

Undecidability of MSO+“ultimately periodic”

Mikołaj Bojańczyk¹, Laure Daviaud¹, Bruno Guillon², Vincent Penelle¹, and A. V. Sreejith¹

¹ University of Warsaw

² University of Milan

Abstract

We prove that MSO on ω -words becomes undecidable if allowing to quantify over sets of positions that are ultimately periodic, *i.e.*, sets X such that for some positive integer p , ultimately either both or none of positions x and $x+p$ belong to X . We obtain it as a corollary of the undecidability of MSO on ω -words extended with the second-order predicate $U_1(X)$ which says that the distance between consecutive positions in a set $X \subseteq \mathbb{N}$ is unbounded. This is achieved by showing that adding U_1 to MSO gives a logic with the same expressive power as MSO+U, a logic on ω -words with undecidable satisfiability.

Keywords and phrases MSO logic, undecidability

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

This paper is about monadic second-order logic (MSO) on ω -words. Büchi’s famous theorem says that given an MSO sentence describing a set of ω -words over some alphabet, one can decide if the sentence is true in at least one ω -word [8]. Büchi’s theorem along with its proof using automata techniques have been the inspiration for a large number of decidability results for variants of MSO, including Rabin’s theorem on the decidability of MSO on infinite trees [10].

One of the themes developed in the wake of Büchi’s result is the question: what can be added to MSO on ω -words so that the logic remains decidable? One direction, studied already in the sixties, has been to extend the logic with predicates such as “*position x is a square number*” or “*position x is a prime number*”. See [2, 9] and the references therein for a discussion on this line of research. Another direction, which is the one taken in this paper, is to study quantifiers which bind set variables. These new quantifiers may for instance talk about the amount of sets satisfying a formula, hence being midway between the universal and the existential quantifiers, like the quantifier “*there exists uncountably many sets*” [1] (which *a posteriori* does not extend the expressivity of MSO on finitely branching trees). Another way is to consider quantifiers that talk about the asymptotic behaviors of infinite sets. This was already the direction followed in [3], where the authors define *Asymptotic* MSO, a logic which talks about the asymptotic behaviors of sequences of integers in a topological flavour. Another example is the quantifier U which was introduced in [5] and that says that some formula $\varphi(X)$ holds for arbitrarily large finite set X . However, in [7] it was shown that this logic has undecidable satisfiability; see [4] for a discussion on MSO+U, namely MSO extended with the quantifier U .

In this work, we study a new quantifier for MSO. Indeed, we consider a quantifier which talks about ultimately periodic sets. Recall that a set of positions $X \subseteq \mathbb{N}$ is called *ultimately periodic* if there is some period $p \in \mathbb{N}$ such that for sufficiently large positions $x \in \mathbb{N}$, either both or none of x and $x+p$ belong to X . We consider the logic MSO+P, *i.e.*, MSO augmented



© Mikołaj Bojańczyk, Laure Daviaud, Bruno Guillon, Vincent Penelle, A. V. Sreejith;
licensed under Creative Commons License CC-BY



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

XX:2 Undecidability of MSO+“ultimately periodic”

with the quantifier P that ranges over ultimately periodic sets:

$\text{PX}\varphi(X)$: “the formula $\varphi(X)$ is true for all ultimately periodic sets X ”.

► **Example 1.** The language of ω -words over $\{a, b\}$ such that the positions labeled by a form an ultimately periodic set is definable in $\text{MSO}+P$ by the following formula:

$$\exists X \quad \underbrace{\forall x \quad (x \in X \iff a(x))}_{X \text{ is the set of positions labeled by } a} \quad \wedge \quad \underbrace{\neg \left(\text{PY} \quad (X \neq Y) \right)}_{X \text{ is ultimately periodic}}.$$

Though our quantifier P extends the expressivity of MSO (the language from Example 1 is not definable in MSO), it is *a priori* not clear whether it has decidable or undecidable satisfiability. The main result of this paper, which answers an open question raised in [6], is that $\text{MSO}+P$ has undecidable satisfiability.

► **Theorem 2.** *Satisfiability over ω -words is undecidable for $\text{MSO}+P$.*

This result is obtained by a reduction from the above-mentioned logic $\text{MSO}+U$, which has undecidable satisfiability over ω -words [7], where the quantifier U is defined by:

$\text{UX}\varphi(X)$: “for all $k \in \mathbb{N}$, $\varphi(X)$ is true for some finite set X of size at least k ”.

The logic $\text{MSO}+U$ was also used in [3, Theorem 13] where it is reduced to Asymptotic MSO ; however, it is not clear how to compare the expressive power of $\text{MSO}+P$ with the one of Asymptotic MSO over ω -words. In our case, the reduction from $\text{MSO}+U$ to $\text{MSO}+P$ is obtained by a sequence of reductions, involving other variants of MSO that are obtained by adding some second-order predicates. We first consider the logic $\text{MSO}+U_1$, namely MSO extended with the second-order predicate U_1 defined by:

$U_1(X)$: “for all $k \in \mathbb{N}$, there exist two consecutive positions of X at distance at least k ”.

► **Example 3 (from [6]).** Consider the language of ω -words of the form $a^{n_1}ba^{n_2}b\dots$ such that $\{n_1, n_2, \dots\}$ is unbounded. This language can be defined in $\text{MSO}+U$ by a formula saying that there are factors of consecutive a 's of arbitrarily large size. It can also be defined in $\text{MSO}+U_1$ saying that the set of the positions labeled by b satisfies the predicate U_1 .

It is easy to see that the predicate U_1 can be defined in the logic $\text{MSO}+U$: a set X satisfies $U_1(X)$ if and only if there exist intervals (finite connected sets of positions) of arbitrarily large size which are disjoint with X . Therefore, the logic $\text{MSO}+U_1$ can be seen as a fragment of $\text{MSO}+U$. Is this fragment proper? The principal technical contribution of this paper, which implies Theorem 2, is showing that actually the two logics are the same:

► **Theorem 4.** *The logics $\text{MSO}+U$ and $\text{MSO}+U_1$ define the same languages of ω -words, and translations both ways are effective.*

Proof of Theorem 2 assuming Theorem 4. Since $\text{MSO}+U$ has undecidable satisfiability [7], the same follows for $\text{MSO}+U_1$ by Theorem 4. Moreover, the predicate U_1 can be defined in terms of ultimate periodicity: a set X of positions satisfies $U_1(X)$ if and only if for all infinite ultimately periodic sets Y , there are at least two positions in Y that are not separated by a position from X . It follows that every sentence of $\text{MSO}+U_1$ can be effectively rewritten into a sentence of $\text{MSO}+P$ which is true on the same ω -words. Hence, Theorem 2 follows. ◀

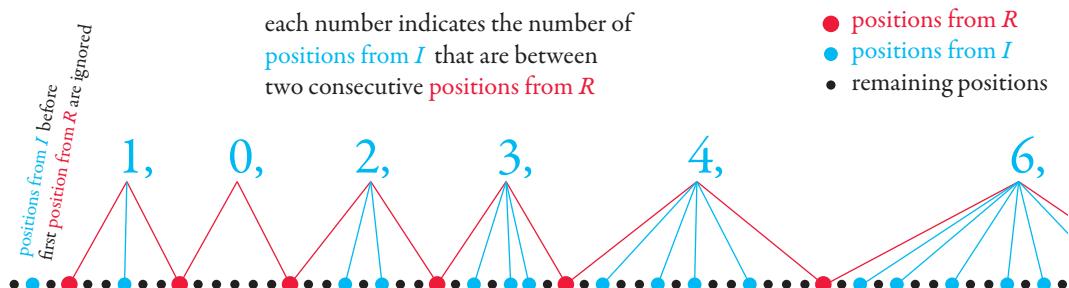
Outline of the paper. The rest of the paper is mainly devoted to the proof of Theorem 4. In Section 2 we introduce an intermediate logic $\text{MSO} + \mathbf{U}_2$. We first prove that $\text{MSO} + \mathbf{U}$ and $\text{MSO} + \mathbf{U}_2$ are effectively equivalent (Section 2.1, Lemma 6). Then, we prove that $\text{MSO} + \mathbf{U}_2$ and $\text{MSO} + \mathbf{U}_1$ are also effectively equivalent (Section 2.2, Lemma 7), assuming a certain property, namely Lemma 8. The proof of this latter lemma is the subject of Section 3. Finally, in Section 4, we discuss the expressive power of $\text{MSO} + \mathbf{P}$ with respect to $\text{MSO} + \mathbf{U}$. We show that the property “ultimately periodic” can be expressed in $\text{MSO} + \mathbf{U}$ if allowing a certain encoding.

2 How $\text{MSO} + \mathbf{U}_1$ talks about vector sequences

In this section, we introduce a new predicate on pairs of sets of positions $\mathbf{U}_2(R, I)$ and consider the logic MSO extended with this predicate: $\text{MSO} + \mathbf{U}_2$. We first prove in Section 2.1, that $\text{MSO} + \mathbf{U}$ has the same expressive power as $\text{MSO} + \mathbf{U}_2$. Then, we show in Section 2.2, that $\text{MSO} + \mathbf{U}_2$ is itself as expressive as $\text{MSO} + \mathbf{U}_1$, assuming a certain property, namely Lemma 8. The proof of this lemma is given in Section 3.

2.1 From quantifier \mathbf{U} to predicate \mathbf{U}_2

We say that a sequence of natural numbers is *unbounded* if arbitrarily large numbers occur in it. For two disjoint sets of positions $R, I \subseteq \mathbb{N}$ with R infinite, we define a sequence of numbers as in Figure 1: the i -th element of this sequence is the number of elements from I between the i -th and the $(i+1)$ -th elements from R . In particular, elements of I in positions smaller than all the positions from R are not relevant.



■ **Figure 1** Two sets of positions $R, I \subseteq \mathbb{N}$ and the sequence in \mathbb{N}^ω that they define. The sequence is only defined when R and I are disjoint, and R is infinite.

We now define the binary predicate \mathbf{U}_2 :

$\mathbf{U}_2(R, I) :$ “the sequence of numbers encoded by R, I is defined and unbounded”.

The difference between the predicate \mathbf{U}_1 , which has one free set variable, and the predicate \mathbf{U}_2 , which has two free set variables, is that the latter is able to ignore some positions. It is illustrated by the following example, which can be easily defined in $\text{MSO} + \mathbf{U}_2$.

► **Example 5.** Let L be the union of the languages of ω -words of the form $(ac^*)^{n_1}b(ac^*)^{n_2}b\dots$ where $\{n_1, n_2, \dots\}$ is unbounded. It can be defined in $\text{MSO} + \mathbf{U}_2$ in the following way: an ω -word belongs to L if and only if it belongs to $((ac^*)^*b)^\omega$ and the sequence of numbers encoded by $R = \{\text{set of positions labeled by } b\}$ and $I = \{\text{set of positions labeled by } a\}$ is unbounded. The language also admits a simple definition in $\text{MSO} + \mathbf{U}$: there exist arbitrarily large finite sets of positions labeled by a such that no two positions from the set are separated

XX:4 Undecidability of MSO+“ultimately periodic”

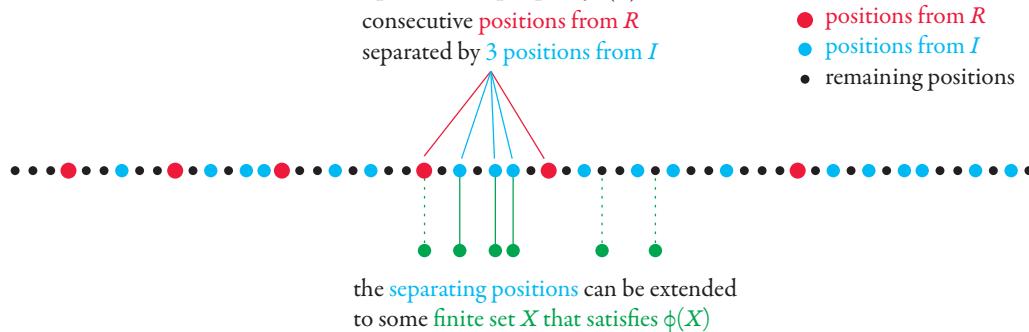
by a b . It is however not straightforward to define it in the logic $\text{MSO}+\text{U}_1$, since in every ω -word, the distance between two successive b 's separated by at least one a can be increased by adding occurrences of c 's, without changing the membership of the word to the language.

The following lemma follows essentially from the same argument as in [6, Lemma 5.5].

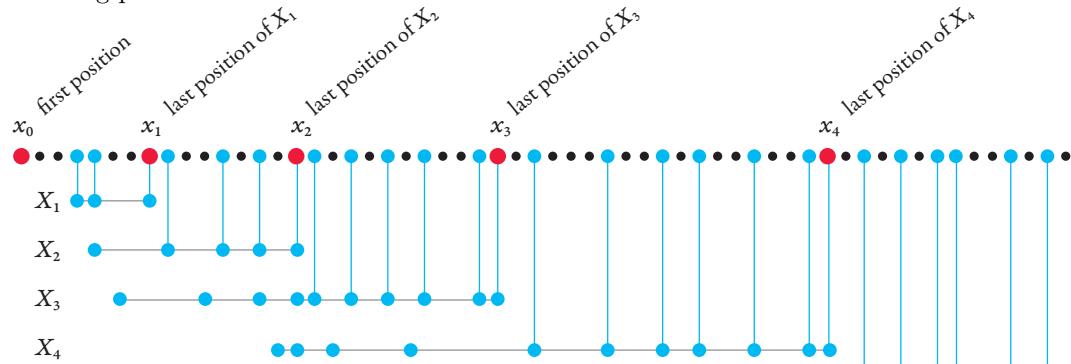
► **Lemma 6.** *The logics $\text{MSO}+\text{U}$ and $\text{MSO}+\text{U}_2$ define the same languages of ω -words, and translations both ways are effective.*

Proof. The predicate U_2 is easily seen to be expressible in $\text{MSO}+\text{U}$: a sequence of numbers encoded by R and I is unbounded if the subsets of elements of I between two consecutive elements of R are of arbitrarily large size, which is expressible in $\text{MSO}+\text{U}$. For the converse implication, we use the following observation. A formula $\text{UX}\varphi(X)$ is true if and only if the following condition holds:

- (*) There exist two sets $R, I \subseteq \mathbb{N}$ which satisfy $\text{U}_2(R, I)$ such that for every two consecutive positions $r, s \in R$, there exists some finite X satisfying $\varphi(X)$ which contains all positions of I between r and s . Here is a picture of property (*):



Condition (*) is clearly expressible in MSO with the predicate U_2 . Hence the lemma will follow once we prove the equivalence: $\text{UX}\varphi(X)$ if and only if (*). The right-to-left implication is easy to see. For the converse implication, we do the following construction. We define a sequence of positions $0 = x_0 < x_1 < \dots$ as follows by induction. Define x_0 to be the first position, *i.e.*, the number 0. Suppose that x_n has already been defined. By the assumption $\text{UX}\varphi(X)$, there exists a set X_{n+1} which satisfies φ and which contains at least n positions after x_n . Define x_{n+1} to be the last position of X_{n+1} . This process is illustrated in the following picture:



Define R to be all the positions x_0, x_1, \dots in the sequence thus obtained, and define I to be the set of positions x such that $x_n < x < x_{n+1}$ and $x \in X_{n+1}$ holds for some n . By construction, the sets I and R thus obtained will satisfy U_2 . ◀

2.2 From predicate U_2 to predicate U_1

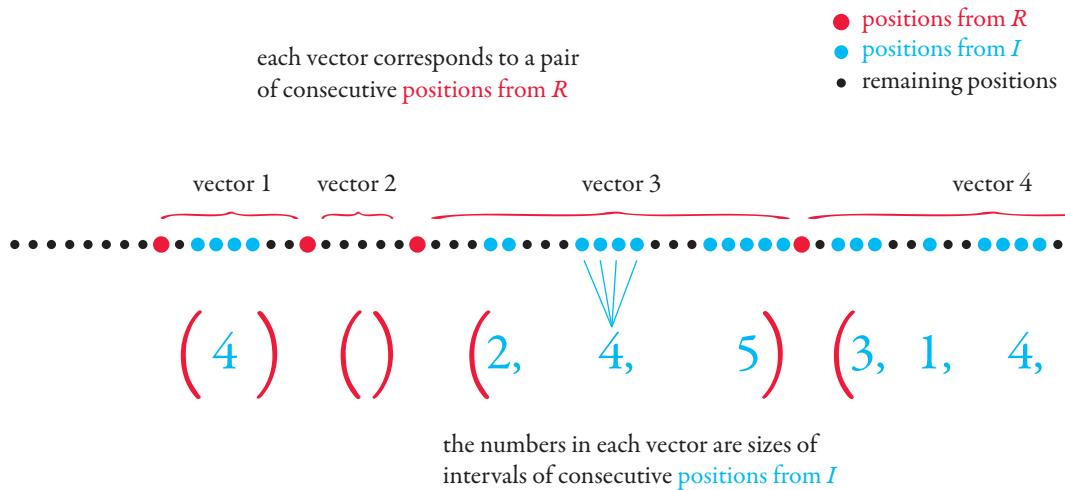
Lemma 6 above states that $\text{MSO} + U$ and $\text{MSO} + U_2$ have the same expressive power, thus giving a first step towards the proof of Theorem 4. The second step, which is the key point of our result, is the following lemma, which states that $\text{MSO} + U_2$ is as expressive as $\text{MSO} + U_1$.

► **Lemma 7.** *The logics $\text{MSO} + U_2$ and $\text{MSO} + U_1$ define the same languages of ω -words, and translations both ways are effective.*

Note that one direction is straightforward, since $U_1(X)$ holds if and only if $U_2(X, \mathbb{N} \setminus X)$ holds. Hence, it remains to show that the predicate U_2 can be defined by a formula of the logic $\text{MSO} + U_1$ with two free set variables.

In our proof, we use terminology and techniques about sequences of vectors of natural numbers that were used in the undecidability proof for $\text{MSO} + U$ [7]. A *vector sequence* is defined to be an element of $(\mathbb{N}^*)^\omega$, i.e., a sequence of possibly empty tuples of natural numbers. For a vector sequence $\mathbf{f} \in (\mathbb{N}^*)^\omega$, we define its *dimension*, denoted by $\dim(\mathbf{f}) \in \mathbb{N}^\omega$, as being the number sequence of the dimensions of the vectors: the i -th element in $\dim(\mathbf{f})$ is the *dimension of the i -th vector in \mathbf{f}* , i.e., the number of coordinates in the i -th tuple in \mathbf{f} . We say that a vector sequence \mathbf{f} *tends towards infinity*, denoted by $\mathbf{f} \rightarrow \infty$, if every natural number appears in finitely many vectors from \mathbf{f} .

Recall Figure 1, which showed how to encode a number sequence using two sets of positions $R, I \subseteq \mathbb{N}$. We now show that the same two sets of positions can be used to define a vector sequence. As it was the case in Figure 1, we assume that R is infinite and disjoint from I . Under these assumptions, we write $\mathbf{f}_{R,I} \in (\mathbb{N}^*)^\omega$ for the vector sequence defined according to the description from Figure 2, i.e., the sequence of vectors whose coordinates are the lengths of intervals in I between two consecutive elements of R . Remark that 0 is not encoded and thus that $\mathbf{f}_{R,I}$ contains possibly empty vectors with only positive coordinates. As for Figure 1, the positions of I smaller than all the positions of R are not relevant.



■ **Figure 2** Two sets of positions $I, R \subseteq \mathbb{N}$ and the vector sequence that they encode. This vector sequence is defined only when I and R are disjoint and R is infinite.

The following lemma states that the unboundedness of the dimensions of a vector sequence can be expressed in $\text{MSO} + U_1$. Its proof will be the subject of Section 3.

XX:6 Undecidability of MSO+“ultimately periodic”

► **Lemma 8.** *There is a formula $\varphi(R, I)$ in $\text{MSO} + U_1$ which is true if and only if the vector sequence $\mathbf{f}_{R,I}$ is defined and satisfies $\dim(\mathbf{f}_{R,I})$ is unbounded.*

Proof of Lemma 7 assuming Lemma 8. As the predicate U_1 is easily definable in $\text{MSO} + U_2$ ($U_1(X)$ holds if and only if $U_2(X, \mathbb{N} \setminus X)$ holds), it remains to prove the converse, i.e., to exhibit a formula $\varphi(X, Y)$ of $\text{MSO} + U_1$ which holds if and only if $U_2(X, Y)$ holds.

By definition of U_2 , it is sufficient to write a formula $\varphi(R, I)$ which holds if and only if $U_2(R, I)$ holds, assuming R and I are disjoint and R is infinite. We fix such R and I , which hence define a vector sequence $\mathbf{f}_{R,I}$ according to the encoding of Figure 2. Furthermore, $U_2(R, I)$ holds if and only if summing all coordinates of each vector from the sequence yields a number sequence which is unbounded. We show that we can construct a $J \subseteq I$ such that $U_2(R, J)$ holds if and only if the dimension of $\mathbf{f}_{R,J}$ is unbounded.

For every non negative integer x , if $2x$ belongs to I as well as either $2x+1$ or $2x-1$, then we remove $2x$ from I . Let J denote the thus obtained set. As at least one position from I over two is kept in J , we have that $U_2(R, I)$ holds if and only if $U_2(R, J)$ holds. Furthermore, J has the property that each of its point is isolated whence the sum of the coordinates of a vector from $\mathbf{f}_{R,J}$ is equal to its dimension. Therefore, $U_2(R, J)$ holds if and only if the dimension of the vector sequence $\mathbf{f}_{R,J}$ is unbounded. Moreover, J is definable in MSO in the sense that there is a MSO -formula with two free set variables I and J which holds if and only if J is obtained from I by the process given above.

We conclude the proof of Lemma 7 by applying Lemma 8. ◀

3 Expressing unboundedness of dimensions

This section is devoted to proving Lemma 8. Our proof has two steps. In Section 3.1, we show that in order to prove Lemma 8, it suffices to show the following lemma.

► **Lemma 9.** *There is a formula $\varphi(R, I)$ in $\text{MSO} + U_1$ which is true if and only if the vector sequence $\mathbf{f}_{R,I}$ is defined, satisfies $\mathbf{f}_{R,I} \rightarrow \infty$ and $\dim(\mathbf{f}_{R,I})$ is unbounded.*

Lemma 9 itself will be proved in Section 3.2.

3.1 From Lemma 9 to Lemma 8

Recall that the vector sequence $\mathbf{f}_{R,I}$ is defined if and only if R is infinite and disjoint from a set I , which is clearly expressible in MSO . So from now on, we will always consider vector sequences $\mathbf{f}_{R,I}$ which are defined.

Given a vector sequence, a *sub-sequence* is an infinite vector sequence obtained by dropping some of the vectors, like this:

$$\begin{array}{ccccccccc} (\color{blue}{1}, \color{blue}{2}) & (\color{red}{1}, \cancel{\color{red}2}, \cancel{\color{red}3}, \color{blue}{4}) & (\color{blue}{2}, \color{blue}{3}) & (\color{blue}{7}, \color{blue}{2}, \color{blue}{3}, \color{blue}{5}) & (\cancel{\color{blue}6}, \cancel{\color{blue}4}, \cancel{\color{blue}2}, \color{blue}{8}, \color{blue}{9}) & (\cancel{\color{blue}1}, \cancel{\color{blue}3}, \cancel{\color{blue}4}, \color{blue}{5}) & (\color{blue}{1}, \color{blue}{3}, \color{blue}{4}, \color{blue}{5}) & \dots \\ (\color{blue}{1}, \color{blue}{2}) & & (\color{blue}{2}, \color{blue}{3}) & (\color{blue}{7}, \color{blue}{2}, \color{blue}{3}, \color{blue}{5}) & & & (\color{blue}{1}, \color{blue}{3}, \color{blue}{4}, \color{blue}{5}) & \dots \end{array}$$

An *extraction* is obtained by removing some of the coordinates in some of the vectors, but keeping at least one coordinate from each vector, like this:

$$\begin{array}{ccccccccc} (\color{blue}{1}, \cancel{\color{blue}2}) & (\cancel{\color{red}1}, \color{red}3, \cancel{\color{red}4}) & (\color{blue}{2}, \color{blue}{3}) & (\color{blue}{7}, \cancel{\color{blue}2}, \color{blue}{3}, \color{blue}{5}) & (\color{red}6, \cancel{\color{red}4}, \color{red}2, \cancel{\color{red}8}, \color{red}9) & (\color{blue}{1}, \color{blue}{3}, \cancel{\color{blue}4}, \color{blue}{5}) & (\color{red}1, \cancel{\color{red}3}, \color{red}4, \color{red}5) & \dots \\ (\color{blue}{1}, \color{blue}{ }) & (\color{red}3, \color{red}{ }) & (\color{blue}{2}, \color{blue}{3}) & (\color{blue}{7}, \color{blue}{3}, \color{blue}{5}) & (\color{red}6, \color{red}{2}, \color{red}{ }) & (\color{blue}{1}, \color{blue}{3}, \color{blue}{5}) & (\color{red}1, \color{red}{4}, \color{red}{5}) & \dots \end{array}$$

In particular a vector sequence containing at least one empty vector has no extraction. An

extraction is called *interval-closed* if the coordinates which are kept in a given vector are consecutive, like this:

$$(1, \cancel{X}) \quad (\cancel{X}, 3, \cancel{X}) \quad (2, 3) \quad (\cancel{X}, \cancel{X}, 3, 5) \quad (\cancel{X}, 4, 2, \cancel{X}, \cancel{X}) \quad (1, 3, \cancel{X}, \cancel{X}) \quad (\cancel{X}, 3, 4, \cancel{X}) \dots \\ (1) \quad (3) \quad (2, 3) \quad (3, 5) \quad (4, 2) \quad (1, 3) \quad (3, 4) \dots$$

A *1-extraction* is an extraction with only vectors of dimension 1. In particular, it is interval-closed. A *sub-extraction* is an extraction of a sub-sequence.

The following lemma gives two conditions equivalent to the unboundedness of the dimension of a vector sequence.

► **Lemma 10.** *Let \mathbf{f} be a vector sequence. Then $\dim(\mathbf{f})$ is unbounded if and only if one of the following conditions is true:*

- (1) *There exists a sub-extraction \mathbf{g} of \mathbf{f} such that $\mathbf{g} \rightarrow \infty$ and $\dim(\mathbf{g})$ is unbounded.*
- (2) *There exists an interval-closed sub-extraction \mathbf{g} of \mathbf{f} such that $\dim(\mathbf{g})$ is unbounded and \mathbf{g} is bounded, i.e., there exists a bound $n \in \mathbb{N}$ such that all the coordinates occurring in the sequence are less than n .*

Proof. The if implication is clear, since admitting a sub-extraction of unbounded dimension trivially implies the unboundedness of the dimension.

We now focus on the only-if implication. Assume that $\dim(\mathbf{f})$ is unbounded and condition (1) is false. Then, for all sub-extractions \mathbf{g} of \mathbf{f} we have that $\mathbf{g} \rightarrow \infty$ implies $\dim(\mathbf{g})$ is bounded. This implies that there exists a constant m such that for all vectors v in \mathbf{f} , the number of coordinates greater than m is less than m . Indeed, if no such m exists, we can exhibit a sub-extraction \mathbf{g} of \mathbf{f} , which tends towards infinity and has unbounded dimension. Let us mark the coordinates smaller than m in every vector of \mathbf{f} . For every integer n , we can find a vector of \mathbf{f} with at least n consecutive marked coordinates. Indeed suppose there is at most n consecutive marked coordinates in every vector. Then, as there are at most m unmarked coordinates, a vector cannot have dimension greater than $n(m+1) + m$ (i.e., $m+1$ blocks of n marked positions plus m unmarked positions separating them). This contradicts that $\dim(\mathbf{f})$ is unbounded.

Thus by keeping for each vector one of the longest block of consecutive marked positions and removing vectors without any marked position, we can construct an interval-closed sub-extraction \mathbf{g} of \mathbf{f} which is bounded by m and has unbounded dimension. ◀

We now proceed to show how Lemma 9 implies Lemma 8. Recall that Lemma 8 says that there is a formula $\varphi(R, I)$ in MSO+U₁ which is true if and only if the vector sequence $\mathbf{f}_{R,I}$ is defined and satisfies $\dim(\mathbf{f}_{R,I})$ is unbounded.

Proof of Lemma 8 assuming Lemma 9. Consider R and I two sets of positions such that $\mathbf{f}_{R,I}$ is defined. We will use Lemma 10 to express in MSO+U₁ that $\dim(\mathbf{f}_{R,I})$ is unbounded. The first condition of the lemma is expressible in MSO+U₁ by Lemma 9, as shown in the next section. However, in order to express the second condition, we need that the gaps between consecutive intervals of I have bounded length. This is ensured by extending the intervals as explained in the following. For any two disjoint sets R and I , let us construct J as follows: for every two consecutive positions $x < y$ in I such that there is no element of R in between, we add to I all the positions between x and y , except $y - 1$, as in Figure 3. The set J is expressible in MSO, in the sense that there is a MSO-formula with three free set variables R , I , J which holds if and only if J is obtained from I by this process. Moreover, it is clear that $\dim(\mathbf{f}_{R,I})$ is unbounded if and only if $\dim(\mathbf{f}_{R,J})$ is unbounded since the dimension in

XX:8 Undecidability of MSO+“ultimately periodic”

both sequences are pairwise equal (see Figure 3). It suffices thus to prove Lemma 8 for sets R and I which satisfy:

- (*) between two consecutive elements of I , there is either an element of R or at most one position not in I .

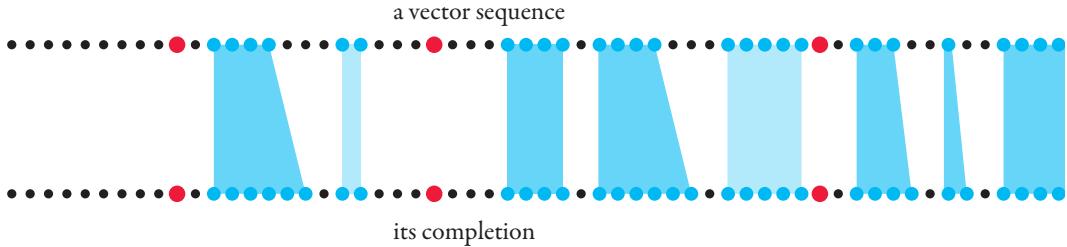


Figure 3 Completion of a vector sequence.

Let R and I be two sets satisfying (*). We want to express that $\dim(\mathbf{f}_{R,I})$ is unbounded. It suffices to show that for each condition in Lemma 10, there exists a formula of $\text{MSO}+\text{U}_1$ which says that the condition is satisfied by $\mathbf{f}_{R,I}$. We treat the two cases separately.

1. We want to say that there exists a sub-extraction \mathbf{g} of $\mathbf{f}_{R,I}$ such that $\mathbf{g} \rightarrow \infty$ and $\dim(\mathbf{g})$ is unbounded. Sub-extraction can be simulated in MSO according to the picture in Figure 4, and the condition “ $\mathbf{g} \rightarrow \infty$ and $\dim(\mathbf{g})$ is unbounded” can be checked using Lemma 9.

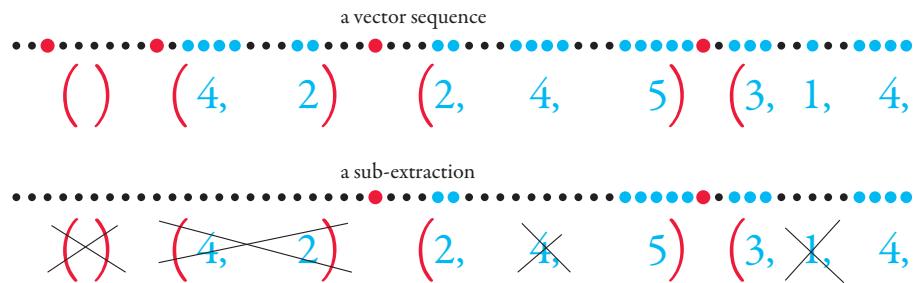


Figure 4 Sub-extraction of a vector sequence.

2. We want to say that there exists an interval-closed sub-extraction \mathbf{g} of $\mathbf{f}_{R,I}$ such that \mathbf{g} is bounded and $\dim(\mathbf{g})$ is unbounded. We use the same approach as in the previous item, *i.e.*, we simulate sub-extraction in the logic using the encoding from Figure 4, thus obtaining R', I' . Additionally, we ensure that the selected sub-extraction is interval-closed. Importantly, taking an interval-closed sub-extraction preserves property (*). It remains therefore to write a formula in $\text{MSO}+\text{U}_1$ which says that sets R', I' satisfy $\mathbf{f}_{R',I'}$ is bounded and $\dim(\mathbf{f}_{R',I'})$ is unbounded.

- The boundedness of $\mathbf{f}_{R',I'}$ is expressed in $\text{MSO}+\text{U}_1$ by the fact that the complement of I' has to satisfy the negation of U_1 . Indeed, this exactly says that the intervals of I' are bounded.
- Now, since $\mathbf{f}_{R',I'}$ is bounded then $\dim(\mathbf{f}_{R',I'})$ is unbounded if and only if the sequence of the number of occurrences of elements of I' between two consecutive elements of R' is unbounded. Consider the set X of all the positions which are either in R' or not adjacent to an element in I' (which is expressible in MSO). Then, since R' and I' satisfy (*), $\dim(\mathbf{f}_{R',I'})$ is unbounded if and only if $\text{U}_1(X)$ holds.

This completes the reduction of Lemma 8 to Lemma 9. ◀

3.2 Unboundedness of dimensions for sequences tending to infinity

We now prove Lemma 9 whence concluding the proof of Theorem 4. The lemma says that there is a formula $\varphi(R, I)$ in $\text{MSO} + \text{U}_1$ which is true if and only if the vector sequence $\mathbf{f}_{R,I}$ is defined and satisfies $\mathbf{f}_{R,I} \rightarrow \infty$ and $\dim(\mathbf{f}_{R,I})$ is unbounded.

The main idea of the proof is to relate the unboundedness of the dimension with a property definable in $\text{MSO} + \text{U}_1$, namely the *asymptotic mix* property, which was already used in [7] to prove undecidability of $\text{MSO} + \text{U}$. This is done in Lemma 11 below, for which we need a couple of notions on vector sequences.

Given an infinite set $S \subseteq \mathbb{N}$ of positions and a number sequence $f \in \mathbb{N}^\omega$, the *restriction of f to S* is the number sequence obtained by discarding elements of index not in S and is denoted $f|_S$. Two number sequences f and g are called *asymptotically equivalent*, denoted by $f \sim g$, if they are bounded on the same sets of positions, *i.e.*, if for every set S , $f|_S$ is bounded if and only if $g|_S$ is bounded. For example, the number sequences $f = 1, 1, 3, 1, 5, 1, 7, 1, 9, \dots$ and $g = 2, 1, 4, 1, 6, 1, 8, 1, 10, \dots$ are asymptotically equivalent, but f and $h = 1, 3, 1, 5, 1, 7, 1, 9, 1, \dots$ are not.

We say that a vector sequence \mathbf{f} is an *asymptotic mix* of a vector sequence \mathbf{g} if for all 1-extraction f of \mathbf{f} , there exists a 1-extraction g of \mathbf{g} such that $f \sim g$. In particular, the empty vector cannot occur in \mathbf{f} or \mathbf{g} . Note that this is not a symmetric relation.

Given two vector sequences \mathbf{f} and \mathbf{g} , we say that \mathbf{g} *dominates* \mathbf{f} , denoted $\mathbf{f} \leq \mathbf{g}$, if for all i , the i -th vectors in both sequences have the same dimension, and the i -th vector of \mathbf{f} is coordinatewise smaller than or equal to the i -th vector of \mathbf{g} . Furthermore, we suppose that all the coordinates in \mathbf{f} and \mathbf{g} are positive.

An extraction \mathbf{g} of \mathbf{f} is said *strict*, if at least one coordinate by vector of \mathbf{f} has been popped in \mathbf{g} , *i.e.*, the dimensions of the vectors of \mathbf{g} are pointwise smaller than those of \mathbf{f} . In particular, a vector sequence with some vectors of dimension less than 2 admit no strict extraction, nevertheless, it may admit strict sub-extractions.

The following lemma relates the notion of asymptotic mix with the unboundedness of the dimension of a vector sequence, providing the vector sequence tends towards infinity.

► **Lemma 11.** *Let \mathbf{f} be a vector sequence such that $\mathbf{f} \rightarrow \infty$. Then, $\dim(\mathbf{f})$ is unbounded if and only if there exists a sub-sequence \mathbf{h} and a strict extraction \mathbf{g} of \mathbf{h} satisfying:*

$(P_{\mathbf{g},\mathbf{h}})$ *for every $\mathbf{h}' \leq \mathbf{h}$, there exists $\mathbf{g}' \leq \mathbf{g}$ such that \mathbf{h}' is an asymptotic mix of \mathbf{g}' .*

Proof. The if direction follows from [7, Lemma 2.2], while the converse is new to this work. More precisely, it is proved in [7, Lemma 2.2] that given two vector sequences \mathbf{h} and \mathbf{g} tending towards infinity, if \mathbf{h} and \mathbf{g} have bounded dimension, then the two following properties are equivalent:

- (1) for infinitely many i 's, the i -th vector of \mathbf{h} has higher dimension than the i -th vector of \mathbf{g} ;
- (2) the negation of $(P_{\mathbf{g},\mathbf{h}})$: there exists $\mathbf{h}' \leq \mathbf{h}$ which is not an asymptotic mix of any $\mathbf{g}' \leq \mathbf{g}$.

Let \mathbf{f} be a vector sequence which tends towards infinity and suppose that it has bounded dimension. Let \mathbf{h} be a sub-sequence of \mathbf{f} and \mathbf{g} be an extraction of \mathbf{h} . By definition, \mathbf{h} and \mathbf{g} satisfy (1), have bounded dimension and tend towards infinity. Therefore, by [7, Lemma 2.2], they satisfy (2), *i.e.*, the negation of $(P_{\mathbf{g},\mathbf{h}})$.

We now prove the only if direction. We suppose that \mathbf{f} tends towards infinity and that $\dim(\mathbf{f})$ is unbounded. There exists a sub-sequence \mathbf{h} of \mathbf{f} such that every vector occurring in \mathbf{h} has dimension at least 2 and $\dim(\mathbf{h})$ tends towards infinity.

XX:10 Undecidability of MSO+“ultimately periodic”

Define \mathbf{g} as being the extraction obtained from \mathbf{h} by dropping the last coordinate in each vector and observe that it is a strict extraction of \mathbf{h} . We prove now a result stronger than required, by exhibiting a universal $\mathbf{g}' \leq \mathbf{g}$ which makes $(P_{\mathbf{g}, \mathbf{h}})$ true whatever the choice of $\mathbf{h}' \leq \mathbf{h}$, hence inverting the order of the universal and the existential quantifiers. Let \mathbf{g}' be the vector sequence defined as follows: for each position i , denoting the i -th vector of \mathbf{h} by (v_1, v_2, \dots, v_k) , the i -th vector of \mathbf{g}' is set to $(\min(v_1, 1), \min(v_2, 2), \dots, \min(v_{k-1}, k-1))$. For instance:

$$\begin{aligned}\mathbf{h} &= (2, 1), (3, 2, 1), (4, 3, 2, 1), (5, 4, 3, 2, 1), (6, 5, 4, 3, 2, 1), \dots \\ \mathbf{g} &= (2), (3, 2), (4, 3, 2), (5, 4, 3, 2), (6, 5, 4, 3, 2), \dots \\ \mathbf{g}' &= (1), (1, 2), (1, 2, 2), (1, 2, 3, 2), (1, 2, 3, 3, 2), \dots\end{aligned}$$

We now prove that every $\mathbf{h}' \leq \mathbf{h}$ is an asymptotic mix of \mathbf{g}' , that is, for every 1-extraction h of \mathbf{h}' , there exists a 1-extraction g of \mathbf{g}' such that $h \sim g$. We fix an arbitrary vector sequence $\mathbf{h}' \leq \mathbf{h}$ and a 1-extraction h of \mathbf{h}' . We define g as the number sequence whose i -th number, for each position i , is the maximal coordinate in the i -th vector of \mathbf{g}' which is less than or equal to the i -th number in h . This coordinate always exists since the first coordinate of every vector in the sequence \mathbf{g}' exists and is equal to 1. For instance, according to the previous example and given \mathbf{h}' and h , the number sequence g is defined as follows:

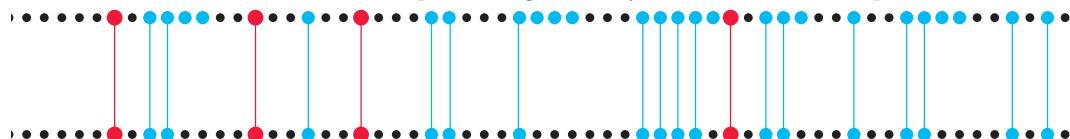
$$\begin{aligned}\mathbf{h}' &= (1, 1), (2, 1, 1), (3, 2, 1, 1), (4, 3, 2, 1, 1), (5, 4, 3, 2, 1, 1), \dots \\ h &= 1, 2, 1, 4, 1, \dots \\ \mathbf{g}' &= (1), (1, 2), (1, 2, 2), (1, 2, 3, 2), (1, 2, 3, 3, 2), \dots \\ g &= 1, 2, 1, 3, 1, \dots\end{aligned}$$

Observe that, by definition, for every position i , the i -th number of g is less than or equal to the i -th number of h . Hence, for every set S of positions, if $h|_S$ is bounded then so is $g|_S$. Suppose now that for some set S of positions, $h|_S$ is unbounded and fix an integer n . We want to prove that some number in $g|_S$ is greater than n . Since both \mathbf{h} and $\dim(\mathbf{h})$ tend towards infinity, we can find a threshold $m \in \mathbb{N}$ such that for every position $i \geq m$, the i -th vector of \mathbf{h} has dimension greater than n and furthermore all its coordinates are greater than n . Thus, for all positions $i \geq m$, the n -th coordinate of the i -th vector of \mathbf{g}' is defined and it is equal to n . Therefore, for all $i \in S$ with $i \geq m$ and such that the i -th number in h is at least n , the i -th number in g is at least n as well. Hence, $g|_S$ is unbounded. \blacktriangleleft

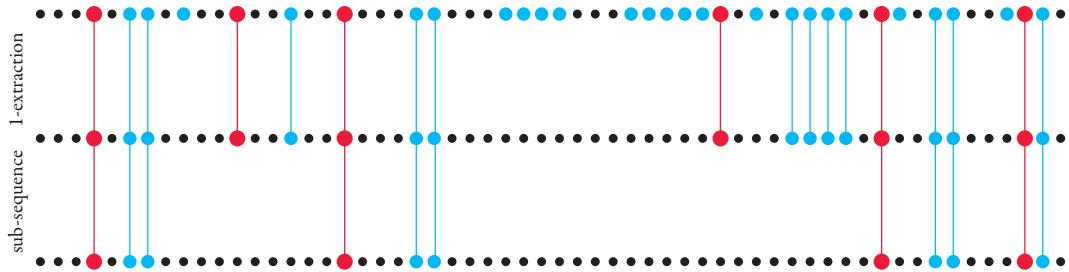
To conclude the proof of Lemma 9, it remains to express in MSO+U₁ the property stated in Lemma 11: given two sets R and I encoding a vector sequence $\mathbf{f}_{R,I}$ which tends towards infinity, there exists a sub-sequence \mathbf{h} of $\mathbf{f}_{R,I}$ and a strict extraction \mathbf{g} of \mathbf{h} satisfying $(P_{\mathbf{g}, \mathbf{h}})$.

Given a vector sequence \mathbf{f} and a set S , we say that an encoding of \mathbf{f} is S -synchronised if it is of the form S, J for some set J . Sub-sequences and strict extractions can be simulated in MSO by defining subsets of the original encoding, according to the picture in Figure 4. In particular, given R and I , one can construct in MSO any sub-sequence \mathbf{h} of $\mathbf{f}_{R,I}$ and any strict extraction \mathbf{g} of \mathbf{h} , such that the encodings of \mathbf{h} and \mathbf{g} are S -synchronised for some $S \subseteq R$.

Now, we express $(P_{\mathbf{g}, \mathbf{h}})$ on \mathbf{h} and \mathbf{g} encoded by S, J and S, J' respectively. Every dominated sequence \mathbf{h}' (resp. \mathbf{g}') of \mathbf{h} (resp. \mathbf{g}) can be obtained by deleting some rightmost consecutive positions of intervals of J (resp. J'), keeping at least one position for each interval. This can be done in MSO, preserving the S -synchronisation, as depicted below:



Finally, we prove that the property saying that \mathbf{h}' is an asymptotic mix of \mathbf{g}' can be expressed in $\text{MSO} + \mathbf{U}_1$ (using the S -synchronisation of the encodings of the two sequences). Indeed, 1-extractions can be simulated in MSO according to the picture in Figure 4, still preserving the S -synchronisation. Consider then K, K' such that S, K encodes a 1-extraction of \mathbf{h}' and S, K' encodes a 1-extraction of \mathbf{g}' , *i.e.*, such that between every two consecutive positions of S , the elements of K (*resp.* K') form an interval. It remains to prove that we can express in $\text{MSO} + \mathbf{U}_1$ that $\mathbf{f}_{S,K}$ and $\mathbf{f}_{S,K'}$, viewed as number sequences, are asymptotically equivalent. We select a subset of indices of the two number sequences $\mathbf{f}_{S,K}$ and $\mathbf{f}_{S,K'}$ by selecting elements of S (and keeping the maximal intervals in K and K' that directly follow the selected elements), and we check thanks to \mathbf{U}_1 that the corresponding sequences of numbers are both bounded or both unbounded (checking that the complements of K (*resp.* K') satisfy \mathbf{U}_1 or not). This process is depicted here:



4 How $\text{MSO} + \mathbf{U}$ may talk about ultimate periodicity

We have already observed that the predicate \mathbf{U}_1 can be expressed in MSO augmented with the quantifier P . It is not clear that the converse is true, and we leave this as an open problem. However, in Theorem 12 below, we show that, up to a certain encoding, all languages expressible in $\text{MSO} + P$ can be expressed in $\text{MSO} + \mathbf{U}$.

Let Σ be a finite alphabet, $\#$ be a symbol not belonging to Σ , and $\Sigma_\#$ denote $\Sigma \cup \{\#\}$. We define $\pi_\Sigma : \Sigma_\#^\omega \rightarrow \Sigma^\omega$ to be the function which erases all appearances of $\#$. This is a partial function, because it is only defined on ω -words over $\Sigma_\#$ that contain infinitely many letters from Σ . We extend π_Σ to languages of ω -words in a natural way.

► **Theorem 12.** *Every language definable in $\text{MSO} + P$ over Σ is equal to $\pi_\Sigma(L)$ for some language $L \subseteq \Sigma_\#^\omega$ definable in $\text{MSO} + \mathbf{U}$.*

The main difficulty in expressing “ultimate periodicity” is to check the existence of a constant, namely the period, which is ultimately repeated. Such a property was already the crucial point in the proof of the undecidability of $\text{MSO} + \mathbf{U}$ [7]. Indeed the authors managed to express that an encoded vector sequence (as in Figure 2) tending towards infinity has “ultimately constant dimension”, *i.e.*, all but finitely many vectors of the sequence have the same dimension (this dimension will represent the period). This is the content of the following lemma which is a direct consequence of [7, Lemma 3.1].

► **Lemma 13.** *There exists a formula $\varphi(R, I)$ in $\text{MSO} + \mathbf{U}$ with two free set variables which holds if and only if $\mathbf{f}_{R,I}$ is defined, tends towards infinity, and $\dim(\mathbf{f}_{R,I})$ is ultimately constant.*

In order to use Lemma 13, we will encode a word w over Σ by adding factors of consecutive $\#$ between every two letters from Σ , such that the lengths of those factors tend towards infinity. Let $w = w_1 w_2 \dots$ be an ω -word over Σ where the w_i 's are letters, and let $f = n_1, n_2, \dots$ be a number sequence. We define the ω -word w_f over $\Sigma_\#$ as $w_1 \#^{n_1} w_2 \#^{n_2} \dots$. In particular

XX:12 Undecidability of MSO+“ultimately periodic”

$\pi_\Sigma(w_f) = w$. Now, given a sentence φ of $\text{MSO}+\text{P}$ over Σ , we consider the language L over $\Sigma_\#$ of all the ω -words of the form w_f for w satisfying φ and f a number sequence tending towards infinity. As we can easily check in $\text{MSO}+\text{U}$ that the sequence of lengths of the factors of $\#$ tends towards infinity, Theorem 12 follows immediately from the following lemma.

► **Lemma 14.** *For every sentence φ in $\text{MSO}+\text{P}$, one can effectively construct a formula $\varphi_\#$ in $\text{MSO}+\text{U}$ such that for every ω -word $w \in \Sigma^\omega$, the following conditions are equivalent:*

1. φ is true in w ;
2. $\varphi_\#$ is true in w_f for some f which tends towards infinity;
3. $\varphi_\#$ is true in w_f for every f which tends towards infinity.

Proof sketch. We proceed by induction on the structure of the formula φ , using the natural mapping from positions of w into positions of w_f to handle free variables. In particular, every free set variable used in the induction hypothesis is supposed to contain only positions not labeled by $\#$. The key point of the induction is the case of a formula $\text{PX}\varphi(X)$, all other formula being translated straightforwardly. We translate it to a formula of the form: for all sets Y which contain only positions not labeled by $\#$ and which are “ultimately periodic when ignoring $\#$ ’s”, the translation obtained for φ by induction must be true for Y .

It remains to express the property “ Y is ultimately periodic when ignoring $\#$ ’s” as a formula of $\text{MSO}+\text{U}$ with one free set variable Y which contains only positions not labeled by $\#$. It can be done in the following way: “ Y is ultimately periodic when ignoring $\#$ ’s” if and only if there exist two sets R and I such that:

- I is exactly the set of all the $\#$, R is infinite and disjoint from I (hence $\mathbf{f}_{R,I}$ is defined);
- $\dim(\mathbf{f}_{R,I})$ is ultimately constant (this constant dimension will represent the period of Y);
- for every infinite set R' disjoint from I , alternating with R (*i.e.*, such that there is exactly one element of R' between two consecutive elements of R) and such that $\dim(\mathbf{f}_{R',I})$ is ultimately constant, we have that all elements of R' are either ultimately in Y or ultimately not in Y .

These properties can be expressed in $\text{MSO}+\text{U}$ thanks to Lemma 13, since the vector sequence $\mathbf{f}_{R,I}$, tends towards infinity. ◀

5 Conclusion

We have considered the extensions of MSO with the following features (two quantifiers and two second-order predicates):

$\text{PX}\varphi(X)$ the formula $\varphi(X)$ is true for all ultimately periodic sets X .

$\text{UX}\varphi(X)$ the formula $\varphi(X)$ is true for sets X of arbitrarily large finite size.

$\text{U}_1(X)$ for all $k \in \mathbb{N}$, there exist two consecutive positions of X at distance at least k .

$\text{U}_2(R, I)$ the sequence of numbers encoded by R, I (Figure 1) is defined and unbounded.

We have proved that $\text{MSO}+\text{U}$, $\text{MSO}+\text{U}_1$ and $\text{MSO}+\text{U}_2$ have the same expressive power, and the translations between these logics are effective (Theorem 4, Lemmas 6 and 7). Furthermore, all three have undecidable satisfiability on ω -words by [7]. Moreover we have seen that U_1 is expressible in $\text{MSO}+\text{P}$, while the converse is also true up to some encoding (Theorem 12). As a consequence, MSO extended with the quantifier P has undecidable satisfiability (Theorem 2).

We believe that $\text{MSO}+\text{U}_1$ can be reduced to many extensions of MSO , in a simpler way than reducing $\text{MSO}+\text{U}$. It is for instance the case for MSO augmented with the quantifier that ranges over periodic sets (and not necessarily ultimately periodic sets).

References

- 1 Vince Bárány, Lukasz Kaiser, and Alexander Rabinovich. Expressing cardinality quantifiers in monadic second-order logic over chains. *Journal of Symbolic Logic*, 76(2):603–619, 2011.
- 2 Alexis Bès. *A Survey of Arithmetical Definability*. Société Mathématique de Belgique, 2002.
- 3 Achim Blumensath, Olivier Carton, and Thomas Colcombet. Asymptotic monadic second-order logic. In *Mathematical Foundations of Computer Science 2014 - 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part I*, pages 87–98, 2014.
- 4 Mikołaj Bojańczyk. U. *ACM SIGLOG News*, 2(4):2–15, 2015.
- 5 Mikołaj Bojańczyk. A bounding quantifier. In *Computer Science Logic, 18th International Workshop, CSL 2004, 13th Annual Conference of the EACSL, Karpacz, Poland, September 20-24, 2004, Proceedings*, pages 41–55, 2004.
- 6 Mikołaj Bojańczyk and Thomas Colcombet. Bounds in ω -regularity. In *21th IEEE Symposium on Logic in Computer Science (LICS 2006), 12-15 August 2006, Seattle, WA, USA, Proceedings*, pages 285–296, 2006.
- 7 Mikołaj Bojańczyk, Paweł Parys, and Szymon Toruńczyk. The MSO+U Theory of (\mathbb{N}, \leq) Is Undecidable. *33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016, February 17-20, 2016, Orléans, France*, pages 21:1–21:8, 2016.
- 8 J. R. Büchi. On a decision method in restricted second order arithmetic. In *Proceedings of the International Congress on Logic, Method, and Philosophy of Science*, pages 1–12, Stanford, CA, USA, 1962. Stanford University Press.
- 9 Dietrich Kuske, Jiamou Liu, and Anastasia Moskvina. Infinite and Bi-infinite Words with Decidable Monadic Theories. In Stephan Kreutzer, editor, *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*, volume 41 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 472–486, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- 10 Michael Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.