The MSO+U Theory of $(\mathbb{N}, <)$ Is Undecidable

Mikołaj Bojańczyk, Paweł Parys*, and Szymon Toruńczyk*

University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland
`{bojan,parys,szymtor}@mimuw.edu.pl`

───── **Abstract** ─────

We consider the logic MSO+U, which is monadic second-order logic extended with the unbounding quantifier. The unbounding quantifier is used to say that a property of finite sets holds for sets of arbitrarily large size. We prove that the logic is undecidable on infinite words, i.e. the MSO+U theory of $(\mathbb{N}, \leq)$ is undecidable. This settles an open problem about the logic, and improves a previous undecidability result, which used infinite trees and additional axioms from set theory.

1 Introduction

A celebrated result of Büchi is that the monadic second-order (MSO) theory is decidable for the structure of natural numbers with order

$$(\mathbb{N}, \leq).$$

Stated differently, satisfiability of MSO is decidable over infinite words. This paper shows that the decidability fails after MSO is extended with the unbounding quantifier. The unbounding quantifier, denoted by

$$\mathsf{U}X.\ \varphi(X),$$

binds a set variable X and says that $\varphi(X)$ holds for arbitrarily large finite sets X. As usual with quantifiers, the formula $\varphi(X)$ might have other free variables beside of X. Denote by MSO+U the extension of MSO by this quantifier. The main contribution of the paper is the following theorem.

▶ **Theorem 1.** *The* MSO+U *theory of* $(\mathbb{N}, \leq)$ *is undecidable.*

A corollary of the main theorem is undecidability of the logic MSO+inf, which is a logic on profinite words defined in [**?**], because decidability of MSO+U reduces to decidability of MSO+inf. Another corollary is that the satisfiability on weighted infinite words is undecidable for the logic AMSO introduced in [**?**], again because of a reduction from decidability of MSO+U.

Background.

The logic MSO+U was introduced in [**?**], where it was shown that satisfiability is decidable for formulas on infinite trees where the U quantifier is only used once and not under the scope of set quantification. A significantly more powerful fragment of the logic, albeit for

---

infinite words, was shown decidable in [**?**] using automata with counters. These automata were further developed into the theory of cost functions initiated by Colcombet in [**?**]. The decidability result from [**?**] entails decidability of the star height problem.

The difficulty of MSO+U comes from the interaction between the unbounding quantifier and quantification over possibly infinite sets. This motivated the study of WMSO+U, which is the variant of MSO+U where set quantification is restricted to finite sets. On infinite words, satisfiability of WMSO+U is decidable, and the logic has an automaton model [**?**]. Similar results hold for infinite trees [**?**]. The results from [**?**] have been used to decide properties of CTL* [**?**]. Currently, the strongest decidability result in this line is about WMSO+U on infinite trees extended with quantification over infinite paths [**?**]. The latter result entails decidability of problems such as the realisability problem for prompt LTL [**?**], deciding the winner in cost parity games [**?**], or deciding certain properties of energy games [**?**].

While the above results showed that fragments MSO+U can be decidable, and can be used to prove results not directly related to the logic itself, it was not known whether the full logic was decidable. The first evidence that MSO+U can be too expressive was given in [**?**], where it was shown that MSO+U can define languages of infinite words that are arbitrarily high in the projective hierarchy from descriptive set theory. This result was used in [**?**], where it was shown that, modulo a certain assumption from set theory (namely V=L), the MSO+U theory of the complete binary tree is undecidable. The result from [**?**] implies that there can be no algorithm which decides MSO+U on the complete binary tree, and which has a correctness proof in the ZFC axioms of set theory. This paper strengthens the result from [**?**] in two ways: first, we use no additional assumptions from set theory, and second, we prove undecidability for words and not trees.

2 Vector sequences

It is clear that extending MSO by the ability to express precise equality of some quantities, like set sizes, immediately leads to undecidability. The idea behind our undecidability proof is to show that, under a certain encoding, MSO+U can express that two vector sequences have the same dimension. We begin by presenting some observations about vector sequences.

Define a *number sequence* to be an element of $\mathbb{N}^\omega$, and define a *vector sequence* to be an element of $(\mathbb{N}^*)^\omega$, i.e. an infinite sequence of vectors of natural numbers of possibly different dimensions. We write $\mathbf{f}, \mathbf{g}$ for vector sequences and f, g for number sequences. If f is a number sequence and $\mathbf{f}$ is a vector sequence, then we write $f \in \mathbf{f}$ if for every position i, the i-th number in the sequence f appears in one of the coordinates of the i-th vector in the vector sequence $\mathbf{f}$. For example, the relationship $f \in \mathbf{f}$ is satisfied by

$$f = 0, 0, 0, \ldots \qquad \mathbf{f} = (0), (1, 0), (2, 1, 0), (3, 2, 1, 0), \ldots$$

Number sequences are called *asymptotically equivalent* if they are bounded on the same sets of positions. For example, the sequence of squares is asymptotically equivalent to every number sequence with infinite lim inf. A vector sequence $\mathbf{f}$ is called an *asymptotic mix* of a vector sequence $\mathbf{g}$ if every $f \in \mathbf{f}$ is asymptotically equivalent to some $g \in \mathbf{g}$. A vector sequence of dimension d is one where all vectors have dimension d.

▶ **Lemma 2.** *Let $d \in \mathbb{N}$. There exists a vector sequence of dimension d which is not an asymptotic mix of any vector sequence of dimension $d - 1$.*

Proof

In the definitions above, a *sequence* is a family indexed by natural numbers – formally, a

function from the indexing set $\mathbb{N}$ to some universe. Since the definition of asymptotic mix does not use the order structure of the indexing set, in the proof of this lemma we allow families to be indexed by other countable sets, namely by vectors of natural numbers. All notions introduced above lift to the setting of families indexed by a fixed countable set. By induction on d, we will prove the following claim about vector families indexed by $\mathbb{N}^d$. We claim that the d-dimensional identity

$$\mathbf{id} : \mathbb{N}^d \to \mathbb{N}^d,$$

is not an asymptotic mix of any vector family

$$\mathbf{g} : \mathbb{N}^d \to \mathbb{N}^{d-1}.$$

The induction base of $d = 1$ is vacuous. Let us prove the claim for dimension d assuming that it has been proved for smaller dimensions.

Toward a contradiction, suppose that the d-dimensional identity is an asymptotic mix of some $\mathbf{g} : \mathbb{N}^d \to \mathbb{N}^{d-1}$. Consider the subset of arguments $\{0\} \times \mathbb{N}^{d-1}$. The first coordinate of the d-dimensional identity is bounded on this subset, namely it is zero, and therefore there must be some $g \in \mathbf{g}$ which is bounded on this set. By permuting the vectors in $\mathbf{g}$, without loss of generality, we assume that the first coordinate of $\mathbf{g}$ is bounded on arguments from $\{0\} \times \mathbb{N}^{d-1}$. Let

$$\mathbf{g}' : \mathbb{N}^d \to \mathbb{N}^{d-2}$$

be the vector family obtained from $\mathbf{g}$ by removing the first coordinate. Let

$$\pi_i : \mathbb{N}^d \to \mathbb{N} \qquad \text{with } i \in \{2, \dots, d\}$$

be the projection onto the i-th coordinate, which satisfies $\pi_i \in \mathbf{id}$. Therefore, each π_i must be asymptotically equivalent to some $g_i \in \mathbf{g}$. Let $X_i \subseteq \mathbb{N}^d$ be the set of arguments x where g_i agrees with the first coordinate of $\mathbf{g}$. In other words, when restricted to arguments outside X_i the projection π_i is asymptotically equivalent to some $g_i \in \mathbf{g}'$. Since the first coordinate of $\mathbf{g}$ is bounded on the set $\{0\} \times \mathbb{N}^{d-1}$, it follows that there is some $c_i \in \mathbb{N}$ such that X_i does not contain any arguments which have zero on the first coordinate and at least c_i on the i-th coordinate. Taking c to be the maximum of all $c_2, \dots, c_d$, we see that none of the sets $X_2, \dots, X_d$ intersects the set

$$X = \{(0, n_2, \dots, n_d) : n_2, \dots, n_d \geq c\}.$$

It is easy to observe that the vector family

$$(0, n_2, \dots, n_d) \in X \qquad \mapsto \qquad (n_2, \dots, n_d) \tag{1}$$

is an asymptotic mix of $\mathbf{g}'$ (restricted to X), which is a vector family of dimension $d - 2$. This contradicts the induction assumption, because the vector family in (1) is the $(d-1)$-dimensional identity, up to reindexing. $\square$

A vector sequence is said to have *bounded dimension* if there is some d such that all vectors in the sequence have dimension at most d. A vector sequence is said to *tend to infinity* if for every n, all but finitely many vectors in the sequence have all entries at least n. We order vector sequences coordinatewise in the following way: we write $\mathbf{f} \leq \mathbf{g}$ if for every i, the i-th vectors in both sequences have the same dimension, and the i-th vector of $\mathbf{f}$ is coordinatewise smaller or equal to the i-th vector of $\mathbf{g}$. A corollary of the above lemma is the following lemma, which characterises dimensions in terms only of boundedness properties.

▶ **Lemma 3.** *Let* $\mathbf{f}_1, \mathbf{f}_2$ *be vector sequences of bounded dimensions which tend to infinity. Then the following conditions are equivalent:*

1. *on infinitely many positions* $\mathbf{f}_1$ *has a vector of higher dimension than* $\mathbf{f}_2$;
2. *there exists some* $\mathbf{g}_1 \leq \mathbf{f}_1$ *which is not an asymptotic mix of any* $\mathbf{g}_2 \leq \mathbf{f}_2$.

Proof

Say that two vector sequences are asymptotically equivalent if they have the same dimension d, and for each coordinate $i \in \{1, \ldots, d\}$ the corresponding number sequences are asymptotically equivalent. Vector sequences that tend to infinity are maximal with respect to asymptotical equivalence in the following sense: if a vector sequence $\mathbf{f}$ of fixed dimension d tends to infinity, then for every vector sequence $\mathbf{h}$ of the same dimension there exists an asymptotically equivalent vector sequence $\mathbf{g} \leq \mathbf{f}$ (to obtain such $\mathbf{g}$, on each coordinate of each position we can take the minimum of the two numbers appearing in this place in $\mathbf{f}$ and $\mathbf{h}$). A corollary of this observation is that if $\mathbf{f}_2$ is a vector sequence of bounded dimension which tends to infinity, then every vector sequence at each (or at each except finitely many) position having dimension smaller or equal to the dimension of $\mathbf{f}_2$ is an asymptotic mix of some $\mathbf{g}_2 \leq \mathbf{f}_2$. This corollary gives the implication 2→1 in the lemma.

For the implication 1→2, we use Lemma 2. Let d_1 be such that on an infinite set $X \subseteq \mathbb{N}$ of positions $\mathbf{f}_1$ has dimension d_1 and $\mathbf{f}_2$ has a smaller dimension. By Lemma 2, there is a vector sequence

$$\mathbf{h} : X \to \mathbb{N}^{d_1}$$

of dimension d_1 which is not an asymptotic mix of any vector sequence of smaller dimension. As we have observed, $\mathbf{h}$ is asymptotically equivalent to some $\mathbf{g}_1 \leq \mathbf{f}_1$ (when restricted to positions from X), because $\mathbf{f}_1$ tends to infinity on all coordinates. Therefore, $\mathbf{g}_1$ is not an asymptotic mix of any $\mathbf{g}_2 \leq \mathbf{f}_2$ on X, since such a vector sequence $\mathbf{g}_2$ has strictly smaller dimension. We can arbitrarily extend $\mathbf{g}_1$ to all positions outside of X, and still it will not be an asymptotic mix of any $\mathbf{g}_2 \leq \mathbf{f}_2$. $\square$

3   Encoding a Minsky machine

We now use the results on vector sequences from the previous section to prove undecidability of MSO+U. To do this, it will be convenient to view an infinite word as a sequence of finite trees of bounded depth, in the following sense. Consider a word

$$w \in \{1, 2, 3, \ldots, n\}^\omega$$

which has infinitely many 1's. We view such a word as an infinite sequence of trees of depth (at most) n, denoted by $\mathsf{tree}(w)$, as described in Figure 1.

The key to the undecidability proof is the following lemma, which says that, in a certain asymptotic sense, degrees can be compared for equality. Here the degree of a tree node is defined to be the number of its children.

▶ **Lemma 4.** *There is an* MSO+U *formula, which defines the set of words*

$$w \in \{1, 2, 3\}^\omega$$

which have infinitely many 1's and such that $\mathsf{tree}(w)$ *has the following properties:*

(a) the degree of depth-2 nodes tends to infinity;
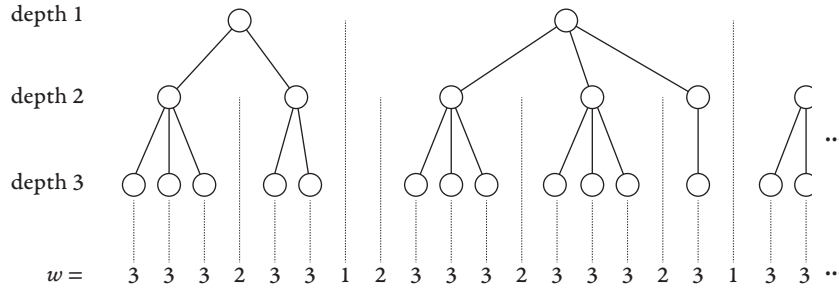
Figure 1 An example of $\mathsf{tree}(w)$ for $n = 3$. Formally speaking, the leaves of $\mathsf{tree}(w)$ are positions with label n, while the tree structure is defined by the following rule. For $1 \leq i < n$, two leaves which correspond to positions x and y with label n have a common ancestor at depth i if and only if there is no position between x and y which has label in $\{1, \ldots, i\}$. In particular, if between x and y there is a position with label 1, then x and y are in different trees of the sequence. Note that the mapping $w \mapsto \mathsf{tree}(w)$ is not one-to-one, e.g. in the picture, the first 2 just after the first 1 could be removed from w without affecting $\mathsf{tree}(w)$.

(b) all but finitely many nodes of depth 1 have the same degree.

Proof
Condition (a) is easily seen to be expressible in MSO+U. One says that for every infinite set of depth-2 nodes, their degrees are unbounded.

Let us focus on condition (b). Fix a word w with infinitely many 1's as in the statement of the lemma. For an infinite set X of depth-1 nodes, define

$$\mathbf{f}_X : \mathbb{N} \to \mathbb{N}^*$$

to be the vector sequence, where the i-th vector is the sequence of degrees of the children of the i-th node from X. Condition (a) says that if X is the set of all depth-1 nodes, then $\mathbf{f}_X$ tends to infinity, which implies that $\mathbf{f}_X$ also tends to infinity for any other infinite set X of depth-1 nodes.

Call two sets X, Y of depth-1 nodes *alternating* if every two nodes in X are separated by a node in Y, and vice versa. Condition (b) is equivalent to saying that
- depth-1 nodes have bounded degree;
- one cannot find infinite alternating sets X, Y of depth-1 nodes such that infinitely often $\mathbf{f}_X$ has strictly bigger dimension than $\mathbf{f}_Y$.

The first condition is clearly expressible in MSO+U. The second condition, thanks to Lemma 3, can be restated as: one cannot find infinite alternating sets X, Y of depth-1 nodes such that there is some $\mathbf{g}_X \leq \mathbf{f}_X$ which is not an asymptotic mix of any $\mathbf{g}_Y \leq \mathbf{f}_Y$. This is expressible in MSO+U (the quantification over vector sequences $\mathbf{g}_X \leq \mathbf{f}_X$ amounts to selecting a subset of depth-3 nodes that are descendants of nodes in X). $\square$

Minsky machines.

To prove undecidability, we reduce emptiness for Minsky machines to deciding MSO+U. By a Minsky machine we mean a (possibly nondeterministic) device which has a finite state space, and two counters that can be incremented, decremented, and tested for zero. It is undecidable whether a given Minsky machine has an accepting run, i.e. one which begins in a designated initial state with zero on both counters, and ends in a designated final state.

Let ρ be a finite run of a Minsky machine of length d. We say that a vector of natural numbers $(n_1, \ldots, n_{2d})$ *describes* the run ρ if, for $i = 1, \ldots, d$, the numbers n_{2i-1}, n_{2i} store the value of the two counters in the i-th configuration of ρ. Note that this description does not specify fully the run ρ, as the state information is missing. The following lemma contains the reduction of Minsky machine emptiness to satisfiability of MSO+U.
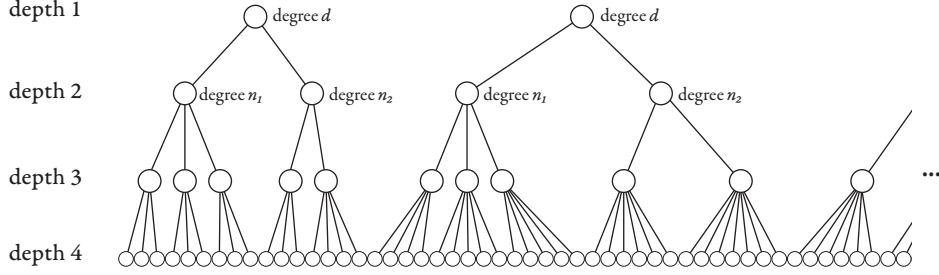
Figure 2 A sequence of trees as in Lemma 5. Here $d = 2$, $n_1 = 3$, and $n_2 = 2$.

▶ **Lemma 5.** *For every Minsky machine, one can compute a formula of* MSO+U *which defines the set of words*

$$w \in \{1, 2, 3, 4\}^\omega$$

which have infinitely many 1's and such that tree(w) *has the following properties, which are illustrated in Figure 2:*

(a) the degree of depth-3 nodes tends to infinity;

(b) all but finitely many depth-1 nodes have the same degree d;

(c) for every $i \in \{1, \ldots, d\}$, all but finitely many depth-2 nodes that are an i-th child have the same degree, call it n_i;

(d) $n_1 - 1, \ldots, n_d - 1$ describe some accepting run of the Minsky machine.

Proof

Condition (a) is clearly expressible in MSO+U.

We say that a sequence of trees of depth 3 is *well-formed* if the degree of depth-2 nodes tends to infinity, and that it has *almost constant degree* if all but finitely many depth-1 nodes have the same degree. Lemma 4 says that MSO+U can express the conjunction of being well-formed and having almost constant degree. We will use this property to define conditions (b), (c) and (d).

Define the *flattening* of tree(w) to be the sequence of depth-3 trees obtained from tree(w) by removing all depth-3 nodes and connecting all depth-4 nodes directly to their depth-2 grandparents. By condition (a), the flattening is well-formed. Since the flattening does not change the degree of depth-1 nodes, condition (b) is the same as saying that the flattening has almost constant degree, and therefore can be expressed in MSO+U thanks to Lemma 4.

Define a *depth-2 selector with offset i* to be a set of nodes X in the tree tree(w) which selects exactly one child for every depth-1 node (and therefore X contains only depth-2 nodes), and all but finitely many nodes in X are an i-th child. A *depth-2 selector*, without i being mentioned, is a depth-2 selector for some i. Being a depth-2 selector is equivalent to saying that one gets a well-formed sequence of almost constant degree if one keeps only nodes from $X_\leftarrow$ and their descendants, where $X_\leftarrow$ is the set of nodes of depth 2 that have

a sibling from X to the right. Therefore, being a depth-2 selector is definable in MSO+U. Condition (c) is the same as saying that for every depth-2 selector X, if one only keeps the nodes from X and their descendants, then the resulting sequence has almost constant degree, which can be expressed in MSO+U thanks to Lemma 4.

We are left with condition (d) about Minsky machines. We say that a depth-2 selector X *represents zero*, if all but finitely many nodes in X have degree one (recall that condition (d) uses $n_i - 1$ to represent a counter value, because a depth-2 node cannot have degree zero). Representing zero is definable in first-order logic. If X, Y are selectors, we say that Y *increments* X if there is some n such that all but finitely many nodes in X have degree n, and all but finitely many nodes in Y have degree $n + 1$. This is equivalent to saying that if one keeps only nodes from $X \cup Y$ and their descendants, and then removes one subtree of every node from Y, then the resulting sequence of depth-3 trees has almost constant degree. Therefore incrementation is definable in MSO+U (technically, one needs to translate the above tree properties to word properties, via the encoding from Figure 1). Using formulas for representing zero and incrementation, it is easy to formalise condition (d) in MSO+U (the formula first guesses the missing state information to fully specify the run ρ, and then verifies its consistency with the Minsky machine). □

In particular, the formula computed in Lemma 5 is satisfiable if and only if the Minsky machine has an accepting run. This yields undecidability of MSO+U on infinite words, which is the same as our main Theorem 1.

Quantifier complexity

Here we examine the quantification structure of the formulas in the undecidability proof. We count the number of blocks of quantifiers of same type. We do not claim that the formulas are optimal.

The more interesting part of the formula in Lemma 4 says: for all sets X, Y and functions $\mathbf{g}_X$ there exists a function $\mathbf{g}_Y$ such that $\mathbf{g}_X$ is an asymptotic mix of $\mathbf{g}_Y$. The condition "$\mathbf{g}_X$ is an asymptotic mix of $\mathbf{g}_Y$" is expressed as: for every $g_X \in \mathbf{g}_X$ there exists $g_Y \in \mathbf{g}_Y$ such that for every set of positions Z either both g_X and g_Y are bounded on Z or none of them is. Thus the entire formula has six blocks of quantifiers, starting from universal quantifiers (where the most internal quantifiers are U and negations of U).

The formula from Lemma 5 says: there exists an infinite word (a labelling of $(\mathbb{N}, \leq)$) and a labelling by states of the Minsky machine, such that for every set X either X is not a depth-2 selector or the children selected by X satisfy appropriate conditions. Saying that X is not a depth-2 selector amounts to using the formula from Lemma 4 negatively, starting from an existential quantifiers. The rest of the condition about X says that all but finitely many nodes in X have the same degree, and that the degree of these nodes is smaller/greater by one than the degree of the left siblings of nodes in X, which is expressible by using the formula from Lemma 4 positively. Concluding, the whole formula uses eight nested blocks of quantifiers: seven blocks of alternating existential and universal quantifiers, ended by quantifiers U and negations of U.

─── **References** ───────────────────────────────

1   Achim Blumensath, Olivier Carton, and Thomas Colcombet. Asymptotic monadic second-
order logic. In Erzsébet Csuhaj-Varjú, Martin Dietzfelbinger, and Zoltán Ésik, editors,
Mathematical Foundations of Computer Science 2014 - 39th International Symposium,

MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part I, volume 8634 of *Lecture Notes in Computer Science*, pages 87–98. Springer, 2014.

2 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.

3 Mikołaj Bojańczyk. Weak MSO with the unbounding quantifier. *Theory Comput. Syst.*, 48(3):554–576, 2011.

4 Mikołaj Bojańczyk. Weak MSO+U with path quantifiers over infinite trees. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, pages 38–49, 2014.

5 Mikołaj Bojańczyk and Thomas Colcombet. Bounds in w-regularity. In *21th IEEE Symposium on Logic in Computer Science (LICS 2006), 12-15 August 2006, Seattle, WA, USA, Proceedings*, pages 285–296, 2006.

6 Mikołaj Bojańczyk, Tomasz Gogacz, Henryk Michalewski, and Michał Skrzypczak. On the decidability of MSO+U on infinite trees. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II*, pages 50–61, 2014.

7 Mikołaj Bojańczyk and Szymon Toruńczyk. Weak MSO+U over infinite trees. In *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, pages 648–660, 2012.

8 Tomás Brázdil, Krishnendu Chatterjee, Antonín Kucera, and Petr Novotný. Efficient controller synthesis for consumption games with multiple resource types. In *CAV*, pages 23–38, 2012.

9 Claudia Carapelle, Alexander Kartzow, and Markus Lohrey. Satisfiability of CTL* with constraints. In *CONCUR 2013 - Concurrency Theory - 24th International Conference, CONCUR 2013, Buenos Aires, Argentina, August 27-30, 2013. Proceedings*, pages 455–469, 2013.

10 Thomas Colcombet. The theory of stabilisation monoids and regular cost functions. In *Automata, Languages and Programming, 36th Internatilonal Collogquium, ICALP 2009, Rhodes, greece, July 5-12, 2009, Proceedings, Part II*, pages 139–150, 2009.

11 Nathanaël Fijalkow and Martin Zimmermann. Cost-parity and cost-Streett games. In *FSTTCS*, pages 124–135, 2012.

12 Szczepan Hummel and Michał Skrzypczak. The topological complexity of MSO+U and related automata models. *Fundam. Inform.*, 119(1):87–111, 2012.

13 Orna Kupferman, Nir Piterman, and Moshe Y. Vardi. From liveness to promptness. *Formal Methods in System Design*, 34(2):83–103, 2009.

14 Szymon Toruńczyk. Languages of profinite words and the limitedness problem. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part II*, pages 377–389, 2012.