# Languages of profinite words and the limitedness problem

Szymon Toruńczyk

University of Warsaw

**Abstract.** We present a new, self-contained proof of the limitedness problem. The key novelty is a description using profinite words, which unifies and simplifies the previous approaches, and seamlessly extends the theory of regular languages. We also define a logic over profinite words, called MSO+inf and show that the satisfiability problem of MSO+$\mathbb{B}$ reduces to the satisfiability problem of our logic.

## 1 Introduction

This paper is an attempt to establish a natural framework for problems related to the limitedness problem. A notable example of such a problem is the decidability of the logic MSO+$\mathbb{B}$.

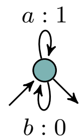



**Fig. 1.** A distance automaton over the input alphabet $\{a, b\}$.

The *limitedness problem* was introduced by Hashiguchi [8] on his way to solving the famous star height problem. In its basic form, it concerns *distance automata*, i.e. nondeterministic automata, whose transitions are additionally labeled by nonnegative, integer weights, such as the one depicted in Figure 1. A distance automaton is *limited* if there exists a bound $n$ such that every accepted word has some accepting run whose sum of weights is bounded by $n$. Thus the *limitedness problem* is a decision problem which asks whether a given automaton is limited. The automaton in the example is not limited: the words $a, a^2, a^3, \ldots$ require accepting runs of ever larger weights.

The *logic MSO*+$\mathbb{B}$ was introduced by Bojańczyk in his dissertation (see also [2]) in relation with a problem concerning modal $\mu$-calculus. It is an extension of the usual MSO logic – over infinite trees or words – by the quantifier $\mathbb{B}$, defined so that the formula $\mathbb{B}X.\varphi(X)$ holds if and only if all the sets of positions $X$ satisfying the formula $\varphi$ in the given model have a commonly bounded size. A typical language of infinite words defined in this logic is:

$$L_B = \{a^{n_1}ba^{n_2}b\ldots : \quad \text{the sequence } n_1, n_2, \ldots \text{ is bounded}\}.$$

Note that this language is not $\omega$-regular, as its complement does not contain any ultimately periodic word. As a far-reaching project (see [3] for a survey), Bojańczyk posed the question of decidability of satisfiability of the logic MSO+$\mathbb{B}$ over infinite trees. Still, it is not even known to be decidable over infinite words.

A syntactic fragment of the logic MSO+$\mathbb{B}$ has been shown decidable in [4]. The key tool used in this paper is a model of automata called *ωB-automata*. Later, the authors discovered that limitedness of distance automata can be easily decided using their results concerning ωB-automata. The link with the limitedness problem has been exploited in [6], where Colcombet defined *B-automata* and developed his theory of regular cost functions and stabilization semigroups. B-automata directly generalize distance automata, by allowing more than one counter which, moreover, can be reset.

*Our contribution* is a theory which we believe to be the appropriate setting for considering limitedness of B-automata, and related problems. As a starting point, we see that B-automata naturally define languages of *profinite* words. The set of profinite words has a rich algebraic and topological structure, which we find very useful in the context of limitedness.

For instance, consider the distance automaton from Figure 1. There is a *profinite* word, denoted $a^\omega$ (not to be confused with the infinite word) which witnesses the fact that the automaton is *not* limited – this word can be defined as the limit of the sequence of finite words $(a^{n!})_{n=1}^{\infty}$. We say that this profinite word does *not* belong to the language of this automaton; the language of this automaton consists of profinite words which only have finitely many $a$'s, such as $b$ or $b^\omega a$.

We call the class of languages of profinite words defined by B-automata *B-regular languages*. Our main result states that this class can be characterized in terms of logic, regular expressions and semigroups. The result generalizes the main results of the papers [11, 13, 9, 1, 4], and implies the main result of [6, 7]. The description in terms of semigroups immediately implies decidability of the limitedness problem for B-automata, which, in our framework is simply the question of language universality. In particular, together with Kirsten's elegant reduction of the star height problem to the limitedness problem, our result gives yet another proof of decidability of the star height problem. The result also implies decidability of a more general problem – limitedness of Boolean combinations of B-automata. The remaining characterizations are primarily of conceptual value, as they manifest both that our framework is appropriate, and that the class of B-regular languages is robust. Note that most of these characterizations are also available in the framework of Colcombet. One exception is a new, finite-index characterization of B-regular languages, à la the Myhill-Nerode theorem; it seems that this result cannot be even phrased in the other frameworks.

Lastly, we show that our framework is suited for dealing with the satisfiability problem for MSO+$\mathbb{B}$ over infinite words – we prove that this problem can be reduced to the satisfiability problem of a new logic MSO+inf over profinite words, which we introduce here. This seems impossible in the other frameworks. In fact, our reduction is very general, and works for very many logics. The proof extends Büchi's ideas, and consists of two key ingredients: convergent Ramsey factorizations of infinite words, and a model of deterministic automata over infinite words with a profinite acceptance condition.

*Related work.* Several proofs of decidability of the limitedness problem exist [8, 11, 13, 9, 1, 6]. Our proof builds on ideas from all of these papers, and simplifies them greatly. Hashigushi's #-expressions acquire a new, concrete meaning in our framework, as simply defining profinite words. We extend Leung's insight of considering the compact topological semigroup of all matrices over the tropical semiring, to considering the profinite semigroup. Also, Leung introduced finite versions of his topological semigroups, which are predecessors of stabilization semigroups of Colcombet. The factorization forests of Simon play a key role in the main technical part of our proof. The proof of Kirsten applies to a model very similar to B-automata, but with a hierarchical constraint on the counter operations. Kirsten generalized Leung's proof, providing further instances of stabilization semigroups; however, the topological insights of Leung disappeared, as he no longer considered compact topological semigroups.

Colcombet used ideas from [4] and of Kirsten in [7], where he developed his theory of regular cost functions. In his theory, a B-automaton defines a *B-regular cost function* – an equivalence class of number-valued functions. These cost functions also have equivalent descriptions in terms of regular expressions, logic and semigroups. The crucial discovery of that paper is the tight two-way correspondence between stabilization semigroups (defined there) and B-automata. Still, the topological insights of Leung remained missing.

On a general level, and also on the level of proof structure, our approach resembles the approach of Colcombet. We outline the key differences. As we deal with languages which are subsets of a topological semigroup, many classical notions naturally lift to our setting – such as recognizable subsets, Myhill-Nerode equivalence, homomorphisms. In Colcombet's framework, cost functions are not sets, and have no apparent algebraic nor topological structure (they only have a lattice structure, corresponding to the lattice ordering of languages). Because of this, the natural notions mentioned above do not exist, or have non-obvious definitions – an example is the complex notion of *compatible mapping* [6], which corresponds to our $\infty$*-homomorphism.* Even the notion of a Boolean combination of cost functions is meaningless. As a result, cost functions are not well-suited for the study of the full logic MSO+$\mathbb{B}$. On a technical level, the proofs in [6, 7] deal with the relative notions of "big" vs. "small" values, and this relativity needs to be carefully controlled in the calculations and proofs. In our more abstract setting, we deal with the absolute notions of infinite vs. finite, and computations involve usual set-theoretic equalities.

*Outline of the paper.* First, we recall the definitions of B- and S-automata, and of profinite words. Next, we show how languages of profinite words can be defined using automata, regular expressions and logic. Then we present our main technical tool – recognition by homomorphisms. In Section 5, we state the central result. Finally, we show a link between languages of infinite words and of profinite words. Due to space limitations, many details are deferred to the appendix.

*Acknowledgements.* I would like to thank Mikołaj Bojańczyk and Thomas Colcombet for many useful comments.

## 2 Preliminaries

Let us fix a finite alphabet $A$; finite words are assumed to be elements of $A$. In the examples, we will more concretely assume the alphabet $A = \{a, b\}$. By $\mathbb{N}$ we denote $\{0, 1, 2, \ldots\}$, and by $\overline{\mathbb{N}}$ we denote $\mathbb{N} \cup \{\omega\}$. We treat $\overline{\mathbb{N}}$ as a compact metric space, in which $d(m, n) = |2^{-m} - 2^{-n}|$ (where $2^{-\omega} = 0$).

**B-automata and S-automata** (implicit in [4], defined in [6]) are nondeterministic automata over finite words, equipped with a finite number of counters. There are two counter operations available for each counter: *inc* increases the current value of the counter by 1 and *reset* sets the value to 0. A transition of a B- or S-automaton may trigger any sequence of operations on its counters. If the operation *reset* is performed in a run $\rho$ on a counter which currently stores a value $n$, then we say that *$n$ is a reset value* in the considered run $\rho$. The two models – B- and S-automata – differ in the semantics of the functions they define.

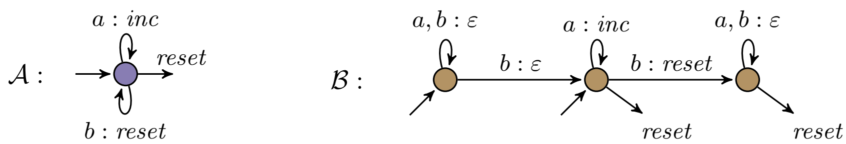
First, consider a B-automaton $\mathcal{A}$. Since $\mathcal{A}$ is nondeterministic, there might be many runs over a single word. For a particular run $\rho$, we define the *value* of $\rho$ as its maximal reset value. Next, the *valuation* $f_{\mathcal{A}}(w)$ of an input word $w$ under the automaton $\mathcal{A}$ is the minimum of the values of all accepting runs $\rho$ over $w$:

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

Note that min ranges only over the accepting runs $\rho$ of $\mathcal{A}$. We assume $\max(\emptyset) = 0$ and $\min(\emptyset) = \omega$, so if $\mathcal{A}$ has no accepting run over $w$, then $f_{\mathcal{A}}(w) = \omega$.

If $\mathcal{A}$ is an S-automaton, the definition of a valuation $f_{\mathcal{A}}(w)$ of an input word $w$ is completely dual – simply swap min with max in the formula above.

*Example 1 (The running example).* Let $\mathcal{A}$ be the B-automaton with one counter which is depicted in the left-hand side of the figure below.




We declare that the automaton resets its counter after reading the entire word – this extra feature can be easily eliminated using nondeterminism. Then,

$$f_{\mathcal{A}}(w) = \max\{n_1, n_2, \ldots, n_k\} \qquad \textit{for } w = a^{n_1} b a^{n_2} \ldots b a^{n_k}.$$

Now consider the S-automaton $\mathcal{B}$ depicted in the right-hand side of the figure. It has one counter, which is also assumed to be reset at the end of the run. The reader can check that each accepting run of $\mathcal{B}$ over an input word $w$ corresponds to a block of $a$'s in $w$, and that $f_{\mathcal{B}}(w)$ is the length of the largest block of $a$'s in $w$. Therefore, $f_{\mathcal{B}}$ and $f_{\mathcal{A}}$ are precisely the same function from $A^+$ to $\overline{\mathbb{N}}$.

*Example 2.* Let $\mathcal{A}$ be a finite nondeterministic automaton. If we view $\mathcal{A}$ as a B-automaton with no counters, the induced function assigns 0 to any word accepted by $\mathcal{A}$ and $\omega$ to any rejected word. Dually, if we treat $\mathcal{A}$ as an S-automaton, the induced function assigns $\omega$ to any accepted word, and 0 to any rejected word.

A B- or S-automaton is said to be *limited* if the function $f_{\mathcal{A}}$ has finite range (it may nevertheless contain the value $\omega$). The *limitedness problem* for B- or S-automata is then to decide whether a given B- or S-automaton is limited. The automata in the example are not limited, since $f_{\mathcal{A}}(a^n) = n$ for any $n \in \mathbb{N}$.

**Profinite words** should be thought of as limits of sequences of finite words, with respect to *all* regular languages. A formal definition follows (see e.g. [12] for more details). We say that an infinite sequence $w_1, w_2, \ldots \in A^+$ of finite (nonempty) words *ultimately* belongs to the regular language $L \subseteq A^+$ if almost all the words $w_1, w_2, \ldots$ belong to $L$. We say that a sequence of words is *convergent*, if for any regular language $L$, the sequence ultimately belongs to $L$ or ultimately belongs to the complement of $L$. Every constant sequence is convergent. The sequence $a, a^{2!}, a^{3!}, \ldots$ is also convergent, as follows from a pumping argument for regular languages. However, the sequence $a, a^2, a^3, \ldots$ is not convergent, since the regular language $(aa)^+$ only contains every other of its elements. Two convergent sequences are *equivalent* if they belong ultimately to precisely the same regular languages. In other words, interleaving one sequence with the other yields a convergent sequence. An equivalence class of convergent sequences is a *profinite word*. A profinite word is uniquely specified by the set of regular languages to which it ultimately belongs. For example, the equivalence class of the convergent sequence $a, a^{2!}, a^{3!}, \ldots$, which is a profinite word denoted $a^\omega$, ultimately belongs to the languages $a^+, (aa)^+, (aaa)^+, \ldots$, and does not ultimately belong to the languages $a^* \cdot b \cdot a^*$ nor $a \cdot (aa)^+$. We denote profinite words by $x, y, \ldots$, and the set of all profinite words by $\widehat{A^+}$. We define $\widehat{A^*} = \widehat{A^+} \cup \{\varepsilon\}$, where $\varepsilon$ is the *empty word*. Note that the set of finite words $A^+$ naturally embeds into the set of profinite words $\widehat{A^+}$, via constant convergent sequences. We call subsets of $\widehat{A^+}$ or of $\widehat{A^*}$ *languages of profinite words*.

The set of profinite words forms a semigroup: if $w_1, w_2, \ldots$ and $v_1, v_2, \ldots$ are two convergent sequences, then the sequence $w_1v_1, w_2v_2, \ldots$ is also convergent. There is another important operation on profinite words, called the *$\omega$-power*. The $\omega$-power of a convergent sequence $w_1, w_2, w_3, \ldots$ is the sequence $w_1^1, w_2^{2!}, w_3^{3!}, \ldots$, which also turns out to be convergent. This operation induces an operation $x \mapsto x^\omega$ defined over profinite words.

The set of profinite words carries a compact metric: the distance between two profinite words $x, y$ is $\frac{1}{n}$, where $n$ is the smallest size – measured as size of the minimal automaton – of a regular language $L$ such that $x$ ultimately belongs to $L$ and $y$ does not. This metric is compatible with the notion of convergence defined above. In particular, the set $A^+$ of finite words is dense in the set of profinite words, $\widehat{A^+}$. Multiplication and the $\omega$-power are continuous mappings over $\widehat{A^+}$. One can prove that $x^\omega = \lim_{n\to\infty} x^{n!}$ for any $x \in \widehat{A^+}$.

The closure $\overline{L}$ in $\widehat{A^+}$ of any regular language $L \subseteq A^+$ turns out to be both closed and open, i.e. *clopen* in $\widehat{A^+}$. Conversely, any clopen subset of $\widehat{A^+}$ is of the form $\overline{L}$ for some regular language $L$, so clopen sets correspond precisely to regular languages. Any open set in $\widehat{A^+}$ is a (possibly infinite) union of clopen sets.

## 3 Languages of profinite words

In this section we discuss several ways of describing languages of profinite words – via automata, regular expressions and logic.

**B- and S-regular languages.** The essential idea underlying our theory is to consider B- and S-automata as processing not only finite words, but also profinite words. Let $\mathcal{A}$ be a B- or S-automaton. The following, simple observation relies on the fact that for each $n \in \mathbb{N}$, the language $\{w \in A^+ : \ f_{\mathcal{A}}(w) < n\}$ is regular.

**Fact 1.** *Let $w_1, w_2, \ldots$ be a convergent sequence of finite words. Then, the sequence $f_{\mathcal{A}}(w_1), f_{\mathcal{A}}(w_2), \ldots$ is convergent in $\overline{\mathbb{N}} = \mathbb{N} \cup \{\omega\}$.*

Therefore, it makes sense to define, for any $x \in \widehat{A^+}$,

$$\widehat{f_{\mathcal{A}}}(x) \quad \stackrel{def}{=} \quad \lim_{n\to\infty} f_{\mathcal{A}}(w_n),$$

where $w_1, w_2, \ldots$ is any sequence of finite words which converges to $x$. This value may happen to be $\omega$. It is straightforward to show that $\widehat{f_{\mathcal{A}}}$ is a well-defined continuous function from $\widehat{A^+}$ to $\overline{\mathbb{N}}$. Moreover, by density of $A^+$ in $\widehat{A^+}$, the continuous extension of $f_{\mathcal{A}}$ to $\widehat{A^+}$ is unique, so we will further identify $f_{\mathcal{A}}$ with the continuous mapping $\widehat{f_{\mathcal{A}}} \colon \widehat{A^+} \to \overline{\mathbb{N}}$.

Similarly to the idea underlying cost functions [6], we do not care about the exact values of the function $f_{\mathcal{A}}$ (this would quickly lead to undecidability, as demonstrated by Krob [10]). What we care about is over which sequences of words, $f_{\mathcal{A}}$ grows indefinitely. By continuity of $f_{\mathcal{A}}$ and compactness of $\widehat{A^+}$, this is encoded in the set

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

This is a closed set as the inverse image of a point under a continuous mapping.

This motivates the following definitions. For an S-automaton $\mathcal{A}$, we define the set $L(\mathcal{A})$ consisting of all profinite words $x$ such that $f_{\mathcal{A}}(x) = \omega$. For a B-automaton $\mathcal{A}$, we define $L(\mathcal{A})$ dually, as the language of all profinite words $x$ such that $f_{\mathcal{A}}(x) < \omega$. In either case, we call $L(\mathcal{A})$ the *language recognized by* $\mathcal{A}$. The reason why the definitions differ is that S-automata try to maximize, while B-automata try to minimize the value of a run. We call a language $L \subseteq \widehat{A^+}$ *B-regular* (respectively, *S-regular*), if it is recognized by a B-automaton (respectively, S-automaton). Note that S-regular languages are closed, and B-regular languages are open subsets of $\widehat{A^+}$. In particular, a language is both B- and S-regular if and only if it is clopen.

*Example 3.* Let $\mathcal{A}$ be the B-automaton from Example 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:

$$L(\mathcal{A}) = \{x \in \widehat{A^+} : \ f_{\mathcal{A}}(x) < \omega\} = \bigcup_{n\in\mathbb{N}} \{x \in \widehat{A^+} : \ x \text{ has no infix } a^n\}.$$

It is not difficult to show (using compactness and continuity of multiplication) that a profinite word has arbitrarily long blocks of $a$'s if and only if it contains

$a^\omega$ as an infix. (We say that $u$ is an *infix* of $v$ if $v = v_1 \cdot u \cdot v_2$ for some, potentially empty, profinite words $v_1, v_2$.) Therefore, if $\mathcal{B}$ is the S-automaton from Example 1 (recall that $f_\mathcal{A} = f_\mathcal{B}$), we deduce that

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

*Limitedness.* Assume that we want to test for limitedness of a B-automaton $\mathcal{A}$. It is easy to reduce the general case to the case when the underlying finite automaton accepts all finite words (to do this, it suffices to consider the disjoint union of $\mathcal{A}$ and $\mathcal{A}'$, where $\mathcal{A}'$ is a B-automaton which maps all words accepted by $\mathcal{A}$ to $\omega$, and the rest to 0). Then, an immediate compactness argument shows:

**Fact 2.** *A B-automaton $\mathcal{A}$ which accepts all finite words is limited iff $L(\mathcal{A}) = \widehat{A^+}$.*

*Closure properties.* As usual with nondeterministic automata, both classes – of B- and S-regular languages – are closed under language projection, and also under union and intersection.They are not, however, closed under complements: the complement of the B-regular language $L(\mathcal{A})$ from the previous example is not B-regular, since it is not an open set. However, this complement is an S-regular language, as it is equal to $L(\mathcal{B})$. More generally, we will prove the difficult result that complements of B-regular languages are S-regular, and vice versa.

**The logic MSO+inf.** We introduce the logic MSO+inf over profinite words. First, we define its base fragment, the logic MSO. A formula of this logic describes a set of profinite words. Usually, in the case of finite or infinite words, one sees such a word as a model whose elements are positions of the word, and so a formula of MSO speaks about sets of positions of the word. However, in profinite words, "positions" are not well-defined. To define the logic MSO over profinite words, we view the constructs of MSO as operations on languages of profinite words. We describe how to interpret the second-order existential quantifier $\exists$; for the other constructs, the idea is even simpler. We view the quantifier $\exists$ as language projection. What language do we project? A formula $\varphi(X)$ beneath a quantifier $\exists$ defines a language $L_\varphi$ over the extended alphabet $A \times \{0,1\}$. For example, $\varphi(X) = a(X) \wedge \text{singleton}(X)$ defines the language $L_\varphi$ of those profinite words over $A \times \{0,1\}$, which contain precisely one symbol $(a,1)$ and no other symbols with a 1 on the second coordinate. We define the language of the formula $\exists X.\varphi(X)$ as the projection of the language $L_\varphi$, forgetting about the second coordinate. Therefore, $\exists X.a(X) \wedge \text{singleton}(X)$ describes the set of profinite words which have precisely one letter $a$.

With similar ideas, it is easy to interpret all the usual constructs of MSO as language operations: the Boolean connectives $\wedge, \vee, \neg$, the binary predicates $<, \in$ and the unary predicates $a(X)$, per each letter $a \in A$. This way, we define the semantic of the MSO logic over profinite words. This logic describes precisely the class of clopen sets. To go beyond that, we add a predicate $\inf(X)$ which holds in a profinite word over $A \times \{0,1\}$ if it has infinitely many 1's on the second coordinate. This is a closed, but not open property of profinite words over the alphabet $A \times \{0,1\}$, so it is not definable in MSO. We denote the logic

MSO extended by the quantifier inf by MSO+inf and distinguish the syntactic fragment MSO+inf$^+$ (resp., MSO+inf$^-$) where the predicate inf appears only under an even (resp. odd) number of negations.

*Example 4.* Consider the S-regular language $L(\mathcal{B})$ from Example 3: "there is an infinite block of $a$'s". It can be described by the following formula of MSO+inf$^+$:

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

This example can be easily extended, yielding the following.

**Proposition 3.** *B-regular languages are definable in* MSO+inf$^-$*, and S-regular languages are definable in* MSO+inf$^+$*. The translations are effective.*

**B- and S-regular expressions.** We consider the usual syntax of regular expressions, except that apart from the usual Kleene star, which corresponds to *unrestricted* iteration, there are two new iteration operations: *finite* iteration, denoted $L^{<\infty}$, and *infinite* iteration, denoted $L^{\infty}$. Formally, we define *profinite sequences* of profinite words, as profinite words over the alphabet $A$ with an additional *separator* symbol †. A profinite word $x \in \widehat{A^+}$ is an *element* of a profinite sequence $\hat{x}$ if $\dagger x \dagger$ is an infix of $\dagger \hat{x} \dagger$. The *concatenation* of $\hat{x}$ is obtained by removing the symbols †. We define $L^{\infty}$ (resp. $L^{<\infty}$ and $L^*$) as concatenations of profinite sequences containing infinitely (resp. finitely, arbitrarily) many separators, and whose elements belong to $L$. *B-regular expressions* can only use the exponents $<\!\infty$ and $*$, while *S-regular expressions* can only use the exponents $\infty$ and $*$.

*Example 5.* The B-regular expression $(a^{<\infty}\, b)^*\, a^{<\infty}$ describes precisely the language accepted by the B-automaton $\mathcal{A}$ from Example 3 – "every block of $a$'s has a finite length". The S-regular expression $(a+b)^*\, a^{\infty}\, (a+b)^*$ describes precisely the complement of $L(\mathcal{A})$, i.e. the language accepted by the S-automaton $\mathcal{B}$.

Mimicking the standard translation from regular expressions to automata we get:

**Proposition 4.** *A language defined by a B-/S-regular expression is B-/S-regular.*

## 4 Recognizable languages

**Syntactic congruence.** Just as multiplication is intimately related with regular languages, multiplication together with the $\omega$-power over $\widehat{A^+}$ turn out to be of central importance for B- and S-regular languages. For notational reasons, we view $(\widehat{A^+}, \cdot, \omega)$ as an algebra over the signature $\langle\, \cdot\, , \# \rangle$, where the $\omega$-power of $\widehat{A^+}$ plays the role of the operation $\#$ of the signature. 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 $\omega$-power, and membership in $L$.

*Example 6.* Let $L = (a^{<\infty}\, b)^*\, a^{<\infty}$ be the language of the B-automaton which computes the maximal length of a block of $a$'s. It is easy to see that the equivalence classes of $\simeq_L$ (and also of $\simeq_K$, for $K = \widehat{A^+} - L$) are:

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

**Stabilization semigroups.** We 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 $\omega$-power in the profinite semigroup. The syntactic algebra also naturally inherits the *quotient* topology from $\widehat{A^+}$, which is usually non-Hausdorff, i.e. there might be singleton sets which are not closed. (However, if $L$ is a closed or open language, then the quotient topology is $T_0$, i.e. if $x \in \overline{\{y\}}$ and $y \in \overline{\{x\}}$ for $x, y \in S_L$, then $x = y$.) 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 the $\omega$-power over $\widehat{A^+}$. Namely, for $s, t, e \in S$:

$$
\begin{aligned}
s \cdot (t \cdot s)^\# &= (s \cdot t)^\# \cdot s & \qquad s^\# \cdot s^\# &= s^\# \\
(s^\#)^\# &= s^\# & e \cdot e^\# &= e^\# \qquad \text{for idemptent } e \\
(s^n)^\# &= s^\# \qquad \text{for } n = 1, 2, 3 \ldots & s^\# &\in \overline{\{s^n :\ n \in \mathbb{N}\}}.
\end{aligned}
$$

A *stabilization semigroup* is a $T_0$ topological space $S$ equipped with two continuous operations $\cdot$ and $\#$ satisfying the above axioms, apart from associativity of $\cdot$.

*Example 7.* Let $S_L$ denote the quotient set induced by the language $L$ from Example 6. As noted there, $S_L$ consists of three equivalence classes, which we denote by $[a]$, $[b]$ and $[a^\omega]$, 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, $[a^\omega]$ is the zero element and $[a]$ is the neutral element; stabilization maps $[a]$ to $[a^\omega]$ and $s$ to $s$ otherwise; $[a^\omega]$ is contained in the closure of $[a]$ and in the closure of $[b]$.

**Recognizability.** We consider an analogue 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 discrete semigroup such that for the induced homomorphism $\hat{\alpha} \colon \widehat{A^+} \to S$ we have $L = \hat{\alpha}^{-1}(F)$ for some $F \subseteq S$.

Instead of semigroups, we deal with finite stabilization semigroups. A *homomorphism* $\hat{\alpha}$ from $\widehat{A^+}$ to a stabilization semigroup $S$ is required to preserve multiplication and map the $\omega$-power in $\widehat{A^+}$ to stabilization in $S$. We use a notion of invariance of $\hat{\alpha}$ under *infinite substitutions*, which intuitively means that if a profinite word $x$ is factorized into a profinite sequence of factors, and each factor $x_i$ is replaced by some other factor $y_i$ with $\hat{\alpha}(x_i) = \hat{\alpha}(y_i)$, then, for the resulting concatenation $y$ of the factors $y_i$, $\hat{\alpha}(x) = \hat{\alpha}(y)$. We say that such a homomorphism $\hat{\alpha} \colon \widehat{A^+} \to S$ is an $\infty$-*homomorphism.* The following result plays a pivotal role in the theory, and its proof is difficult comparing to the classical case.

**Theorem 5.** *Let $\alpha \colon A \to S$ be any mapping from a finite alphabet $A$ to a finite stabilization semigroup $S$. Then there exists a unique $\infty$-homomorphism $\hat{\alpha} \colon \widehat{A^+} \to S$ extending $\alpha$. The mapping $\hat{\alpha}$ is continuous. Its image is the subset of $S$ generated from $\alpha(A)$ by the operations $\langle\,\cdot\,,\#\rangle$.*

Note that the extension $\hat{\alpha}$ is not necessarily the *unique* continuous homomorphic extension of $\alpha$. We call $\hat{\alpha}$ the $\infty$-homomorphism *induced* by $\alpha$. We say that a language $L \subseteq \widehat{A^+}$ is *recognized* by $\hat{\alpha}\colon \widehat{A^+} \to S$ if $L = \hat{\alpha}^{-1}(F)$ for some $F \subseteq S$; if additionally $F$ is closed (resp. open) in $S$, we say that $L$ is $\downarrow$*-recognizable* (resp. $\uparrow$*-recognizable*). Note that a recognizable set is described in a finite manner by $\alpha\colon A \to S$ and $F \subseteq S$. It is crucial that the image of $\hat{\alpha}$ can be computed from $\alpha$.

*Example 8.* Let $S$ be the stabilization semigroup $\widehat{A^+}/{\simeq_L}$ from the previous example, whose elements are $[a], [b], [a^\omega]$. Let $\alpha\colon A \to S$ map $a$ to $[a]$ and $b$ to $[b]$. We will check that the quotient mapping $\alpha_L\colon \widehat{A^+} \to S$ is the $\infty$-homomorphism induced by $\alpha$. We argue that $\alpha_L$ is invariant under infinite substitutions. Consider a profinite word $x$, and choose some factorization of $x$. Replace each factor by some other factor, with the same image under $\alpha_L$. Schematically:

$$\begin{array}{cccccccc} x = & aaa & aaba & aaa & \cdots & ab^\omega a & baaab \\ & \downarrow & \downarrow & \downarrow & \cdots & \downarrow & \downarrow \\ y = & aaaaa & (ab)^\omega & aaaaaaa & \cdots & aaaaabaaaa & aaaaabaaa \end{array}$$

Intuitively, it is clear that if the original word $x$ contains no infinite block of $a$'s, then no such block can appear in the resulting word $y$ either. Hence, $\alpha_L(y) = \alpha_L(x)$.

The proof of Theorem 5 extends the idea of Simon's factorization trees to profinite words and stabilization semigroups, which we shortly describe. Start with any profinite word $x$. We want to determine the *type* of $x$, i.e. $\hat{\alpha}(x)$. If $x$ is a single letter $a$, then its type is $\alpha(a)$. If not, we try to factorize $x$ into a profinite sequence of factors, for which the type can be determined. We use three rules:

- If $x = x_1 \cdot x_2$, and $\hat{\alpha}(x_1) = s_1$, $\hat{\alpha}(x_2) = s_2$, then $\hat{\alpha}(x) = s_1 \cdot s_2$,
- If $x$ factorizes into finitely many factors, each of idempotent type $e$, then $\hat{\alpha}(x) = e$,
- If $x$ factorizes into infinitely many factors, each of idempotent type $e$, then $\hat{\alpha}(x) = e^\#$.

We prove by induction on $|S|$ that in a finite number of steps, depending only on $|S|$, using the above three rules, any profinite word $x$ can be iteratively split into single letters. Moreover, we prove that the resulting type does not depend on the chosen "factorization tree". The proof of existence of factorization trees is similar to the proof of Simon's theorem, and proceeds by induction on the size of $S$. The proof of uniqueness requires the use of the axioms of stabilization semigroups. It is similar to a proof of analogous statement in [7]. An important difference is that there, only finite words have factorization trees, and their output is unique only in an asymptotic way.

The standard Cartesian-product construction yields several closure properties for recognizable languages. For closure under projection, we use two enhanced variants of the powerset construction, similar to constructions from [7].

**Proposition 6.** *Recognizable languages are closed under Boolean combinations. $\downarrow$-recognizable (resp. $\uparrow$-recognizable) languages are closed under unions and intersections. Complements of $\downarrow$-recognizable languages are $\uparrow$-recognizable and vice versa. $\downarrow$-recognizable and $\uparrow$-recognizable languages are closed under projections.*

By inductively applying the above to formulas of MSO+inf, we get:

**Corollary 1.** *Languages definable in* MSO+inf$^-$ *are $\downarrow$-recognizable, and languages definable in* MSO+inf$^+$ *are $\uparrow$-recognizable. The translations are effective.*

## 5 The main results

The main theorem collects the notions and results listed above, proving the equivalence of several characterizations. The last one is a finite-index characterization of B-automata. Up to our knowledge, such a characterization has not been – and perhaps cannot be – phrased in the remaining frameworks.

**Theorem 7.** *Let $L \subseteq \widehat{A^+}$ and $K = \widehat{A^+} - L$ be its complement. The following conditions 1-9 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+inf$^-$*,*
4. *$L$ is $\uparrow$-recognizable,*
5. *$K$ is defined by an S-regular expression,*
6. *$K = L(\mathcal{B})$ for some S-automaton $\mathcal{B}$,*
7. *$K$ is definable in* MSO+inf$^+$*,*
8. *$K$ is $\downarrow$-recognizable,*
9. *The $\langle\,\cdot\,,\#\rangle$-syntactic congruence of $K$ has finite index and $K = \overline{K \cap A^{\langle\,\cdot\,,\omega\rangle}}$.*

In the last characterization, $A^{\langle\,\cdot\,,\omega\rangle}$ is the set of profinite words which can be generated from $A$ by applying multiplication and the $\omega$-power – they are analogues of ultimately periodic words in the theory of $\omega$-regular languages. It follows that a B- or S-regular language is determined by its elements contained in $A^{\langle\,\cdot\,,\omega\rangle}$, similarly as an $\omega$-regular language is determined by its ultimately periodic words.

By the last part of Theorem 5, the image of an $\infty$-homomorphism to a finite stabilization semigroup can be computed using a fixed point calculation. Hence, emptiness of recognizable languages is decidable. This proves the following.

**Theorem 8.** *Emptiness of Boolean combinations of B-regular languages is decidable. In particular, the limitedness problem is decidable for B-automata.*

The above result extends the decidability results of Hashiguchi and Kirsten. As emptiness of Boolean combinations reduces to inclusion testing, it is equivalent to the main result of [7] – that the domination relation is decidable for B-automata.

## 6 From infinite words to profinite words

We describe a connection between $\omega$-words (i.e. mappings from $\mathbb{N}$ to $A$) and profinite words. Recall that any $\omega$-regular language can be presented as a finite union of languages of the form $U \cdot V^\omega$, where $U, V \subseteq A^+$ are regular languages of finite words. We generalize this observation, and provide a meta-reduction between the satisfiability problems for logics over $\omega$-words to corresponding logics over profinite words. The proof resembles Büchi's original proof of decidability of MSO. Instead of the usual Ramsey lemma, we use the following observation (originating from [5]): For any $\omega$-word $w \in A^\omega$ there is a factorization $w = u_0 \cdot u_1 \cdot u_2 \cdots$ such that the sequence $u_0, u_1, u_2, \ldots$ is convergent to some $u_\infty \in \widehat{A^+}$. The proof is an easy, repeated application of the usual Ramsey lemma.

Let $V \subseteq \widehat{A^+}$ be a language of profinite words, and $\varepsilon > 0$ a real number. Consider the following language of infinite words $V_\varepsilon^\omega \subseteq A^\omega$:

$$V_\varepsilon^\omega \quad \stackrel{def}{=} \quad \{v_1 \cdot v_2 \cdot v_3 \cdots :\ \exists v_\infty \in V^* :\ \lim_{n\to\infty} v_n = v_\infty \ \textit{and}\ \forall_n d(v_n, v_\infty) < \varepsilon\}.$$

For a regular language $U \subseteq A^+$ of finite words, we say that the expression $U \cdot V^\omega$ is *well-formed* if the language $U \cdot V_\varepsilon^\omega$ does not depend on the choice of $0 < \varepsilon \leq 1/n$, where $n$ is the size of the minimal automaton recognizing $U$. In this case, we define the language $U \cdot V^\omega$ as $U \cdot V_\varepsilon^\omega$. For example, the expression $(a+b)^* \cdot (a^{<\infty}b)^\omega$ is well-formed and describes the language $L_B$ from the introduction. For a class $\mathcal{L}$ of languages of profinite words, let $\omega\mathcal{L}$ denote the class of all finite unions of languages defined by well-formed expressions $U \cdot V^\omega$ with $U \subseteq A^+$ regular and $V \in \mathcal{L}$.

In the following theorem, by REGULAR, B-REGULAR, S-REGULAR, MSO+inf, we denote the corresponding classes of languages of profinite words, and to each we apply the map $\mathcal{L} \mapsto \omega\mathcal{L}$ as described above, yielding classes of languages of infinite words. The proof of the theorem is very general. It generalizes Büchi's proof of decidability of MSO over infinite words.

**Theorem 9.** *Every $\omega$-regular language is in $\omega$*REGULAR*. Every $\omega$B-regular language is in $\omega$*B-REGULAR*. Every $\omega$S-regular language is in $\omega$*S-REGULAR*. Every* MSO+$\mathbb{B}$ *definable language is in $\omega$*MSO+inf*. The translations are effective.*

The reduction described above allows to transfer results from profinite words to $\omega$-words. For instance, the main results of [4] (concerning $\omega$B- and $\omega$S-regular languages) follow from the results in our paper. More importantly, we get:

**Corollary 2.** *The satisfiability problem for the logic* MSO+$\mathbb{B}$ *over $\omega$-words reduces to the satisfiability problem for the logic* MSO+inf *over profinite words.*

We mention that by refining our Theorem 9, Skrzypczak [14] proved that a language of infinite words which is both $\omega$B-regular and $\omega$S-regular must in fact be $\omega$-regular – reflecting the immediate, analogous fact for profinite words.

*Conclusion.* We presented a new proof and framework for the limitedness problem. We rise the question of decidability of the logic MSO+inf over profinite words.

## References


1. P. A. Abdulla, P. Krcál, and W. Yi. R-automata. In *CONCUR*, pages 67–81, 2008.
2. Mikołaj Bojańczyk. A bounding quantifier. In *Computer Science Logic*, volume 3210 of *Lecture Notes in Computer Science*, pages 41–55, 2004.
3. Mikołaj Bojańczyk. Beyond $\omega$-regular languages. In *STACS*, pages 11–16, 2010.
4. Mikołaj Bojańczyk and Thomas Colcombet. Bounds in $\omega$-regularity. In *Logic in Computer Science*, pages 285–296, 2006.
5. Mikołaj Bojańczyk and Eryk Kopczyński. Ramsey's theorem for colors from a metric space. Submitted, 2010.
6. 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.
7. Thomas Colcombet. Regular cost functions, part i: Logic and algebra over words. submitted, 2011.
8. Kosaburo Hashiguchi. Limitedness theorem on finite automata with distance functions. *Journal of Computer and System Sciences*, 24:233–244, 1982.
9. Daniel Kirsten. Distance desert automata and the star height problem. *Theoretical Informatics and Applications*, 39(3):455–511, 2005.
10. 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.
11. Hing Leung. On the topological structure of a finitely generated semigroup of matrices. *Semigroup Forum*, 37:273–278, 1988.
12. Jean-Éric Pin. Profinite methods in automata theory. In Susanne Albers and Jean-Yves Marion, editors, *STACS*, volume 3 of *LIPIcs*, pages 31–50. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2009.
13. Imre Simon. On semigroups of matrices over the tropical semiring. *ITA*, 28(3-4):277–294, 1994.
14. Michał Skrzypczak. Separation property for $\omega$B- and $\omega$S-regular languages. Submitted, January 2012.

# Appendix

The appendix is split into several sections.


## A. MSO+inf logic

We give a precise definition of the logic MSO+**inf**. .................... 14

## B. Finite topologies

We recall some basic facts about finite topological spaces. ................ 18

## C. Syntactic algebra

We recall some standard notions and properties of of abstract topological algebras, specialized to the case of topological $\langle\,\cdot\,,\#\rangle$-algebras, such as the syntactic congruence, and the quotient topological and algebraic structure. In Section C.2 we analyze quotients of the free profinite semigroup. .......................................... 20

## D. Profinite words and sequences

We pass some intuitions on profinite words and profinite numbers (profinite words over the unary alphabet). In Sections D.3 and D.5 we prove two properties of profinite words, which will be used later on: the amalgamation property, and the uniform lifting property for open and closed sets. .......................................... 27

## E. Proof of Theorem 5

We prove Theorem 5. In Section E.1 we define precisely the notion of an $\infty$-homomorphism. In Section E.2 we define the key tool, namely factorization trees, and state the Factorization Theorem. Using this theorem, in Section E.3 we prove the Thoeorem 5. In Section E.4 we prove a useful property of $\infty$-homomorphisms, called infinite continuity. .. 34

## F. Proof of the Factorization Theorem

We prove the Factorization Theorem. .............................. 40

## G. From homomorphisms to B/S-regular expressions

We prove the implications $\mathbf{4} \to \mathbf{1}$ and $\mathbf{8} \to \mathbf{5}$ of Theorem 7, namely that a $\downarrow$-recognizable language is definable by an S-regular expression, and that a $\uparrow$-recognizable language is definable by a B-regular expression. ........ 47

## H. The powerset constructions

We present two variants of the powerset construction for $\infty$-homomorphisms, and complete the proof of Proposition 6, by proving that $\downarrow$-recognizable and $\uparrow$-recognizable languages are closed under projections. .......................................... 50

## I. Proof of the main theorem

In this section, we finish the proof of Theorem 7 ...................... 56

## J. From $\omega$-words to profinite words

We prove Theorem 9, by providing a reduction from logics over infinite words to logics over finite words. .................................. 57

# A MSO+inf logic

We give a precise definition of the logic MSO+inf.

Following the tradition initiated by Büchi, we present a logic over profinite words, called MSO+inf, which describes languages recognized by B- or S-automata. The logic is an extension of the MSO logic over profinite words, which we also 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 [7]; in these extensions, a formula defines a cost function. However, cost functions cannot describe all of the languages in MSO+inf.

*Monadic Second Order Logic* (MSO) over profinite words is analogous to MSO over finite or $\omega$-words. In fact, the syntax of the logic is exactly the same as for finite or $\omega$-words – we allow first and second order quantification, Boolean operations, set inclusion tests, linear order tests and label tests. Remarkably, a formula $\varphi$ 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 $\varphi$.

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"* |

***Fig. 2.*** *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 $\inf(X)$ which – informally – tests whether $X$ is infinite. A bit more precisely, the formula $\inf(X)$ 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 $\inf(X)$, 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 2). 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+inf$^-$. 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 $\lambda, \lambda'$. The $A$-component of the letter $\lambda$ is denoted $\lambda_A$, and its

$X$-component is denoted $\lambda_X$. Given a profinite word $x$ over the alphabet $A$, an *$\mathscr{X}$-valuation* over $x$ is a profinite word $\nu$ over the $\mathscr{X}$-valuation alphabet, such that its projection onto the $A$-component is $x$.

We interpret an $\mathscr{X}$-valuation $\nu$ 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 $\lambda$ in $\nu$, if $\lambda$ 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 $\lambda$ of the $\mathscr{X}$-valuation alphabet to the letter obtained by omitting its $X$-coordinate. The resulting letter is a letter of a valuation alphabet over the set of variables $\mathscr{X} - \{X\}$. The mapping $\mathit{erase}_X$ naturally extends to a homomorphism from profinite words to profinite words, which maps an $\mathscr{X}$-valuation $\nu$ over $x$ to a $(\mathscr{X} - \{X\})$-valuation $\mathit{erase}_X(\nu)$ over $x$.

We define, for each formula $\varphi$ of MSO+inf with free variables $\mathscr{X}$, and each valuation $\nu$ over $x$, the expression "$x$ satisfies $\varphi[\nu]$", denoted $x \models \varphi[\nu]$, according to the table in Figure 3.

| | | | |
|---|---|---|---|
| $(\vee)$ | $x \models (\varphi \vee \psi)[\nu]$ | 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 $\lambda$ appearing in $\nu$, if $\lambda_X = 1$ then $\lambda_A = a$ |
| $(\subseteq)$ | $x \models (X \subseteq Y)[\nu]$ | iff | for every letter $\lambda$ appearing in $\nu$, if $\lambda_X = 1$ then $\lambda_Y = 1$ |
| $(<)$ | $x \models (X < Y)[\nu]$ | iff | for every pair of letters $\lambda, \lambda'$ such that $\nu$ 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 | $\nu$ contains infinitely many letters $\lambda$ such that $\lambda_X = 1$ |
| $(\exists)$ | $x \models (\exists X.\varphi)[\nu]$ | iff | for some $(\mathscr{X} \cup \{X\})$-valuation $\nu'$ such that $\mathit{erase}_X(\nu') = \nu$, $x \models \varphi[\nu']$ |
| $(\forall)$ | $x \models (\forall X.\varphi)[\nu]$ | iff | for every $(\mathscr{X} \cup \{X\})$-valuation $\nu'$ such that $\mathit{erase}_X(\nu') = \nu$, $x \models \varphi[\nu']$ |

**Fig. 3.** The semantic of the logic MSO+inf

For a formula $\varphi$ 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 $\varphi$ 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 $\varphi$.

*Remark 1.* 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.* 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 2 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.

One can further define a language which is not even a Boolean combination of open sets. First note that the language

$$K = ((a^{\infty} + b^{<\infty})c)^*$$

can be defined in MSO+inf: the formula says that every $a$-block is infinite and that every $b$-block is finite, and that between two consecutive $c$'s there is either an $a$-block, or a $b$-block. Therefore the projection $\pi(K)$ of $K$, where $\pi$ identifies $a$ with $b$, is also definable in MSO+inf. One can also describe $\pi(K)$ as follows. Let $A = \{a, c\}$. Then

$$\pi(K) = \{x \in \widehat{A^+} : \ \#\{n \in \mathbb{N} : \ ca^n c \text{ is an infix of } x\} < \infty\}.$$

The language $\pi(K)$ is not a Boolean combination of open sets. The language $\pi(K)$ described above corresponds to the language constructed in [4] as an example of a language which is $\omega$BS-regular, but whose complement is not.

It is not difficult to see that the following proposition holds. We leave it without a proof, as it is not used in this paper.

**Proposition 10.** *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* $b^*(ab^*)^{\infty}$.

## B Finite topologies

We recall some basic facts about finite topological spaces.

Our approach to analyzing B- and S-regular subsets of $\widehat{A^+}$ is to represent them as subsets of finite semigroups. However, there is a crucial topological information in $\widehat{A^+}$, which we do not want to loose: for instance, the B-regular language $a^*$ is strictly larger than the B-regular language $a^{<\infty}$, yet any element of $a^*$ can be approximated by elements from $a^{<\infty}$. This can be precisely formulated in terms of topology:

$$a^* \subseteq \overline{a^{<\infty}}.$$

We want our finite semigroups representing $\widehat{A^+}$ to be capable of determining when such approximations are possible. For this, we consider finite topological spaces. A finite set can be only equipped with finitely many topologies. In this section, we present a standard way of defining a topology over a finite set $X$ using a preorder, called the *specialization preorder* over $X$.

A *finite topological space* is a topological space with finitely many points, i.e. a finite set $X$ equipped with a topology. Throughout Appendix B we assume that $X$ is finite, and the 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 singleton sets 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 $T_1$.

*Example 10.* The *Sierpiński space* consists of two elements, 1 and $\omega$, such that $\omega$ is the only closed singleton set. Therefore, the closed subsets of the Sierpiński space are $\emptyset, \{\omega\}, \{1, \omega\}$.

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 11.* In the Sierpiński space, the induced preorder results in $\omega < 1$. Hence, it is $T_0$.

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 a topological space $X$ is closed iff it is downward-closed with respect to the specialization preorder. Conversely, if $(X, \leq)$ is a finite partially preordered set, then we may define a topology on $X$, 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 $X$, for which $\leq$ is the specialization preorder.

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 $X$. Recall that a mapping between topological spaces is continuous if and only if the inverse image of a closed set is closed. 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.

**Corollary 3.** *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. 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 $Y = \{y\}$. We denote the smallest upward-closed set

# C Syntactic algebra

We recall some standard notions and properties of of abstract topological algebras, specialized to the case of topological $\langle\,\cdot\,,\#\rangle$-algebras, such as the syntactic congruence, and the quotient topological and algebraic structure. In Section C.2 we analyze quotients of the free profinite semigroup.

## C.1 Syntactic congruence

We will define a congruence induced by a language $L \subseteq \widehat{A^+}$, which respects multiplication and the $\omega$-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 $\omega$-power in $\widehat{A^+}$), and arbitrary elements of $\widehat{A^+}$ as constants (i.e. in a leaf of the term). Note that any such term $\tau$ 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 $\omega$.

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 $\alpha_L$ denote the canonical projection from $\widehat{A^+}$ to $S_L$.

*Example 12.* Let $A = \{a, b\}$ and let $L$ denote set those profinite words over $\{a, b\}$ which contain infinitely many $a$'s. Then $L$ is a closed set.

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.

**Algebraic structure**

**Lemma 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

$$\tau \in \mathit{Terms}(\widehat{A^+}, \cdot, \#) \qquad \mathit{and} \qquad \tau(x \cdot y) \in L.$$

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 $\alpha_L$ an associative operation in $S_L$, which we also denote $s, t \mapsto s \cdot t$. Similarly, the $\omega$-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 $\alpha_L$ 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 $\alpha_L$ becomes a continuous homomorphism.

**Proposition 11.** *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 $\alpha_L$ satisfies the second condition of the proposition.

*2 ⇒ 1.* Let $\alpha, S, F$ 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 $\alpha$ 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 $\tau$ induces a term $\alpha_*(\tau)$ over $(S, \cdot, \#)$, obtained by applying $\alpha$ to the leaves of the term $\tau$, and interpreting the operations $\cdot$ and $\omega$ of $\widehat{A^+}$ as the operations $\cdot$ and $\#$ of $S$. The term $\alpha_*(\tau)$ has one free variable. Since $\alpha$ 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 $x \simeq_L y$.

**Topological structure** We will define a topology over $S_L$ for which $\alpha_L$ becomes a continuous mapping. We use the notion of a *specialization preorder* described in Appendix B.

Note that usually, when the discrete topology is considered over $S_L$, $\alpha_L$ is *not* continuous, as the following example demonstrates.

*Example 13.* Consider the quotient mapping $\alpha_L \colon \widehat{A^+} \to S_L$ from Example 12. Let us denote by $0, 1, \omega$ the elements of $S_L$ corresponding to the $\simeq_L$-equivalence classes $L_0, L_1, L_\omega$, 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 $a^\omega$ of $L_\omega$. Therefore, $\alpha_L$ 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 $\alpha_L$ 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 $\omega$. 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 $L_\omega$ and that $L_0$ is closed.

In general, it is natural to define the topology of $S_L$ as the *quotient topology* induced by $\alpha_L$. 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 $\alpha_L$ 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 $\alpha_L$ 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 12.** *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 $\alpha_L$ are continuous mappings, and $\alpha_L$ 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 13.** *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 $\varphi$ here). Let $\mu$ be right-multiplication by $s_0$ and $\nu$ be right-multiplication by $t_0$. Note that $\mu$ is a continuous mapping from $S$ to $S$. The mappings $\mu, \nu$ 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:

$$\varphi^{-1}\left(\nu^{-1}(U)\right) = \mu^{-1}\left(\varphi^{-1}(U)\right). \tag{4}$$

Since $U$ is open and both $\mu$ and $\varphi$ are continuous, we deduce that the set in the formula (4) is an open subset of $S$. We therefore conclude that $\nu^{-1}(U)$ is open, since its inverse image under $\varphi$ is open by (4). 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-multiplication, both factors of any disjunct in the above union are open sets. Consequently, $V$ is open as a finite union of products of open sets.

This proves that multiplication is continuous in $T$. A similar, but simpler argument also works for stabilization. Therefore, $T$ equipped with the operations of multiplication and stabilization becomes a topological $\langle\,\cdot\,,\#\rangle$-algebra, and $\varphi$ is a continuous mapping of such algebras.

The second part of Theorem 12 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 $\alpha_L$ and $\alpha_K$ are the same mappings, so they induce the same topology over $S_L = S_K$.

**Lemma 2.** *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,

$$\begin{aligned}\downarrow x &= \{y : \ y \preceq_L x\} \\ &= \big\{y : \ \forall \tau. \quad \tau(x) \in L \implies \tau(y) \in L\big\} \\ &= \bigcap_{\tau:\ \tau(x)\in L} \{y : \tau(y) \in L\}.\end{aligned}$$

In the above formulas, $\tau$ ranges over all elements in $\mathit{Terms}(\widehat{A^+},\,\cdot\,,\#)$.

Since any term $\tau$ 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 $\alpha_L$, 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 $\alpha_L$.

This finishes the proof of Theorem 12.

### C.2 Stabilization semigroups

As we have seen, in general, there might be several distinct topologies over $S_L$ for which the quotient mapping $\alpha_L$ 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 topological $\langle\,\cdot\,,\#\rangle$-algebra which is a continuous homomorphic image of any profinite semigroup.

**Proposition 14.** *Let $\alpha$ 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.*

$$s \cdot (t \cdot s)^{\#} = (s \cdot t)^{\#} \cdot s \tag{S1}$$

$$(s^n)^{\#} = s^{\#} \qquad \textit{for } n = 1, 2, 3 \ldots \tag{S2}$$

$$(s^{\#})^{\#} = s^{\#} \tag{S3}$$

$$s^{\#} \cdot s^{\#} = s^{\#} \tag{S4}$$

$$e \cdot e^{\#} = e^{\#} \qquad \textit{if } e = e \cdot e. \tag{S5}$$

*Moreover, for the specialization preorder $\leq$ over $S$,*

$$e^{\#} \leq e \qquad \textit{if } e = e \cdot e. \tag{S6}$$

*Proof.* The first four equalities are an immediate consequence of the corresponding equalities in $\tilde{S}$ and the fact that $\alpha$ is surjective.

Now assume that $e = e^{\#}$, and let $x$ be such that $e = \alpha(x)$. 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 $x^{\omega}$ 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 $\alpha$. In particular, $x^n \in \alpha^{-1}(F)$ for $n = 1, 2, \ldots$. Since $\alpha^{-1}(F)$ is closed, and $x^{\omega}$ 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 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 1.** *We call a* stabilization semigroup *a topological semigroup $S$ endowed with a continuous operation $\#$ which satisfies the axioms* (S1)-(S6) *listed above, where $\leq$ is the specialization preorder of $S$.*

*Remark 3.* The notion of a stabilization semigroup was first introduced by Colcombet in [6]. The definition of Colcombet differs from ours slightly. 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) takes a different form, more suited for idempotents. The remaining axioms are similar. Despite these slight differences, there is a bijective correspondence between our topological $T_0$ stabilization semigroups and the stabilization semigroups of Colcombet. Because of that, we use the name “stabilization semigroup”.

# D Profinite words and sequences

We pass some intuitions on profinite words and profinite numbers (profinite words over the unary alphabet). In Sections D.3 and D.5 we prove two properties of profinite words, which will be used later on: the amalgamation property, and the uniform lifting property for open and closed sets.

Profinite words have certain features which make them similar with finite words, and certain features which make them quite different from finite words. We will describe these features shortly.

A profinite word has a first position, a second position, or a last position, or a penultimate position. A finite profinite word is just a usual finite word. On the other hand, it is useful to imagine that an infinite profinite word is like a very long word, from the point of view of an automaton. An automaton can see subsequent letters of an input word, but if the word is very long, then it cannot distinguish two positions which have similar surroundings. This is formalized by the pumping lemma for automata, and for sufficiently long words. A similar property holds for infinite profinite words: if $x$ is an infinite profinite word, then there exist $u, y, v \in \widehat{A^+}$ such that

$$x = uyv = uy^2v = uy^3v = \dots$$

Note that the above property obviously does not hold for finite words.

Finite words over the unary alphabet form a semigroup, isomorphic to the natural numbers. As we will see, however, infinite profinite words form a set which is uncountable. Moreover, this set has a group structure. These properties distinguish profinite words from finite words. However, we will be mostly using two properties of profinite words which obviously holds in the context of finite words. These properties are described in Section D.3 and Section D.5.

## D.1 Compactness property of profinite words

The set of profinite words forms a compact space, but also each profinite word alone has a certain "compactness" property, described by the following lemma.

**Lemma 3.** *Let $x \in \widehat{A^+}$ be a profinite word, and let $u_1, u_2, u_3, \dots$ be a convergent sequence of infixes of $x$, whose limit is $u_\infty \in \widehat{A^+}$. Then, $u_\infty$ is also an infix of $x$.*

*Proof.* By assumption, for each $n$, there exist $y_n, z_n \in \widehat{A^+}$ such that

$$x = y_n \cdot u_n \cdot z_n.$$

By restricting to a subsequence if necessary, by compactness of the set of profinite words, we may assume that for some $y_\infty, z_\infty \in \widehat{A^+}$,

$$y_\infty = \lim_{n\to\infty} y_n$$
$$z_\infty = \lim_{n\to\infty} z_n.$$

Then, by continuity of multiplication,

$$x = \lim_{n \to} (y_n \cdot u_n \cdot z_n) = y_\infty \cdot u_\infty \cdot z_\infty.$$

In particular, $u_\infty$ is an infix of $x$.

## D.2 Profinite numbers

Let us consider for a moment a unary alphabet $A = \{a\}$. Finite profinite words over $A$ are

$$a, aa, aaa, aaaa, \ldots.$$

What are the infinite profinite words like? To answer this question, we ask: what are the regular languages over $A$ like? Regular languages over a unary alphabet have a very simple description. For two natural numbers $0 \leq k < n$, let us write

$$|w| \mod n = k$$

if the length of $w$ is equal to $k$ modulo $n$. We denote by $L_{n,k}$ the set of words with the above property. Basically, a regular language over $A$ can be of two forms:

- $\{a^n\}$, for some number $n \in \mathbb{N}$
- $L_{n,k}$, for some numbers $0 \leq k < n$.

In general, any regular language over $A$ is a finite union of languages of the above form.

For any profinite word $x \in \widehat{A^+}$ and number $n$, there is exactly one $0 \leq k < n$ such that $x \in L_{n,k}$. We write then that $|x| \mod n = k$. Recall that any profinite word $x$ is uniquely specified by the family of regular languages it ultimately belongs to.

From the description of unary regular languages, it follows that a profinite word $x$ over $A$ is:

- Either a finite word of the form $a^n$ for some $n \in \mathbb{N}$
- Or is infinite, and uniquely specified by the sequence

$$k_1 = (|x| \mod 1), \quad k_2 = (|x| \mod 2), \quad k_3 = (|x| \mod 3), \quad \ldots \quad k_n = (|x| \mod n), \ldots.$$

Therefore, an infinite profinite word $x$ describes a sequence $k_1, k_2, k_3, \ldots$ such that $0 \leq k_n < p_n$, and $x$ is uniquely specified by this sequence. However, the sequence $k_1, k_2, \ldots$ satisfies some additional constraints. It cannot be the case, for instance, that $|x| \mod 2 = 0$ and $|x| \mod 4 = 1$. More generally:

> If $|x| \mod n = k$ and $m \geq 1$, then $|x| \mod (mn) = k + mi$ for some $i \in \mathbb{N}$.

We call a sequence $k_1, k_2, k_3, \ldots$ such that $0 \leq k_n < n$ which satisfies the above constraint *consistent*. Observe that there are uncountable many consistent sequences, since we can choose $|x| \mod p$ for each prime number $p$ independently.

We now show that any consistent sequence $k_1, k_2, \ldots$ can be obtained in the above way from some profinite word $x$. It follows from the Chinese remainder theorem that for any $r \in \mathbb{N}$, there exists some finite number $m_r$ such that $m \bmod n = k_n$ for $n = 1, 2, \ldots, k_r$. Let $w_r$ be the corresponding word, i.e.

$$w_r = a^{m_r}.$$

Then, the sequence of finite words $w_1, w_2, w_3, \ldots$ is a convergent sequence, and its equivalence class $x$ satisfies the property that $|x| \mod n = k_n$ for every $n \in \mathbb{N}$.

It follows that the set of infinite profinite words is uncountable and isomorphic to a subset $S$ of the infinite Cartesian product

$$\mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_5 \ldots,$$

where $\mathbb{Z}_n$ is the cyclic group of order $n$. Moreover, the described isomorphism can be seen to be a semigroup isomorphism. Therefore, $S$ is a subsemigroup of the above product of finite groups. It follows easily that $S$ must be itself a group, and that the described isomorphism is a group isomorphism. Therefore, the set of infinite profinite words over the unary alphabet has a groups structure. The neutral element of this group is the profinite word $a^\omega$, which is the limit of the sequence $a^{1!}, a^{2!}, a^{3!}, \ldots$, and thus has the property that $|a^\omega| \mod n = 0$ for every number $n$. This group contains the integer numbers as a subgroup: this is the subgroup generated by $a^{\omega+1} \stackrel{def}{=} a^\omega \cdot a$. The inverse of $a^\omega$ is an element denoted $a^{\omega-1}$.

We call the profinite words over the unary alphabet *profinite numbers*. As profinite words over the unary alphabet are often written using exponents, like $a^1, a^2, a^3, a^\omega, a^{\omega-1}, a^{\omega+1}$, when we refer to them as profinite numbers, we will denote them simply $1, 2, 3, \omega, \omega-1, \omega+1$, etc. Note that as there are uncountably many profinite numbers, not every profinite number is listed in this list.

### D.3 The amalgamation property of profinite words

We will be relying on a property of profinite words, which is obvious in the case of finite words. Let us first illustrate with an example in the finite case. Let $A_0 = \{a, b\}$, $A_1 = A_0 \cup \{\spadesuit, \clubsuit\}$ and $A_2 = A_0 \cup \{\heartsuit, \diamondsuit\}$. Consider a word $w_0$ over the alphabet $A_0$, such as

$$w_0 = baba.$$

Now, assume that we are given two extensions of $w_0$, over the extended alphabets $w_1$ and $w_2$, respectively:

$$\begin{aligned} w_1 &= \clubsuit b \clubsuit a \spadesuit b \clubsuit \spadesuit a \clubsuit \\ w_2 &= \diamondsuit b \heartsuit a \diamondsuit b \heartsuit \diamondsuit a. \end{aligned}$$

Note that $w_1$ can be obtained from $w_0$ by inserting some letters from $A_1 - A_0$, and $w_2$ can be extended from $w_1$ by inserting some letters from $A_2 - A_0$. Then, there exists a word over the alphabet $A_1 \cup A_2$, such as

$$w = \clubsuit\diamondsuit b \clubsuit \heartsuit a \spadesuit \diamondsuit b \clubsuit \spadesuit \heartsuit \diamondsuit a \clubsuit,$$

which is both an extension of $w_1$, and of $w_2$.

We will use an analogous property of profinite words. Let $A \subseteq B$ be two alphabets. There exists a natural mapping $\hat{\pi}$ from $\widehat{B^*}$ to $\widehat{A^*}$ which ignores all the letters from outside of $A$. To define it formally, first, let $\pi\colon B \to \widehat{A^*}$ be defined by

$$\pi(b) = \begin{cases} b & \text{if } b \in A \\ \varepsilon & \text{if } b \notin A. \end{cases}$$

Then, we define $\hat{\pi}\colon B \to \widehat{A^*}$ as the unique continuous homomorphism extending $\pi$.

**Lemma 4.** *Let $A_1$ and $A_2$ are two alphabets, and let $A_0 = A_1 \cap A_2$, and $A = A_1 \cup A_2$. Let $x_0 \in \widehat{A_0}^*$, $x_1 \in \widehat{A_1}^*$ and $x_2 \in \widehat{A_2}^*$ be such that both $x_1$ and $x_2$ are extensions of $x_0$. Then there exists a word $x \in \widehat{A^+}$ which is both an extension of $x_1$ and of $x_2$.*

### D.4 Profinite sequences

Let $\sigma$ be a profinite sequence of words over the alphabet $A$, where the separator symbol is $\dagger$. By erasing in $\sigma$ all letters of $A$, we obtain a profinite word $x$ over the unary alphabet. The *length* of a profinite sequence is the profinite number of occurrences of $\dagger$ in the word $x\dagger$. For instance, the length of the sequence $ab \dagger ba$ is the profinite number 2, and the length of the profinite sequence $(a\dagger)^\omega a$ is the profinite number $\omega + 1$. To imitate the notation for finite sequences, by

$$x_1, x_2, \ldots, x_\alpha,$$

we denote a profinite sequence of length $\alpha$, whose first element is $x_1$, second element is $x_2$, last element is $x_\alpha$.

Let $\sigma$ be a profinite sequence. The *concatenation* of $\sigma$, denoted $\Pi(\sigma)$ is obtained from $\sigma$ by removing the separator symbols $\dagger$. We will also say that $\sigma$ is a *factorization* of $\Pi(\sigma)$. We will call elements of $\sigma$ *factors* of the factorization. Suppose we are given two factorizations $\sigma, \tau$ of the same profinite word $x$. We may assume that in the factorization $\sigma$, the separator symbol is $\dagger_1$, and that in the factorization $\tau$, the separator symbol is $\dagger_2$. Then, both $\sigma$ and $\tau$ are extensions of $x$. By Lemma 4, there exists a common extension, call it $\psi$, of $\sigma$ and $\tau$.

Let us fix $\psi$ as described above. Because both $\sigma$ and $\tau$ are "embedded" in the same word $\psi$, we may refer to properties of factors of $\sigma$, which relate them with factors of $\tau$. More precisely, in the context of $\psi$, a *factor* of $\sigma$ is a word $x$ such that $\dagger_1 x \dagger_1$ is an infix of $\psi$. The word $x$, itself may contain some symbols $\dagger_2$, so we may view it as a further factorization. Let $\lambda$ be the length of this factorization. We then say that the factor $x$ of $\sigma$ *intersects* $\lambda$ factors of $\tau$.

### D.5 The uniform lifting property

Let $A, B$ be two finite alphabets. Let † be a separator symbol not belonging to $A \cup B$. For a set $L \subseteq \widehat{A^+}$, let

$$L^\dagger$$

denote the set of all profinite sequences whose all elements belong to $L$. Let

$$\pi \quad : \quad A \quad \longrightarrow \quad B^*$$

be any "substitution" (note that $\pi$ can also erase letters). We also implicitly assume that $\pi(\dagger) = \dagger$. Then, $\pi$ extends in a unique way to a homomorphism which we denote with the same symbol:

$$\pi \quad : \quad (\widehat{A^*})^\dagger \quad \longrightarrow \quad (\widehat{B^*})^\dagger\,.$$

We say that $L \subseteq \widehat{A^+}$ has the *uniform lifting property* if

$$\pi(L)^\dagger = \pi(L^\dagger)$$

for every $\pi \colon A \to B^*$. Note that the right-to-left inclusion above always holds. It is the left-to-right inclusion which might fail.

The uniform lifting property has some resemblance with the axiom of choice. Namely, $L$ has the uniform lifting property if and only if for every $\pi \colon A \to B^*$ and for every profinite sequence $\hat{y}$:

$$y_1, y_2, \ldots, y_\gamma$$

if for each element $y_i$ of $\hat{y}$ there exists some $x_i \in L$ such that $\pi(x_i) = y_i$, then there exists a profinite sequence $\hat{x}$:

$$x_1, x_2, \ldots, x_\gamma$$

such that $\hat{x}$ maps to $\hat{y}$.

It might seem that all sets should have the uniform lifting property.

**Proposition 15.** *Closed languages and open languages have the uniform lifting property.*

However, the above proposition is as much as we can get: we will see an example of a union of a closed set with an open set which does not have the uniform lifting property.

*Example 14.* Let $A = \{a, b\}$ and let $B = \{a\}$, and let

$$\pi \quad : \quad A \quad \longrightarrow \quad B$$

map $a$ and $b$ to $b$. Let

$$L = a^{<\infty} \cup b^{\infty}.$$

Then,

$$\pi(L) = a^*.$$

Let $\hat{x} \in \pi_A(L)^\dagger$ be a profinite sequence such that every element of $a^*$ is an element of $\hat{x}$.

Then,

$$\hat{x} \in \pi_A(L)^\dagger.$$

However, we will see that

$$\hat{x} \notin \pi_A(L^\dagger).$$

Indeed, assume the opposite, i.e. that there is a profinite sequence $\hat{y} \in L^\dagger$ which projects to $\hat{x}$ under $\pi_A$. Then, $\hat{y}$ must contain the elements

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

From Lemma 3 it follows that $\hat{y}$ contains the element $a^\omega$ as well. But $a^\omega \notin L$, contradicting the assumption that $\hat{y} \in L^\dagger$. Therefore,

$$\hat{x} \in \pi_A(L)^\dagger - \pi_A(L^\dagger).$$

We now prove Proposition 15, that both closed and open sets have the uniform lifting property. First we prove this for sets which are both closed and open.

**Lemma 5.** *Clopen sets have the uniform lifting property.*

*Proof.* This follows from the fact that an analogous property obviously holds for regular languages of finite words, and hence for clopen sets.

**Lemma 6.** *Closed sets have the uniform lifting property.*

*Proof.* Let $L \subseteq \widehat{A^+}$ be a closed set. We will prove that

$$\pi(L)^\dagger \subseteq \pi(L^\dagger).$$

There exists a descending sequence $L_1 \supseteq L_2 \supseteq L_3 \supseteq \ldots$ of clopen sets such that

$$L = \bigcap_{n=1}^{\infty} L_n.$$

Let $\hat{y} \in \pi(L)^\dagger$. In particular, $\hat{y} \in \pi(L_n)^\dagger$ for $n = 1, 2, \ldots$. By the previous lemma, this implies that $\hat{y} \in \pi(L_n^\dagger)$ for $n = 1, 2, \ldots$.

Let $F = \pi^{-1}(\hat{y})$. Then, $F$ is a closed subset of $(\widehat{A^+})^\dagger$, such that for each $n$, $L_n^\dagger \cap F$ is nonempty. Therefore, the we have a descending sequence of closed, nonempty sets:

$$L_1^\dagger \cap F \quad \supseteq \quad L_2^\dagger \cap F \quad \supseteq \quad L_3^\dagger \cap F \quad \supseteq \quad \ldots.$$

By compactness, the intersection of the above sequence,

$$\bigcap_{n=1}^{\infty} L_n^{\dagger} \cap F$$

is nonempty. Observe that

$$\bigcap_{n=1}^{\infty} L_n^{\dagger} = L^{\dagger}.$$

Hence $L^{\dagger} \cap F$ is nonempty, which shows that there exist some $\hat{x} \in L^{\dagger}$ such that $\pi(\hat{x}) = \hat{y}$.

**Lemma 7.** *Open sets have the uniform lifting property.*

*Proof.* Assume that $L \subseteq \widehat{A^+}$ is an open set. Let us fix a sequence $L_1, L_2, \ldots$ of clopen subsets of $\widehat{A^+}$, such that

$$L = \bigcup L_n.$$

Let $\hat{y} \in \pi(L)^{\dagger}$. Consider the set $F \subseteq \widehat{B^+}$

$$F = \{y : \ y \text{ is an element of } \hat{y}\}.$$

Since $\hat{y} \in \pi(L)^{\dagger}$, it follows that

$$F \subseteq \pi(L) = \bigcup_{k=1}^{\infty} \pi(L_k). \tag{1}$$

By Lemma 3, the set $F$ is closed in $\widehat{B^+}$, hence compact. On the other hand, each of the sets $\pi(L_n)$ is open. From compactness of $F$, it follows that $F$ is covered by finitely many sets $\pi(L_k)$, i.e. for some $n \in \mathbb{N}$,

$$F \subseteq \bigcup_{k=1}^{n} \pi(L_k) = \pi(\bigcup_{k=1}^{n} L_k).$$

Let

$$K = \bigcup_{k=1}^{n} L_k \subseteq L.$$

Then, $K$ is a clopen set, as a finite union of clopen sets. Moreover, by (1), the elements of $\hat{y}$ are contained in $\pi(K)$, i.e.

$$\hat{y} \in \pi(K)^{\dagger}.$$

Applying the already proved case for clopen sets, Lemma 5 to $K$, we get that

$$\hat{y} \in \pi(K)^{\dagger} = \pi(K^{\dagger}) \subseteq \pi(L^{\dagger}).$$

This concludes the lemma, ending the proof of Proposition 15.

# E Proof of Theorem 5

We prove Theorem 5. In Section E.1 we define precisely the notion of an $\infty$-homomorphism. In Section E.2 we define the key tool, namely factorization trees, and state the Factorization Theorem. Using this theorem, in Section E.3 we prove the Thoeorem 5. In Section E.4 we prove a useful property of $\infty$-homomorphisms, called infinite continuity.

## E.1 Invariance under infinite substitutions

The formal definition of *invariance under infinite substitutions* is slightly technical. Let $\beta\colon \widehat{A^+} \to S$ be any mapping to a finite set $S$.

A *substitution scheme* is a profinite sequence $\sigma$ of triples of the form $(x_i, s_i, y_i)$, where:

- $x_i \in \widehat{A^+}$,
- $s_i \in S$,
- $y_i \in \widehat{A^+}$.

By "reading" just the $x_i$'s, i.e. by replacing each triple $(x_i, s_i, y_i)$ by $x_i$, we obtain a profinite sequence $x_1, x_2, \ldots, x_\gamma$, called the *factorized source* of $\sigma$. The concatenation of the profinite sequence $x_1, x_2, \ldots, x_\gamma$ is called the *source* of $\sigma$. Similarly, by reading just the $y_i$'s, we obtain a profinite sequence $y_1, y_2, \ldots, y_\gamma$, called the *factorized target* of $\sigma$, whose concatenation is the *target* of $\sigma$.

The $s_i$'s serve for testing consistency of a substitution scheme. A substitution scheme is *consistent* with the mapping $\beta\colon \widehat{A^+} \to S$ if for every triple $(x_i, s_i, y_i)$ in $\sigma$,

$$\beta(x_i) = s_i = \beta(y_i).$$

We say that $\beta$ is *invariant under infinite substitutions*, if for any substitution scheme $\sigma$ which is consistent with $\beta$,

$$\beta(x) = \beta(y),$$

where $x$ is is the source of the scheme $\sigma$ and $y$ is its target.

If $S$ is a semigroup and $\beta\colon \widehat{A^+} \to S$ is a homomorphism which is invariant under infinite substitutions, then we also say that it is an $\infty$-*homomorphism.*

**Lemma 8.** *Let $\beta\colon \widehat{A^+} \to S$ be an $\infty$-homomorphism to a finite $\langle \cdot, \# \rangle$-algebra. Then, for each idempotent $e \in S$*

$$\text{if} \quad x \in \beta^{-1}(e)^\infty \quad \text{then} \quad \beta(x) = e^\#.$$

*Proof.* Assume that $\beta$ is a semigroup homomorphism which is invariant under infinite substitutions. Let $L = \beta^{-1}(e)$, and assume that $x \in L^\infty$. We must show that $\beta(x) = e^\#$. Let us choose any $x_0 \in L$. Since $x \in L^\infty$, there exists a profinite sequence $\hat{x} \in L^\dagger$ such that $\Pi(\hat{x}) = x$. Out of $\hat{x}$ we produce a substitution scheme $\sigma$, as illustrated below:

$$\begin{array}{llll} \hat{x} = & x_1, & x_2, & x_3, \quad \ldots \\ & & & \\ \sigma = & (x_1, e, x_0), & (x_2, e, x_0), & (x_3, e, x_0), \ldots \end{array}$$

It is easy to see that the source and target of $\sigma$ are $x$ and $x_0^\gamma$, respectively, where $\gamma$ some infinite profinite exponent, i.e. a profinite word over the unary alphabet.

Since $\beta$ is an $\infty$-homomorphism, it follows that $\beta(x) = \beta(x_0^\gamma)$. Since infinite profinite exponents form a group, we may write $\gamma = \gamma^{-1} \cdot \gamma \cdot \gamma$, and accordingly write

$$x_0^\gamma = x_0^{\gamma^{-1}} \cdot x_0^\gamma \cdot x_0^\gamma.$$

Now, using an appropriate substitution scheme, we prove that $\beta(x_0^\gamma) = \beta(x_0^\gamma)^2$ – indeed, we may use a substitution scheme which "duplicates" each occurrence of $x_0$, and therefore preserves $\beta$, since $\beta(x_0) = e = e^2 = \beta(x_0 x_0)$. Reassuming, we have:

$$\begin{aligned} &\beta(x) = \beta(x_0^\gamma) = \\ &\quad \beta(x_0^{\gamma^{-1}} \cdot x_0^\gamma \cdot x_0^\gamma) = \beta(x_0^{\gamma^{-1}}) \cdot \beta(x_0^\gamma)^2 = \beta(x_0^{\gamma^{-1}}) \cdot \beta(x_0^\gamma) = \\ &\quad \beta(x_0^{\gamma^{-1}} \cdot x_0^\gamma) = \beta(x_0^\omega) = \beta(x_0)^\# = e^\#. \end{aligned}$$

### E.2 Factorization trees for stabilization semigroups

In this section we formulate a version of Simon's Factorization Forest Theorem for profinite words and stabilization semigroups.

Let us fix a mapping $\alpha\colon A \to S$ from a finite alphabet to a finite stabilization semigroup. For each $h \geq 1$, a factorization tree $f$ of height $h$ is a profinite sequence

$$s, f_1, f_2, \ldots, f_\gamma \tag{1}$$

consisting of:

- The *output* $s$ of $f$, which is an element of $S$, and
- The sequence of *child trees*, which is a nonempty profinite sequence $f_1, f_2, \ldots, f_\gamma$ of factorization trees of height at most $h - 1$. The length $\gamma$ of this sequence is the *rank* of the root of $f$.

The factorization tree $f$ is moreover subdue to the following restrictions.

**Base rule.** If there is only one child tree, then it is a single letter $a \in A$ (thought of as a degenerated factorization tree of height 0). Then the output of $f$ is equal to $\alpha(a)$, and $f$ has height 1.

**Binary rule.** If there are only two child trees $f', f''$ and their outputs are $s', s'' \in S$ respectively, then the output of $f$ is equal to $s' \cdot s''$.

**Idempotent rule.** Otherwise, there is an idempotent $e \in S$ such that each child tree has output $e$. If the rank of the root of $f$ is finite, then the output of $f$ is equal to $e$. If the rank of the root of $f$ is infinite, then the output of $f$ is equal to $e^{\#}$.

Formally, factorization trees of height $h$ are profinite words over an alphabet $B_h$ defined inductively:

- $B_0 = A$
- $B_h = B_{h-1} \cup \{\dagger_h\} \cup S$ for $h \geq 1$, where $\dagger_h \notin B_{h-1}$ is a separator symbol, used for encoding the profinite sequence (1).

The *input* of a factorization tree $f$ is the profinite word $x \in \widehat{A^+}$ obtained by erasing all the symbols from outside of $A$. Formally, we consider the image of $f$ under the unique continuous homomorphism

$$\pi_A \quad : \quad \widehat{B_h^+} \quad \longrightarrow \quad \widehat{A^+}$$

which maps a letter $b \in B_h$ to $b$ if $b \in A$ and to $\varepsilon$ otherwise. We will also say that $f$ is a factorization tree *over* $x$.

A factorization tree can be visualized as a tree, where the base rule corresponds to leaves, the binary rule corresponds to nodes of outdegree two, and the idempotent rule corresponds to nodes of a degree which might be any profinite number.

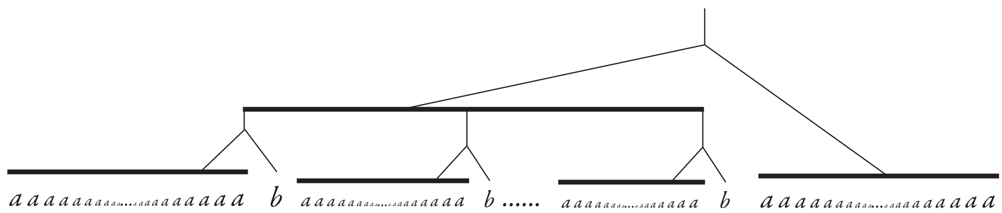


**Fig. 4.** A factorization tree of height 4 and input $(a^\omega b)^\omega a^\omega$

*Remark 4.* The set $T_s^h$ of all factorization trees of height $h$ and output $s$ is a language of profinite words over the alphabet $B_h$, which can be described by a formula of MSO+inf of the form

$$\forall X.(\mathrm{fin}(X) \implies \varphi(X)) \quad \wedge \quad \forall X.(\mathrm{inf}(X) \implies \psi(X)),$$

where $\varphi$ and $\psi$ are formulas of FO. The set of all profinite words over the alphabet $A$ which have a factorization tree of height $h$ and output $s$ is then the projection of $T_s^h$ under the mapping $\pi_A$.

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 $\widehat{A^+}$ 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, and that for a given input, the output does not depend on the choice of the factorization tree. This is stated below.

**Factorization Theorem 16.** *For any finite stabilization semigroup $S$, $\|S\|$ is finite. Any two factorization trees over the same input word have the same output.*

We relegate the proof of the theorem to Appendix F.

### E.3 Proof of Theorem 5

Assuming the Factorization Theorem, we show how to prove Theorem 5.

Let $\alpha$ be as in the statement above. We list three claims, which follow easily from the Factorization Theorem, and together yield the theorem above.

*Claim.* There is at most one $\infty$-homomorphism $\hat{\alpha}$ which extends $\alpha$.

Indeed, assume that $\hat{\alpha}(x)$ is some $\infty$-homomorphism from $\widehat{A^+}$ to $S$. Then, for any factorization tree $f$ with input $x$, the output of $f$ is necessarily equal to $\hat{\alpha}(x)$. This follows by induction on the height of $f$, and from Lemma 8.

It still remains to prove that there exists *some* $\infty$-homomorphism $\hat{\alpha}$ extending $\alpha$. The definition of $\hat{\alpha}(x)$ for $x \in \widehat{A^+}$ is natural: let $\hat{\alpha}(x)$ be the output of some chosen factorization tree with input $x$. Clearly, $\hat{\alpha}$ extends $\alpha$.

*Claim.* The mapping $\hat{\alpha}$ is an $\infty$-homomorphism.

We show that $\hat{\alpha}$ is a homomorphisms which is invariant under infinite substitutions. To prove that $\hat{\alpha}$ preserves multiplication, let

- $f$ be a factorization tree with input $x$ and output $\hat{\alpha}(x)$,
- $g$ be a factorization tree with input $x$ and output $\hat{\alpha}(y)$,
- $h$ be a factorization tree with input $xy$ and output $\hat{\alpha}(xy)$.

Then, $(f, g)$ is a factorization tree with input $xy$. Hence, by the second part of the Factorization Theorem, the output of $(f, g)$ is equal to the output of $h$.

To prove that $\hat{\alpha}$ is invariant under infinite substitutions, we proceed very similarly, but deal with infinite factorizations of the input word.

The last part of Theorem 5 also follows easily. Let $\alpha(A)^{\langle \cdot, \# \rangle} \subseteq S$ be the smallest subset of $S$ which contains $\alpha(A)$, and is closed under multiplication and stabilization. Then $\alpha(A)^{\langle \cdot, \# \rangle}$ is a stabilization semigroup.

*Claim.*

$$\hat{\alpha}(\widehat{A^+}) = \alpha(A)^{\langle \cdot, \# \rangle}.$$

First the left-to-right inclusion. The mapping $\alpha\colon A \to S$ has its image in the stabilization semigroup $\alpha(A)^{\langle\,\cdot\,,\#\rangle}$. Therefore, we may consider the induced mapping $\hat{\alpha}\colon \widehat{A^+} \to \alpha(A)^{\langle\,\cdot\,,\#\rangle}$. It is the unique $\infty$-homomorphism extending $\alpha$, and clearly its image is contained in $\alpha(A)^{\langle\,\cdot\,,\#\rangle}$.

For the right-to-left inclusion, let $t \in \alpha(A)^{\langle\,\cdot\,,\#\rangle}$. Then, $t$ can be obtained by evaluating a term $\mathcal{T}$ using stabilization and multiplication, where the constants are elements of $\alpha(A)$. Without loss of generality, stabilization is applied only to idempotent elements (since idempotents can be reached already by using multiplication). Replacing in the term $\mathcal{T}$ stabilization by the $\omega$-power and for each $a \in A$, the constant $\alpha(a)$ by $a$, we obtain a term $\mathcal{T}'$ using $\omega$-power which evaluates to some profinite word $x$ in $\widehat{A^+}$. Clearly $\hat{\alpha}(x) = t$, i.e. $t$ is contained in the image of $\hat{\alpha}$. This shows the right-to-left inclusion.

This proves Theorem 5. We prove the Factorization Theorem in Section F.

### E.4 Infinite continuity

We prove a useful property of the induced mapping $\hat{\alpha}$, which we call *infinite continuity*. We will use this property in the proof that $\downarrow$- and $\uparrow$-recognizable languages are B- and S-regular, and in the proof that they are closed under projections. We now define this term precisely.

Let $S$ be a finite topological stabilization semigroup. Consider the identity mapping $\beta\colon S \to S$, and the induced $\infty$-homomorphism

$$\hat{\beta} \quad : \quad \widehat{S^+} \quad \longrightarrow \quad S\,.$$

We call $\hat{\beta}$ the *profinite product* in $S$, since it represents computation of profinite products in $S$. For instance, for any finite sequence $s_1, s_2, \ldots, s_n \in S$,

$$\hat{\beta}(s_1, s_2, \ldots, s_n) = s_1 \cdot s_2 \cdots s_n \in S.$$

Also, if $e \in S$ is idempotent, and $\hat{e} \in \widehat{S^+}$ is an infinite profinite sequence of $e$'s, then

$$\hat{\beta}(\hat{e}) = e^{\#}.$$

We will prove one useful property of the profinite product in $S$. By assumption on continuity of multiplication, it follows that if $s_1, s_2$ and $s_1', s_2'$ are such that

$$s_1 \leq s_1',$$
$$s_2 \leq s_2',$$

then also

$$s_1 \cdot s_2 \leq s_1' \cdot s_2.$$

We will see that this extends to profinite sequences.

**Lemma 9.** *Let $s_1, s_2, \ldots, s_\gamma$ and $s'_1, s'_2, \ldots, s'_\gamma$ be two aligned profinite sequences of elements of $S$, such that for each $i$,*

$$s_i \leq s'_i.$$

*Then*

$$\hat{\beta}(s_1 s_2 \ldots s_\gamma) \leq \hat{\beta}(s'_1 s'_2 \ldots s'_\gamma).$$

*Proof.* The proof proceeds by induction on the height $h$ of the smallest factorization tree for the sequence

$$s_1, s_2, \ldots, s_\gamma.$$

The base case $h = 1$ is trivial.

For the inductive step, we consider the binary case, and the finite and infinite idempotent cases. The binary case follows from continuity of multiplication in $S$, which says precisely that if $s_1 \leq s'_1$ and $s_2 \leq s'_2$, then

$$s_1 \cdot s_2 \leq s'_1 \cdot s'_2.$$

The idempotent cases follow from the following lemma.

**Lemma 10.** *Let $g$ be a factorization tree of a profinite word $\Pi(\hat{x})$ where $\hat{x}$ is a profinite sequence $\hat{x} = x_1, x_2 \ldots, x_\gamma$, such that $\hat{\alpha}(x_i) \geq e$ for all $i$. Then the output $s$ of $g$ satisfies:*

$$\begin{aligned} s &\geq e && \text{if the sequence } \hat{x} \text{ is finite} \\ s &\geq e^{\#} && \text{if the sequence } \hat{x} \text{ is infinite.} \end{aligned}$$

*Proof.* The proof proceeds by induction on the structure of $g$. The inductive base is obvious. The most interesting case is the idempotent case of infinite branching. In this case, $g$ is a profinite sequence of trees $g_1, g_2, \ldots, g_\gamma$, each of which has the same output $d$. By assumption, $d \geq e$. Then, the output of $g$ is $d^{\#}$, and

$$d^{\#} \geq e^{\#}$$

by continuity of $\#$ in $S$.

## F Proof of the Factorization Theorem

We prove the Factorization Theorem.

**Factorization Theorem 17.** *For any finite stabilization semigroup $S$, $\|S\|$ is finite. Any two factorization trees over the same input word have the same output.*

The proof of the first part of the above statement is very similar to the proof of the standard factorization theorem of Simon (see e.g. a survey by Bojańczyk, "Factorization Forests"). The proof proceeds by induction on the structure of the stabilization semigroup $S$. We skip the proof in this appendix, and only prove the second part of the statement.

We will say that two factorization trees $f, g$ are *equivalent* if they have the same inputs and outputs.

**Proposition 18.** *Any two factorization trees with the same input are equivalent, i.e. have the same output.*

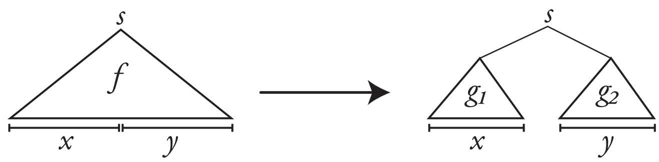
The rest of this section is devoted to proving the above proposition.

We first illustrate the problem in the case of factorization trees which use only the binary rule. The input of such a tree necessarily needs to be a finite word. Proving equivalence of two trees amounts to showing equalities in $S$ like

$$((st)(ts))(((ts)t)s) = (s(t(t(s(t(s(ts))))))).$$

For this, we need to rearrange the corresponding evaluation trees until they coincide, using the associativity axiom for semigroups. The problem becomes much more complicated when stabilization and infinite profinite words come into play.

**Lemma 11.** *Let $f$ be a factorization tree with input $x \cdot y$, where $x, y \in \widehat{A^+}$. Then there exists (see figure below) an equivalent factorization tree $g = (g_1, g_2)$, such that the input of $g_1$ is $x$ and the input of $g_2$ is $y$. Moreover, we may assume that the heights of $g_1$ and $g_2$ do not exceed the height of $f$.*




*Proof.* The proof proceeds by induction on the structure of $f$. If $f$ is a single letter, then there is nothing to prove, since it is impossible that $f$ has input $xy$ with $x, y \in \widehat{A^+}$.

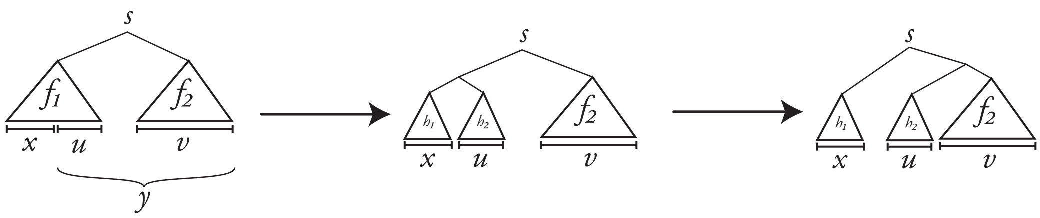



**Fig. 5.** Binary case

*Binary case* Suppose that $f = (f_1, f_2)$. Then, the input of $f$ has two factorizations: $in(f_1) \cdot in(f_2)$ and $x \cdot y$. By the amalgamation property for profinite words, we may assume that the factors $f_1$ and $f_2$ are in one of three positions with respect to the factors $x, y$ of $x \cdot y$, described below. The first position implies the existence of a $u \in \widehat{A^*}$ such that:

$$in(f_1) = x \cdot u \qquad \text{and} \qquad y = u \cdot in(f_2);$$

the second position implies the existence of a $u \in \widehat{A^*}$ such that:

$$x = in(f_1) \cdot u \qquad \text{and} \qquad in(f_2) = u \cdot y.$$

Finally, in the third position, $in(f_1) = x$ and $in(f_2) = y$, so there is nothing to prove. Assume that the first case holds (the second case is symmetric). This is illustrated in the left-hand side of Figure 5. Then, using the inductive assumption for $f_1$, we can find a factorization tree $(h_1, h_2)$ equivalent to $f_1$, and such that the input of $h_1$ is $x$ and the input of $h_2$ is $u$ (see center of Figure 5). Then, we take $g = (g_1, g_2)$, where $g_1 = h_1$ and $g_2 = (h_2, f_2)$ (see right-hand side of Figure 5). The output of $g$ is equal to the output of $f$ by associativity. Moreover, we may assume that the height of $h_1$ does not exceed the height of $f_1$ – therefore the height of $g_1$ doesn't exceed the height of $f$ – and that the height of $h_2$ does not exceed the height of $f_1$ – then the height of $g_2 = (h_2, f_2)$ is not larger than the height of $f = (f_1, f_2)$.

*Idempotent/stabilization case* This case is similar, but slightly more involved than the binary case. We illustrate the reasoning in Figure 6. Assume that $f$ is a profinite sequence of factorization trees with outputs $e$. The sequence $f$ can be either finite or infinite. By the amalgamation property, $f$ factorizes as

$$f = f_1 h f_2,$$

where

- $f_1$ is a prefix of the profinite sequence $f$, and is a factorization tree with some input $v_1$ and output $e$ or $e^{\#}$, depending on whether the sequence $f_1$ is finite or not

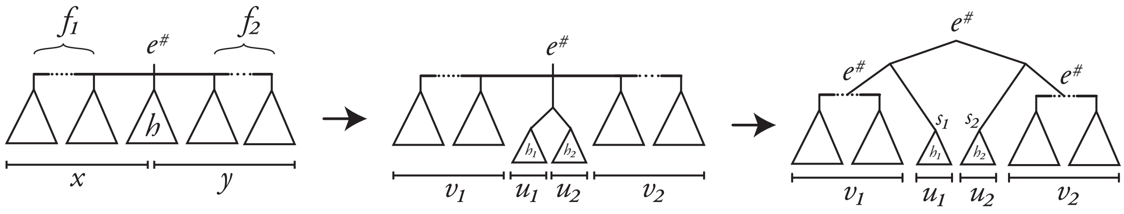



**Fig. 6.** Idempotent/stabilization case

- $f_2$ is a suffix of the profinite sequence $f$, and is a factorization tree with some input $v_2$ and output $e$ or $e^\#$, depending on whether the sequence $f_1$ is finite or not
- $h$ is an element of the profinite sequence $f$, and is a factorization tree of lower height than $f$, with input $u_1u_2$ and output $e$, and

$$v_1u_1 = x \qquad \textit{and} \qquad u_2v_2 = y.$$

Note that the outputs of $f, f_1$ and $f_2$ may be either $e$ or $e^\#$, depending on whether the sequences are finite or infinite. However, the profinite sequence $f$ is infinite if and only if $f_1$ or $f_2$ is an infinite sequence. In the illustration, we assume that both $f_1$ and $f_2$ are infinite, so both of them have output $e^\#$, but the other cases are similar. Taking into account that $e = ee$ and

$$e^\# e = e^\# = ee^\# = e^\# e^\#,$$

it follows that the output of $f$ is equal to the product of the outputs of $f_1$ and $f_2$.

By the inductive assumption, we may replace $h$ by an equivalent tree $(h_1, h_2)$, where the input of $h_1$ is $u_1$, and the input of $h_2$ is $u_2$. Let $s_1$ and $s_2$ denote the outputs of $h_1$ and $h_2$. Then we have that $s_1s_2 = e$. Then, we consider the tree $g = (g_1, g_2)$, where $g_1 = (f_1, h_1)$ and $g_2 = (h_2, f_2)$, as illustrated at the right-hand side of Figure 6. Using associativity and the fact that $e^\# e = e^\# = ee^\#$ in stabilization semigroups, we see that the output of $g$ is:

$$out(f_1) \cdot s_1 \cdot s_2 \cdot out(f_2) = out(f_1) \cdot e \cdot out(f_2) = out(f_1)out(f_2) = out(f).$$

It is easy to see that if the trees $h_1$ and $h_2$ are chosen so that their height does not exceed the height of $h$, then the heights of $g_1$ and $g_2$ do not exceed the height of $f$. This finishes the proof of the lemma.

*Proof (Proof of Proposition 18).* Assume that $f$ and $g$ are two factorization trees over the same input word $x$. We must show that $f$ and $g$ are equivalent. By the amalgamation property, we may assume that $f$ and $g$ are inscribed in a single profinite word. The proof proceeds by a double induction, first on the structure of $f$, and then on the structure of $g$. The case when $f$ or $g$ is a single letter is trivial, since any factorization tree over a letter $a$ has output $\alpha(a)$.

*f has binary root* Suppose that $f = (f_1, f_2)$. Apply Lemma 11, and replace $g$ by an equivalent tree $(g_1, g_2)$, such that

$$in(f_1) = in(g_1) \qquad \text{and} \qquad in(f_2) = in(g_2).$$

Use the inductive assumption for $f_1$ and $g_1$ (note that the height of $f_1$ is smaller than the height of $f$), and conclude that $f_1$ and $g_1$ are equivalent. Similarly, $f_2$ and $g_2$ are equivalent. Therefore, $f = (f_1, f_2)$ is equivalent to $(g_1, g_2)$, which is equivalent to $g$.

*f has idempotent root and g has binary root* Suppose now that $f$ is a profinite sequence of factorization trees, all of which have output $e$, and that $g$ has a binary root, i.e. $= (g_1, g_2)$. Let the inputs of $g_1$ and $g_2$ be $x_1$ and $x_2$, respectively.

Then, by the factorization lemma for profinite sequences, we can find factorization trees $f_1, f_2, f_0$ of heights $h, h$ and $h - 1$, respectively (see Figure 7) so that

$$f = f_1 f_0 f_2,$$

and that the input $y$ of $f_0$ can be split into $y_1, y_2$ satisfying

$$x_1 = in(f_1) \cdot y_1 \qquad \text{and} \qquad x_2 = y_2 \cdot in(f_2).$$

If $y_1$ or $y_2$ is the empty word, then the reasoning trivializes, so we assume they

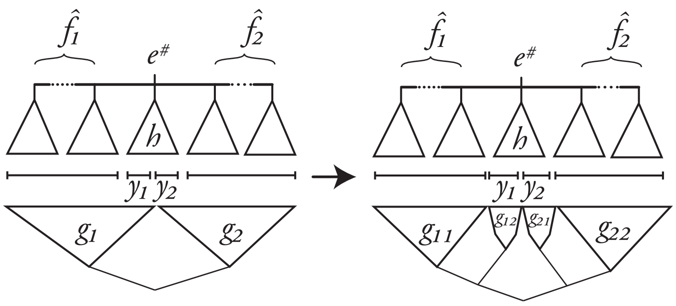



**Fig. 7.** Idempotent/stabilization case

are both nonempty. Apply Lemma 11 to three trees: $g_1$ and $g_2$ (see right-hand side part of Figure 7), replacing:

- $g_1$ by a tree $(g_{11}, g_{12})$, such that the input of $g_{12}$ is $y_1$,
- $g_2$ by a tree $(g_{21}, g_{22})$, such that the input of $g_{12}$ is $y_1$.

Now apply the inductive assumption to:

- $f_1$ and $g_{11}$,
- $f_2$ and $g_{22}$,
- $f_0$ and $(g_{12}, g_{21})$,

and conclude that the pairs of trees listed above are equivalent. Note that $g_{11}$ and $g_{22}$ are of smaller height than $g$, and that $f_0$ is of smaller height than $f$, so the inductive assumption can be applied.

Since the output of the tree $g$ is the product of the outputs of $g_{11}, g_{12}, g_{21}$ and $g_{21}$, and on the other hand, the output of $f$ is equal to the product of the outputs of $f_1$, $f_0$ and $f_2$, we deduce that the output of $f = \hat{f}$ is equal to the output of $g$.

*$f$ and $g$ have idempotent root* Finally, we consider the case when both $f$ and $g$ use the idempotent rule at the topmost node. We assume that both $f$ and $g$ have infinitely many factors (otherwise, the reasoning is simpler, or can be deduced from the binary case which we already proved).

Let us assume that the first factor of $g$ is contained in the first factor of $f$, and that the last factor of $g$ is contained in the last factor of $f$. In general, there are three other configurations, but the argument is very similar in the other cases.

We say that a factor of $f$ is a *split* factor if it intersects at least two factors of $g$. The split factors of $f$ form a profinite sequence, which we denote $(f_1, f_2, \ldots, f_\gamma)$.

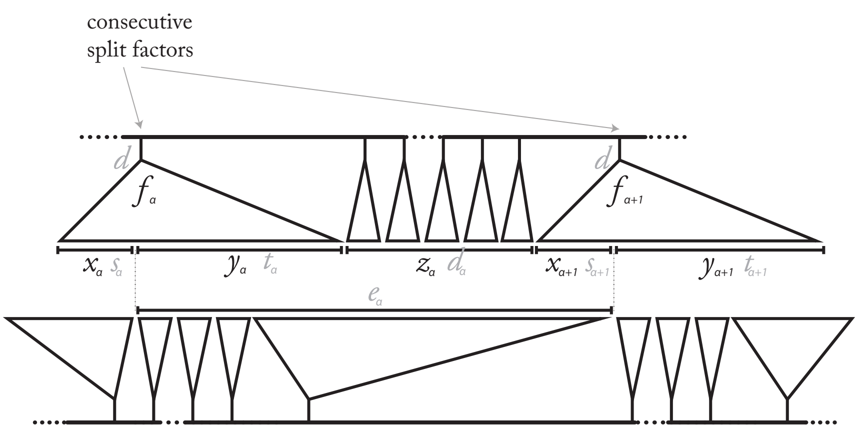



**Fig. 8.** Split factors and the profinite words $x_\alpha, y_\alpha, z_\alpha$. The grey symbols represent the outputs of some factorization trees over the corresponding profinite words

Let $f_\alpha$ be a split factor. In fact, $f_\alpha$ is split into two parts (see Figure 8) – by the first separator of $g$ within $f_\alpha$. Even though these parts might not be well-formed factorization trees, still we can consider the underlying profinite words over the alphabet $A$ – call them $x_\alpha$ and $y_\alpha$ – by ignoring all letters from outside the alphabet $A$. Then, $x_\alpha \cdot y_\alpha$ is the input of the factor $f_\alpha$. Consider two consecutive split factors of $f$, $f_\alpha$ and $f_{\alpha+1}$. Let $z_\alpha$ denote the profinite word between $f_\alpha$ and $f_{\alpha+1}$. This word might be empty – then we say that its factorization tree has output $\varepsilon$. Otherwise, $z_\alpha$ has a factorization tree with output $d$ or $d^\#$ (see Figure 8). Moreover:

- $x_\alpha \cdot y_\alpha$ has a factorization tree $f_\alpha$ of height at most $h-1$, with output $d$
- $y_\alpha \cdot z_\alpha \cdot x_{\alpha+1}$ has a factorization tree with output $e$ or $e^\#$

By Lemma 11, there exist $s_\alpha, t_\alpha, d_\alpha, s_{\alpha+1}, t_{\alpha+1} \in S$ such that (see grey elements in Figure 8):

- $x_\alpha$ has some factorization tree with output $s_\alpha$ and height at most $h-1$,
- $y_\alpha$ has some factorization tree with output $t_\alpha$ and height at most $h-1$,
- $z_\alpha$ has some factorization tree with output $d_\alpha \in \{\varepsilon, d, d^\#\}$,
- $x_{\alpha+1}$ has some factorization tree with output $s_{\alpha+1}$ and height at most $h-1$,
- $y_{\alpha+1}$ has some factorization tree with output $t_{\alpha+1}$ and height at most $h-1$, and

$$s_\alpha \cdot t_\alpha = d = s_{\alpha+1} \cdot t_{\alpha+1}, \tag{1}$$
$$t_\alpha \cdot d_\alpha \cdot s_{\alpha+1} = e \ \mathit{or} \quad e^\#. \tag{2}$$

Moreover, by the inductive assumption, the elements $s_\alpha, t_\alpha, s_{\alpha+1}, t_{\alpha+1}$ do not depend on the choices of the factorization trees.

*Claim.* In the situation described above,

$$d^\# = s_1 e^\# t_\gamma. \tag{3}$$

Note that by our assumption on how the first and last factors of $f$ interact with the first and last factors of $g$, we have :

$$s_1 = e,$$
$$t_\gamma = e.$$

From this, and from the claim above, we obtain the desired result that $d^\# = e^\#$.

We now proceed to proving the claim. Define a set of pairs $P \subseteq S^2$

$$P \quad \overset{def}{=} \quad \{(t_\alpha, s_{\alpha+1}) : \text{ for some pair of consecutive factors } f_\alpha, f_{\alpha+1} \text{ of } f\}.$$

We prove the claim by induction on the size of $P$.

*$P$ contains one pair* The inductive base is when $P$ contains only one pair, $(t, s)$. Let $d' \in \{\varepsilon, d, d^{\#}\}$ be the element $d_\alpha$ corresponding to some split factor $f_\alpha$ (for instance, $d' = d_1$). By equations (1) and (2),

$$s \cdot t = d \qquad \textit{and} \qquad t \cdot d' \cdot s = e.$$

Then,

$$d^{\#} = dd'(dd')^{\#}d = std'(std')^{\#}st = s(td's)^{\#}td'st = se^{\#}et = se^{\#}t.$$

Above, we applied the axiom $u(vu)^{\#} = (uv)^{\#}u$ to the case $u = td'$ and $v = s$. Also, we used the axiom that $(d^{\#})^{\#} = d^{\#}$, since possibly $dd' = d^{\#}$.

*$P$ contains many pairs* Now the inductive step. Assume that $P$ has more than one element. Let $(t, s) \in P$. Assume that $(t, s) = (t_\alpha, s_\alpha)$ for some split factor $f_\alpha$. Let $d' = d_\alpha \in \{\varepsilon, d, d^{\#}\}$, so by equation (2) we get that

$$t \cdot d' \cdot s \in \{e, e^{\#}\}.$$

We can split the factorization tree $f$ into parts in which the pair $(t, s)$ no longer appears. Formally, we construct out of $f$ a profinite sequence $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_\delta$ of factorization trees, for which the corresponding set of pairs is contained in $P - \{(t, s)\}$. Apply the inductive assumption to $f_1, f_2$ and $f_\delta$. We get:

$$d^{\#} = s_1 e^{\#} t = s e^{\#} t = s e^{\#} t_\gamma.$$

Therefore,

$$d^{\#} = d^{\#} d' d^{\#} d' d^{\#} = s_1 e^{\#} \overbrace{t \cdot d' \cdot s}^{e/e^{\#}} e^{\#} \overbrace{t \cdot d' \cdot s}^{e/e^{\#}} e^{\#} t_\gamma = s_1 e^{\#} t_\gamma.$$

This finishes the proof of the claim, ending the proof of the last case of Proposition 18, and thus finishing the proof of the Factorization Theorem.

# G From homomorphisms to B/S-regular expressions

We prove the implications $4 \to 1$ and $8 \to 5$ of Theorem 7, namely that a $\downarrow$-recognizable language is definable by an S-regular expression, and that a $\uparrow$-recognizable language is definable by a B-regular expression.

We first present a proof for $\uparrow$-recognizable sets. Afterwards, we will point out the minor differences in the proof for $\downarrow$-recognizable.

## G.1 From $\downarrow$-recognizability to B-regular expressions

Let

$$\hat{\alpha} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad S$$

be an $\infty$-homomorphism to a finite topological stabilization semigroup. We will show the following.

**Proposition 19.** *For any $s \in S$, the set $\hat{\alpha}^{-1}(\uparrow s)$ is definable by a B-regular expression.*

This proposition implies that for any upward-closed set $F \subseteq S$, $\hat{\alpha}^{-1}(F)$ is definable by a B-regular expression, as $F$ is a finite union of sets of the form $\uparrow s$, and B-regular expressions are closed under finite unions.

Recall that every $x \in \widehat{A^+}$ has some factorization tree of height at most $\|S\|$. The output of any factorization tree over $x$ is called the *type* of $x$, and denoted $\hat{\alpha}(x)$. This does not depend on the choice of the factorization tree.

First, let us describe the set of all elements $x \in \widehat{A^+}$ which have some factorization tree of height $h$ and output $s$. For each $s \in S$ and each natural $h \geq 1$, let us define a set $L_{s,h} \subseteq \widehat{A^+}$ by induction on $h$, by the following formula:

$$\begin{aligned} L_{s,1} &\stackrel{def}{=} \{a \in A : \ \alpha(a) = s\} \\ L_{s,h+1} &\stackrel{def}{=} \bigcup_{\substack{u,v \in S \\ u \cdot v = s}} L_{u,h} \cdot L_{v,h} \quad \cup \quad \bigcup_{\substack{e \in S \\ e = e^2 = s}} (L_{e,h})^{<\infty} \quad \cup \quad \bigcup_{\substack{e \in S \\ e = e^2, e^{\#} = s}} (L_{e,h})^{\infty}. \end{aligned}$$

The following lemma follows immediately from the definition of factorization trees.

**Lemma 12.** *For each $s \in S$ and $h \geq 1$, $L_{s,h}$ is precisely the set of those $x \in \widehat{A^+}$ which have a factorization tree of height $h$ and output $s$.*

The above lemma implies in particular that the set of all $x \in \widehat{A^+}$ which have output equal to $s$, i.e. the set $L_{s,\|S\|}$, can be described by a BS-regular expression – this expression is provided by the recursive definition of $L_{s,h}$. Recall that our aim is to get a B-regular expression, i.e. one not using infinite iteration. The set of profinite words of type $s$ can not always be described by a B-regular expression. However, the set of profinite words of type *at least* $s$ can be described by a B-regular expression, as we will see. The idea is that the infinite iteration in the definition of $L_{s,h}$ can be replaced by unrestricted iteration. This potentially increases the set $L_{s,h}$ by allowing elements which do not have type $s$, but these elements actually turn out to have type at least $s$.

For each $s \in S$ and each natural $h \geq 1$, let us define a set $L_{\uparrow s,h} \subseteq \widehat{A^+}$ by induction on $h$. The recursive definition differs from the definition of $L_{s,h}$ only in that infinite iteration in the last term is replaced by unrestricted iteration:

$$L_{\uparrow s,1} \stackrel{def}{=} \{a \in A : \ \alpha(a) = s\}$$

$$L_{\uparrow s,h+1} \stackrel{def}{=} \underbrace{\bigcup_{\substack{u,v \in S\\ u \cdot v = s}} L_{\uparrow u,h} \cdot L_{\uparrow v,h}}_{(A)} \quad \cup \quad \underbrace{\bigcup_{\substack{e \in S\\ e=e^2=s}} (L_{\uparrow e,h})^{<\infty}}_{(B)} \quad \cup \quad \underbrace{\bigcup_{\substack{e \in S\\ e=e^2, e^{\#}=s}} (L_{\uparrow e,h})^{*}}_{(C)}.$$

The following claim is obvious.

*Claim.* For each $h \geq 1$ and $s \in S$, the set $L_{\uparrow s,h}$ is definable by a B-regular expression.

Let $L_{*,h}$ denote the set of all profinite words which have a factorization tree of height at most $h$, and any output.

**Lemma 13.** *For each $h \geq 1$,*

$$L_{*,h} \quad \cap \quad \hat{\alpha}^{-1}(\uparrow s) \qquad \subseteq \qquad L_{\uparrow s,h} \qquad \subseteq \qquad \hat{\alpha}^{-1}(\uparrow s).$$

*Proof.* The first inclusion is easy: $L_{*,h} \cap \hat{\alpha}^{-1}(\uparrow s)$ is equal to $L_{s,h}$, and $L_{s,h} \subseteq L_{\uparrow s,h}$ because the recursive definition of $L_{\uparrow s,h}$ is more permissive than the definition of $L_{s,h}$.

Now we prove the second inclusion. We proceed by induction on $h$. The case $h = 1$ is trivial. Assume that $h \geq 2$. We will show that each of the terms (A), (B) and (C) from the definition of $L_{\uparrow s,h}$ is contained in $\hat{\alpha}^{-1}(\uparrow s)$. The most interesting case is the case of the term (C). For the terms (A) and (B) ones the reasoning is similar. Let $e \in S$ be such that $e = e^2$ and $s = e^{\#}$. We use the inductive assumption that

$$L_{e,h} \subseteq \hat{\alpha}^{-1}(\uparrow e).$$

Therefore, any element $x$ of $(L_{e,h})^*$, is a concatenation of a profinite sequence of elements

$$x_1, x_2, \ldots, x_\gamma,$$

such that $\hat{\alpha}(x_i) \geq e$ for each $i$. From invariance under infinite substitutions and from Lemma 9 we deduce that

$$\hat{\alpha}(x) \geq e^{\#} = s \qquad \textit{if } \gamma \textit{ is infinite}$$

$$\hat{\alpha}(x) \geq e \geq e^{\#} = s \qquad \textit{if } \gamma \textit{ is finite}$$

Therefore, $x \in \hat{\alpha}^{-1}(\uparrow s)$, what is what we needed to prove. This ends the proof of the lemma.

Taking $h = \|S\|$ in the above lemma we have $L_{*,h} \cap \hat{\alpha}^{-1}(\uparrow s) = \hat{\alpha}^{-1}(\uparrow s)$, and hence

$$\hat{\alpha}^{-1}(\uparrow s) = L_{\uparrow s,h}.$$

Therefore, by Claim G.1, $\hat{\alpha}^{-1}(\uparrow s)$ is definable by a B-regular expression. This finishes the proof of Proposition 19.

## G.2 From $\uparrow$-recognizability to S-regular expressions

The proof for $\uparrow$-recognizable sets is virtually identical. We briefly describe it. Again, it suffices to prove the following.

**Proposition 20.** *For any $s \in S$, the set $\hat{\alpha}^{-1}(\downarrow s)$ is definable by an S-regular expression.*

This time, we define languages $L_{\downarrow s,h}$ recursively, by replacing finite iteration in the second term of the definition of $L_{s,h}$ by unrestricted iteration:

$$\begin{aligned} L_{\downarrow s,1} &\stackrel{def}{=} \{a \in A : \alpha(a) = s\} \\ L_{\downarrow s,h+1} &\stackrel{def}{=} \underbrace{\bigcup_{\substack{u,v \in S \\ u \cdot v = s}} L_{\downarrow u,h} \cdot L_{\downarrow v,h}}_{(A)} \quad \cup \quad \underbrace{\bigcup_{\substack{e \in S \\ e = e^2 = s}} (L_{\downarrow e,h})^*}_{(B)} \quad \cup \quad \underbrace{\bigcup_{\substack{e \in S \\ e = e^2, e^{\#} = s}} (L_{\downarrow e,h})^\infty}_{(C)} . \end{aligned}$$

Obviously, we have:

*Claim.* For each $h \geq 1$ and $s \in S$, the set $L_{\downarrow s,h}$ is definable by an S-regular expression.

Analogously as before, we also have the following.

**Lemma 14.** *For each $h \geq 1$,*

$$L_{*,h} \quad \cap \quad \hat{\alpha}^{-1}(\downarrow s) \qquad \subseteq \qquad L_{\downarrow s,h} \qquad \subseteq \qquad \hat{\alpha}^{-1}(\downarrow s).$$

The proof of this lemma is completely dual to the proof of Lemma 13.

*Proof (Sketch of proof).* Again, the first inclusion is obvious, and the second inclusion is proved by induction on $h$. This time, the most interesting case is the case of the term (B). Let $e \in S$ be such that $e = e^2$ and $s = e^{\#}$. We use the inductive assumption that

$$L_{\downarrow e,h} \subseteq \hat{\alpha}^{-1}(\downarrow e).$$

Therefore, any element $x$ of $(L_{\downarrow e,h})^*$, is a concatenation of a profinite sequence of elements

$$x_1, x_2, \ldots, x_\gamma,$$

such that $\hat{\alpha}(x_i) \leq e$ for each $i$. From invariance under infinite substitutions and from Lemma 9 we get that

$$\begin{aligned} \hat{\alpha}(x) \leq e^{\#} \leq e = s &\qquad \textit{if } \gamma \textit{ is infinite} \\ \hat{\alpha}(x) \leq e = s &\qquad \textit{if } \gamma \textit{ is finite} \end{aligned}$$

Therefore, $x \in \hat{\alpha}^{-1}(\downarrow s)$, what we wanted to show.

Plugging $h = \|S\|$ in Lemma 14, we get Proposition 20.

# H The powerset constructions

We present two variants of the powerset construction for $\infty$-homomorphisms, and complete the proof of Proposition 6, by proving that $\downarrow$-recognizable and $\uparrow$-recognizable languages are closed under projections.

Let

$$\hat{\alpha} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad S$$

be an $\infty$-homomorphism to a finite stabilization semigroup. Let

$$P_{\downarrow}(S) \quad \stackrel{def}{=} \quad \{X \subseteq S : X \ = \downarrow X\} \quad \subseteq \quad P(S)$$

denote the set of downward-closed (or topologically closed) subsets of $S$, and let

$$P_{\uparrow}(S) \quad \stackrel{def}{=} \quad \{X \subseteq S : X = \uparrow X\} \quad \subseteq \quad P(S)$$

denote the upward-closed (or topologically open) subsets of $S$.

Let $\pi\colon A \to B$ be a mapping of finite alphabets, and let

$$\hat{\pi} \quad : \quad \widehat{A^+} \quad \longrightarrow \quad \widehat{B^+}$$

be the induced mapping of profinite words.

We define two mappings,

$$P_{\downarrow}\hat{\alpha} \quad : \quad \widehat{B^+} \quad \longrightarrow \quad P_{\downarrow}S \tag{1}$$

$$P_{\uparrow}\hat{\alpha} \quad : \quad \widehat{B^+} \quad \longrightarrow \quad P_{\uparrow}S \tag{2}$$

as follows. For $y \in \widehat{B^+}$,

$$P_{\downarrow}\hat{\alpha}(y) \quad \stackrel{def}{=} \quad \downarrow\{\hat{\alpha}(x) : \ \hat{\pi}(x) = y\} \tag{3}$$

$$P_{\uparrow}\hat{\alpha}(y) \quad \stackrel{def}{=} \quad \uparrow\{\hat{\alpha}(x) : \ \hat{\pi}(x) = y\}. \tag{4}$$

The following lemma is straightforward.

**Lemma 15.** *Let $L \subseteq \widehat{A^+}$ be $\downarrow$-recognized by $\hat{\alpha}\colon \widehat{A^+} \to S$ and a closed set $F \subseteq S$:*

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

*Then the projection $\hat{\pi}(L)$ satisfies:*

$$\hat{\pi}(L) = (P_{\uparrow}\hat{\alpha})^{-1}(F_{\uparrow}) \qquad \textit{where} \quad F_{\uparrow} = \{X \in P_{\uparrow}S : \ X \cap F \neq \emptyset\}.$$

*Dually, let $K \subseteq \widehat{A^+}$ be $\uparrow$-recognized by $\hat{\alpha}\colon \widehat{A^+} \to S$ and an open set $U \subseteq S$:*

$$K = \hat{\alpha}^{-1}(U).$$

*Then the projection $\hat{\pi}(K)$ satisfies:*

$$\hat{\pi}(K) = (P_{\downarrow}\hat{\alpha})^{-1}(U_{\downarrow}) \qquad \textit{where} \quad U_{\downarrow} = \{X \in P_{\downarrow}S : \ X \cap U \neq \emptyset\}.$$

The above lemma delineates our plan for proving Proposition 6. We will show:

**Proposition 21.** *Let $\hat{\alpha}$, $P_{\downarrow}\hat{\alpha}$, $P_{\uparrow}\hat{\alpha}$, $F_{\uparrow}$, $U_{\downarrow}$ be as above. Then the images* $\mathrm{Im}(P_{\downarrow}\hat{\alpha}) \subseteq P_{\downarrow}S$ *and* $\mathrm{Im}(P_{\uparrow}\hat{\alpha}) \subseteq P_{\uparrow}S$ *can be equipped with a structure such that:*

1. $\mathrm{Im}(P_{\downarrow}\hat{\alpha})$ *and* $\mathrm{Im}(P_{\uparrow}\hat{\alpha})$ *are topological stabilization semigroups,*
2. *The mappings $P_{\downarrow}\hat{\alpha}$ and $P_{\uparrow}\hat{\alpha}$ are $\infty$-homomorphisms onto their images,*
3. *The set $F_{\uparrow}$ is closed in* $\mathrm{Im}(P_{\downarrow}\hat{\alpha})$ *and the set $U_{\downarrow}$ is open in* $\mathrm{Im}(P_{\uparrow}\hat{\alpha})$.

Together with Lemma 15, the above proposition proves that the projection of a $\downarrow$-recognizable set is $\downarrow$-recognizable, and that the projection of a $\uparrow$-recognizable set is $\uparrow$-recognizable.

First we prove the following lemma which is mostly related to the second item in the above proposition. Notice however that it does not need to talk about any structure of $P_{\downarrow}S$ or $P_{\uparrow}S$.

**Lemma 16.** *The mappings $P_{\downarrow}\hat{\alpha}$ and $P_{\uparrow}\hat{\alpha}$ are invariant under infinite substitutions.*

*Proof.* We consider the mapping $P_{\downarrow}\hat{\alpha}$. The argumentation for $P_{\uparrow}\hat{\alpha}$ is dual.

Let $\sigma$ be a substitution scheme

$$\sigma \quad : \quad (y_1, S_1, y_1'), (y_2, S_2, y_2'), \ldots, (y_\gamma, S_\gamma, y_\gamma'),$$

where for all $i$, $y_i, y_i' \in \widehat{B^+}$ and $S_i \in P_{\uparrow}S$. Assume that $\sigma$ is compatible with $P\hat{\alpha}$. Let $y$ denote the concatenation of $y_1, y_2, \ldots, y_\gamma$, and similarly let $y'$ denote the concatenation of $y_1', y_2', \ldots, y_\gamma'$. We want to show that

$$P_{\downarrow}\hat{\alpha}(y) \;=\; P_{\downarrow}\hat{\alpha}(y'). \tag{5}$$

It suffices to show the left-to-right inclusion of the above equality, since then the other inclusion follows from symmetry.

Let $s \in P_{\downarrow}\hat{\alpha}(y)$. This means that there exists an $x \in \widehat{A^+}$ such that:

$$\hat{\pi}(x) = y,$$
$$\hat{\alpha}(x) \geq s.$$

The intuition is simple. The following steps describe this intuition.

1. Find a profinite factorization

   $$x_1, x_2, \ldots, x_\gamma$$

   of $x$ which is consistent with the factorization of $y$ given by $\sigma$, i.e. each $x_i$ is projected to a corresponding factor $y_i$ under $\hat{\pi}$.
2. Therefore, each factor $x_i$ has a type which belongs $P_{\downarrow}\hat{\alpha}(y_i)$.
3. Use the assumption that $P_{\downarrow}\hat{\alpha}(y_i) = P_{\downarrow}\hat{\alpha}(y_i')$, and conclude that there exists some $x_i'$ which projects to $y_i'$ under $\hat{\pi}$ and with type at least as big as the type of $x_i$.

4. Concatenate all the $x'_i$, yielding a profinite word $x'$ which projects to $y'$.
5. The profinite word has type at least as big as the type of $x$, by the inequalities and some continuity property of concatenation. Therefore, the type of $x'$ is at least $\hat{\alpha}(x) \geq s$. This proves that $s \in P_{\downarrow}\hat{\alpha}(y)$, demonstrating the left-to-right inclusion of (5).

*Remark 5.* In a formal proof we need to be very careful. All the above steps should be done uniformly on an entire profinite sequence. To do this, we will rely on the uniform lifting property of closed and open sets. In particular, we will have to make sure that the properties which we consider are either closed or open. It is here where we rely on the fact that $P_{\downarrow}\hat{\alpha}$ only talks about downward closed sets, and that in the five steps above we only use estimates such as "at least as big" rather than "equal to". These precautions are necessary. On the level sketched in the above five steps, it might seem that the proof should also work for the "usual" powerset mapping

$$P\hat{\alpha} \quad : \quad \widehat{B^+} \quad \longrightarrow \quad P(S)$$

defined similarly as $P_{\downarrow}\hat{\alpha}$, but without applying the downward closure. However, the mapping $P\hat{\alpha}$ is usually *not* invariant under infinite substitutions.

In fact, in order for the proof to work, we will split the third step described above into three small steps – it might be impossible to find uniformly both the profinite words $x_i$ and the $x'_i$ as described in the third step. Because of this issue, we will have substeps:

3.1. "store" an $s_i \in S$ which bounds from above the type of $x_i$,
3.2. "forget" the $x_i$,
3.3. "store" an $x'_i$ whose type bounds from above the element $s_i$.

In substep 3.1, we are first dealing with a closed property: "the type of $x_i$ is bounded from above by $s_i$" and then, in substep 3.3, separately with an open property "the type of $x'_i$ is bounded from below by $s_i$". If we tried to proceed as described in the Step 3 directly, we would have to deal with the property: "the type of $x_i$ is bounded from above by the type of $x'_i$", which is neither closed nor open (but an intersection of such). As a side remark, note that the property "the type of $x_i$ is equal to $s_i$" is also neither closed nor open.

We now describe the above steps in detail.

*Step 1: Factorize $x$.* By the amalgamation property for profinite words, there exists a combined profinite sequence of quadruples

$$(x_1, y_1, S_1, y'_1), (x_2, y_2, S_2, y'_2), \ldots, (x_\gamma, y_\gamma, S_\gamma, y'_\gamma)$$

such that:

- The product $x_1 x_2 \ldots x_\gamma$ is equal to $x$,
- The product $y_1 y_2 \ldots y_\gamma$ is equal to $y$,
- The product $y'_1 y'_2 \ldots y'_\gamma$ is equal to $y'$,
- $\hat{\pi}(x_i) = y_i$ for each $i$,
- $P_{\downarrow}\hat{\alpha}(y_i) = S_i = P_{\downarrow}\hat{\alpha}(y'_i)$ for each $i$.

*Step 2: Bound the types of the factors of $x$.* It follows that

$$\hat{\alpha}(x_i) \in P_\downarrow \hat{\alpha}(y_i) = P_\downarrow \hat{\alpha}(y'_i),$$

i.e. for each $i$, there exists some $s_i \in S$ such that

$$\hat{\pi}(x_i) = y_i, \tag{6}$$
$$s_i \in S_i \tag{7}$$
$$\hat{\alpha}(x_i) \leq s_i. \tag{8}$$

*Step 3.1: Uniformly bound the types of the factors of $x$.* Note that (8), (7), (6) define a closed property of quintuples $(x_i, s_i, y_i, S_i, y'_i)$ – this is because $\hat{\alpha}^{-1}(\downarrow s_i)$ is a closed set. Let us call this property $P$. By the uniform lifting property of closed sets, this means that we can choose the $s_i$ in a uniform way, i.e. there exists a profinite sequence of quintuples

$$(x_1, s_1, y_1, S_1, y'_1), (x_2, s_2, y_2, S_2, y'_2), \ldots, (x_\gamma, s_\gamma, y_\gamma, S_\gamma, y'_\gamma)$$

such that for each $i$, the properties (8), (7), (6) hold.

*Step 3.2: Forget the factors of $x$.* Now we project out the $x_i$'s from the sequence, yielding a profinite sequence of quadruples:

$$(s_1, y_1, S_1, y'_1), (s_2, y_2, S_2, y'_2), \ldots, (s_\gamma, y_\gamma, S_\gamma, y'_\gamma).$$

*Step 3.3: Find factors of $x'$.* We invert Step 3.1. Consider the properties

$$\hat{\pi}(x'_i) = y'_i, \tag{9}$$
$$s_i \in S_i \tag{10}$$
$$\hat{\alpha}(x'_i) \geq s_i. \tag{11}$$

The above property is an open property of quintuples $(x'_i, s_i, y_i, S_i, y'_i)$. Therefore, we can uniformly find $x'_i$ which satisfy the above property.

*Step 4: Concatenate the $x_i$'s.* This step is easy. We define $x'$ as the concatenation of the profinite sequence $x'_1, x'_2, \ldots, x'_\gamma$.

*Step 5: Conclude.* We will prove:

$$\hat{\pi}(x') = y' \tag{12}$$
$$\hat{\alpha}(x') \geq \hat{\alpha}(x) \geq s. \tag{13}$$

This in turn implies that $s \in P_\downarrow \hat{\alpha}(y')$, ending the left-to-right inclusion of (5).

The equality (12) is obvious – it follows from (9) and from commutativity of (profinite) concatenation and projection (recall that profinite concatenation is simply the removal of separating symbols, while projection is just relabeling of letters).

It remains to prove the inequality (13). Consider the profinite sequence of quintuples considered in Step 3.1. Consider the type $t$ of the profinite sequence

$$s_1, s_2, \ldots, s_\gamma.$$

Formally, we consider the identity mapping $\beta\colon S \to S$ and the induced $\infty$-homomorphism

$$\hat{\beta} \quad : \quad \widehat{S^+} \quad \longrightarrow \quad S,$$

and define

$$t \quad \stackrel{def}{=} \quad \hat{\beta}(s_1, s_2, \ldots, s_\gamma).$$

Then the inequality (13) follows from the following claim.

*Claim.*

$$\hat{\alpha}(x) \leq t \leq \hat{\alpha}(x').$$

Both inequalities above are consequences of the inequalities (8), (11) of Lemma 9.

This ends the proof that the map $P_\downarrow\hat{\alpha}$ is invariant under infinite substitutions. For the mapping $P_\uparrow\hat{\alpha}$, the proof proceeds dually, by changing the directions of the inequalities.

We now come back to the proof of Proposition 21, which says that not only $P_\uparrow\hat{\alpha}$ and $P_\downarrow\hat{\alpha}$ are invariant under infinite substitutions, but they are in fact $\infty$-homomorphisms to suitably defined topological stabilization semigroups.

*Proof (Proof of Proposition 21).* We will prove the statement for the mapping

$$P_\downarrow\hat{\alpha} \quad : \quad \widehat{B^+} \quad \longrightarrow \quad T,$$

where $T \subseteq P_\downarrow S$ denotes the image of $P_\downarrow\hat{\alpha}$. For $P_\downarrow\hat{\alpha}$ the argument is completely analogous.

We begin with defining the appropriate structure of a topological stabilization semigroup over the set $T$. Since $P_\downarrow\hat{\alpha}$ is invariant under infinite substitutions, it in particular preserves multiplication in $\hat{B}$. This means that the image $T$ can be equipped in a unique way with a semigroup structure, such that the mapping $P_\downarrow\hat{\alpha}$ becomes a homomorphism of semigroups. Similarly, $T$ can be equipped in a unique way with a stabilization mapping $\#$ so that for all $y \in \widehat{B^+}$,

$$\alpha(y^\omega) = \alpha(y)^\#.$$

This way, $T$ becomes a $\langle\, \cdot\,, \#\rangle$-algebra, and $P_\downarrow\hat{\alpha}$ becomes a homomorphism of such algebras. Finally, we equip $T$ with the quotient topology induced by the mapping $P_\downarrow\hat{\alpha}$. By Proposition 13 and Proposition 14, $T$ is then a topological stabilization semigroup, and $P_\downarrow\hat{\alpha}$ is a homomorphism of topological stabilization semigroups. It remains to check that the set $U_\downarrow$ is open in $T$. By definition of the quotient topology, this is equivalent to

$$(P_\downarrow\hat{\alpha})^{-1}(U_\downarrow) = \hat{\pi}(L)$$

being a closed set. This is clearly true, since the mapping $\hat{\pi}$ maps open sets to open sets, and $L$ is open (the mapping $\hat{\pi}$ also maps closed sets to closed sets, which is needed in the dual proof).

This finishes the proof that $\downarrow$-recognizable and $\uparrow$-recognizable sets are closed under projection, ending the proof of Proposition 6.

# I Proof of the main theorem

In this section, we finish the proof of Theorem 7

The translations $1 \to 2 \to 3 \to 4$ and $5 \to 6 \to 7 \to 8$ follow from Propositions 4 and 3 and Corollary 1, and are effective. The equivalence of the conditions 3 and 7 is obvious. The translations $4 \to 1$ and $8 \to 5$ where proved in Appendix G. It remains to prove the equivalence $8 \Leftrightarrow 9$.

*8 ⇒ 9.* The class $\mathcal{K}$ of $\downarrow$-recognizable languages in $\widehat{A^+}$ satisfies the following conditions:

1. $\mathcal{K}$ consists of closed languages,
2. Every language $K \in \mathcal{K}$ contains an element of $A^{\langle\,\cdot\,,\omega\rangle}$,
3. $\mathcal{K}$ is closed under intersections with clopen sets, i.e. if $K \in \mathcal{K}$ and $U \subseteq \widehat{A^+}$ is clopen, then $K \cap U \in \mathcal{K}$.

The first property holds for S-regular languages, which are the same as $\downarrow$-recognizable languages. The second property follows from the description of the image of $\infty$-homomorphisms given by Theorem 5. The third property follows from the fact that $\downarrow$-recognizable languages are closed under intersections by Proposition 6 and that clopen sets are $\downarrow$-recognizable. Then the implication 8⇒9 in the theorem follows from a simple topological lemma:

**Lemma 17.** *Let $\mathcal{K}$ be a class of languages of profinite words over $A$ which satisfies the properties $1, 2, 3$ listed above. Then, $K = \overline{K \cap A^{\langle\,\cdot\,,\omega\rangle}}$ for every $K \in \mathcal{K}$.*

*9 ⇒ 8.* We now proceed to the reverse implication. Let $\alpha_K \colon \widehat{A^+} \to S_K$ be the syntactic homomorphism. By Theorem 12 and Proposition 14 of Appendix C, $S_K$ is a finite stabilization semigroup and $\alpha_K$ is a continuous homomorphism recognizing $K$. Let $F \subseteq S_K$ be such that $K = \alpha_K^{-1}(F)$. Let $\hat{\alpha} \colon \widehat{A^+} \to S_K$ be the mapping induced by the restriction of $\alpha_K$ to $A$ . We do not prove that $\hat{\alpha} = \alpha_K$, but rather show that $\hat{\alpha}^{-1}(F) = K$. Indeed, since $\hat{\alpha}$ and $\alpha_K$ agree over $A^{\langle\,\cdot\,,\omega\rangle}$, we have that

$$K \cap A^{\langle\,\cdot\,,\omega\rangle} = \alpha_K^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle} = \hat{\alpha}^{-1}(F) \cap A^{\langle\,\cdot\,,\omega\rangle}.$$

By assumption, the closure of the left-hand-side set above is $K$. On the other hand, the closure of the right-hand side set above is equal to $\hat{\alpha}^{-1}(F)$ by the already proved implication 8⇒9 applied to the $\downarrow$-recognizable language $\hat{\alpha}^{-1}(F)$. Therefore, $K = \hat{\alpha}^{-1}(F)$, i.e. $K$ is $\downarrow$-recognizable.

# J From $\omega$-words to profinite words

We prove Theorem 9, by providing a reduction from logics over infinite words to logics over finite words.

There are two key ideas of our reduction:

**A Ramsey-type lemma** allowing to split an infinite word into a convergent sequence of finite words; and

**A model of deterministic automata** with *limitary acceptance condition.* In this model, an acceptance condition is specified by a language of profinite words $L$, and a word is accepted if and only if the run of the automaton has a factorization which is convergent to an element of $L$.

We now present these notions.

## J.1 Ramsey lemma for profinite words

Let $A$ be a finite alphabet.

**Lemma 18.** *Let $w \in A^\omega$ be an infinite word over the alphabet $A$. Then there exists a profinite word $w_\infty \in \widehat{A^+}$, and a factorization*

$$w = w_0 \cdot w_1 \cdot w_2 \cdots \tag{1}$$

*of $w$ into finite words $w_0, w_1, w_2, \ldots \in A^+$, such that*

$$\lim_{n\to\infty} w_n = w_\infty,$$

*where convergence is in the profinite topology over $\widehat{A^+}$.*

*Proof.* First, let $L$ be a fixed regular language, and let

$$w = u_0 \cdot u_1 \cdot u_2 \cdots$$

be any factorization of $w$ into finite words. Then, from the usual Ramsey theorem it follows that $w$ has a coarser factorization

$$w = u'_0 \cdot u'_1 \cdot u'_2 \cdots$$

such that

$u'_1, u'_2, \ldots$ either all belong to $L$, or all belong to the complement of $L$.

Let $L_1, L_2, L_3, \ldots$ be an enumeration of all regular languages over the alphabet $A$. The lemma then follows from a diagonal argument, by applying the above reasoning for $L_1, L_2, L_3, \ldots$, starting from the factorization

$$w = a_1 \cdot a_2 \cdot a_3 \cdots$$

of $w$ into single letters.

## J.2 Deterministic automata with limitary acceptance condition

First we define a notion of deterministic automaton over infinite words, with an abstract acceptance condition. Then we define limitary acceptance conditions.

A *deterministic automaton* $\mathcal{A}$ with states $Q$, over the alphabet $A$, is described by a transition function $\delta\colon Q \times A \to Q$. Its *acceptance condition* is a distinguished language $F \subseteq Q^\omega$. If $w \in A^\omega$ is an infinite word, then by $\delta(w)$ we denote the sequence of states of $\mathcal{A}$ when ran over $w$. We say that $\mathcal{A}$ *accepts* $w$ if $\delta(w) \in F$. We also say that $\mathcal{A}$ *recognizes* the language $\{w : \ \delta(w) \in F\}$. An *$F$-language* is a finite union of languages accepted by deterministic automata with states $Q$ and acceptance condition $F \subseteq Q^\omega$.

We say that an acceptance condition $F \subseteq Q^\omega$ is *prefix-independent*, if for all $v \in Q^*$,$w \in Q^\omega$,

$$w \in F \qquad \Longleftrightarrow \qquad vw \in F.$$

We will be mostly interested in prefix-independent acceptance conditions.

*Subset construction for $F$-automata* Let $\mathcal{A}$ be a deterministic automaton with input alphabet $A \times \{0,1\}$, states $Q$ and initial state $q_0 \in Q$, and transition function

$$\delta \quad : \quad Q \times A \times \{0,1\} \quad \longrightarrow \quad Q\,.$$

Let $F \subseteq A^\omega$ be the acceptance condition of $\mathcal{A}$, and let $L \subseteq (A \times \{0,1\})^\omega$ be the language accepted by $\mathcal{A}$. We will construct a deterministic automaton $\exists\mathcal{A}$ which accepts the language

$$\exists L \quad \stackrel{def}{=} \quad \{w \in A^\omega : \ \exists X \subseteq \mathbb{N} : \ w \otimes X \in L\}.$$

The automaton $\exists\mathcal{A}$ has transition function $\exists\delta$, states $P(Q) \times A$ and input alphabet $A$:

$$\exists\delta \quad : \quad (P(Q) \times A) \times A \quad \longrightarrow \quad P(Q) \times A\,.$$

The state transitions of $\exists\delta$ are as in the usual subset construction on the $P(Q)$ coordinate, and the $A$ coordinate serves for remembering the last seen input letter. The initial state is the pair $(\{q_0\}, a_0)$ where $q_0$ is the initial state of $\mathcal{A}$, and $a_0 \in A$ is some fixed letter (the choice of $a_0$ doesn't matter).

We state two obvious properties of $\exists\delta$. Let $w = b_1 b_2 b_3 \ldots \in A^\omega$ be an input word, and let $(\exists\delta)(w) = (Q_0, a_0)(Q_1, a_1)(Q_2, a_2)\ldots$ be the sequence of states of $\exists\mathcal{A}$ when ran over $w$. Then, $b_1 = a_1, b_2 = a_2, \ldots$ and the sequence $(\exists\delta)(w)$ satisfies the following condition $(*)$:

$$Q_n = \{\delta(q, a_n, i) : \ q \in Q_{n-1}, i \in \{0,1\}\} \quad \textit{for every } n = 1, 2, 3, \ldots$$

Moreover, we have the following.

*Claim.* $w \in \exists L$ if and only if $(\exists\delta)(w) = (Q_0, a_0)(Q_1, a_1)(Q_2, a_2)\ldots$ satisfies the condition $(**)$:

*There exists a sequence* $q_1, q_2, \ldots \in Q$ *and a sequence* $i_1, i_2, i_3, \ldots \in \{0,1\}$ *such that for every* $n = 1, 2, \ldots,$

- $q_n \in Q_n$
- $q_n = \delta(q_{n-1}, a_n, i_n)$ *(with* $q_0$ *being the initial state of* $\mathcal{A}$*)*
- $q_0 q_1 q_2 q_3 \ldots \in F$

We define the acceptance condition of $\exists\mathcal{A}$ as the language $\exists F$ of all sequences

$$(Q_0, a_0), (Q_1, a_1), \ldots \in (P(Q) \times A)^\omega$$

which share an infinite suffix with some sequence which satisfies the conditions $(*)$ and $(**)$. By definition, the acceptance condition $\exists F$ is prefix-independent.

*Claim.* If $L \subseteq (A \times \{0,1\})^\omega$ is accepted by an $F$-automaton and $F$ is prefix-independent, then $\exists L \subseteq A^\omega$ is accepted by the $\exists F$-automaton $\exists\mathcal{A}$ described above.

*Proof.* It suffices to observe that for a given input word $w \in A^\omega$, $w \in \exists L$ if and only if $(\exists\delta)(w)$ shares an infinite suffix with some sequence which satisfies the conditions $(*)$ and $(**)$.

*Limitary acceptance conditions* Let $F$ be an acceptance condition, i.e. simply some language $F \subseteq Q^\omega$. We say that $F$ is *limitary*, if it satisfies:

*For every word* $w = w_0 w_1 w_2 \ldots$ *with* $\lim_{n\to\infty} w_n = w_\infty \in \widehat{Q^+}$*, whether* $w$ *belongs to* $F$ *depends only on* $w_\infty$*.*

Clearly, a limitary acceptance condition is prefix-independent. We say that a language of profinite words $F' \subseteq \widehat{Q^+}$ is a *limit* of an acceptance condition $F \subseteq Q^\omega$, if the following holds.

*For any* $w \in Q^\omega$*,* $w_\infty \in \widehat{Q^+}$*,* $w_0, w_1, w_2 \ldots \in Q^+$ *such that*

$$\lim_{n\to\infty} w_n = w_\infty,$$
$$w_0 \cdot w_1 \cdot w_2 \cdots = w,$$

*the following equivalence holds:*

$$w \in F \qquad \textit{iff} \qquad w_\infty \in F'.$$

It is easy to see that an acceptance condition $F$ is limitary if and only if it has a limit.

*Example 15.* The Büchi acceptance condition is limitary. If $\lim w_n = x \in \widehat{\{0,1\}^+}$, then $w_0 w_1 w_2 \dots$ belongs to *Büchi* if and only if $x$ contains the letter 1. Therefore, the corresponding language of profinite words is the language of words which contain at least one letter 1.

The B-acceptance condition,

$$B = \{a^{n_1} b a^{n_2} b \dots : \quad \textit{the sequence } n_1, n_2, \dots \textit{ is bounded}\}$$

is limitary – $w_0 w_1 w_2 \dots$ belongs to $B$ if and only if the limit $x$ has only $a$-blocks of bounded size.

### J.3 $\mathcal{L}$-limitary languages

Let $\mathcal{L}$ be a class of languages of profinite words (the languages need not be over a unique alphabet). An *$\mathcal{L}$-limitary language* is a language $L \subseteq A^\omega$ which is accepted by an deterministic automaton $\mathcal{A}$ with a limitary acceptance condition $F$, whose limit $F'$ is in $\mathcal{L}$.

For $F' \subseteq \widehat{P^+}$, $G' \subseteq \widehat{Q^+}$, let us define:

$$\begin{aligned}
\neg F' &\stackrel{def}{=} \widehat{P^+} - F' \\
F' \oplus G' &\stackrel{def}{=} \{x \in \widehat{P \times Q^+} : \pi_P(x) \in F' \vee \pi_Q(x) \in G'\} \\
F' \otimes G' &\stackrel{def}{=} \{x \in \widehat{P \times Q^+} : \pi_P(x) \in F' \wedge \pi_Q(x) \in G'\}
\end{aligned}$$

where $\pi_P, \pi_Q$ denote the two projections from $\widehat{P \times Q^+}$ to $\widehat{P^+}$ and $\widehat{Q^+}$, respectively.

**Proposition 22.** *If $\mathcal{L}$ is closed under $\neg$, then $\mathcal{L}$-limitary languages are closed under complements. If $\mathcal{L}$ is closed under $\oplus$, then $\mathcal{L}$-limitary languages are closed under unions. If $\mathcal{L}$ is closed under $\otimes$, then $\mathcal{L}$-limitary languages are closed under intersections. If $\mathcal{L}$ contains regular languages, is closed under $\otimes$ and under projection, then $\mathcal{L}$-limitary languages are closed under projection.*

*Proof.* We prove closure under intersections, assuming that $\mathcal{L}$ is closed under $\otimes$. Let $K, L$ be two $\mathcal{L}$-limitary languages, and let $\mathcal{A}$, $\mathcal{B}$ be the corresponding automata, and $F, G$ their respective acceptance conditions. Consider the Cartesian product automaton $\mathcal{C}$, whose state space is the Cartesian $P \times Q$ product of the state spaces of $\mathcal{A}$ and $\mathcal{B}$, respectively. It is easy to see that setting

$$F \otimes G \quad \stackrel{def}{=} \quad \{\rho \in (P \times Q)^\omega : \ \pi_P(\rho) \in F \wedge \pi_Q(\rho) \in G\}$$

as the acceptance condition for $\mathcal{C}$ yields an automaton which accepts the language $K \cap L$. It remains to see that $F \otimes G$ has limit $F' \otimes G'$. Take any $\rho \in (P \times Q)^\omega$, and a convergent factorization

$$\rho = \rho_0 \cdot \rho_1 \cdots, \qquad \text{where } \lim_{n\to\infty} \rho_n = \rho_\infty \in \widehat{(P \times Q)}^+.$$

By continuity of the projections $\pi_P$ and $\pi_Q$, it follows that

$$\lim_{n\to\infty} \pi_P(\rho_n) = \pi_P(\rho_\infty)$$
$$\lim_{n\to\infty} \pi_Q(\rho_n) = \pi_Q(\rho_\infty).$$

We show that $\rho \in F \otimes G$ if and only if $\rho_\infty \in F' \otimes G'$.

For the left-to-right implication, assume $\rho \in F \otimes G$. This means that $\pi_P(\rho) \in F$ and $\pi_Q(\rho) \in G$. Hence, $\pi_P(\rho_\infty) \in F'$ and $\pi_Q(\rho_\infty) \in G'$, i.e. $\rho_\infty \in F' \otimes G'$. The other implication goes along the same lines.

Therefore, the language $K \cap L$ recognized by $\mathcal{C}$ is $\mathcal{L}$-limitary, i.e. $\mathcal{L}$-limitary languages are closed under intersections. The proof for unions is very similar, and for complements its even simpler.

The last part of the proposition follows from the lemma below.

**Lemma 19.** *Let $F$ be a limitary acceptance condition. Then the acceptance condition $\exists F$ is also limitary, and is a projection of a language of the form $F' \otimes R$, where $R$ is a regular language.*

*Proof.* Let $F$ be a limitary acceptance condition with limit $F'$. We will construct a limit $\exists F'$ for the acceptance condition $\exists F$. Let $x$ be a profinite word over the alphabet $(P(Q) \times A)$. Then, $x \in \exists F'$ if and only if there is a profinite word $\hat{x}$ over the alphabet $(P(Q) \times A \times Q)$ such that

- The projection of $\hat{x}$ onto the first two coordinates yields $x$
- The projection of $\hat{x}$ onto the last coordinate yields a word in $F'$
- For every two consecutive letters $(P, a)$, $(P', a')$ of $x \cdot x$, $P' = \{\delta(q, a, i) : q \in P, i \in \{0, 1\}\}$
- For every letter $(P, a, q)$ of $\hat{x}$, $q \in P$
- For every two consecutive letters $(P, a, q), (P', a', q')$ of $\hat{x} \cdot \hat{x}$, $q' = \delta(q, a, 0)$ or $q' = \delta(q, a, 1)$

It is easy to see that $\exists F'$ is the limit of the acceptance condition $\exists F$.

### J.4 $\mathcal{L}$-limitary languages $\subseteq \omega\mathcal{L}$.

We will show that under some mild assumptions on $\mathcal{L}$, $\mathcal{L}$-limitary languages are contained in the class $\omega\mathcal{L}$ which appears in the statement of Theorem 9.

Let $\mathcal{A}$ be a deterministic automaton, and let $q_0 \in Q$ be its initial state. For a given word $w = a_1 a_2 a_3 \ldots a_n$, suppose that we run the automaton $\mathcal{A}$ and the obtained run is

$$q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} q_3 \cdots q_{n-1} \xrightarrow{a_n} q_n.$$

We define

$$\delta(a_1 a_2 \cdots a_n) = q_1 q_2 \cdots q_n.$$

The mapping $\delta$ extends uniquely to a continuous mapping

$$\delta \quad : \quad \widehat{A^+} \quad \longrightarrow \quad \widehat{Q^+}.$$

**Definition 2.** *We say that* $\mathcal{L}$ *is* closed under inverse images of deterministic transducers *if for every* $V \in \mathcal{L}$ *and every deterministic automaton* $\mathcal{A}$,

$$\delta^{-1}(V) \in \mathcal{L},$$

*where* $\delta$ *is the transition function of* $\mathcal{A}$.

**Proposition 23.** *Suppose that*

- $\mathcal{L}$ *is closed under inverse images of deterministic transducers,*
- $\mathcal{L}$ *is closed under intersecting with languages of the form:* $\{x \in \widehat{A^+} : x \text{ ends with letter } a\}$,
- $\mathcal{L}$ *is closed under* $\oplus$.

*Then, every* $\mathcal{L}$*-limitary language* $L \subseteq A^\omega$ *belongs to the class* $\omega\mathcal{L}$. *The translation is effective.*

*Proof.* Let $\mathcal{A}$ be an $F$-automaton, and let $\delta\colon Q \times A \to Q$ be its transition function. Let $q_0$ be the initial state of $\mathcal{A}$.

Let $w \in A^\omega$ be some input word. First we show that the following conditions are equivalent.

1. $w$ is accepted by $\mathcal{A}$
2. for some state $q \in Q$ and factorization

$$w = w_0 \cdot w_1 \cdot w_2 \ldots,$$

*i)* $\delta^{q_0}(w_0) = q$
*ii)* for each $n$, $\delta^q(w_n) = q$,
*iii)* the sequence of words $w_0, w_1, w_2, \ldots$ is convergent to some profinite word $w_\infty \in \widehat{Q^+}$,
*iv)* $\delta^q(w_\infty)$ belongs to $F'$.

We present the top-down implication. Let

$$\delta(w) = q_0 q_1 q_2 \ldots,$$

and let

$$w' = (a_1, q_1), (a_2, q_2), \ldots$$

be the combined word over the alphabet $A \times Q$, which includes both $w$ and the run of $\mathcal{A}$ over $w$. Let

$$\begin{array}{lcccc} \pi_Q & : & \widehat{(A \times Q)^+} & \longrightarrow & \widehat{Q^+} \\ \pi_A & : & \widehat{(A \times Q)^+} & \longrightarrow & \widehat{A^+} \end{array}$$

be projections, forgetting about the $A$- and $Q$-coordinate, respectively.

Apply the Ramsey lemma for infinite words to $w'$ to find a profinite word $w'_\infty \in \widehat{A^+}$ and factorization of $w'$

$$w' = w'_0 \cdot w'_1 \cdot w'_2 \cdots$$

such that the sequence of words $w'_0, w'_1, w'_2, \ldots$ is convergent to $w'_\infty$. Thanks to convergence, we may assume that the words $w'_1, w'_2, \ldots$ end with the same letter, say $(a, q)$, for some $q \in Q$ and $a \in A$.

It follows from the definition of acceptance by $\mathcal{A}$ that

$$\pi_q(w'_\infty) \in F'.$$

For $n = 0, 1, 2, \ldots$, let

$$w_n = \pi_A(w'_n) \in A^+,$$

and let

$$w_\infty = \pi_A(w_\infty) \in \widehat{A^+}.$$

By continuity of projection,

$$\lim_{n\to\infty} w_n = w_\infty.$$

Moreover, by determinism,

$$\pi_Q(w'_\infty) = \delta^q(w_\infty).$$

It follows that $q$ and the factorization

$$w = w_0 \cdot w_1 \cdot w_2 \ldots,$$

satisfy the conditions $i), ii), iii)$.

The bottom-up implication is similar.

For $p, q \in Q$, let

$$L_{p,q} \quad \stackrel{def}{=} \quad \{w \in A^+ :\ \delta^p(w) \text{ ends with } q\}$$

denote the regular language of finite words $w$ such that $\mathcal{A}$ reaches state $q$, after processing $w$, when starting from state $p$.

From the equivalence of the two conditions, it is easy to deduce that if $L$ is the language accepted by $\mathcal{A}$, then

$$L = \bigcup_{q\in Q} L_{q_0,q} \cdot V_q^\omega,$$

where

$$V_q = \{x \in \widehat{A^+} :\ \delta^q(x) \in (F' \cap \{\textit{profinite words ending with } q\})\}.$$

Of course, in the above union, we might as well consider only states $q$ which are reachable from $q_0$ – otherwise $L_{q_0,q}$ is the empty language. By assumption on $\mathcal{L}$, $V_q \in \mathcal{L}$. Together with the following fact, this proves that $L \in \omega\mathcal{L}$.

**Fact 24.** *Assume that $q$ is reachable from $q_0$. The expression $L_{q_0,q} \cdot V_q^\omega$ is well-formed.*

*Proof.* Let $N$ be the size of the minimal automaton for the language $L_{q_0,q}$. Observe that the minimal automaton for the language $L_{q,q}$ has size at most $N$. This is because the Myhill-Nerode equivalence of $L_{q_0,q}$ is contained in the Myhill-Nerode equivalence of $L_{q,q}$, as is easy to check.

Let $\varepsilon, \varepsilon'$ be two real positive numbers smaller than $\frac{1}{N}$. Let

$$w \in L_{q_0,q} \cdot (V_q)^{\omega}_{\varepsilon}.$$

We will show that

$$w \in L_{q_0,q} \cdot (V_q)^{\omega}_{\varepsilon'}.$$

By assumption, there exists a sequence $v_0, v_1, v_2, \ldots$ of finite words such that

- $v_0 \in L_{q_0,q}$
- $\lim_{n\to\infty} v_n = v_\infty \in V_q$
- $d(v_n, v_\infty) < \varepsilon < \frac{1}{N}$ for each $n$.

In particular, since $v_\infty \in V_q \subseteq \overline{L_{q,q}}$, it follows from the third item above that for each $n$, $v_n \in L_{q,q}$.

Therefore,

$$v_0 \cdot v_1 \cdots v_n \in L_{q_0,q} \qquad \text{for every } n.$$

We may choose $n$ large enough, so that

$$d(v_m, v_\infty) < \varepsilon' \qquad \text{for all } m \geq n.$$

Then, $w$ can be decomposed as

$$w = (v_0 \cdots v_n) \cdot (v_{n+1} \cdot v_{n+2} \cdots),$$

and

- $(v_0 \cdots v_n) \in L_{q_0,q}$,
- $\lim_{n\to\infty} v_n = v_\infty \in V_q$
- $d(v_n, v_\infty) < \varepsilon'$ for each $n$.

This proves that

$$w \in L_{q_0,q} \cdot (V_q)^{\omega}_{\varepsilon'},$$

ending the proof of the fact.

This ends the proof of the proposition.

### J.5 Proof of Theorem 9

In Theorem 9 we deal with classes $\mathcal{L}$ which satisfy the assumptions of Proposition 23. Hence, in each of the cases, $\omega\mathcal{L}$ contains the class of $\mathcal{L}$-limitary languages.

We prove that $\mathcal{L}$-limitary languages, in turn, contain the respective classes defined by various logics. We consider only the case when $\mathcal{L}$ is the class of languages definable in MSO+inf. For the other classes we proceed similarly, by using the logical characterizations of the respective classes of languages of infinite words:

- $\omega$-regular languages = languages definable by MSO,
- $\omega$B-regular languages = languages definable by MSO+$\mathbb{B}^+$,
- $\omega$S-regular languages = languages definable by MSO+$\mathbb{B}^-$,

and analogous characterizations of classes of languages of profinite words:

- regular languages = languages definable by MSO,
- B-regular languages = languages definable by MSO+$\inf^-$,
- S-regular languages = languages definable by MSO+$\inf^+$.

**Lemma 20.** *Let $\mathcal{L}$ be the class of languages definable in* MSO+inf*. Then languages definable in* MSO+$\mathbb{B}$ *are $\mathcal{L}$-limitary.*

*Proof.* Observe that the class $\mathcal{L}$ has all the good properties: closure under projection, complementation, $\oplus$, inverse images of deterministic transducers.

Let $L$ be a language defined by a formula $\varphi$ of MSO+$\mathbb{B}$. The proof proceeds by induction on the structure of the formula $\varphi \in$ MSO+$\mathbb{B}$. The inductive step follows from Proposition 22. It remains to prove the inductive base. The inductive base considers three basic languages, over the alphabet $\{a, b\} \times \{0, 1\}$:

- The set of infinite words in which every 1-labeled position has also label $a$
- The set of infinite words in which all 1-labeled positions are before all $a$-labeled positions
- The set of infinite words over the alphabet $\{a, b\}$ in which the lengths of the $a$ blocks are bounded.

The first two languages are $\omega$-regular, in particular, they are of the form $U \cdot V^\omega$, where $V$ is regular. The last language is equal to

$$(a + b)^* \cdot (a^{<\infty} b)^\omega.$$

This finishes the inductive proof, showing that every language definable in MSO+$\mathbb{B}$ belongs to $\omega\mathcal{L}$.