton is deterministic, by labeling each transition with a different label. Because of that, deciding emptiness of S-automata is much easier than deciding limitedness of distance or B-automata (limitedness of B-automata corresponds to their universality).

We will sketch two proofs of decidability of the emptiness problem for S-automata. The first proof we will present uses the difficult result proved in Part I. This approach might seem ridiculously overcomplicated – we're using decidability of limitedness for B-automata to deduce decidability of limitedness for S-automata, while the approach in the paper [BC06] is exactly converse. However, we present this proof, as it is quick. A direct proof will be sketched later, and another proof can be extracted from [BC06].

The idea of the quick-and-dirty proof we present now is that a deterministic S-automaton can be simulated by a nondeterministic B-automaton – indeed, the valuation of a word under a deterministic S-automaton is the minimal reset value for the unique run, and this minimum can be evaluated by a B-automaton, by using nondeterminism.

Lemma 10.6. *For any deterministic S-automaton $\mathcal{A}$ there exists a nondeterministic B-automaton $\mathcal{B}$ such that $f_{\mathcal{B}} = f_{\mathcal{A}}$.*

Sketch of proof. To simulate $\mathcal{A}$, the automaton $\mathcal{B}$ has one counter, and stores the following information in its states:

- The current state of $\mathcal{A}$
- An indication distinguishing one of the following $|C| + 2$ possibilities: (1) no counter of $\mathcal{A}$ was yet active; (2) a counter c of $\mathcal{A}$ is currently active (for each $c \in C$); (3) some counter of $\mathcal{A}$ was previously active (does not matter which one)

The automaton $\mathcal{B}$ simulates the automaton $\mathcal{A}$, and nondeterministically decides that some counter c becomes active. This can be done at any moment when $\mathcal{A}$ performs a *reset* on the counter c or at the beginning of the run, provided that no counter was active previously. Next, $\mathcal{B}$ simulates in its only counter all the counter operations performed by $\mathcal{A}$ on the active counter, until it is reset. Then, $\mathcal{B}$ also resets its counter, and states that "some counter of $\mathcal{A}$ was previously active" (and no other counter will become active in the same run). This way, using nondeterminism, $\mathcal{B}$ is capable of choosing the minimal reset value of any counter. □

For an S-automaton $\mathcal{A}$, to determine if $L(\mathcal{A})$ is empty, we need to decide whether there is a profinite word x such that $f_{\mathcal{A}}(x) = \omega$. This is equivalent to testing whether the B-automaton $\mathcal{B}$ is not limited.

Corollary 10.7. *Emptiness of S-automata is decidable.*

Direct proof

We now sketch a more straightforward proof for deciding emptiness of S-automata. The idea is to mimic Kleene's construction of a regular expression from a finite automaton, by reducing inductively the automaton, in each step removing a state, thus decomposing into smaller automata. However, we need to compute a piece of information about the smaller automata, which says more than just whether they are empty or not. Roughly, this information should tell us whether there exists an accepting run with a certain behavior specified for each counter separately. This behavior is captured by a semigroup $\overline{\mathbb{S}}$, dual to the semigroup $\overline{\mathbb{B}}$ considered in Chapter 7 of Part I of this thesis.

The semigroup of counter operations $\overline{\mathbb{S}}$ We define the semigroup $\overline{\mathbb{S}}$ as the set $\overline{\mathbb{N}} \cup \overline{\mathbb{N}}^2 \cup \overline{\mathbb{N}}^3$ with the semigroup structure described by the table in Figure 10.1. For the purpose of simplifying the definition of multiplication in $\overline{\mathbb{S}}$, we write elements of the form $(k, l) \in \overline{\mathbb{S}}$ as triples of the form $(k, \perp, l)$, where $\perp > \omega$. Therefore, $\overline{\mathbb{S}} \cong \overline{\mathbb{N}} \cup (\overline{\mathbb{N}} \times \mathcal{T} \times \overline{\mathbb{N}})$, where $\mathcal{T} = \overline{\mathbb{N}} \cup \{\perp\}$.

The interpretation is that $(n) \in \mathbb{S}$ represents the sequence inc^n, $(k, l) = (k, \perp, l)$ represents the sequence $inc^k reset\ inc^l$, and (k, n, l) represents a sequence of the form

$$inc^k reset\ (inc + reset)^* reset\ inc^l,$$

in which the number of increments in the shortest infix of the form $reset\ inc^*\ reset$ is equal to n. Note that, unlike in $\overline{\mathbb{B}}$, $(\omega, \omega, 10)$ or $(10, 5, \omega)$ represent distinct limits of operations.

$$\begin{aligned}
(k) \cdot (l) &= (k + l) \\
(k) \cdot (l, n, m) &= (k + l, n, m) \\
(l, n, m) \cdot (k) &= (l, n, n + k) \\
(l, n, m) \cdot (l', n', m') &= (l, \min(n, n', m + l'), m')
\end{aligned}$$

FIGURE 10.1: *The multiplication table in* $\mathbb{S}$. *We assume* $k, l, l', m, m' \in \overline{\mathbb{N}}$ *and* $n, n' \in \overline{\mathbb{N}} \cup \{\perp\}$.

The semigroup $\overline{\mathbb{S}}$ carries a natural compact metric topology, as it is defined as a union of three disjoint metric compact spaces. This topology is totally disconnected, so $\overline{\mathbb{S}}$ has a structure of a profinite semigroup. The ω-power in $\overline{\mathbb{S}}$ is described by the table in Figure 10.2.

$$\begin{aligned}
(0)^\omega &= (0) \\
(m)^\omega &= \ \omega \qquad \text{for } m \geq 1 \\
(n_1, n_2, n_3)^\omega &= (n_1, n_2, n_3)^2 = (n_1, \min(n_2, n_1 + n_3), n_3)
\end{aligned}$$

FIGURE 10.2: *The ω-power in* $\overline{\mathbb{S}}$.

The semigroup of transitions Let C be a finite set of counters and Q be a finite set of states. Then we consider the set $\overline{\mathbb{S}}^C$ endowed with the coordinatewise semigroup structure, and

$$Q \times \overline{\mathbb{S}}^C \times Q$$

endowed with a partial associative mapping $\cdot$, defined by

$$(p, \tau, q) \cdot (q, \tau', r) = (p, \tau \cdot \tau', r).$$

This way, $Q \times \overline{\mathbb{S}}^C \times Q$ has a structure of a partial profinite semigroup. The ω-power is defined for elements of the form (q, τ, q), and yields (q, τ', q), where $\tau'[c] = (\tau[c])^\omega$ for every counter $c \in C$.

The approximative versions We consider the *approximative semigroup*

$$\overline{\mathbb{S}}_{/1=2} = \{0,1,\omega\} \cup \{0,1,\omega\}^2 \cup \{0,1,\omega\}^3,$$

defined analogously as $\overline{\mathbb{S}}$ via the table in Figure 10.1, but in which addition is performed only up to the threshold 1, i.e. $1+1=1$. Let $\alpha_{(1=2)} \colon \overline{\mathbb{S}} \to \overline{\mathbb{S}}_{/1=2}$ be the natural homomorphism, replacing finite, positive values by 1. By # we denote the operation in $\overline{\mathbb{S}}_{/1=2}$ defined by the table in Figure 10.2, but in which addition is only up to threshold 1. Then, it is clear that

$$\alpha_{(1=2)}(s^\omega) = \alpha_{(1=2)}(s)^\#.$$

We will be interested in the homomorphism

$$\alpha_{(1=2)} \quad : \quad Q \times \overline{\mathbb{S}}^C \times Q \quad \longrightarrow \quad Q \times \overline{\mathbb{S}}_1^C \times Q,$$

which is defined naturally via a Cartesian product construction, and which is a homomorphism of semigroups, which maps the ω-power to the operation #.

For considering the emptiness problem, those are all the data structures we need to define. Note that we do not need to consider sets of transitions, as it was in the case of the limitedness problem for B-automata, nor min-semirings, as it was the case in distance automata, since our automata are assumed to be deterministic, so there is always precisely *one* run for a given finite word. This is another way of explaining why limitedness of S-automata is simpler than limitedness of distance automata.

Let $\mathcal{A}$ be an S-automaton with counters C and transition relation

$$\delta \subseteq Q \times A \times (\{inc, reset\}^*)^C \times Q.$$

We define a set $\Delta \subseteq Q \times \overline{\mathbb{S}}^C \times Q$ induced by δ as follows.

$$\Delta \overset{def}{=} \{(p,\tau,q) : \quad (p,a,op,q) \in \delta \textit{ and } \tau \in \overline{\mathbb{S}}^C \textit{ represents the sequence of counter operations in } op\}.$$

In the end, we are interested in determining whether the set $\overline{\Delta^+}$ contains an element (p, τ, q), such that

- p is an initial state,
- q is an accepting state,
- for every $c \in C$, $\tau[c]$ is of the form (ω, ω, n) or (ω, n) or (n), where $n \in \overline{\mathbb{N}}$.

Lemma 10.8. *Let $\Delta \subseteq Q \times \overline{\mathbb{S}}^C \times Q$ be a finite set. Then,*

$$\alpha_{(1=2)}(\overline{\Delta^+}) = \alpha_{(1=2)}(\Delta)^{\langle\,\cdot\,,\#\rangle}, \tag{6}$$

where $X^{\langle\,\cdot\,,\#\rangle}$ denotes closure under multiplication and stabilization in $Q \times \overline{\mathbb{S}}_1^C \times Q$.

The above lemma shares a strong resemblance with the characterization by H. Leung for distance automata, stated in Theorem 5.2, Part I of this thesis. However, as we mentioned, its proof is much simpler.

Sketch of proof. The proof proceeds by induction on the number of states, $|Q|$, imitating Kleene's construction of a regular expression from an automaton.

The harder part is the left-to-right inclusion. In the inductive step, we choose any state $q_0 \in Q$. Then, we consider a profinite accepting run ρ, and split it into factors according to the appearance of the state q_0. We moreover group the factors into chunks of bounded size so that each chunk has the same "type" with respect to $\alpha_{(1=2)}$ (the bound on the size depends only on Q and C). We then use the inductive assumption for the automaton with the state q_0 removed. □

Corollary 10.9. *If $L(\mathcal{A})$ is nonempty, then it contains a profinite word from the set $A^{\langle\,\cdot\,,\omega\rangle}$.*

This corollary also follows immediately, using Lemma 10.6, from a difficult fact that the complement of a language accepted by a B-automaton contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$, which will be proved in Proposition 11.13.

Proposition 10.10. *Let $\mathcal{A}$ be an S-automaton and let $L = L(\mathcal{A})$. Then*

$$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}. \tag{7}$$

Proof. The class of languages accepted by S-automata satisfies the assumptions of Lemma 8.3. □

CHAPTER 11

Syntactic algebra

In this chapter, we develop the algebraic counterpart of the theory, by linking B/S-regular languages with the ω-power in the profinite semigroup. In Section 11.1 we define in a natural way a syntactic congruence induced by a language of profinite words with respect to the operations of multiplication and the ω-power, and in Section 11.2 we analyze the quotient algebraic structure and extract the notion of an abstract stabilization semigroup. Those correspond precisely to the stabilization semigroups discovered by Colcombet [Col09].

11.1 Syntactic congruence

We will define a congruence induced by a language $L \subseteq \widehat{A^+}$, which respects multiplication and the ω-power. All the definitions and properties established in this section can be easily generalized to abstract topological algebras over arbitrary signatures, but we will restrain ourselves from such generalizations.

Let $\mathit{Terms}(\widehat{A^+}, \,\cdot\,, \#)$ denote the set of all the terms with one free variable which may appear only once in the term, and where the terms use the binary symbol $\cdot$ of multiplication, the unary symbol # (interpreted as the ω-power in $\widehat{A^+}$), and arbitrary elements of $\widehat{A^+}$ as a constants (i.e. in a leaf of the term). Note that any such term τ defines a mapping which maps a profinite word x to the profinite word $\tau(x)$. Moreover, this function is continuous, as it is a composition of the continuous functions $\cdot$ and ω.

Let $L \subseteq \widehat{A^+}$ be any set. For $x, y \in \widehat{A^+}$, we write $x \preceq_L y$ if for every term $\tau \in \mathit{Terms}(\widehat{A^+}, \,\cdot\,, \#)$

$$\tau(y) \in L \implies \tau(x) \in L.$$

It is clear that $\preceq_L$ is a partial preorder. We define $\simeq_L$ to be the equivalence relation induced by $\preceq_L$, i.e. $x \simeq_L y$ iff $x \preceq_L y$ and $y \preceq_L x$. We call $\simeq_L$ the *$\langle \,\cdot\,, \#\rangle$-syntactic congruence* of L. Note that $\simeq_L$ saturates the set L, meaning that L is a union of $\simeq_L$-equivalence classes. We say that L has *finite $\langle \,\cdot\,, \#\rangle$-index* if $\simeq_L$ has finitely many equivalence classes. Let $S_L = \widehat{A^+}/_{\simeq_L}$ denote the set of equivalence classes of the congruence $\simeq_L$ and let α_L denote the canonical projection from $\widehat{A^+}$ to S_L.

Example 11.1. Let $A = \{a, b\}$ and let L denote the closed subset of $\widehat{A^+}$ which corresponds to the function computing the number of occurrences of the letter a. Therefore, L is the set of those profinite words over $\{a, b\}$ which contain infinitely many a's.

The equivalence $\simeq_L$ can be easily seen to have three equivalence classes:

– L_0, the set of profinite words containing no letter a,

– L_1, the set of profinite words containing a finite, nonzero number of a's,

– $L_\omega = L$, the set of profinite words containing infinitely many a's.

11.1.1 Algebraic structure

Lemma 11.1. *The operations in $\langle\ \cdot\ , \#\rangle$ preserve the relation $\preceq_L$. More precisely, if $x' \preceq_L x$ and $y' \preceq_L y$ then*

$$x' \cdot y' \preceq_L x \cdot y, \tag{1}$$
$$(x')^\omega \preceq_L x^\omega. \tag{2}$$

Proof. The proof is standard universal algebra. We present a proof for the multiplication operation, and for the operation $\# = \omega$ the proof is completely analogous.

First we show that if $x' \preceq_L x$ then $x' \cdot y \preceq_L x \cdot y$ for any $y \in \widehat{A^+}$. Indeed, assume that

$$\tau \in \mathit{Terms}(\widehat{A^+},\ \cdot\ , \#) \qquad \textit{and} \qquad \tau(x \cdot y) \in L.$$

For $z \in \widehat{A^+}$, let

$$\sigma(z) = \tau(z \cdot y).$$

Then we can treat $\sigma(z)$ as a term with free variable z, i.e. $\sigma \in \mathit{Terms}(\widehat{A^+},\ \cdot\ , \#)$. Since

$$\sigma(x) = \tau(x \cdot y) \in L$$

and $x' \preceq_L x$, it follows that $\sigma(x') \in L$. Therefore, we have shown that whenever $\tau(x \cdot y) \in L$, then also $\tau(x' \cdot y) = \sigma(x') \in L$. This proves that $x' \cdot y \preceq_L x \cdot y$.

By symmetry, if $y' \preceq_L y$ then $x \cdot y' \preceq_L x \cdot y$ for any x. Combining these two implications together, we obtain that $x' \cdot y' \preceq_L x \cdot y$. □

From the above lemma it follows that the $\langle\ \cdot\ , \#\rangle$-syntactic congruence preserves the operations in the signature $\langle\ \cdot\ , \#\rangle$. Therefore, the associative operation $x, y \mapsto x \cdot y$ of $\widehat{A^+}$ induces via α_L an associative operation in S_L, which we also denote $s, t \mapsto s \cdot t$. Similarly, the ω-power of $\widehat{A^+}$ induces a unary operation in S_L, which we call *stabilization*, and denote it by $s \mapsto s^\#$. This way, S_L becomes an algebra over the signature $\langle\ \cdot\ , \#\rangle$, and the mapping α_L becomes a homomorphism of $\langle\ \cdot\ , \#\rangle$-algebras. We call $\alpha_L \colon \widehat{A^+} \to S_L$ the $\langle\ \cdot\ , \#\rangle$-*syntactic homomorphism* induced by L. We will later on equip S_L with a suitable topology, for which α_L becomes a continuous homomorphism.

Proposition 11.2. *Let $L \subseteq \widehat{A^+}$. Then the following conditions are equivalent.*

1. *L has finite $\langle\, \cdot\, , \#\rangle$-index,*
2. *There is a finite $\langle\, \cdot\, , \#\rangle$-algebra $(S, \cdot, \#)$ with a distinguished subset F and a homomorphism $\alpha\colon \widehat{A^+} \to S$ of $\langle\, \cdot\, , \#\rangle$-algebras, such that $L = \alpha^{-1}(F)$.*

Proof. 1 ⇒ 2. If L has finite $\langle\, \cdot\, , \#\rangle$-index, then the $\langle\, \cdot\, , \#\rangle$-syntactic homomorphism α_L satisfies the second condition of the proposition.

2 ⇒ 1. Let α, S, F be as in the second condition of the proposition. It suffices to show that if $\alpha(x) = \alpha(y)$, then $x \simeq_L y$. Since α has a finite image, this will prove that $\simeq_L$ has finitely many equivalence classes.

To this end, let $\tau \in \mathit{Terms}(\widehat{A^+}, \,\cdot\, , \#)$ be any term with one free variable appearing once. The term τ induces a term $\alpha_*(\tau)$ over $(S, \,\cdot\, , \#)$, obtained by applying α to the leaves of the term τ, and interpreting the operations $\cdot$ and ω of $\widehat{A^+}$ as the operations $\cdot$ and # of S. The term $\alpha_*(\tau)$ has one free variable. Since α is a homomorphism of $\langle\, \cdot\, , \#\rangle$-algebras, for any $z \in \widehat{A^+}$,

$$\alpha_*(\tau)(\alpha(z)) = \alpha(\tau(z)). \tag{3}$$

Since $\tau(z) \in L$ if and only if $\alpha(\tau(z)) \in F$, from (3) we deduce that membership of $\tau(z)$ to L depends only on $\alpha(z)$. In particular, $\alpha(x) = \alpha(y)$ implies that $x \simeq_L y$. □

11.1.2 Topological structure

We would like to define a topology over S_L for which α_L becomes a continuous mapping. Note that usually, when the discrete topology is considered over S_L, α_L is *not* continuous, as the following example demonstrates.

Example 11.2. Consider the quotient mapping $\alpha_L\colon \widehat{A^+} \to S_L$ from Example 11.1. Let us denote by $0, 1, \omega$ the elements of S_L corresponding to the $\simeq_L$-equivalence classes L_0, L_1, L_ω, respectively. Note that the equivalence class L_1 is not closed, since the sequence $a, a^{2!}, a^{3!}, \ldots$ of its elements converges to the element a^ω of L_ω. Therefore, α_L is not continuous for the discrete topology over S_L (otherwise $\alpha_L^{-1}(\{1\})$ would be closed).

Apart from the degenerated topology (consisting of $\emptyset$ and S_L), there are precisely two topologies over S_L for which α_L is a continuous mapping. We describe them by their specialization preorders. In the first topology, we have $0 > 1 > \omega$. In the second topology, we have $1 > \omega$ and 0 incomparable with neither 1 nor ω. Note that the first preorder corresponds precisely to the partial order over S_L induced from $\preceq_L$. The second preorder corresponds to the quotient topology over S_L, i.e. reflects the fact that in $\widehat{A^+}$ the closure of L_1 contains L_ω and that L_0 is closed.

In general, it is natural to define the topology of S_L as the *quotient topology* induced by α_L. This is a notion from general topology – the quotient topology induced by $\alpha_L\colon \widehat{A^+} \to S_L$ is the strongest topology over S_L for which the mapping α_L is continuous. Equivalently, $F \subseteq S_L$ is closed if and only if the inverse image $\alpha_L^{-1}(F)$ is a closed subset of $\widehat{A^+}$. The advantage of considering the quotient topology is that the mapping α_L obviously becomes a continuous mapping from $\widehat{A^+}$ to S_L. Moreover, as the following results show, multiplication and stabilization are continuous in S_L.

Theorem 11.3. *Let $L \subseteq \widehat{A^+}$ be any set of finite $\langle\, \cdot\, ,\#\rangle$-index, and let $\alpha_L\colon \widehat{A^+} \to S_L$ be the induced $\langle\, \cdot\, ,\#\rangle$-syntactic homomorphism, and S_L be equipped with the quotient topology. Then multiplication and stabilization induced via α_L are continuous mappings, and α_L is a continuous homomorphism of topological $\langle\, \cdot\, ,\#\rangle$-algebras. If, moreover, L is a closed or open subset of $\widehat{A^+}$, then S_L is a T_0-topological space.*

The first part of the theorem follows from the following proposition.

Proposition 11.4. *Let $\varphi\colon S \to T$ be a surjective homomorphism of $\langle\, \cdot\, ,\#\rangle$-algebras from a topological $\langle\, \cdot\, ,\#\rangle$-algebra to a finite $\langle\, \cdot\, ,\#\rangle$-algebra. Let T be equipped with the quotient topology, i.e. such that F is closed in T iff $\varphi^{-1}(F)$ is closed in S. Then multiplication and stabilization in T are continuous.*

Proof. First we show that for any fixed $t_0 \in T$, right-multiplication by t_0, i.e. the mapping

$$t \mapsto t \cdot t_0$$

is a continuous mapping from T to T. We will then deduce that two-sided multiplication from $T \times T$ to T is continuous.

Let us fix $t_0 \in T$ and any $s_0 \in S$ such that $\varphi(s_0) = t_0$ (we use surjectivity of φ here). Let μ be right-multiplication by s_0 and ν be right-multiplication by t_0. Note that μ is a continuous mapping from S to S. The mappings μ, ν are linked via the following commuting diagram.

$$\begin{tikzcd}
S \arrow[r, "\mu"] \arrow[d, "\varphi"'] & S \arrow[d, "\varphi"] \\
T \arrow[r, "\nu"] & T
\end{tikzcd}$$

Let U be an open subset of T. We must show that $\nu^{-1}(U)$ is an open subset of T. By commutativity of the diagram, we have:

$$\varphi^{-1}\left(\nu^{-1}(U)\right) = \mu^{-1}\left(\varphi^{-1}(U)\right). \tag{1}$$

Since U is open and both μ and φ are continuous, we deduce that the set in the formula (1) is an open subset of S. We therefore conclude that $\nu^{-1}(U)$ is open, since its inverse image under φ is open by (1). This proves that right-multiplication is a continuous mapping in T. By repeating the above proof for left-multiplication by any fixed t_0, or for stabilization, we deduce that both these mappings are continuous mappings from T to T.

We now conclude that two-sided multiplication is a continuous mapping. For this, we use finiteness of T. Let U be any open subset of T, and let

$$V = \{(t, t') \in T \times T :\ t \cdot t' \in U\}.$$

We need to show that V is an open set. Since T is finite, V can be written as a finite union

$$V = \bigcup_{(t_0,t'_0)\in U} \{t' : \ t_0 \cdot t' \in U\} \times \{t : \ t \cdot t'_0 \in V\}.$$

Now, by continuity of left- and right-multiplicaiton, both factors of any disjunct in the above union is an open set. Consequently, V is finite as a finite union of products of open sets.

Therefore, T equipped with the operations of multiplication and stabilization becomes a topological $\langle \cdot, \# \rangle$-algebra, and φ is a continuous mapping of such algebras. □

The second part of Theorem 11.3 follows from the following lemma. It applies only to closed sets, but in case of an open set L, we may consider its complement K instead, and then α_L and α_K are the same mappings, so they induce the same topology over $S_L = S_K$.

Lemma 11.5. *Assume that $L \subseteq \widehat{A^+}$ is closed and let $x \in \widehat{A^+}$. Then $\downarrow x = \{y : \ y \preceq_L x\}$ is a closed subset of $\widehat{A^+}$. As a consequence, the quotient topology on S_L is T_0.*

Proof. By the chosen definitions,

$$\begin{aligned} \downarrow x &= \{y : \ y \preceq_L x\} \\ &= \{y : \ \forall \tau. \quad \tau(x) \in L \implies \tau(y) \in L\} \\ &= \bigcap_{\tau:\ \tau(x)\in L} \{y : \tau(y) \in L\}. \end{aligned}$$

In the above formulas, τ ranges over all elements in *Terms*$(\widehat{A^+}, \cdot, \#)$.

Since any term τ induces a continuous mapping from $\widehat{A^+}$ to itself, it follows that each of the sets $\{y : \ \tau(y) \in L\}$ is a closed subset of $\widehat{A^+}$ (here we use the assumption that L is closed). Therefore, $\downarrow x$ is an intersection of closed sets, so it is closed itself.

To prove that S_L is a T_0-topological space, by surjectivity of α_L, it suffices to show that if $x, y \in \widehat{A^+}$ are two points which are not equivalent with respect to $\simeq_L$, then $\alpha_L(x)$ and $\alpha_L(y)$ can be separated by a closed subset of S_L. If x and y are not $\simeq_L$-equivalent, then, by definition, either $\downarrow x$ does not contain y, or $\downarrow y$ does not contain x. In either case, we have that $\alpha_L(x)$ and $\alpha_L(y)$ can be separated by a closed set – either the image of $\downarrow x$, or the image of $\downarrow y$ under α_L. □

This finishes the proof of Theorem 11.3.

11.1.3 Order topology

In this section, we make a digression into a different way of introducing a topology over S_L. The definitions and results presented in Section 11.1.3 will not be used elsewhere in this thesis.

Since S_L is defined as the quotient by the equivalence relation induced by $\preceq_L$, it follows that S_L possesses an induced partial order, also denoted $\preceq_L$. We can then consider the topology over S_L for which $\preceq_L$ is the specialization preorder, i.e. such that the closed subsets are precisely the downward-closed subsets with respect to $\preceq_L$. We call this the *order topology* over S_L. Usually, the order topology differs from the quotient topology, as we saw in Example 11.2.

Remark 11.1. The definition of the relation $\preceq_L$ is chosen so that it is consistent with the specialization preorder of the order topology. As a consequence, in the chosen definition of the relation $\preceq_L$, the order is the inverse to the order which is usually considered in the theory of ordered semigroups. However, this discrepancy should not lead to confusion, since apart from Section 11.1.3 we will not be using the partial order $\preceq_L$, but the specialization order.

The order topology has the following properties, some of which can be seen as drawbacks.

Lemma 11.6. *Let $L \subseteq \widehat{A^+}$ be any subset and S_L be equipped with the order topology. Then:*

1. *S_L is a T_0 topological space*
2. *The set $\alpha_L(L)$ is closed in S_L*
3. *S_L is a topological algebra over the signature $\langle\,\cdot\,,\#\rangle$*
4. *$\alpha_L\colon \widehat{A^+} \to S_L$ is a homomorphism of $\langle\,\cdot\,,\#\rangle$-algebras which, in general, is not continuous.*

Proof. Note that since $\preceq_L$ is a partial order over S_L, the order topology is automatically a T_0 topology. Also, since L is downward-closed with respect to $\preceq_L$, it follows that $\alpha_L(L) \subseteq S_L$ is closed in the order topology. This implies that whenever L is not a closed subset of $\widehat{A^+}$, then α_L is not a continuous mapping with respect to the order topology, since $L = \alpha_L^{-1}(\alpha_L(L))$. (In particular, in this case, the order topology differs from the quotient topology.)

From Lemma 11.1 it follows that multiplication and stabilization in S_L preserve the relation $\preceq_L$ over S_L, so they are continuous mappings with respect to the order topology. Therefore, S_L becomes a topological algebra over the signature $\langle\,\cdot\,,\#\rangle$. By definition of multiplication and stabilization in S_L, α_L is a homomorphism of $\langle\,\cdot\,,\#\rangle$-algebras, but in general, it is not continuous. □

As we have seen, the fact that $\alpha_L(L)$ is always a closed set in the order topology is a drawback, which restricts the applicability of this topology only to the case when L is a closed subset of $\widehat{A^+}$, since otherwise α_L is not continuous. Conversely, we will see that if $L \subseteq \widehat{A^+}$ is a closed set and moreover S_L is finite, then α_L is a continuous mapping. As an advantage of the order topology, in this case, we have a certain universal property.

Proposition 11.7. *Let $L \subseteq \widehat{A^+}$ be a closed subset of finite $\langle\,\cdot\,,\#\rangle$-index, and let $\alpha_L\colon \widehat{A^+} \to S_L$ by its $\langle\,\cdot\,,\#\rangle$-syntactic homomorphism. Let S_L be equipped with the order topology. Then S_L is a T_0 topological $\langle\,\cdot\,,\#\rangle$-algebra, and α_L is a continuous homomorphism of such algebras, which has the following universal property.*

> *For any continuous homomorphism $\alpha\colon \widehat{A^+} \to S$ such that $L = \alpha^{-1}(F)$ for some closed $F \subseteq S$, there exists a unique continuous $\langle\,\cdot\,,\#\rangle$-homomorphism φ such that the following diagram commutes.*

$$\begin{tikzcd}
\widehat{A^+} \arrow[r, "\alpha"] \arrow[dr, "\alpha_L"'] & S \arrow[d, dashed, "\varphi"] \\
 & S_L
\end{tikzcd}$$

Sketch of proof. Any closed subset of S_L is a finite union of sets of the form $\downarrow s$, and $\downarrow s$ is closed in S_L by Lemma 11.5. This proves continuity of α_L.

It remains to prove the universal property. Its proof is standard universal algebra reasoning, involving the consideration of terms over the signature $\langle \cdot, \# \rangle$ in each of the algebras $\widehat{A^+}$, S and S_L, very similarly as in the proof of Proposition 11.2. This reasoning is extended by a simple continuity argument, similar to the one used in Lemma 11.5. We skip this proof, as Proposition 11.7 will not be used anywhere in this thesis. □

Corollary 11.8. *The order topology is the weakest topology on S_L for which α_L is a continuous mapping and L is a closed set.*

11.2 Stabilization semigroups

As we have seen, in general, there might be several distinct topologies over S_L for which the quotient mapping α_L is continuous. However, we may prove some properties of S_L which hold independently of the chosen topology. In fact, the following properties hold in any $\langle \cdot, \# \rangle$-algebra which is a continuous homomorphic image of any profinite semigroup.

Proposition 11.9. *Let α be a continuous, surjective homomorphism of topological $\langle \cdot, \# \rangle$-algebras from a profinite semigroup $(\tilde{S}, \cdot, \omega)$, to $(S, \cdot, \#)$. Then, S satisfies the following identities.*

$$s \cdot (t \cdot s)^\# = (s \cdot t)^\# \cdot s \tag{S1}$$

$$(s^n)^\# = s^\# \qquad \text{for } n = 1, 2, 3 \ldots \tag{S2}$$

$$(s^\#)^\# = s^\# \tag{S3}$$

$$s^\# \cdot s^\# = s^\# \tag{S4}$$

$$e \cdot e^\# = e^\# \qquad \text{if } e = e \cdot e. \tag{S5}$$

Moreover, for the specialization preorder $\leq$ over S,

$$e^\# \leq e \qquad \text{if } e = e \cdot e. \tag{S6}$$

Proof. The first four equalities are an immediate consequence of the corresponding equalities in $\tilde{S}$ (see Proposition 4.2 in the Preliminaries) and the fact that α is surjective.

Now assume that $e = e^\#$, and let x be such that $s = \alpha(x)$. Then x and x^2 have the same image in S, so also $x^{\omega-1} \cdot x$ and $x^{\omega-1} \cdot x^2$ have the same image in S. Since the first element is equal to x^ω and the latter is equal to $x \cdot x^\omega$, this proves the equality (S5).

We now prove the inequality (S6). Let $x \in \tilde{S}$ be such that $\alpha(x) = e$, and let F be the closure of the set $\{e\}$ in S. We show that F contains $e^\#$, proving that $e^\# \leq e$ with respect to the specialization preorder. Indeed, by idempotency of e we have that $x, x^2, x^3, \ldots$ are all mapped to e by α. In particular, $x^n \in \alpha^{-1}(F)$ for $n = 1, 2, \ldots$. Since $\alpha^{-1}(F)$ is closed, and x^ω is the limit point of the sequence $x^{n!}$, it follows that $x^\omega \in \alpha^{-1}(F)$. Therefore, $e^\# \in F$, proving that $e^\# \leq e$. □

Remark 11.2. The inequality (S6) could be replaced by a condition:

> *For any $s \in S$, $s^{\#}$ is the limit of the sequence $s, s^{2!}, s^{3!}, \ldots$*

Indeed, the above sequence is ultimately equal to the unique idempotent power e of s, so the above condition could be further rephrased:

> *For any idempotent $e \in S$, $e^{\#}$ is the limit of the sequence $e, e, e, \ldots$*

By definition of convergence, this is equivalent to saying that $e^{\#} \leq e$ for every idempotent $e \in S$.

Definition 4. We call a *topological stabilization semigroup* a finite topological semigroup S endowed with a continuous operation # which satisfies the axioms (S1)-(S6) listed above, where $\leq$ is the specialization preorder of S. In Part II of this thesis, we will only consider topological stabilization semigroups, so we will skip the attribute "topological". Moreover, unless explicitly stated otherwise, we will always assume that the topology is T_0, i.e. that the specialization preorder $\leq$ is a partial order.

11.2.1 Equivalence with stabilization semigroups of Colcombet

There exists already a different definition of stabilization semigroups, introduced by Colcombet [Col09]. We will see that there is a straightforward correspondence between the stabilization semigroups of Colcombet and the ones defined above (with the implicit T_0 separation axiom).

The definition of Colcombet differs in the following ways. There, a stabilization semigroup is assumed to be a partially ordered semigroup, and # is only assumed to be defined for idempotents. Because of that, the axiom (S1) does not make sense for all s and t, but only for s and t such that both $s \cdot t$ and $t \cdot s$ are idempotent. However, the definition of Colcombet does not include this restricted form of the axiom (S1), but a stronger one which, as we will see, encompasses both (S1) and (S5):

$$s \cdot (t \cdot s)^{\#} \cdot t = (s \cdot t)^{\#} \qquad \text{if } s \cdot t \text{ and } t \cdot s \text{ are idempotent.} \tag{S1$'$}$$

Colcombet also assumes the axioms (S3) and (S4), both restricted to the case of idempotent s, the axiom (S6) and that $e \leq f \implies e^{\#} \leq f^{\#}$ for idempotent e, f, which is equivalent to our assumption on continuity of #. Those are all the axioms of a stabilization semigroup in the sense of Colcombet. The axiom (S2) is trivially satisfied for idempotent s, so it is redundant.

Note that if S is a stabilization semigroup in the sense of Colcombet, then there is only one sensible way of extending # to all elements of S, so that the axiom (S2) holds: we must define $s^{\#} = e^{\#}$, where $e = s^k$ is the idempotent power of s.

Proposition 11.10. *Let S be a T_0 stabilization semigroup. Restricting # to idempotents yields a stabilization semigroup in the sense of Colcombet. Conversely, if S is a stabilization semigroup in the sense of Colcombet, then by extending # to S in the unique way, we obtain a T_0 stabilization semigroup.*

Proof. For the first part of the statement, we only need to prove that in stabilization semigroups, (S1′) holds for s, t such that $s \cdot t$ and $t \cdot s$ are idempotent. This is indeed the case, since

$$s \cdot (t \cdot s)^{\#} \cdot t \overset{\text{(S1)}}{=} s \cdot t \cdot (s \cdot t)^{\#} \overset{\text{(S5)}}{=} (s \cdot t)^{\#}.$$

Now we turn to the the second part of the statement. Assume that # is defined only over idempotents and satisfies the axioms (S3),(S4),(S6) for idempotent s and the axiom (S1′). Let $\omega \in \mathbb{N}$ denote the idempotent power of the semigroup S. We set

$$s^{\#} = (s^{\omega})^{\#} \quad \text{for all } s \in S.$$

We now show that this extension of # to S satisfies the axioms (S1)-(S6). Also, note that since in Colcombet's stabilization semigroups multiplication and stabilization preserve the order $\leq$,

$$s \leq t \quad \Longrightarrow \quad s^{\omega} \leq t^{\omega} \quad \Longrightarrow \quad s^{\#} = (s^{\omega})^{\#} \leq (t^{\omega})^{\#} = t^{\#},$$

so the extension of # respects the order on S, i.e. is a continuous mapping over S.

The identities (S2),(S3),(S4),(S6) are trivially satisfied. To prove (S1), choose any $s, t \in S$ and define

$$\begin{aligned} a &\overset{def}{=} (s \cdot t)^{\omega} \cdot s, \\ b &\overset{def}{=} (t \cdot s)^{\omega - 1} \cdot t \end{aligned}$$

Note that $a \cdot b = (s \cdot t)^{\omega}$ and $b \cdot a = (t \cdot s)^{\omega}$ are both idempotent. Then:

$$\begin{aligned} s \cdot (t \cdot s)^{\#} &= s \cdot ((t \cdot s)^{\omega})^{\#} \\ &= s \cdot (b \cdot a)^{\#} \\ &= s \cdot b \cdot (a \cdot b)^{\#} \cdot a && \text{(S1′)} \\ &= (s \cdot t)^{\omega} (s \cdot t)^{\#} \cdot (s \cdot t)^{\omega} \cdot s \\ &= (s \cdot t)^{\#} \cdot s. && \text{(S1′)} \end{aligned}$$

Finally, to prove (S5), choose an idempotent e and note that

$$\begin{aligned} e \cdot e^{\#} &= e \cdot (e \cdot e^{\#} \cdot e) && \text{(S1′)} \\ &= e \cdot e^{\#} \cdot e \\ &= e^{\#}. && \text{(S1′)} \end{aligned}$$

□

11.3 Semigroups of transformations of B- and S-automata

Recall that in the classical theory of finite automata, a nondeterministic automaton induces a semigroup of its transformations. A similar situation occurs with B- and S-automata, but the transformation semigroups are stabilization semigroups. We describe those constructions, proving the following.

Proposition 11.11. *Let $\mathcal{A}$ be a B- or S-automaton. Then there exists a finite T_0 stabilization semigroup S and a continuous homomorphism of $\langle\,\cdot\,,\#\rangle$-algebras*

$$\hat{\alpha} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad S,$$

such that $L\left(\mathcal{A}\right) = \hat{\alpha}^{-1}(F)$ for some $F \subseteq S$, where F is open in the case of B-automata and F is closed in the case of S-automata. The stabilization semigroup S, the set F, and the restriction $\hat{\alpha}|_A$ are computable from $\mathcal{A}$.

Moreover,

$$K = \overline{K \cap A^{\langle\,\cdot\,,\omega\rangle}},$$

where $K = \widehat{A^+} - L\left(\mathcal{A}\right)$ in the case of B-automata and $K = L\left(\mathcal{A}\right)$ in the case of S-automata.

In the rest of Section 11.3 we prove the above proposition. First we deal with B-automata, and then with S-automata.

11.3.1 Transformation semigroup for B-automata

In Part I, Chapter 7, we constructed a profinite semigroup

$$P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$$

of transformations for B-automata with states Q and counters C. Moreover, we considered a homomorphism

$$\alpha_{(1=2)} \quad : \quad P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q) \quad \longrightarrow \quad P_{\uparrow}(Q \times \overline{\mathbb{B}}^C_{/1=2} \times Q) \quad \overset{def}{=} \quad \mathbb{P}_1$$

onto a finite semigroup which we denote here by $\mathbb{P}_1$, and the above homomorphism induced the structure of a $\langle\,\cdot\,,\#\rangle$-algebra on $\mathbb{P}_1$. Moreover, the homomorphism is order-preserving, where both sets are ordered by inclusion $\subseteq$. Let us consider the topology over $\mathbb{P}_1$ for which $\subseteq$ is the specialization preorder. (We could also consider the quotient topology, and this wouldn't affect the reasonings too much.) Clearly, this is a T_0 topology, since $\subseteq$ is a partial order.

Lemma 11.12. *The mapping $\alpha_{(1=2)}$ is continuous homomorphism of topological $\langle\,\cdot\,,\#\rangle$-algebras.*

Sketch of proof. We show show that any upward-closed (with respect to $\subseteq$) subset of

$$\mathbb{P} \quad \overset{def}{=} \quad P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$$

is open. The topology of $\mathbb{P}$ is such that every mapping

$$\alpha_{(N=\omega)} \quad : \quad \mathbb{P} \quad \longrightarrow \quad \mathbb{P}_{/N=\omega}$$

(using notations from Chapter 7) is continuous, where the codomain is equipped with the discrete topology. As a consequence of Dickson's lemma, just as in the proof of Lemma 7.2 in Chapter 7, any upward-closed subset of $\mathbb{P}$ has a finite number of minimal elements. Therefore, for any upward-closed subset U of $\mathbb{P}$ there is a number N such that U is an inverse image under $\alpha_{(N=\omega)}$ of some upward-closed set in $\mathbb{P}_{/N=\omega}$. But any upward-closed set in $\mathbb{P}_{/N=\omega}$ is open. This proves that any upward-closed set in $\mathbb{P}$ is open. As a consequence, the mapping $\alpha_{(1=2)}$ is continuous.

We also need to show that multiplication and stabilization in $\mathbb{P}_1$ are continuous. This follows from the fact that multiplication and the ω-power are order-preserving mappings over $\widehat{A^+}$. □

Therefore, the homomorphism $\alpha_{(1=2)}$ is a surjective homomorphism of topological $\langle \cdot , \# \rangle$-algebras, and $P_\uparrow(Q \times \overline{\mathbb{B}}^C_{/1=2} \times Q)$ becomes a T_0 stabilization semigroup by Proposition 11.9. In Chapter 7 we furthermore defined the mapping

$$\delta_{\mathcal{A}} \quad : \quad A^+ \quad \longrightarrow \quad P_\uparrow(Q \times \overline{\mathbb{B}}^C \times Q),$$

induced by the transition relation of $\mathcal{A}$. By the universal property of profinite semigroups, the mapping $\delta_{\mathcal{A}}$ extends to a unique continuous homomorphism $\hat{\delta}_{\mathcal{A}}$ defined over $\widehat{A^+}$. Then the composition

$$\hat{\alpha} \quad \overset{def}{=} \quad \alpha_{(1=2)} \circ \hat{\delta}_{\mathcal{A}} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad P_\uparrow(Q \times \overline{\mathbb{B}}^C_{/1=2} \times Q),$$

is a continuous homomorphism of stabilization semigroups, from the free profinite semigroup to a finite T_0 stabilization semigroup.

It follows from the characterization given in Section 7.1.1 in Part I of this thesis, that the ω-set of the function $f_{\mathcal{A}}$ determined by $\mathcal{A}$ is the inverse image under α of a downward-closed subset of

$$P_\uparrow(Q \times \overline{\mathbb{B}}^C_{/1=2} \times Q).$$

Consequently, the set $L(\mathcal{A})$ is an inverse image under $\hat{\alpha}$ of an upward-closed (hence open) set F. The set F can be explicitly described as the set of all $x \in P_\uparrow(Q \times \overline{\mathbb{B}}^C_{/1=2} \times Q)$ with the following property.

> *There exists a transition τ in x from an initial state to an accepting state such that $\tau[c] < \omega$ for every counter $c \in C$.*

Clearly, the above property is preserved by taking supersets, i.e. if x satisfies the above property and $x \subseteq y$, then also y satisfies the above property. This corresponds to the fact that F is an upward-closed set.

We prove an important property of complements of languages accepted by B-automata.

Proposition 11.13. *Let $\mathcal{A}$ be a B-automaton, and let $K = \widehat{A^+} - L(\mathcal{A})$ be the ω-set of the function $f_{\mathcal{A}}$. Then,*

$$K = \overline{K \cap A^{\langle\,\cdot\,,\omega\rangle}}.$$

Proof. We show that the class of complements of languages accepted by B-automata satisfies the assumptions of Lemma 8.3. Indeed, since B-automata accept open sets, their complements are closed sets. Moreover, an intersection of the complement of a language accepted by a B-automaton with a clopen set is again a complement of a language accepted by a B-automaton, which follows from Proposition 9.3.

It remains to prove that if the complement K of a language accepted by a B-automaton is nonempty, then it contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$. This result is difficult, but it follows from the results shown in Part I (in fact, it easily implies that the limitedness problem for B-automata is decidable), as we now describe.

We show that if K is nonempty then it contains a profinite word y from the set $A^{\langle\,\cdot\,,\omega\rangle}$. Let B be the set of transitions induced by $\mathcal{A}$:

$$B \quad \overset{def}{=} \quad \delta_{\mathcal{A}}(A) = \{\delta_{\mathcal{A}}(a) : a \in A\}.$$

Then, the image of $\hat{\delta}_{\mathcal{A}}$ is equal to $\overline{B^+}$ – this follows from the universal property of profinite semigroups, stated in Proposition 4.7 of the Preliminaries. Since B is finite and $P_{\uparrow}(Q \times \overline{\mathbb{B}}^C \times Q)$ is $\langle\,\cdot\,,\omega\rangle$-locally closed by Theorem 7.1 of Part I, it follows that

$$\operatorname{Im}\hat{\delta}_{\mathcal{A}} = \overline{B^+} = B^{\langle\,\cdot\,,\omega\rangle} = \hat{\delta}_{\mathcal{A}}(A^{\langle\,\cdot\,,\omega\rangle}). \tag{1}$$

The last equation is a consequence of the fact that $\hat{\delta}_{\mathcal{A}}$ preserves multiplication and the ω-power.

Assume that $x \in K$, i.e. x is a profinite word which does not belong to $L(\mathcal{A})$. By equation (1), there is a profinite word $y \in A^{\langle\,\cdot\,,\omega\rangle}$ such that $\hat{\delta}_{\mathcal{A}}(x) = \hat{\delta}_{\mathcal{A}}(y)$. But membership of a profinite word to $L(\mathcal{A})$ depends only on its image under $\hat{\delta}_{\mathcal{A}}$. Therefore, since x does not belong to $L(\mathcal{A})$, neither does y, so K contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$.

We have therefore shown that for any B-automaton $\mathcal{A}$, the closed set $\widehat{A^+} - L(\mathcal{A})$ either is empty, or contains an element from $A^{\langle\,\cdot\,,\omega\rangle}$. This shows that the class of complements of languages accepted by B-automata satisfies the assumptions of Lemma 8.3. □

11.3.2 Transformation semigroup for S-automata

We only sketch the analogous construction for S-automata, since it is completely dual. We construct the profinite semigroup

$$P_{\downarrow}(Q \times \overline{\mathbb{S}}^C \times Q)$$

of transformations for S-automata with counters C and states Q, where $\overline{\mathbb{S}}$ is the profinite semigroup defined in Section 10.3, ordered coordinatewisely, and $P_{\downarrow}$ denotes downward-closed subsets. Next, $\alpha_{(1=2)} : \overline{\mathbb{S}} \to \overline{\mathbb{S}}_{/1=2}$ induces the homomorphism

$$\alpha_{(1=2)} \quad : \quad P_{\downarrow}(Q \times \overline{\mathbb{S}}^{C} \times Q) \quad \longrightarrow \quad P_{\downarrow}(Q \times \overline{\mathbb{S}}_{1}^{C} \times Q)$$

to the finite semigroup of downward-closed subsets of $Q \times \overline{\mathbb{S}}_{1}^{C} \times Q$. This homomorphism induces the structure of a stabilization semigroup on its image. We consider the topology over the image for which the specialization preorder is the *inverse* of subset inclusion.

Lemma 11.14. *The mapping $\alpha_{(1=2)}$ is a continuous homomorphism of topological $\langle\, \cdot\, , \#\rangle$-algebras.*

Sketch of proof. This time, it suffices to prove that any *downward*-closed subset of $P_{\downarrow}(Q \times \overline{\mathbb{S}}^{C} \times Q)$ is open. The proof is exactly the same as the proof of Lemma 11.12, but we use the fact that any downward-closed set is determined by a finite number of elements (which also follows from Dickson's lemma). □

An S-automaton with counters C and states Q induces a mapping

$$\delta_{\mathcal{A}} \quad : \quad A \quad \longrightarrow \quad P_{\downarrow}(Q \times \overline{\mathbb{S}}^{C} \times Q),$$

defined by

$$\delta_{\mathcal{A}}(a) \quad \overset{def}{=} \quad \downarrow\{(p, o, q) : \quad (p, a, \tilde{o}, q) \in \delta, \quad \tilde{o} \textit{ induces the operation } o \in \overline{\mathbb{S}}^{C})\}.$$

The mapping $\delta_{\mathcal{A}}$ extends to a unique continuous homomorphism $\hat{\delta}_{\mathcal{A}}$ defined over $\widehat{A^{+}}$. We denote by $\hat{\alpha}$ the composition of $\hat{\delta}_{\mathcal{A}}$ with $\alpha_{(1=2)}$. This is a homomorphism of stabilization semigroups, from the free profinite semigroup to a finite stabilization semigroup. Then, $L(\mathcal{A})$ is the inverse image of an upward-closed (hence closed) subset F of $P_{\downarrow}(Q \times \overline{\mathbb{S}}_{1}^{C} \times Q)$, which consists of all $x \in P_{\downarrow}(Q \times \overline{\mathbb{S}}_{1}^{C} \times Q)$ with the following property.

> *There exists a transition τ in x from an initial state to an accepting state such that $\tau[c] = (\omega)$ or $\tau[c] = (\omega, n)$ or $\tau[c] = (\omega, \omega, n)$ for every counter $c \in C$.*

Recall that by Proposition 10.10,

$$L(\mathcal{A}) = \overline{L(\mathcal{A}) \cap A^{\langle\, \cdot\, , \omega\rangle}}.$$

This completes the proof of Proposition 11.11.

CHAPTER 12

The homomorphic extension property

Recall that in the case of usual semigroups, in order to specify uniquely a continuous homomorphism from $\widehat{A^+}$ to a finite semigroup, it suffices to define its values over the elements of A. The situation is different for stabilization semigroups. Unfortunately, in general, there might be many continuous homomorphisms from $\widehat{A^+}$ to a finite stabilization semigroup, which agree over the elements of A. However, precisely one of these mappings is canonical, in the sense that it has several desired properties. In this chapter, we will prove the existence of canonical extensions.

A *canonical* mapping (or homomorphism) is a continuous homomorphism from $\widehat{A^+}$ to a finite stabilization semigroup which satisfies any of the equivalent conditions of the following proposition.

Proposition 12.1. *Let $\hat{\alpha}\colon \widehat{A^+} \to S$ be a continuous homomorphism to a finite stabilization semigroup. Then, the following conditions are equivalent.*

1. *$\hat{\alpha}$ is the "largest", in the following sense. For any continuous homomorphism $\beta\colon \widehat{A^+} \to S$, if β and $\hat{\alpha}$ agree over A (i.e. $\beta|_A = \hat{\alpha}|_A$) then for all $x \in \widehat{A^+}$, $\beta(x) \leq \hat{\alpha}(x)$ in the specialization preorder over S,*

2. *For any closed subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is an S-regular language,*

3. *For any open subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is a B-regular language,*

4. *For any closed subset $F \subseteq S$,*
$$\hat{\alpha}^{-1}(F) = \overline{\hat{\alpha}^{-1}(F) \cap A^{\langle \cdot ,\omega \rangle}}.$$

Moreover, if $\hat{\alpha}$ satisfies any of the above conditions, then its image is equal to the $\langle \ \cdot \ ,\#\rangle$-subalgebra generated by $\hat{\alpha}(A)$ in S.

From the characterization given by the first item in the above proposition, it follows that if the canonical extension of α exists, then it is unique, as there is at most one largest mapping. The following, crucial theorem says that any mapping from a finite alphabet to a finite stabilization semigroup can indeed be extended to a canonical mapping.

Theorem 8.1. *Let $\alpha\colon A \to S$ be any mapping from a finite alphabet A to a finite T_0 stabilization semigroup S. Then there exists a unique canonical continuous homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ extending α.*

The rest of this chapter is devoted to proving Theorem 8.1 and Proposition 12.1. Let us fix a finite stabilization semigroup S and a mapping $\alpha\colon A \to S$ from a finite alphabet. We will define the mapping $\hat{\alpha}\colon \widehat{A^+} \to S$ via factorization trees, and we will later prove that this mapping has the properties required in Theorem 8.1, i.e. that it is a canonical homomorphism. We will also prove the equivalent statements of Proposition 12.1

There is some link between Theorem 8.1 above and the theorem of existence and uniqueness of compatible mappings of Colcombet [Col09]. However, it is easier to prove our theorem directly than by deriving it from the theorem of Colcombet (still, we use a lemma of Colcombet).

Proof strategy

We sketch the strategy of proving Theorem 8.1 by presenting an overcomplicated proof of a trivial fact – that for any finite semigroup S and any mapping α from A to S, there exists a unique homomorphism $\tilde{\alpha}$ from A^+ to S extending α. The standard construction of such a mapping is by defining $\tilde{\alpha}(a_1a_2\ldots a_n)$ as $\alpha(a_1)\alpha(a_2)\ldots\alpha(a_n)$ for any word $a_1a_2\ldots a_n$. Note that we use the assumption that multiplication in S is associative to make sure that $\tilde{\alpha}$ is a homomorphism – for instance, $\tilde{\alpha}(abcd) = \tilde{\alpha}(ab)\tilde{\alpha}(cd)$ thanks to the fact that we can "rearrange" the product $((ab)c)d$ into $(ab)(cd)$. The above definition of $\tilde{\alpha}$ relies on the fact that elements of A^+ are finite words, while in Theorem 8.1 we will have to deal with infinite profinite words. Therefore, we describe another way of constructing the homomorphism $\tilde{\alpha}\colon A^+ \to S$, by using factorization trees. The advantage of this approach will be that the value $\tilde{\alpha}(w)$ will be determined in a number of steps which does not depend on the length of w, as it will be bounded by a bound depending only on S. This will be very helpful when dealing with infinite words.

Recall that in the classical version of the factorization theorem of Simon (i.e. the case of Theorem 6.4 of Part I, in which stabilization is equal to the idempotent power), for any word in A^+, one can construct a factorization tree with respect to S, whose height is bounded by some constant $H = \|S\|$ dependent only S. Moreover, in the classical case, the output of a factorization tree over input word w is unique, i.e. does not depend on the tree. One can prove this by using the associativity axiom to "rearrange" one factorization tree into another one. This is a rather complicated way of proving uniqueness, since we could also simply notice that the output of a factorization tree over w is in fact equal to $\tilde{\alpha}(w)$, but we pretend not to know this while presenting this proof strategy.

By the factorization theorem, any finite word w has some factorization tree of height at most H. We define the value $\tilde{\alpha}(w)$ as the output of such a factorization tree. The definition makes sense, thanks to uniqueness of the output. Moreover, we can think that it takes only H steps to compute the value of $\tilde{\alpha}(w)$. The defined mapping $\tilde{\alpha}$ is homomorphism extending α. Homomorphicity follows from the fact that if $\tilde{\alpha}(w) = s$ and $\tilde{\alpha}(v) = t$, then wv has a factorization tree of height $H + 1$ with output st. Since the output of a factorization tree is unique, it follows that wv also has a factorization tree of height H with output st, so $\alpha(wv) = st$.

Our strategy for proving Theorem 8.1 resembles the above construction of $\tilde{\alpha}$ with the use of factorization trees. We will define the notion of *factorization tables*, which are like factorization trees, but suited for profinite words. It turns out that any profinite word x has some factorization table of height bounded by some value H, and, crucially, that its output is unique, i.e. does not depend on the choice of the factorization table. We will then define $\hat{\alpha}(x)$ as the output of the factorization table over x.

Colcombet's factorization trees

We first define the notion of a *δ-factorization tree* of a finite word $w \in A^+$ with respect to $\alpha\colon A \to S$, which is a generalization suggested by Colcombet [Col10a] of the notion of factorization trees which were considered in Chapter 6 of Part I. The novelty is that a δ-factorization tree t has an additional parameter $\delta \in \mathbb{N}$, called the *degree* of t. The base rule and the binary rule are the same as before, but the stabilization rule requires at least as many children as the degree of t. If there are less children, we may use the *idempotent rule*, which is similar to the stabilization rule, but does not apply stabilization to the output. A formal definition follows.

Let $\delta \in \mathbb{N}$. A *δ-factorization tree* t is tied to the following objects: the mapping α, an input word $w \in A^+$, and an output $s \in S$.

- *Base rule.* Each letter $a \in A$ is a δ-factorization tree, with input a and output $\alpha(a)$.

- *Binary rule.* Suppose that f, g are δ-factorization trees f, g with inputs $v, w \in A^+$ and outputs $s, t \in S$, respectively. Then (f, g) is a δ-factorization tree with input vw and output st.

- *Idempotent rule.* Suppose that the δ-factorization trees $f_1, \ldots, f_n$ have inputs $v_1, \ldots, v_n$, but the same idempotent $e \in S$ on output, and that $n < \delta$. Then $\#(f_1, \ldots, f_n)$ is a δ-factorization tree with input $v_1 \cdots v_n$ and output e.

- *Stabilization rule.* Suppose that the δ-factorization trees $f_1, \ldots, f_n$ have inputs $v_1, \ldots, v_n$, but the same idempotent $e \in S$ on output, and that $n \geq \delta$. Then $\#(f_1, \ldots, f_n)$ is a δ-factorization tree with input $v_1 \cdots v_n$ and output $e^\#$.

A δ-factorization tree can be seen as a tree, where the base rule corresponds to leaves, the binary rule corresponds to nodes of outdegree two, the idempotent rule corresponds to nodes of outdegree at most δ, and the stabilization rule corresponds to nodes of outdegree at least δ. The factorization trees considered in Part I can be seen as a special case, in which $\delta = 1$. In Chapter 12, when we say *factorization tree*, we mean a δ-factorization tree for some $\delta \in \mathbb{N}$ which is not necessarily equal to 1.

The formulations of the following two lemmas are due to T. Colcombet [Col10a].

Lemma 12.2. *Let S be a finite stabilization semigroup. There exists a constant $\|S\|$, such that for every degree $\delta \in \mathbb{N}$ and every $w \in A^+$ there exists a factorization tree over w of height $\|S\|$ and degree δ.*

We will call factorization trees of height $\|S\|$ *small factorization trees*. The actual value of $\|S\|$ will not be important to us (it can be bounded by $3|S|$). The above lemma says that any word has a small factorization tree of any given degree δ. It can be proved identically as our Theorem 6.4 of Part I.

Another lemma that we will use, also due to Colcombet, relates the outputs of two different factorization trees over the same input.

Lemma 12.3. *Let $h \in \mathbb{N}$. For any degree $\delta_1 \in \mathbb{N}$ there exists a degree δ_2 such that for any factorization trees t_1, t_2 over the same word and of height h, if t_1 has degree at most δ_1 and t_2 has degree at least δ_2, then*

$$\mathit{output}(t_1) \leq \mathit{output}(t_2)$$

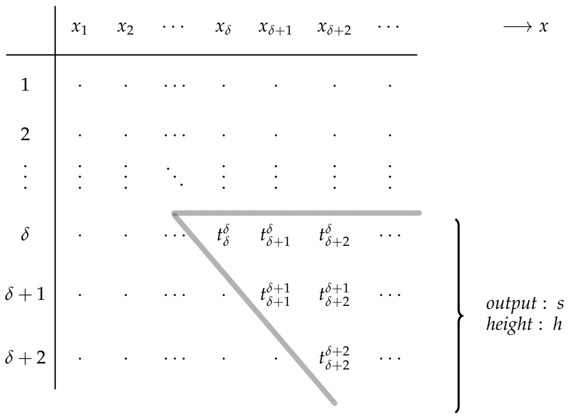
with respect to the partial order of S.

We do not prove this lemma here. It involves "rearranging" the trees t_1 and t_2 and using the axioms of stabilization semigroups.

12.1 Factorization tables and the induced mapping

Let $x \in \widehat{A^+}$, $s \in S$ and $h \in \mathbb{N}$. A *factorization table* of the profinite word x with respect to $\alpha \colon A \to S$ is described by the following ingredients:

- The *output* s, an element of S,
- The *height* h, a natural number,
- The *degree* δ, a natural number,
- The *base*, a sequence $x_1, x_2, \ldots$ of finite words over A which converges to x,
- For every $n, d \in \mathbb{N}$ such that $\delta \leq d \leq n$ (see Figure 12.1) a factorization tree t_n^d over x_n of degree d, height at most h and output s.



	x_1	x_2	$\cdots$	x_δ	$x_{\delta+1}$	$x_{\delta+2}$	$\cdots$	$\longrightarrow x$
1	$\cdot$	$\cdot$	$\cdots$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
2	$\cdot$	$\cdot$	$\cdots$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
δ	$\cdot$	$\cdot$	$\cdots$	t_δ^δ	$t_{\delta+1}^\delta$	$t_{\delta+2}^\delta$	$\cdots$	
$\delta+1$	$\cdot$	$\cdot$	$\cdots$	$\cdot$	$t_{\delta+1}^{\delta+1}$	$t_{\delta+2}^{\delta+1}$	$\cdots$	
$\delta+2$	$\cdot$	$\cdot$	$\cdots$	$\cdot$	$\cdot$	$t_{\delta+2}^{\delta+2}$	$\cdots$	

FIGURE 12.1: *A factorization table*

We identify the factorization table with the family of factorization trees $\{t_n^d\}_{\delta \leq d \leq n}$, since all the other ingredients of the table can be extracted from this family.

Remark 12.1. Let us be given a factorization table of a profinite word x. Then, we can restrict the factorization table to any infinite subsequence of its base, resulting in a new factorization table which has the same height, degree and output. Moreover, if x has a factorization table of degree δ, height h and output s, then it also has a factorization table of degree δ', height h' and output s for any $\delta' \geq \delta$ and $h' \geq h$.

We show that any profinite word has some factorization table. As far as two factorization trees of a given finite input word may yield distinct outputs, this discrepancy does not occur with factorization tables over profinite words.

Proposition 12.4. *Let $x \in \widehat{A^+}$. Then x has a factorization table of height $\|S\|$. Moreover, any two factorization tables of x have the same output.*

Using Lemmas 12.2 and 12.3 we will construct a factorization table for x. Uniqueness will crucially depend on the fact that we are dealing with convergent sequences of words, representing profinite words. A formal proof follows.

Proof. Let $h = \|S\|$ be the constant from Lemma 12.2. Using Lemma 12.3 for $h = \|S\|$, we define inductively an infinite sequence $d_1, d_2, \ldots$, by setting $d_1 = 1$, and for $n = 1, 2, \ldots$, we define d_{n+1} to be the value δ_2 granted by Lemma 12.3 when δ_1 is equal to $\max(d_n, n)$.

We will use the typeface $\mathbf{x}$ to denote an entire infinite sequence of finite words, which converges to x.

Let k be a fixed number. We will use an operation sift_k which maps a sequence $\mathbf{x}$ to its subsequence $\mathbf{y}$. We only require that this operation results in an infinite sequence of words which all have a small factorization tree of degree d_k with the *same* output.

Such an operation can be described as follows. Assume that $\mathbf{x} = x_1, x_2, \ldots$ By Lemma 12.2, for each n, there is a small factorization tree of the word x_n of degree d_k, and of some output $s_n \in S$. Since S is finite, there must be some $s \in S$ and some infinite sequence $n_1 < n_2 < \ldots$ such that $s_{n_i} = s$ for all $i = 1, 2, \ldots$ We choose $\mathit{sift}_k(\mathbf{x})$ to be the sequence $x_{n_1}, x_{n_2}, \ldots$. We call s the *output* of the resulting sequence $\mathit{sift}_k(\mathbf{x})$.

Let us start with any sequence $\mathbf{x}_0$ of words, which converges to x. For $n = 1, 2, \ldots$, we inductively define $\mathbf{x}_n = \mathit{sift}_n(\mathbf{x}_{n-1})$. Let s_n be the output of the sequence $\mathbf{x}_n$.

Since for each $n = 1, 2, \ldots$, the sequence $\mathbf{x}_n$ is a subsequence of the sequence $\mathbf{x}_{n-1}$, it follows that any element of the sequence $\mathbf{x}_n$ has a small factorization tree of degree d_{n-1} of output s_{n-1}, as well as a small factorization tree of degree d_n of output s_n. From Lemma 12.3 and from the definition of d_n it follows that $s_{n-1} \leq s_n$. Therefore, the sequence $s_1 \leq s_2 \leq \ldots$ must stabilize, i.e. there exists an element $s \in S$ and number N such that $s_k = s$ for all $k \geq N$.

For $n = 1, 2, \ldots$, let x_n denote the first element of the sequence $\mathbf{x}_{n+N}$. We have thus constructed a sequence $(x_n)_{n=1}^{\infty}$ with the following property:

If $n, k \in \mathbb{N}$ are such that $N \leq k \leq n + 1$, then the word x_n has a small factorization tree of degree d_k and output s. $(*)$

The sequence $x_1, x_2, \ldots$ will form the base of our factorization table of x. The output of the factorization table will be s, its height h and its degree $\delta = d_{N+1}$. It remains to show that for each $d, n \in \mathbb{N}$ such that $\delta \leq d \leq n$, the word x_n has a small factorization tree t_n^d of degree d and output s.

Let $\delta \leq d \leq n$. Choose any small factorization tree of x_n of degree d and some output t. By the property $(*)$, the word x_n has a small factorization table of degree d_N and output s, and since $d_{N+1} = \delta \leq d$, it follows from Lemma 12.3 that $s \leq t$. Similarly, x_n has a small factorization table of degree d_{n+1} and output s. Since x_n has a small factorization tree of output t and degree d, and $d \leq n \leq d_{n+1}$ it follows that $t \leq s$. This proves that $t = s$, so x_n has a small factorization tree of degree d and output s.

Therefore, the profinite word x has a factorization table with output s and of height $h = \|S\|$. This proves the first part of the proposition.

To prove the second part, assume that x has two factorization tables, $\mathcal{F} = \{p_n^d\}_{d \leq n}$ and $\mathcal{G} = \{q_n^d\}_{d \leq n}$, with outputs s and u, respectively. We will show that $s = u$. We may assume that each of the two factorization tables has the same height h and degree δ (see Remark 12.1). Let $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ be the bases of the factorization tables $\mathcal{F}$ and $\mathcal{G}$, respectively.

First we will consider the special case when both factorization tables have the same base, i.e. $x_n = y_n$ for all n. Let h be the height of the two tables, let $\delta_1 = \delta$ and let δ_2 be the value granted by Lemma 12.3. Pick any word x_n, where $n \geq \delta_2$. It has a factorization tree of height h, degree δ_1 and output s and a factorization tree of height h, degree δ_2 and output u. From Lemma 12.3, we deduce that $s \leq u$. By symmetry, $u \leq s$. This proves that $s = u$.

Now, we will reduce the general case to the special case which was just solved. For any $s \in S$ and number $d \in \mathbb{N}$, let $L_{s,d}$ be the set consisting of all finite words which have a factorization tree of height h, degree d and output s.

Lemma 12.5. *Let $t \in S$ and $d \in \mathbb{N}$. Then $L_{s,d} \subseteq A^+$ is a regular language.*

Proof. One way of proving this lemma is to construct a nondeterministic automaton which tries to build a factorization tree of degree δ and height h for a given input word – it needs to store at most h outputs of factorization trees of heights $1, 2, 3, \ldots, h$, and have as many counters which count only up to d.

We will present a different proof, via regular expressions. It proceeds by induction on the height h. For a number $h \in \mathbb{N}$, let $L_{s,d}^h$ be the set of all finite words which have a factorization tree of height h, degree d and output s.

In the base case, when $h = 1$, $L_{s,d}^h$ is the regular expression which describes the finite set of letters in A which are mapped to s by α. If $h > 1$, then

$$L_{s,d}^h = \bigcup_{\substack{u,v \in S \\ u \cdot v = s}} L_{u,d}^{h-1} \cdot L_{v,d}^{h-1} \quad \cup \quad \bigcup_{\substack{e \in S \\ e = e^2 = s}} (L_{e,d}^{h-1})^{<d} \quad \cup \quad \bigcup_{\substack{e \in S \\ e = e^2 \\ e^{\#} = s}} (L_{e,d}^{h-1})^{\geq d}. \qquad (1)$$

This finishes the proof of the inductive step. Hence, $L_{s,d}^h$ is a regular language for any $h \in \mathbb{N}$. □

Note that for any fixed $d \geq \delta$, almost all the elements of the sequence $x_1, x_2, \ldots$ belong to the regular language $L_{s,d}$. Since the sequence $y_1, y_2, \ldots$ is convergent to the same limit, it follows that almost all of its elements also belong to the language $L_{s,d}$. For each $d \geq \delta$, let k_d be any number such that $y_k \in L_{s,d}$ for all $k \geq k_d$. In particular, we have that

$$y_{k_n} \in \bigcap_{\delta \leq d \leq n} L_{s,d}.$$

This implies that there is a factorization table of x with base $y_{k_1}, y_{k_2}, \ldots$ height h, degree δ and output s. But we also know that there is a factorization table of x with base $y_{k_1}, y_{k_2}, \ldots$ and output u (see Remark 12.1). From the special case which was solved previously, we conclude that $s = u$. This finishes the proof of the proposition. □

Definition 5. Let $\alpha \colon A \to S$ be a mapping from an alphabet to a finite stabilization semigroup. Let us denote by $\hat{\alpha}(x)$ the output of any factorization table of x. This definition is correct thanks to Proposition 12.4. We call the mapping $\hat{\alpha} \colon \widehat{A^+} \to S$ the *mapping induced by* α.

Lemma 12.6. *The mapping $\hat{\alpha} \colon \widehat{A^+} \to S$ is a homomorphism of $\langle \cdot, \# \rangle$-algebras which extends α.*

The idea is simply that out of two factorization tables of height h, for two profinite words x and y, we can naturally construct a factorization table of height $h+1$ for $x \cdot y$, and similarly for x^ω. Below we describe the construction in detail.

Proof. It is clear that $\hat{\alpha}$ extends α, since for any letter $a \in A$, a has a very simple factorization table of height 1 and output $\alpha(a)$. Therefore, $\hat{\alpha}(a) = \alpha(a)$.

We now prove that $\hat{\alpha}$ preserves multiplication. Let x, y have a factorization table be two profinite words with factorization tables $\{p_n^d\}_{\delta \leq d \leq n}$ and $\{q_n^d\}_{\delta \leq d \leq n}$, respectively. We may assume that both tables have a height bounded by some number h. Let $x_1, x_2, \ldots$ be the base of the first factorization table and $y_1, y_2, \ldots$ be the base of the second factorization table, and let s, t be their respective outputs. Then $s = \hat{\alpha}(x)$ and $t = \hat{\alpha}(y)$.

For each $\delta \leq d \leq n$, let r_n^d be the factorization tree obtained from p_n^d and q_n^d by joining both trees via one binary node at the root. Then r_n^d is a factorization tree over the word $x_n \cdot y_n$, of height $h+1$, degree d and with output $s \cdot t$. Moreover, by continuity of multiplication, the sequence $(x_n \cdot y_n)_n$ converges to $x \cdot y$. Therefore, $\{r_n^d\}_{\delta \leq d \leq n}$ is a factorization table of the profinite word $x \cdot y$, and its output is $s \cdot t$. This shows that $\hat{\alpha}(x \cdot y) = \hat{\alpha}(x) \cdot \hat{\alpha}(y)$.

The proof for stabilization is similar. The only novelty here is that we must also control the degrees of the constructed factorization trees. Let x be a profinite word, $\{p_n^d\}_{\delta \leq d \leq n}$ its factorization table of height h, output $\hat{\alpha}(x) = s$ and base $x_1, x_2, \ldots$. We first consider the case when $s = e$ is an idempotent.

We construct a factorization table $\{r_n^d\}_{\delta \leq d \leq n}$ of height $h+1$ and base

$$(x_1)^{1!}, (x_2)^{2!}, (x_3)^{3!}, \ldots \tag{2}$$

in the natural way. Namely, for any $\delta \leq d \leq n$, r_n^d is the tree obtained by joining $n!$ copies of the tree p_n^d by adding one idempotent node at the root. Since $\delta \leq n!$, we obtain a legitimate factorization tree of $(x_n)^{n!}$ of height $h+1$, degree d and output $e^\#$. Moreover, the base (2) converges to x^ω. Therefore, we have shown that if $\hat{\alpha}(x) = e$ is idempotent, then $\hat{\alpha}(x^\omega) = e^\#$.

Now consider the general case, when $\hat{\alpha}(x) = s$ is not necessarily idempotent. Let N be the idempotent power of the semigroup S. Then $\hat{\alpha}(x)^N$ is an idempotent and is equal to $\hat{\alpha}(x^N)$ by what we have proven for multiplication, so:

$$\hat{\alpha}(x^\omega) = \hat{\alpha}\big((x^N)^\omega\big) = \hat{\alpha}(x^N)^\# = \big(\hat{\alpha}(x)^N\big)^\# = \hat{\alpha}(x)^\#. \qquad \square$$

12.1.1 Relationship with B- and S-regular languages

We will see that an inverse image of a closed set under an induced mapping is an S-regular language, and that an inverse image of an open set under an induced mapping is a B-regular language. A similar conversion – to regular cost functions, and not relying on factorization tables, but on δ-factorization trees – was suggested in [Col10a].

For each element $s \in S$ and number $d \in \mathbb{N}$, we define the set $L_{\downarrow s,d} \subseteq A^+$ as the regular language consisting of all those words $w \in A^+$, for which there exists a factorization tree of height $\|S\|$ and degree d, whose output t satisfies $t \leq s$. Next, let us define the set $L_{\downarrow s} \subseteq \widehat{A^+}$ by:

$$L_{\downarrow s} = \bigcap_{d \in \mathbb{N}} \overline{L_{\downarrow s,d}}.$$

Dually, for each $s \in S$ and $d \in \mathbb{N}$, let $L_{\uparrow s,d} \subseteq A^+$ be the regular language consisting of all those words $w \in A^+$, for which there exists a factorization tree of height $\|S\|$ and degree d, whose output t satisfies $t \geq s$. Let $L_{\uparrow s} \subseteq \widehat{A^+}$ be defined by:

$$L_{\uparrow s} = \bigcup_{d \in \mathbb{N}} \overline{L_{\uparrow s,d}}.$$

Remark 12.2. It follows similarly as in the proof of Lemma 12.5 that for each $s \in S$, the language $L_{\downarrow s}$ is S-regular and the language $L_{\uparrow s}$ is B-regular. We only sketch how to derive this conclusion for $L_{\downarrow s}$. For any $h \geq 1$, we define the language $L^h_{\downarrow s,d}$ analogously to $L_{\downarrow s,d}$, but we require the height to be equal to h. Then, clearly, an analogous formula to (1) holds for $L^h_{\downarrow s,d}$, but below each of the unions, the equalities "$\ldots = s$" should be changed to inequalities "$\ldots \leq s$". Moreover, we can safely replace the exponent $<d$ by the Kleene star $*$. This modification does not add new elements, because if $w \in (L^{h-1}_{\downarrow e,d})^*$ and $e \leq s$, then either $w \in (L^{h-1}_{\downarrow e,d})^{<d}$, or otherwise $w \in (L^{h-1}_{\downarrow e,d})^{\geq d}$, and moreover $e^\# \leq s$ since $e^\# \leq e \leq s$. Using this observation, we see that $L_{\downarrow s,d}$ can be defined by a single regular expression, using the exponent $\geq d$, but not the exponent $<d$. Therefore, it follows that $L_{\downarrow s}$ is defined by the analogous S-regular expression, in which the exponent $\geq d$ is replaced by ∞.

For showing that the language $L_{\uparrow s}$ is defined by a B-regular expression, we proceed completely dually (i.e. we put the inequalities "$\ldots \geq s$" under the unions in the formula (1), and we replace the exponent $\geq d$ by the Kleene star $*$, and finally the exponent $<d$ by $<\infty$).

The following proposition describes the relationship between the mapping $\hat{\alpha}$ and the languages $L_{\downarrow s}$ and $L_{\uparrow s}$.

Proposition 12.7. *Let $\alpha\colon A \to S$ be a mapping from the alphabet A to a finite stabilization semigroup S and let $\hat{\alpha}\colon \widehat{A^+} \to S$ be the induced mapping. Then, for any $s \in S$,*

$$\hat{\alpha}^{-1}(\downarrow s) = L_{\downarrow s}, \tag{1}$$

$$\hat{\alpha}^{-1}(\uparrow s) = L_{\uparrow s}, \tag{2}$$

$$\hat{\alpha}^{-1}(\{s\}) = L_{\downarrow s} \cap L_{\uparrow s}. \tag{3}$$

As a consequence, for any set $F \subseteq S$,

- *If F is closed, then $\hat{\alpha}^{-1}(F)$ is an S-regular language,*
- *If F is open, then $\hat{\alpha}^{-1}(F)$ is a B-regular language,*
- *In general, $\hat{\alpha}^{-1}(F)$ is a positive Boolean combination of B- and S-regular languages.*

Proof. First we show how the equations (1) and (2) imply the remaining statements of the proposition. The equation (3) follows immediately from the previous two equations, since $\{s\} = (\downarrow s) \cap (\uparrow s)$ and inverse images commute with intersection.

Now, let F be a closed subset of S. Then

$$F = \bigcup_{s \in F} \downarrow s,$$

so

$$\hat{\alpha}^{-1}(F) = \bigcup_{s \in F} \hat{\alpha}^{-1}(\downarrow s) = \bigcup_{s \in F} L_{\downarrow s}.$$

Since the language $L_{\downarrow s}$ is S-regular and S-regular languages are closed under union, $\hat{\alpha}^{-1}(F)$ is also S-regular.

The proof for open sets $F \subseteq S$ is similar, with the only difference that the arrow $\downarrow$ should be replaced by the arrow $\uparrow$. For an arbitrary $F \subseteq S$, $\hat{\alpha}^{-1}(F)$ is a union of sets of the form (3).

It remains to prove the equations (1) and (2). The equation (1) is equivalent to the following

Claim 1. *Let $x \in \widehat{A^+}$ and $s \in S$. Then,*

$$\hat{\alpha}(x) \le s \iff x \in L_{\downarrow s}. \tag{4}$$

Let $\mathcal{F}$ be a small factorization table of x. Let $x_1, x_2, \ldots$ be the base, δ the degree and $u \in S$ the output of $\mathcal{F}$.

Assume that $\hat{\alpha}(x) = u \le s$. We prove the left-to-right implication of (4), by showing that $x \in \overline{L_{\downarrow s,d}}$ for any given $d \in \mathbb{N}$. First, we show that for any large enough n,

$$x_n \in L_{\downarrow u,d}.$$

We consider two cases. If $d \ge \delta$, then by definition of a factorization table with output u, for any $n \ge d$, x_n has a factorization tree of degree d and output u, so $x_n \in L_{\downarrow u,d}$. If $d < \delta$ and $n \ge \delta$, then x_n has a small factorization tree of degree 1 and some output r, and a small factorization tree of degree δ and output u. We can safely assume that δ is big enough so that we can apply Lemma 12.3 to conclude that $r \leq u$ (we assume that $\delta \geq \delta_2$, where $\delta_1 = 1$). Therefore, $x_n \in L_{\downarrow u,d}$.

We have thus shown that for any sufficiently large n,

$$x_n \in L_{\downarrow u,d} \subseteq L_{\downarrow s,d}.$$

Since $x = \lim_{n\to\infty} x_n$, it follows that $x \in \overline{L_{\downarrow s,d}}$. Since $d \in \mathbb{N}$ is arbitrary,

$$x \in \bigcap_{d\in\mathbb{N}} \overline{L_{\downarrow s,d}} = L_{\downarrow s},$$

proving the left-to-right implication of (4).

Now we will prove the right-to-left implication of (4). Assume that $x \in L_{\downarrow s}$, i.e. for all $d \in \mathbb{N}$, $x \in \overline{L_{\downarrow s,d}}$. Let $\delta_1 = \delta$ and δ_2 be the number granted by Lemma 12.3. Fix any number $d \geq \delta_2$. Then, since $\overline{L_{\downarrow s,d}}$ is an open set containing x, there is an index $n \geq d$ such that $x_n \in L_{\downarrow s,d}$. Hence, the word x_n has a small factorization tree of degree d and output t which satisfies $t \leq s$. Moreover, since x_n is in the base of the factorization table $\mathcal{F}$, it follows that x_n also has a small factorization tree of degree δ_1 and output u. From Lemma 12.3 we deduce that $u \leq t$. Together with $t \leq s$ this proves that $\hat{\alpha}(x) \leq s$. This finishes the proof of the claim.

The equation (2) of Proposition 12.7, is equivalent to the following

Claim 2. *Let $x \in \widehat{A^+}$ and $s \in S$. Then,*

$$\hat{\alpha}(x) \geq s \iff x \in L_{\uparrow s}. \tag{5}$$

Assume that $\hat{\alpha}(x) = u \geq s$. If $\mathcal{F}$ is a factorization table over x with base $x_1, x_2, \ldots$, output $u \geq s$ and degree δ, it follows that for all n, $x_n \in L_{\uparrow u,\delta} \subseteq L_{\uparrow s,\delta}$. Since $\overline{L_{\uparrow s,\delta}}$ is an open set, we conclude that $x \in \overline{L_{\uparrow s,\delta}} \subseteq L_{\uparrow s}$.

To demonstrate the implication in the other direction, assume that $x \in L_{\uparrow s}$. By definition, this implies that there exists a number d such that $x \in \overline{L_{\uparrow s,d}}$, and, consequently, that almost all elements $x_1, x_2, \ldots$ belong to the regular language $L_{\uparrow s,d}$. In particular, there is an index $n \geq \delta_2$, where δ_2 is the number from Lemma 12.3 for $\delta_1 = d$, such that x_n has a small factorization tree of degree d and with output $t \geq s$. But since $\delta_2 \leq n$ and x_n is in the base of the table $\mathcal{F}$, the word x_n also has a small factorization tree of degree δ_2 and output u. By Lemma 12.3, we have that $u \geq t$. Altogether, this proves that $\hat{\alpha}(x) = u \geq t \geq s$, which is what we needed.

This ends the proof of Proposition 12.7. □

12.2 Canonicity

In this section, we prove Proposition 12.1 which gives different characterizations of canonical mappings, and we show that the induced mapping is a canonical mapping and vice-versa. Observe that the induced mapping is continuous, since by Proposition 12.7, the inverse image of a closed set is S-regular, so in particular, closed. We prove the following.

Lemma 12.8. *Let $\hat{\alpha}\colon \widehat{A^+} \to S$ be any continuous homomorphism to a finite stabilization semigroup. Then, the following conditions are equivalent.*

1. *$\hat{\alpha}$ is the mapping induced by its restriction α to A,*
2. *For any closed subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is an S-regular language,*
3. *For any open subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is a B-regular language,*
4. *For any closed subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is accepted by an S-automaton,*
5. *For any open subset $F \subseteq S$, the inverse image $\hat{\alpha}^{-1}(F)$ is accepted by a B-automaton,*
6. *For any closed subset $F \subseteq S$,*
$$\hat{\alpha}^{-1}(F) = \overline{\hat{\alpha}^{-1}(F) \cap A^{\langle\, \cdot\, ,\omega\rangle}},$$
7. *$\hat{\alpha}$ is the "largest", in the following sense. For any continuous homomorphism $\beta\colon \widehat{A^+} \to S$, if β and $\hat{\alpha}$ agree over A, then for all $x \in \widehat{A^+}$, $\beta(x) \leq \hat{\alpha}(x)$ in the specialization preorder over S.*

Moreover, if a mapping $\hat{\alpha}$ satisfies any of the above conditions, then its image is equal to the $\langle\, \cdot\, ,\#\rangle$-subalgebra of S generated by A.

Proof. The implications *1⇒2* and *1⇒3* follow from Proposition 12.7. The implications *2⇒4* and *3⇒5* follow from Proposition 10.1. The implications *4⇒6* and *5⇒6* follow from Propositions 10.10 and 11.13, respectively.

We now prove the implication *6⇒7*. Let $\beta\colon \widehat{A^+} \to S$ be any continuous homomorphism, such that β and $\hat{\alpha}$ agree over A. We will prove that for all $s \in S$,

$$\hat{\alpha}^{-1}(\downarrow s) \subseteq \beta^{-1}(\downarrow s), \tag{6}$$

which in turn implies that for all $x \in \widehat{A^+}$, $\beta(x) \leq \hat{\alpha}(x)$.

Since both $\hat{\alpha}$ and β are homomorphisms which coincide over A, it follows that they must coincide over the set $A^{\langle\, \cdot\, ,\omega\rangle}$. Therefore,

$$\hat{\alpha}^{-1}(\downarrow s) \cap A^{\langle\, \cdot\, ,\omega\rangle} = \beta^{-1}(\downarrow s) \cap A^{\langle\, \cdot\, ,\omega\rangle} \subseteq \beta^{-1}(\downarrow s).$$

We apply the closure to the extreme parts of the above inclusion. Since β is continuous, $\beta^{-1}(\downarrow s)$ is closed, so we get:

$$\overline{\hat{\alpha}^{-1}(\downarrow s) \cap A^{\langle\, \cdot\, ,\omega\rangle}} \subseteq \beta^{-1}(\downarrow s).$$

By the assumption on $\hat{\alpha}$, the left-hand side of the above inclusion is precisely $\hat{\alpha}^{-1}(\downarrow s)$. This proves the inclusion (6), ending the proof of the implication *6⇒7*.

Observe that the implication *7⇒1* follows. Indeed, assume that $\hat{\alpha}$ is the "largest", and let α be its restriction to A. Then, the mapping induced by α is also the "largest", by the implication *1⇒7* which follows from what was already shown. Therefore, $\hat{\alpha}$ is the mapping induced by α.

It remains to prove the last statement of the lemma, talking about the image of a canonical mapping. We may assume that $\hat{\alpha}$ is an induced mapping. We show that if $s = \hat{\alpha}(x)$ for some $x \in \widehat{A^+}$, then there exists a term $y \in A^{\langle \cdot, \omega \rangle}$ such that $s = \tilde{\alpha}(y)$. Since s is the output of a factorization table of $\hat{\alpha}(x)$, it follows that there exists a factorization tree over some word in A^+ whose output is s. This implies that s can be generated by a term which uses multiplication, stabilization and elements of $\alpha(A)$ as constants. We can correspondingly construct an element $y \in A^{\langle \cdot, \omega \rangle}$, using multiplication, the ω-power and elements of A as constants. By homomorphicity of $\tilde{\alpha}$ it follows that $\tilde{\alpha}(y) = s$, and that s is in the $\langle \cdot , \# \rangle$-subalgebra of S generated by $\hat{\alpha}(A)$. This ends the proof of Lemma 12.8. □

Lemma 12.8 immediately implies Proposition 12.1, and also that the induced mapping $\hat{\alpha}$ is canonical, proving Theorem 8.1.

Corollary 12.9. *Let $\hat{\alpha}\colon \widehat{A^+} \to S$ and $\hat{\beta}\colon \widehat{A^+} \to T$ be two canonical mappings. Then their Cartesian product*

$$\hat{\alpha} \times \hat{\beta} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad S \times T$$

is also a canonical mapping.

Proof. This follows from the condition 7 of the lemma, since a product of largest mappings is again a largest mapping, as is easy to verify. □

CHAPTER 13

Main results

In this chapter, we prove the main results of this dissertation, concerning equivalences of many of the various classes of languages of profinite words described so far. Many of these equivalences can be carried out effectively, as demonstrated in Proposition 13.1. The main theorem, Theorem 8.2, furthermore gives descriptions of these classes in terms of abstract properties, such as a finite index property. The results of this chapter follow easily from the theory developed in the previous chapters.

Effective translations Most of the classes of languages of profinite words which were considered – B/S-regular languages, languages accepted by B/S-automata, languages defined by formulas of MSO+fin$^+$ or MSO+inf$^+$– can be given effective descriptions – namely, in terms of regular expressions, automata, and formulas. Proposition 13.1 below states that there are effective translations among those classes. In particular, the proposition implies that B-regular languages are equivalent to B-automata, but most importantly, it implies the difficult complementation result stated in the overview, that B-regular languages are precisely complements of S-regular languages, and that the translation is effective. Note that this implies decidability of the limitedness problem for B-automata – in order to test if a B-automaton is limited, we compute the S-regular expression describing its complement, and check for its emptiness. Testing for emptiness of regular expressions is trivial.

An important means of effectively specifying a language of profinite words is by use of canonical homomorphisms, as we now describe. We say that a language L is *specified* by a canonical mapping $\hat{\alpha}\colon \widehat{A^+} \to S$, if we are given a description of a finite stabilization semigroup S and its subset F, and of a mapping $\alpha\colon A \to S$, such that $L = \hat{\alpha}^{-1}(F)$ for the induced mapping $\hat{\alpha}$. This description is unambiguous, thanks to Theorem 8.1 and Proposition 12.1. Note that languages specified by canonical mappings are closed under Boolean combinations. Closure under unions and intersections follows from Corollary 12.9, and closure under negation follows from replacing the recognizing set $F \subseteq S$ by its complement $S - F$.

Proposition 13.1. *The following classes of languages in $\widehat{A^+}$ coincide, and all translations are effective.*

1. *Languages defined by B-regular expressions,*

2. *Languages accepted by B-automata,*

3. *Languages specified by canonical mappings, as inverse images of open subsets of stabilization semigroups,*

4. *Complements of languages defined by S-regular expressions,*

5. *Complements of languages accepted by S-automata,*

6. *Languages specified by canonical mappings, as inverse images of closed subsets of stabilization semigroups.*

Moreover, B-regular languages effectively correspond to languages definable in MSO+fin$^+$ *and S-regular languages effectively correspond to languages definable in* MSO+inf$^+$.

Proof. We describe the translations in a circular fashion.

1 ⇒ *2*. This translation follows from Proposition 10.1.

2 ⇒ *3*. Let $\mathcal{A}$ be a B-automaton. Construct the homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ to the finite transformation semigroup S of the B-automaton $\mathcal{A}$, as described in Proposition 11.11. Moreover, there is an open set $F \subseteq S$ such that $\hat{\alpha}^{-1}(F) = L(\mathcal{A})$. It is possible to show that $\hat{\alpha}$ is a canonical mapping (one needs to show that the inverse image of any open set is accepted by a B-automaton). We proceed differently. Let $\hat{\beta}$ be the mapping induced by the restriction of $\hat{\alpha}$ to A. Then, $\hat{\beta}$ and $\hat{\alpha}$ agree over $A^{\langle\,\cdot\,,\omega\rangle}$. In particular, for $K = S - F$,

$$\big(\widehat{A^+} - L(\mathcal{A})\big) \cap A^{\langle\,\cdot\,,\omega\rangle} = \hat{\alpha}^{-1}(K) \cap A^{\langle\,\cdot\,,\omega\rangle} = \hat{\beta}^{-1}(K) \cap A^{\langle\,\cdot\,,\omega\rangle}.$$

The closure of the left-hand side set above is precisely $\widehat{A^+} - L(\mathcal{A})$, by Proposition 11.11. The closure of the right-hand side set above is the set $\hat{\beta}^{-1}(K)$ by Proposition 12.1. It follows that $\hat{\beta}^{-1}(F) = L(\mathcal{A})$, so $\hat{\beta}$ is a canonical mapping recognizing $L(\mathcal{A})$.

3 ⇒ *4*. It follows from Proposition 12.1, that if $\hat{\alpha}\colon \widehat{A^+} \to S$ is a canonical mapping and $F \subseteq S$ is an open set, then $\hat{\alpha}^{-1}(S - F)$ is an S-regular language.

Therefore, we have shown the translations *1*⇒*2*⇒*3*⇒*4*. Completely dually, we can follow the translations *4*⇒*5*⇒*6*⇒*1*. We simply repeat the above translations, by swapping "B" with "S" and "closed" with "open", and consider the set $K = F$ instead of the set $K = S - F$ in the translation *2*⇒*3*.

Finally, we prove the equivalence with logic. A B-automaton can be effectively translated into a formula of MSO+fin$^+$ and an S-automaton can be effectively translated into a formula of MSO+inf$^+$, as described in Proposition 10.4.

Conversely, we prove by induction on the structure of a formula φ of MSO+inf that if φ uses the predicate inf only positively, then it defines an S-regular language, and if φ uses the predicate inf only negatively, then it defines a B-regular language. Formally, the considered formula φ can have free variables, so we define $L(\varphi)$ as the language of valuations ν such that $x \models \varphi[\nu]$ for some profinite word x, as defined in Section 9.4.

The inductive base is trivial, since all the predicates define clopen languages, except for the predicate inf, which defines an S-regular language. If φ is a disjunction or conjunction of two formulas, then we use the fact that both B-regular languages and S-regular languages are closed under disjunctions and conjunctions. If φ is a negation of a formula, then we use the result that complements of B-regular languages are S-regular languages and vice-versa. Finally, if φ is obtained from ψ by using an existential quantifier, then we use the fact that B-automata and S-automata are closed under projections, as usual when dealing with nondeterministic automata. □

Corollary 13.2. *B-regular languages and S-regular languages are closed under intersections. Any Boolean combination of B- and S-regular languages is equivalent to a language specified by a canonical mapping.*

Proposition 13.3. *Emptiness of languages specified by canonical mappings can be effectively tested.*

Proof. Let $L = \hat{\alpha}^{-1}(F)$, where $\hat{\alpha}$ is the mapping induced by a mapping $\alpha\colon A \to S$ to a finite T_0 stabilization semigroup and F is any subset of S. To determine emptiness of L, it is sufficient to test if the image of $\hat{\alpha}$ intersects F.

By Proposition 12.1, the image of $\hat{\alpha}$ is the $\langle\,\cdot\,,\#\rangle$-subalgebra of S generated by all elements of the form $\alpha(a)$, for $a \in A$. Since A and S are finite, this subalgebra can be effectively computed by a simple fixed-point algorithm. Then, L is nonempty iff the subalgebra intersects F. □

Corollary 13.4. *Emptiness of a Boolean combination of* MSO+inf^{+} *formulas can be efficiently tested.*

Main theorem

We recall and prove the main theorem stated in the overview. It extends the statement of Proposition 13.1 by including the descriptions in terms of a finite index property. Also, the conditions *4* and *4′* below are weaker than the corresponding conditions *3* and *6* of Proposition 13.1, as we shall explain.

Theorem 8.2. *Let $L \subseteq \widehat{A^+}$. The following conditions are equivalent:*

1. *L is defined by an S-regular expression,*
2. *$L = L(\mathcal{A})$ for some S-automaton $\mathcal{A}$,*
3. *L is definable in* MSO+inf^{+},
4. *L is closed and is recognized by some canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ to some finite T_0 stabilization semigroup,*
5. *$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}$ and its $\langle\,\cdot\,,\#\rangle$-syntactic congruence has finite index.*

Dually, the following conditions are equivalent:

1′. *L is defined by a B-regular expression,*

2′. *$L = L(\mathcal{A})$ for some B-automaton $\mathcal{A}$,*

3′. *L is definable in* MSO+fin^{+}, *4′. L is open and is recognized by some canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ to some finite T_0 stabilization semigroup,*

5′. The complement $\widehat{A^+} - L$ satisfies either of the conditions 4-3.

Moreover, L is recognized by a canonical mapping to a finite T_0 stabilization semigroup if and only if it is a Boolean combination of languages satisfying either of the above conditions.

Proof. Let us introduce an additional condition to the formulation of the theorem.

6. There is a canonical mapping $\hat{\alpha}\colon A \to S$ to a finite stabilization semigroup, and a closed set $F \subseteq S$, such that $L = \hat{\alpha}^{-1}(F)$.

The above condition corresponds precisely to the third condition in Proposition 13.1. Notice the subtle difference between the conditions *4* and *6*: the condition *6* is stronger, since it requires that L is an inverse image of a closed set, whereas the condition *4* only requires that L is closed itself. In particular, the implication *6⇒4* is immediate. In the other way, we show the implications *4⇒5⇒6*.

4 ⇒ 5. Assume that L is closed and is recognized by some canonical mapping $\hat{\alpha}\colon \widehat{A^+} \to S$ to a finite stabilization semigroup, i.e. that for some $F \subseteq S$,

$$\hat{\alpha}^{-1}(F) = L.$$

However, we do not know whether the set F is closed in S; we only know that $L = \hat{\alpha}^{-1}(F)$ is closed and, by Proposition 11.2, it has a finite index. We prove that

$$L = \overline{L \cap A^{\langle\,\cdot\,,\omega\rangle}}. \tag{1}$$

We show that the class of closed languages recognized by canonical mappings satisfies the assumptions of the Lemma 8.3. Clearly, this class consists of closed sets, contains clopen sets and is closed under intersections, by Corollary 12.9. Moreover, any language recognized by a canonical mapping contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$ – this follows from the last part of Proposition 12.1. Therefore, by Lemma 8.3, any closed language recognized by a canonical mapping satisfies the equation (1).

5 ⇒ 6. Let $\alpha_L\colon \widehat{A^+} \to S_L$ be the syntactic homomorphism. By Theorem 11.3 and Proposition 11.9, S_L is a finite stabilization semigroup and α_L is a continuous homomorphism recognizing L. Let $F \subseteq S_L$ be such that $L = \alpha_L^{-1}(F)$. Let $\hat{\alpha}\colon \widehat{A^+} \to S_L$ be the mapping induced by the restriction of α to A . Similarly as in the proof of Proposition 13.1, we avoid proving that $\hat{\alpha} = \alpha_L$, but we show that $\hat{\alpha}^{-1}(F) = L$. Indeed, since $\hat{\alpha}$ and α_L agree over $A^{\langle\,\cdot\,,\omega\rangle}$, we have that

$$L \cap A^{\langle\,\cdot\,,\omega\rangle} = \alpha_L^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle} = \hat{\alpha}^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle}.$$

By assumption, the closure of the left-hand-side set above is L, and by Proposition 12.1, the closure of the right-hand-side set above is $\hat{\alpha}^{-1}(F)$. This proves *6*.

The equivalences *6*$\Leftrightarrow$*1*$\Leftrightarrow$*2*$\Leftrightarrow$*3* follow from Proposition 13.1. This proves the equivalence of all the first five conditions in the statement of the Theorem. The equivalence of the remaining five conditions follows by complementation and from Proposition 13.1.

The last part of the theorem states that L is recognizable by a canonical mapping if and only if it is a Boolean combination of B-regular and S-regular languages. The right-to-left part of this implication is stated in Corollary 13.2. The left-to-right implication follows from Proposition 12.7. □

CHAPTER 14

Conclusion

Summary We have developed a framework which is suitable for investigating various decision problems for B-automata and S-automata, such as limitedness. This framework extends, in many aspects, the framework of regular languages. The key idea is to switch from the realms of finite words to profinite words, and develop a suitable algebraic theory there.

We end this thesis with an overview from a broader perspective, indicating similarities to other existing theories. Many statements are only sketchy and left without proof. In Section 14.1 we give an overview of the theory from a categorical-like perspective. Our framework is very much influenced by the framework of T. Colcombet. We draw a comparison between the two theories in Section 14.2. There are also many resemblances with the theory of ω-regular languages. We describe those similarities in Section 14.3. A natural question to ask is whether it would make more sense to consider not languages of profinite words, but only subsets of the set $A^{\langle\, \cdot\, ,\omega\rangle}$. We describe some difficulties in this approach in Section 14.3.3. Finally, we end with some concluding questions.

14.1 Categorical description

Recall that by Proposition 12.1, canonical mappings are the same as induced mappings. We mention a property which completes the picture by linking quotient mappings with canonical mappings.

Proposition 14.1. *Let $L \subseteq \widehat{A^+}$ be an S-regular language. Then, if $S_L = \widehat{A^+}/\simeq_L$ is considered with the order topology, the quotient mapping $\alpha_L \colon \widehat{A^+} \to S_L$ becomes a canonical homomorphism.*

Remark 14.1. It would be nice to know whether S_L equipped with the quotient topology also has the property stated in the proposition. However, I was unable to answer that question, i.e. to determine whether the syntactic homomorphism of a B- or S-regular language is a canonical mapping, when the quotient topology is considered.

We summarize the general picture from a categorical perspective. Our objects are the free profinite semigroup and finite T_0 stabilization semigroups. Morphisms from $\widehat{A^+}$ to finite stabilization semigroups are the canonical mappings, and morphisms between finite stabilization semigroups are continuous homomorphisms. Composing a canonical mapping with a continuous homomorphism of finite stabilization semigroups yields a canonical mapping, so morphisms are closed under composition. Inverse images of closed (respectively, open) subsets under canonical mappings are S-regular languages (respectively, B-regular languages). Moreover, any such language L has a minimal recognizer (i.e. a terminal recognizing object), namely the quotient semigroup S_L equipped with the order topology.

14.2 Comparison with the framework of Colcombet

Recently, T. Colcombet [Col09, Col10b] has created his theory of regular cost functions. Many notions of our theory have their roots in the theory of Colcombet.

Below we give a quick overview of the similarities and differences between the two theories.

Equivalence relations In the low-level perspective, the differences between the two theories flow from the fact that T. Colcombet deals with arbitrary infinite sequences of finite words, while we deal with infinite sequences of words, which are moreover convergent in the profinite topology. Dealing with convergent sequences has the advantage that one can consider the equivalence classes of convergent sequences, and treat them as single elements (i.e. profinite words). Therefore, from the very beginning, we can forget about equivalence classes and think of profinite words. However, arbitrary sequences of finite words do not admit a suitable equivalence relation. As a result, instead of factoring by an equivalence relation at the outset, T. Colcombet needs to deal with equivalence classes of various types of objects (cost functions, compatible mappings) all throughout his theory.

Cost functions – profinite languages As mentioned in Section 9.2.1, there is a correspondence between the class of closed languages of profinite words and the class of continuous cost functions, i.e. cost functions which have some uniformly continuous representative. Regular cost functions are covered by this correspondence, as they are continuous cost functions.

In general, however, cost functions and languages of profinite words are incomparable. For instance, the cost function which maps a non-regular language to 0 and its complement to ω has no analogue in the class of languages of profinite words. Another example is the following function f which maps a word $w \in \{a,b\}^+$ to the largest number n such that w has n distinct infixes ba^nb.

The other way around, the correspondence only relates closed languages of profinite words with cost functions, and cannot be extended to other languages. As a consequence of this last problem, in the theory of cost functions, B-automata and S-automata recognize the same classes of objects, while in our theory, they recognize either open languages, or closed languages. While for open languages, one can easily define the corresponding cost function by considering their complements instead, some languages do not seem to have a corresponding cost function. Consider for instance the language $L \subseteq \widehat{\{a,b\}^+}$:

$$L = \{x : \textit{ x has an infix } ba^nb \textit{ for infinitely many } n \in \mathbb{N}\}, \tag{1}$$

which is an infinite intersection of B-regular languages of the form

$$L_k = \{x : \ x \textit{ has an infix of the form } ba^n b \textit{ for some } n \geq k\}.$$

Even though the set L seems to be somehow related with the function f considered above, the correspondence which we have been considering cannot capture this.

Compatible mappings – homomorphisms As noticed in Section 11.2.1, there is a straightforward correspondence between the stabilization semigroups considered in this thesis, and the stabilization semigroups defined by T. Colcombet. Moreover, in both theories, a mapping $\alpha\colon A \to S$ from a finite alphabet to a finite stabilization semigroup induces some sort of a canonical mapping. In our theory, this mapping is the canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$. In the theory of T. Colcombet, the induced mapping is a *compatible mapping*. The definition of a compatible mapping is quite technical and long; however, it is easily visible, that compatible mappings share many similarities with homomorphisms. One of the technical difficulties when dealing with compatible mappings is that the axioms are not stated in terms of equalities (like associativity), but in terms of an equivalence of functions.

Cost MSO – MSO+inf Colcombet defines an extension of MSO logic over finite words, called *cost* MSO by a unary predicate $|X| < N$, where X is a free second-order variable, and N is a formal symbol. In cost MSO, the predicate $|X| < N$ is required to appear either positively A closed formula φ of cost MSO, in which the predicate $|X| < N$ appears positively, defines a function $f_\varphi\colon A^+ \to \overline{\mathbb{N}}$ via

$$f_\varphi(w) = \min\{n \in \mathbb{N} : \quad \varphi_{N:=n} \textit{ holds in } w\},$$

where $\varphi_{N:=n}$ is now interpreted as a normal formula of MSO with counting capabilities, by replacing the symbol N by the number n. Similarly, if the predicate $|X| < N$ appears only negatively, then the semantic of f_φ is defined dually, by replacing min with max. As in the case of B- or S-automata, in the theory of Colcombet, the object actually defined by a formula is not the function itself, but the cost function it represents.

There is a straightforward correspondence between cost MSO and the fragments MSO+fin$^+$ and MSO+inf$^+$: we simply replace the predicate $|X| < N$ by the predicate $\mathrm{fin}(X)$. It is easy to see that a formula φ of cost MSO with positive occurrence of the predicate $|X| < N$ defines a cost function which is bounded if and only if the corresponding formula of MSO+fin$^+$ is satisfied for every $x \in \widehat{A^+}$.

The advantage of cost MSO over MSO+inf is its simplicity, as it is defined in terms of usual MSO logic. Therefore, the definition makes sense for other classes of models, for instance graphs, while the semantic of MSO+inf is not immediately obvious for other classes of profinite objects.

On the other hand, the advantage of MSO+inf over MSO is that *any* formula of MSO+inf has a semantic, and not only those formulas, for which the predicate inf appears positively or negatively.

14.3 Two connections with the theory of ω-regular languages

14.3.1 Connection via Wilke algebras

Relatively free pro-V semigroups A *pre-variety* of semigroups is a family $\mathbf{V}$ of finite semigroups which is closed under homomorphic images, taking subsemigroups, and finite Cartesian products. Let A be a finite alphabet. One defines the *relatively free pro-*$\mathbf{V}$ *semigroup*, which we denote $\widehat{A^+_{\mathbf{V}}}$. There are several equivalent definitions of $\widehat{A^+_{\mathbf{V}}}$. One definition is as the inverse limit of the projective system of all A-generated semigroups in $\mathbf{V}$. Another, equivalent definition, is analogous to the definition of $\widehat{A^+}$ considered in the preliminaries, as the sets of convergent sequences with respect to a suitable metric, which takes into account only distinguishability by semigroups in $\mathbf{V}$. The last equivalent definition is by taking the quotient of $\widehat{A^+}$ by the equivalence $\sim_{\mathbf{V}}$, such that $x \sim_{\mathbf{V}} y$ if and only if the image of x is equal to the image of y under any homomorphism from $\widehat{A^+}$ to a semigroup from $\mathbf{V}$.

The pre-variety of definite semigroups Let $\mathbf{DF}$ denote the pre-variety of definite semigroups, i.e. finite semigroups which satisfy the equation

$$s^\omega \cdot t = s^\omega. \tag{1}$$

This pre-variety of semigroups corresponds to the variety of languages for which membership of any given word is determined by some prefix of length depending only on the language.

It is not difficult to check that for two finite words $v, w \in A^+$, the smallest semigroup from $\mathbf{DF}$ which distinguishes v from w has size

$$r_{\mathbf{DF}}(v,w) \quad \overset{def}{=} \quad \max\{|u| : u \text{ is a common prefix of } v \text{ and } w\}.$$

Therefore, for the "distinguishability" metric over A^+

$$d_{\mathbf{DF}}(v,w) \quad \overset{def}{=} \quad 2^{-r(v,w)},$$

the completion of A^+ is homeomorphic to the space $A^\infty = A^+ \cup A^\omega$ of finite and ω-words, with the Cantor topology. We may thus identify $\widehat{A^+_{\mathbf{DF}}}$ with A^∞. The ω-power in $\widehat{A^+_{\mathbf{DF}}}$ can be easily seen to map a finite word $w \in A^+$ to the infinite word $w \cdot w \cdot w \cdots$, and an infinite word $w \in A^\omega$ to itself. Therefore, the set $A^{\langle \cdot, \omega\rangle} \subseteq \widehat{A^+_{\mathbf{DF}}}$ is precisely the set of finite or ultimately periodic words. Furthermore, any term $\tau \in Terms(\widehat{A^+_{\mathbf{DF}}}, \cdot, \#)$ is equivalent to one of the form

$$\tau(x) = u \cdot x \cdot w,$$

where $u \in A^*$ and $w \in A^\infty \cup \{\varepsilon\}$, or to one of the form

$$\tau(x) = u \cdot (x \cdot w)^\omega,$$

where $u \in A^*$ and $w \in A^*$. As a result, it can be easily seen that the relation $\simeq_L$ coincides with the Arnold congruence induced by L.

Wilke algebras Recall that a *Wilke algebra* (see e.g. [Wil93, PP04]) is a two-sorted finite structure (S_+, S_ω), equipped with the following operations: (1) an associative product over S_+, denoted $s, t \mapsto s \cdot t$; (2) a mixed product $S_+ \times S_\omega \to S_\omega$, also denoted $s, t \mapsto s \cdot t$; and (3) an iteration mapping from S_+ to S_ω, which we denote $s \mapsto s^\#$. Moreover, a Wilke algebra is required to satisfy the following axioms, for $s, t \in S_+$ and $u \in S_\omega$.

$$s \cdot (t \cdot s)^\# = (s \cdot t)^\#, \tag{W1}$$

$$(s^n)^\# = s^\# \qquad \text{for } n = 1, 2, 3 \ldots, \tag{W2}$$

$$(s \cdot t) \cdot u = s \cdot (t \cdot u). \tag{W3}$$

There is a striking similarity between the axioms of a Wilke algebra and of a stabilization semigroup. We now explain this similarity.

Consider a language $L \subseteq A^\infty$. It induces, similarly as described in Chapter 11, a $\langle \cdot, \# \rangle$-syntactic congruence $\simeq_L$, and a quotient stabilization semigroup S_L. Because the equation (1) is satisfied in $A^\infty = \widehat{A^+_{\mathbf{DF}}}$, the $\langle \cdot, \# \rangle$-syntactic algebra S_L of L satisfies the equation

$$x^\# \cdot y = x^\#. \tag{W}$$

We may view a (not necessarily topological) stabilization semigroup satisfying (W) as a Wilke algebra, in which the elements of the form $s \cdot t^\#$ constitute the second sort and the remaining elements constitute the first sort of the Wilke algebra. Then, the axioms of a stabilization semigroup listed in Section 11.2 boil down to the first two axioms of Wilke algebras (the third axiom is an immediate consequence of associativity in S).

In the other way around, a Wilke algebra (S_+, S_ω) defines a stabilization semigroup $S = S_+ \cup S_\omega$, such that elements of the second sort act as left-zeroes by multiplication, and # is extended from S_+ to S as an identity mapping over S_ω. It is easy to check that the resulting structure $(S, \cdot, \#)$ is a stabilization semigroup, which clearly satisfies (W).

The correspondence outlined above can be easily extended also to homomorphisms. We therefore conclude the following.

Proposition 14.2. *The category of non-topological stabilization semigroups satisfying* (W) *is isomorphic to the category of Wilke algebras.* □

From a Wilke algebra we can also get a topological stabilization semigroup, by considering any topology for which the resulting stabilization semigroup is a topological $\langle \cdot, \# \rangle$-algebra, i.e. both $\cdot$ and # are continuous, and such that the axiom (S6) holds, i.e. $e^\# \leq e$ for every idempotent e, with respect to the specialization preorder on S. Such a topology always exists – consider the trivial antidiscrete topology – but usually, it cannot be made T_0.

We say that a homomorphism $\hat{\alpha}\colon A^\infty \to S$ is *invariant under infinite substitutions* if whenever $u_1, v_1, u_2, v_2, \ldots$ are such that $u_i, v_i \in A^+$ and $\hat{\alpha}(u_i) = \hat{\alpha}(v_i)$ for $i = 1, 2, \ldots$, then

$$\hat{\alpha}(u_1 u_2 u_3 \ldots) = \hat{\alpha}(v_1 v_2 v_3 \ldots).$$

The following theorem can be seen as a formulation of the theorem of Wilke [Wil93].

Theorem 14.3. *Let $\alpha\colon A \to S$ be a mapping from a finite alphabet to a topological stabilization semigroup satisfying* (W)*. Then there exists a unique extension $\hat{\alpha}\colon A^\infty \to S$ which is a homomorphism of $\langle\, \cdot\,, \#\rangle$-algebras, which is invariant under infinite substitutions. This homomorphism is moreover continuous.*

Proof. For a finite word $u = a_1 a_2 \ldots a_n$, we obviously define

$$\hat{\alpha}(u) = \alpha(a_1) \cdot \alpha(a_2) \cdots \alpha(a_n).$$

For an infinite word $w = a_1 a_2 a_3 \ldots$, we define

$$\hat{\alpha}(w) = s \cdot e^{\#},$$

where $s, e \in S$ are such that w can be factorized into infinitely many factors from A^+

$$w = u_0 u_1 u_2 \ldots,$$

with $\hat{\alpha}(u_0) = s$ and $\hat{\alpha}(u_i) = e$ for $i = 1, 2, \ldots$. The existence of such a factorization follows from Ramsey's theorem. The correctness of the definition of $\hat{\alpha}$ follows from the axioms of a Wilke algebra (or, equivalently, of a stabilization semigroup, together with the equation (W)). Similarly we prove that $\hat{\alpha}$ is a homomorphism of $\langle\, \cdot\,, \#\rangle$-algebras. Also, it is clear that if $\hat{\alpha}$ is to be invariant under infinite substitutions, then it must conform to the above definition.

To prove continuity, assume that $w_1, w_2, \ldots$ is a convergent sequence of elements of A^∞. Let w be its limit, and let $w = u_0 \cdot u_1 \cdot u_2 \cdots$ be its factorization such that for some $s, e \in S$, $\hat{\alpha}(u_0) = s$ and $\hat{\alpha}(u_i) = e$ for all $i = 1, 2, \ldots$ Then $\hat{\alpha}(w) = s \cdot e^{\#}$.

Since the sequence $w_1, w_2, \ldots$ converges to w, all but finitely many of its elements begin with the prefix $u_0 \cdot u_1$. Therefore, for almost every n, there exists $t_n \in S$ such that

$$\hat{\alpha}(w_n) = s \cdot e \cdot t_n.$$

Then, for almost all n, we have:

$$\begin{aligned} \hat{\alpha}(w) &= s \cdot e^{\#} && (2) \\ &= s \cdot e^{\#} \cdot t_n && \text{(by (W))} \\ &\leq s \cdot e \cdot t_n = \hat{\alpha}(w_n). && \text{(by (S6))} \end{aligned}$$

Therefore, we have shown that whenever $w = \lim_{n\to\infty} w_n$, then $\hat{\alpha}(w) \leq \hat{\alpha}(w_n)$ for almost all n. From this, continuity of $\hat{\alpha}$ follows easily. □

Comparison Theorem 14.3 should be compared with Theorem 8.1. We compare both the assumptions and the conclusions of the two theorems.

Assumptions. Clearly, Theorem 8.1 is more general, in the aspect that it does not restrict to stabilization semigroups which satisfy the equation (W). On the other hand, Theorem 14.3 does not restrict only to T_0 stabilization semigroups. Restricting only to T_0 Wilke algebras would correspond to considering ω-regular languages which are Boolean combinations of open sets, in the Cantor topology of A^∞. As we know, there are ω-regular languages which are not Boolean combinations of open sets.

Conclusions. Theorem 8.1 concludes that there exists a canonical homomorphism extending α. This notion of canonicity, defined by several equivalent conditions in Proposition 12.1, is either specific to dealing with T_0 stabilization semigroup (as is the condition of Proposition 12.1 which says that $\hat{\alpha}$ is the "largest" with respect to specialization order) or specific to dealing with B- or S-regular languages. On the other hand, it seems that the notion of invariance under infinite substitutions is specific to dealing with languages in A^∞.

This leads to the following question: is there a common generalization of Theorem 8.1 and Theorem 14.3?

14.3.2 Connection via limitary behavior

Let $L \subseteq A^\infty$ be an ω-regular language. These languages can be characterized as inverse images $\hat{\alpha}^{-1}(F)$ of subsets of Wilke algebras under a mapping $\hat{\alpha}$, as described in Theorem 14.3.

Then, it follows easily from Theorem 14.3 that L can be written as a finite union of sets of the form

$$U \cdot V^\omega,$$

where $U, V \subseteq A^*$ are regular languages of finite words (if $V \subseteq \{\varepsilon\}$ then the above expression evaluates to $U \subseteq A^\infty$). Any word w in the language $U \cdot V^\omega$ has a factorization of the form

$$w = u_0 \cdot u_1 \cdot u_2 \cdots, \tag{1}$$

where $u_0 \in U$ and $u_i \in V$ for $i = 1, 2, 3, \ldots$. Notice that moreover, when constructing the sets U and V Theorem 14.3, we can assume that $V \cdot V \subseteq V$, so $u_i \cdot u_{i+1} \cdot u_{i+2} \cdots u_j \in V$ for any $1 \leq i \leq j$.

We use the following easy generalization of Ramsey's theorem to compact semigroups (see [BK10]).

Proposition 14.4. *Let S be a compact metric topological semigroup, and let $s_1, s_2, \ldots$ be an infinite sequence of its elements. Then, there exists a factorization*

$$\underbrace{(s_1 s_2 \cdots s_{i_1 - 1})}_{u_1} \underbrace{(s_{i_1} s_{i_1+1} \cdots s_{i_2-1})}_{u_2} \underbrace{(s_{i_2} s_{i_2+1} \cdots s_{i_3-1})}_{u_3} \cdots$$

such that the sequence of factors $u_1, u_2, \ldots$ is convergent to an idempotent element $u_\infty \in S$.

We call a factorization as in (1) *convergent*, if $u_1, u_2, u_3, \ldots$ converges to an idempotent profinite word u_∞. We will moreover say that it is ε-convergent for a given number $\varepsilon > 0$, if $d(u_n, u_\infty) < \varepsilon$ for every $n = 1, 2, \ldots$, where d is the profinite metric.

It follows from Proposition 14.4 that we can assume that the factorization (1) is ε-convergent for a given $\varepsilon > 0$, and that the limit u_∞ belongs to the clopen set $\overline{V}$.

It not difficult to conclude the following.

Proposition 14.5. *Let $L \subseteq A^\infty$ be an ω-regular language. Then there exists a number $\varepsilon > 0$ such that for any word $w \in A^\infty$ and any ε-convergent factorization of the form* (1)*, with limit u_∞, membership of w to L depends only on u_0 and u_∞.*

Enhancing slightly the above proposition, it is possible to relate ω-regular languages with clopen subsets of $\widehat{A^+}$ via the limitary behavior of convergent factorizations. We believe that similarly, one can relate ωB-regular languages in A^∞ with B-regular languages in $\widehat{A^+}$, or ωS-regular languages with S-regular languages. More generally, it seems plausible that in a similar fashion one can link languages of ω-words definable in MSO+$\mathbb{B}$ with languages of profinite words definable in MSO+inf, and perhaps prove a two-way reduction between satisfiability of MSO+$\mathbb{B}$ and satisfiability MSO+inf via this connection.

14.3.3 Restricting to terms

In the case of ω-regular languages, a recognizable language is determined uniquely by the set of its ultimately periodic elements (see e.g. [Tho97, PP04]). The situation is analogous in the case of B- and S-regular languages in $\widehat{A^+}$: the condition 5 of Theorem 8.2 says that an S-regular language L is the closure of the set of its elements, which are also elements of $A^{\langle \cdot, \omega \rangle}$. Elements of $A^{\langle \cdot, \omega \rangle}$ can be seen as analogues of ultimately periodic words.

Moreover, similarly to the case of ω-words, it is possible to establish some results when restricting the considerations to profinite words which are elements of $A^{\langle \cdot, \omega \rangle}$.

We treat $A^{\langle \cdot, \omega \rangle}$ as a topological $\langle \cdot, \# \rangle$-algebra, where the ω-power plays the role of stabilization, and the topology is the subspace topology originating from $\widehat{A^+}$. Theorem 8.1 implies the following.

Proposition 14.6. *Let $\alpha \colon A \to S$ be a mapping from a finite alphabet to a finite T_0 stabilization semigroup. Then there exists a* unique *homomorphism $\tilde{\alpha} \colon A^{\langle \cdot, \omega \rangle} \to S$ of $\langle \cdot, \# \rangle$-algebras which extends α. This extension is moreover continuous.*

Proposition 14.7. *Let L be a closed subset of $A^{\langle \cdot, \omega \rangle}$. The following conditions are equivalent.*

1. *The closure of L in $\widehat{A^+}$ is S-regular,*
2. *There is a mapping $\tilde{\alpha} \colon A^{\langle \cdot, \omega \rangle} \to S$ to a finite T_0 stabilization semigroup S, which is a homomorphism of $\langle \cdot, \# \rangle$-algebras, and such that $L = \alpha^{-1}(F)$ for some closed subset $F \subseteq S$.*

Proof. Let K denote the closure of L in $\widehat{A^+}$.

1 ⇒ 2. Since K is S-regular, there exists a homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ onto a finite stabilization semigroup S, and a closed set F such that $\hat{\alpha}^{-1}(F) = K$. Let $\tilde{\alpha}$ be the restriction of $\hat{\alpha}$ to $A^{\langle\,\cdot\,,\omega\rangle}$. Then $\tilde{\alpha}$ is also a homomorphism of stabilization semigroups, and

$$\tilde{\alpha}^{-1}(F) = \hat{\alpha}^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle} = K \cap A^{\langle\,\cdot\,,\omega\rangle} = L.$$

The last equality follows from the fact that L is closed in $A^{\langle\,\cdot\,,\omega\rangle}$.

2 ⇒ 1. Let $\tilde{\alpha}\colon A^{\langle\,\cdot\,,\omega\rangle} \to S$ be as in the second condition of the statement of the proposition. By Theorem 8.1, there exists the canonical homomorphism $\hat{\alpha}\colon \widehat{A^+} \to S$ which extends $\tilde{\alpha}$. Moreover, by Proposition 12.1,

$$\hat{\alpha}^{-1}(F) = \overline{\hat{\alpha}^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle}} = \overline{\tilde{\alpha}^{-1}(F)} = K,$$

and so K is the inverse image of a closed set under a canonical mapping, so K is S-regular by Proposition 12.1. □

Remark 14.2. It is a natural question to ask, why do we consider at all the set of all profinite words, $\widehat{A^+}$, instead of the just the set of terms, $A^{\langle\,\cdot\,,\omega\rangle}$. This is analogous to the situation in the theory of ω-regular languages, where one can consider the set of all infinite words, or restrict to the set of ultimately periodic words. However, in the setting of profinite words, we do not know how to define properly the syntactic congruence, starting from a set $L \subseteq A^{\langle\,\cdot\,,\omega\rangle}$, so that sets of finite index correspond to regular cost functions. The following example illustrates this problem.

Example 14.1. For a set $L \subseteq A^{\langle\,\cdot\,,\omega\rangle}$, let $\simeq'_L$ be the coarsest congruence over $A^{\langle\,\cdot\,,\omega\rangle}$ with respect to multiplication $\cdot$, stabilization # equal to the ω-power, and membership to L. It can be defined as follows:

$$x \simeq'_L y \qquad \text{iff} \qquad \tau(x) \in L \iff \tau(y) \in L \quad \textit{for every } \tau \in \textit{Terms}(A^{\langle\,\cdot\,,\omega\rangle},\ \cdot\,,\#).$$

Consider the set $L \subseteq A^{\langle\,\cdot\,,\omega\rangle}$ from Example 8.9, i.e. $L = \{a^\omega\}$. Clearly, L is a closed set. It is easy to check that $\simeq'_L$ has two equivalence classes: L and $A^{\langle\,\cdot\,,\omega\rangle} - L$. Therefore, $\simeq'_L$ has finite index but L does not seem to correspond to any S-regular language.

14.3.4 Questions

The aim of this thesis is to lay the foundations of a theory in which one could approach the problem of decidability of the logic MSO+𝔹. We are dealing with languages of profinite words instead of languages of infinite words. Therefore, one question which arises is how to connect the two worlds. This can be done in some cases, via a Ramsey-type theorem for compact spaces, as mentioned in Section 14.3.2. The next question which arises is: what is the class of languages of profinite words which corresponds to the logic MSO+𝔹? A natural candidate would be the class of languages defined by MSO+inf. The problem when dealing with this logic is just as in the case of MSO+$\mathbb{B}$ – we do not know how to deal with projections of Boolean combinations of languages defined by B- and S-automata. An example of such a language is the language L defined in Section 14.2. It is a projection of the intersection of a B-regular language with an S-regular language. The language L is neither a closed nor an open language, nor a Boolean combination of such. On the side of stabilization semigroups, the problem which arises is that the resulting stabilization semigroups, although finite, are equipped with a topology that is not even T_0. Therefore, a natural idea would be to extend our theory to one which deals with stabilization semigroups equipped with a topology which is not T_0. For this, it would be useful to generalize Theorem 8.1, as mentioned in Section 14.3.1. Perhaps, for finding such a generalization, one would need to extend the signature $\langle \cdot, \# \rangle$ by other operations than the ω-power in the profinite semigroup, which might turn out to be relevant for the languages definable in MSO+inf.

Bibliography

[Alm05] Jorge Almeida. Profinite semigroups and applications. In *NATO Sci. Ser. II Math. Phys. Chem.*, volume 207, pages 1–45. Springer, Dordrecht, 2005. Cited on pages 13, 32, and 38.

[AM09] Susanne Albers and Jean-Yves Marion, editors. *26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009, February 26-28, 2009, Freiburg, Germany, Proceedings*, volume 3 of *LIPIcs*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2009. Cited on pages 161 and 163.

[BC06] Mikołaj Bojańczyk and Thomas Colcombet. Bounds in ω-regularity. In *Logic in Computer Science*, pages 285–296, 2006. Cited on pages iii, 6, 7, 8, 93, 103, 104, and 115.

[BK10] Mikołaj Bojańczyk and Eryk Kopczyński. Ramsey's theorem for colors from a metric space. Submitted, 2010. Cited on page 157.

[Boj04] Mikołaj Bojańczyk. A bounding quantifier. In *Computer Science Logic*, volume 3210 of *Lecture Notes in Computer Science*, pages 41–55, 2004. Cited on page 5.

[Boj09a] Mikołaj Bojańczyk. Factorization forests. In Volker Diekert and Dirk Nowotka, editors, *Developments in Language Theory*, volume 5583 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2009. Cited on pages 4 and 61.

[Boj09b] Mikołaj Bojańczyk. Weak MSO with the unbounding quantifier. In Albers and Marion [AM09], pages 159–170. Cited on page 6.

[Boj10] Mikołaj Bojańczyk. Beyond ω-regular languages. In *STACS*, pages 11–16, 2010. Cited on page 6.

[BT09] Mikołaj Bojańczyk and Szymon Toruńczyk. Deterministic automata and extensions of Weak MSO. In Ravi Kannan and K. Narayan Kumar, editors, *FSTTCS*, volume 4 of *LIPIcs*, pages 73–84. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2009. Cited on page 6.

[Büc62] Julius R. Büchi. On a decision method in restricted second-order arithmetic. In *Proc. 1960 Int. Congr. for Logic, Methodology and Philosophy of Science*, pages 1–11, 1962. Cited on page 9.

[Col09] Thomas Colcombet. The theory of stabilisation monoids and regular cost functions. In Susanne Albers, Alberto Marchetti-Spaccamela, Yossi Matias, Sotiris E. Nikoletseas, and Wolfgang Thomas, editors, *ICALP (2)*, volume 5556 of *Lecture Notes in Computer Science*, pages 139–150. Springer, 2009. Cited on pages 7, 8, 55, 89, 93, 119, 126, 134, and 152.

[Col10a] Thomas Colcombet. Private communcation, 2010. Cited on pages 135 and 140.

[Col10b] Thomas Colcombet. Regular cost functions, part i: Logic and algebra over words. unpublished, 2010. Cited on pages 7, 72, 105, and 152.

[DDG$^+$10] Aldric Degorre, Laurent Doyen, Raffaella Gentilini, Jean-François Raskin, and Szymon Toruńczyk. Energy and mean-payoff games with imperfect information. In Anuj Dawar and Helmut Veith, editors, *CSL*, volume 6247 of *Lecture Notes in Computer Science*, pages 260–274. Springer, 2010. Cited on page 7.

[Egg63] Lawrence C. Eggan. Transition graphs and the star height of regular events. *Michigan Math. J.*, 10:385–397, 1963. Cited on page 1.

[Eng89] Ryszard Engelking. *General Topology*. Heldermann Verlag, 1989. Cited on page 13.

[Gre51] J. A. Green. On the structure of semigroups. *Annals of Mathematics*, 54(1):163–172, July 1951. Cited on page 26.

[Has82] Kosaburo Hashiguchi. Limitedness theorem on finite automata with distance functions. *Journal of Computer and System Sciences*, 24:233–244, 1982. Cited on pages iii and 3.

[Has88] Kosaburo Hashiguchi. Algorithms for determining relative star height and star height. *Inf. Comput.*, 78(2):124–169, 1988. Cited on page 1.

[HST10] Szczepan Hummel, Michał Skrzypczak, and Szymon Toruńczyk. On the topological complexity of MSO+U and related automata models. In Petr Hlinený and Antonín Kucera, editors, *MFCS*, volume 6281 of *Lecture Notes in Computer Science*, pages 429–440. Springer, 2010. Cited on page 8.

[Kir05] Daniel Kirsten. Distance desert automata and the star height problem. *Theoretical Informatics and Applications*, 39(3):455–511, 2005. Cited on pages iii, 4, and 54.

[Kro92] Daniel Krob. The equality problem for rational series with multiplicities in the tropical semiring is undecidable. In Werner Kuich, editor, *ICALP*, volume 623 of *Lecture Notes in Computer Science*, pages 101–112. Springer, 1992. Cited on page 8.

[Kru72] Joseph B. Kruskal. The theory of well-quasi-ordering: A frequently discovered concept. *Journal of Combinatorial Theory, Series A*, 13(3):297 – 305, 1972. Cited on page 71.

[Leu88] Hing Leung. On the topological structure of a finitely generated semigroup of matrices. *Semigroup Forum*, 37:273–278, 1988. Cited on page 3.

[Leu98] Hing Leung. The topological approach to the limitedness problem on distance automata. *Idempotency*, pages 88–111, 1998. Cited on page 53.

[Num57] Katsumi Numakura. Theorems on compact totally disconnetced semigroups and lattices. In *Proc. Amer. Math. Soc.*, volume 8, pages 626–626, 1957. Cited on page 37.

[Pin98] Jean-Éric Pin. Tropical semirings. In *Idempotency*, pages 50–69. Cambridge University Press, 1998. Cited on pages 3 and 51.

[Pin09] Jean-Éric Pin. Profinite methods in automata theory. In Albers and Marion [AM09], pages 31–50. Cited on page 13.

[PP04] Dominique Perrin and Jean-Éric Pin. *Infinite Words*. Elsevier, 2004. Cited on pages 105, 155, and 158.

[Ree40] David Rees. On semigroups. *Mathematical Proceedings of the Cambridge Philosophical Society*, 36:387–400, 1940. Cited on page 26.

[Sim78] Imre Simon. Limited subsets of a free monoid. *Foundations of Computer Science, Annual IEEE Symposium on*, 0:143–150, 1978. Cited on page 2.

[Sim88] Imre Simon. Recognizable sets with multiplicites in the tropical semiring. In *Mathematical Foundations of Computer Science*, volume 324 of *Lecture Notes in Computer Science*, pages 107–120, 1988. Cited on page 3.

[Sim90] Imre Simon. Factorization forests of finite height. *Theoretical Computer Science*, 72:65–94, 1990. Cited on page 60.

[Sim94] Imre Simon. On semigroups of matrices over the tropical semiring. *ITA*, 28(3-4):277–294, 1994. Cited on pages 4, 5, 59, and 60.

[ST11] Luc Segoufin and Szymon Toruńczyk. Automata based verification over linearly ordered data domains. In *STACS*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2011. To appear. Cited on page 7.

[Sus28] Anton K. Suschkewitsch. Über die endlichen gruppen ohne das gesetz der eindeutigen umkehrbarkeit. *Mathematische Annalen*, 99:30–50, 1928. Cited on page 26.

[Tho90] Wolfgang Thomas. Automata on infinite objects. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 133–192. Elsevier and MIT Press, 1990. Cited on page 9.

[Tho97] Wolfgang Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Language Theory*, volume III, pages 389–455. Springer, 1997. Cited on pages 91, 105, and 158.

[Wil70] Stephen Willard. *General Topology*. Addison-Wesley, 1970. Cited on pages 13 and 38.

[Wil93] Thomas Wilke. An algebraic theory for languages of finite and infinite words. *Inf. J. Alg. Comput.*, 3:447–489, 1993. Cited on pages 56, 155, and 156.