

On the theory of order over real numbers

Bernard Boigelot^{1*}, Pascal Fontaine^{1*} and Baptiste Vergain^{1*}

^{1*}Montefiore Institute, B28, University of Liège, 4000 Liège, Belgium.

*Corresponding author(s). E-mail(s): Bernard.Boigelot@uliege.be;
Pascal.Fontaine@uliege.be; BVergain@uliege.be;

Abstract

We consider the first-order theory of order over real numbers extended with monadic uninterpreted predicates. It has long been established that this theory is decidable, by different approaches such as a reduction to linear-time temporal logic over the reals, expressing the problem as a particular case of deciding the first-order theory of all linear orderings, or reasoning about restrictions of the monadic second-order theory of order. The main contribution of this work is to provide a complete self-contained proof of decidability for this logic. Additionally, our proof develops a symbolic notation system for models of satisfiable formulas, which can be seen as a first step towards an actual implementation of a decision procedure for the theory of order over the reals.

Keywords: Theory of order, real numbers, first-order logic, satisfiability

MSC Classification: Primary 03B25; Secondary 03C64

1 Introduction

Logics related to the theory of order have been extensively studied. In [1], Büchi shows that the monadic second-order logic of one successor S1S is decidable, by introducing his eponymous automata. This logic is equivalent to the monadic second-order theory of order over natural numbers $\text{MSO}(\mathbb{N}, <)$, since the successor relation can be constructed from the order relation, and vice versa. This result is generalized to the monadic second-order theory of countable ordinals in [2], and of all ordinals in [3].

A proof that the monadic first-order theory of linear orderings $\text{MFO}(<)$ is decidable is sketched in [4], and another one is described in more detail in [5]. The latter proof relies on a technique called the *composition method*, originally introduced in [6, 7] and

then refined in [8, 9]. A survey of results related to the composition method and its application to the decision of the monadic theory of ordinals is given in [10].

It is established in [8] that the weak monadic second-order theory of linear orderings is decidable, where *weak* means that the second-order variables are restricted to finite sets. In [11], it is shown that the monadic second-order theory of countable linear orderings is decidable, which entails that $\text{MSO}(\mathbb{Q}, <)$ is decidable as well. Interestingly, this result does not generalize to real numbers, and a proof that $\text{MSO}(\mathbb{R}, <)$ is undecidable is developed in [9].

There are deep connections between the theory of order and temporal logic. Kamp’s theorem [12, 13] states that Linear-time Temporal Logic (LTL) with the modalities *Since* and *Until* has the same expressive power as the monadic first-order theory of order restricted to Dedekind-complete linear orderings. This result therefore applies to the domains $\mathbb{N}$ and $\mathbb{R}$, but not to $\mathbb{Q}$ that does not satisfy the completeness requirement. It is generalized to all linear orderings in [14], after enriching LTL with additional modalities *Since_{Stavi}* and *Until_{Stavi}*. The decidability of LTL over $\mathbb{R}$ is proved in [15]. By Kamp’s theorem, this implies the decidability of the monadic first-order theory of order over real numbers $\text{MFO}(\mathbb{R}, <)$. The result developed in [15] is actually stronger, and establishes the decidability of the universal fragment of the monadic second-order theory of order over real numbers, which entails the decidability of the existential fragment as well.

The complexity of deciding LTL is investigated in [16–18], where it is established in particular that this problem is PSPACE-complete for LTL with only the modality *Until* over all linear orderings, as well as for LTL with *Since* and *Until* over $\mathbb{R}$. A simpler proof of those results appears in [19]. Note that this does not lead to a PSPACE decision procedure for the monadic first-order theory of order over $\mathbb{R}$, since translating a formula of this logic into LTL may incur a non-elementary blowup [13].

The main contribution of this paper is to provide a complete self-contained proof of decidability for the monadic first-order theory of order over real numbers $\text{MFO}(\mathbb{R}, <)$. The approach is similar to [5], but particularized to the special case of the real domain. Our construction also describes explicitly how to manage the uninterpreted monadic predicates involved in the logic, which is only hinted at in [5]. It relies on elements of the composition method, that are mainly borrowed from [10]. A corollary of the decidability result is that satisfiable formulas admit models that can be constructed in a simple way, by applying finitely many times composition operators to elementary sets. As a secondary contribution, we develop a symbolic notation system for such models, which can be seen as a first step towards an actual implementation of a decision procedure for the theory of order over the reals.

Our long term goal is to develop practical techniques to handle quantifiers within Satisfiability Modulo Theory solvers (see e.g., [20, 21] for a general introduction to SMT), in a context mixing arithmetic and uninterpreted predicates. We are convinced that a deep understanding of decision procedures for limited (but decidable) fragments will be helpful in the future to design such practical techniques. For instance, in a recent paper [22], automata-based decision procedures prove to be of help to improve quantified linear integer arithmetic reasoning in SMT solvers.

This paper is organized as follows. In Section 2, we introduce the logic $\text{MFO}(\mathbb{R}, <)$ as well as the notion of *chain*, corresponding to an interpretation of a formula in that logic. In Section 3, we recall elements of model theory, in particular the definition of k -elementary equivalence relations characterizing the chains that cannot be distinguished by formulas of a given quantifier depth k . We also introduce the concept of k -type, which provides a precise representation for the k -elementary equivalence classes. Knowing the k -type of a chain makes it possible to decide whether this chain satisfies arbitrary formulas of quantifier depth k . In Section 4, we describe a mechanism for building sets of chains that we call *constructible*, by starting from elementary sets and introducing composition operators that can be recursively applied in order to obtain more complex sets. We introduce a symbolic notation for such sets of chains, and establish that for any k , the k -type of the chains that belong to a set described in this notation is unique and can effectively be computed. In Section 5, we show that every satisfiable formula of $\text{MFO}(\mathbb{R}, <)$ admits a constructible model that can be discovered by a finite and bounded enumeration, which leads to a decision procedure for this logic.

2 Preliminaries

2.1 Logical language

Formulas of the monadic first-order logic of order over the reals $\text{MFO}(\mathbb{R}, <)$ are constructed from the following elements:

- *First-order variables* $x_1, x_2, \dots$
- *Second-order variables* $X_1, X_2, \dots$
- *Atoms* of the form $\top$, $\perp$, $x_i = x_j$, $x_i < x_j$, and $x_i \in X_j$.
- *Boolean connectives* $\vee$, $\wedge$, and $\neg$.
- *First-order quantifiers* $\exists x_i$ and $\forall x_i$.

The semantics is defined in the usual way. The first-order variables take their value in a domain D (hence the first-order quantifiers range over D), and the second-order variables in 2^D . Our main interest is the domain $D = \mathbb{R}$, but we will show in Section 4.1 that choosing D to be equal to any non-empty open interval (a, b) of $\mathbb{R}$ does not affect satisfiability, i.e., a formula is satisfiable in $\text{MFO}(\mathbb{R}, <)$ iff it is satisfiable in $\text{MFO}((a, b), <)$. We will exploit this property in order to simplify some developments.

2.2 Basic definitions

For $n, m \geq 0$, a (n, m) -chain over the domain D is a tuple $(p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$, where $p_i \in D$ for each $i \in [1, n]$, and $P_i \subseteq D$ for each $i \in [1, m]$. The pair (n, m) is called the *dimension* of the chain.

A formula $\varphi(x_1, x_2, \dots, x_n, X_1, X_2, \dots, X_m)$ with the free first-order variables $x_1, \dots, x_n$ and the free second-order variables $X_1, \dots, X_m$ is *satisfied* by the (n, m) -chain $\mathcal{A} = (p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$, which is denoted by $\mathcal{A} \models \varphi$, if φ evaluates to true when each x_i takes the corresponding p_i as value, and each X_i takes the corresponding P_i as value. A formula φ is *satisfiable* if there exists an (n, m) -chain $\mathcal{A}$ of appropriate

dimension (i.e., matching the numbers of free first-order and second-order variables in φ) such that $\mathcal{A} \models \varphi$.

The *quantifier depth* of a formula φ is equal to its number of nested quantifiers. If φ is in prenex normal form, in other words if it is of the form

$$Q_1 x_{i_1} Q_2 x_{i_2} \cdots Q_k x_{i_k} \psi(x_1, x_2, \dots, x_n, X_1, X_2, \dots, X_m),$$

where ψ is quantifier-free and $Q_i \in \{\exists, \forall\}$ for all $i \in [1, k]$, then its quantifier depth is k .

3 Elements of model theory

The content of this section is mainly inspired by [10].

3.1 k -elementary equivalence

Two (n, m) -chains $\mathcal{A}$ and $\mathcal{B}$, with $n, m \geq 0$, are *k -elementarily equivalent* for some $k \geq 0$, which is denoted by $\mathcal{A} \equiv_k \mathcal{B}$, if for every formula $\varphi(x_1, x_2, \dots, x_n, X_1, X_2, \dots, X_m)$ of quantifier depth less than or equal to k , one has

$$\mathcal{A} \models \varphi \Leftrightarrow \mathcal{B} \models \varphi.$$

In other words, $\mathcal{A}$ and $\mathcal{B}$ cannot be distinguished by formulas with a quantifier depth at most equal to k .

Two (n, m) -chains $\mathcal{A}$ and $\mathcal{B}$ are *elementarily equivalent*, which is denoted by $\mathcal{A} \equiv \mathcal{B}$, if they are such that $\mathcal{A} \equiv_k \mathcal{B}$ for all $k \geq 0$.

3.2 k -types

A *k -type* for (n, m) -chains, with $k, n, m \geq 0$, is an equivalence class of the relation $\equiv_k$. In other words, two (n, m) -chains share the same k -type iff they cannot be distinguished by formulas with a quantifier depth that is at most equal to k .

A precise description of k -types can be constructed by induction on k . For the base case $k = 0$, we have to consider the equivalence classes of the relation $\equiv_0$. For given values of n and m , there are only finitely many distinct atoms over the first-order variables $x_1, x_2, \dots, x_n$ and the second-order variables $X_1, X_2, \dots, X_m$. Two (n, m) -chains $\mathcal{A}$ and $\mathcal{B}$ are such that $\mathcal{A} \equiv_0 \mathcal{B}$ iff the set of all atoms satisfied by $\mathcal{A}$ is exactly equal to the set of all atoms satisfied by $\mathcal{B}$. Indeed, in such a case, $\mathcal{A}$ and $\mathcal{B}$ cannot be distinguished by quantifier-free formulas, and otherwise one can easily build a quantifier-free formula satisfied by $\mathcal{A}$ and not by $\mathcal{B}$. More precisely, every 0-type τ for (n, m) -chains is described without ambiguity by its *characteristic formula* φ_τ , also known as *Hintikka formula* [10, 23], that is satisfied by a (n, m) -chain iff this chain belongs to τ . This formula takes the form

$$\varphi_\tau(x_1, \dots, x_n, X_1, \dots, X_m) = \bigwedge_{\psi \in \Lambda_\tau} \psi \wedge \bigwedge_{\psi \in \Lambda'_\tau} \neg \psi,$$

where Λ_τ and Λ'_τ are the sets of all atoms involving the first-order variables $x_1, x_2, \dots, x_n$ and the second-order variables $X_1, X_2, \dots, X_m$ that are respectively satisfied and not satisfied by the chains in τ . Note that the formula φ_τ is entirely characterized by the set of atoms Λ_τ ; this set thus provides a compact and unambiguous representation of the 0-type τ . Another important property is that, for given n and m , there are only finitely many 0-types for (n, m) -chains.

For the case $k > 0$, we have to characterize the equivalence classes of the relation $\equiv_k$. We show inductively that for every given n and m , each such equivalence class, i.e., each k -type τ for (n, m) -chains, is described without ambiguity by a Hintikka formula φ_τ , such that the (n, m) -chains that belong to τ are exactly those that satisfy φ_τ . Furthermore, for given k, n and m , there are finitely many such Hintikka formulas, thus there only exist finitely many k -types for (n, m) -chains.

Two (n, m) -chains $\mathcal{A}$ and $\mathcal{B}$ are such that $\mathcal{A} \equiv_k \mathcal{B}$ iff they cannot be distinguished by formulas of the form

$$Qx_{n+1} \varphi(x_1, x_2, \dots, x_{n+1}, X_1, X_2, \dots, X_m), \quad (1)$$

where $Q \in \{\exists, \forall\}$ and the quantifier depth of φ is $k - 1$. By inductive hypothesis, all $(n + 1, m)$ -chains $(p_1, p_2, \dots, p_{n+1}, P_1, P_2, \dots, P_m)$ are members of finitely many $(k - 1)$ -types. Let $T = \{\tau_1, \tau_2, \dots, \tau_\ell\}$ be the set of these $(k - 1)$ -types, and for every $i \in [1, \ell]$, let φ_i be the Hintikka formula describing τ_i .

For a (n, m) -chain $\mathcal{A} = (p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$, each $(n + 1, m)$ -chain $(p_1, p_2, \dots, p_{n+1}, P_1, P_2, \dots, P_m)$ that can be obtained by extending $\mathcal{A}$ with an additional first-order value $p_{n+1} \in D$ satisfies some Hintikka formula φ_i , with $i \in [1, \ell]$. (Note that we arbitrarily place the additional value p_{n+1} after the existing ones.) Let τ be the k -type of $\mathcal{A}$, and I_τ be the set of all $i \in [1, \ell]$ such that $(p_1, p_2, \dots, p_{n+1}, P_1, P_2, \dots, P_m)$ satisfies φ_i for some $p_{n+1} \in D$. The (n, m) -chains belonging to τ are then exactly those that satisfy the Hintikka formula

$$\begin{aligned} \varphi_\tau(x_1, \dots, x_n, X_1, \dots, X_m) = & \bigwedge_{i \in I_\tau} \exists x_{n+1} \varphi_i(x_1, \dots, x_{n+1}, X_1, \dots, X_m) \\ & \wedge \forall x_{n+1} \bigvee_{i \in I_\tau} \varphi_i(x_1, \dots, x_{n+1}, X_1, \dots, X_m), \end{aligned}$$

in which the first term expresses that each φ_i with $i \in I_\tau$ is satisfied for at least one value of p_{n+1} , and the second term imposes that for every value of p_{n+1} , there is some φ_i with $i \in I_\tau$ that is satisfied.

The Hintikka formula φ_τ provides a unambiguous representation of the k -type τ , since it admits the quantifier depth k and it is satisfied by a (n, m) -chain iff this chain belongs to τ . Remark that the formula φ_τ is entirely characterized by the set I_τ , or equivalently by the set of $(k - 1)$ -types T . This set T thus provides a compact and unambiguous representation of the k -type τ . In other words, a k -type, for $k > 0$, can uniquely be represented as a set of $(k - 1)$ -types, of which there exist only finitely many instances. We have just established the following theorem.

Theorem 1. *For given values of $k, n, m \geq 0$, there exist finitely many k -types for (n, m) -chains.*

From now on, we will assimilate k -types with their descriptions, i.e., we will consider that a k -type corresponds indifferently to a set of chains belonging to some equivalence class of $\equiv_k$, to a Hintikka formula representing this class, or to a set of either atoms (if $k = 0$), or $(k - 1)$ -types (if $k > 0$). In most technical developments, we will rely on the representation of types as sets.

Note that a k -type is also a $(k - 1)$ -type, and the description of this $(k - 1)$ -type as a set can easily be computed recursively. Indeed, for $k = 1$, the 0-type corresponding to a given 1-type τ is simply the same 0-type obtained by suppressing in any 0-type belonging to τ all atoms referring to the last first-order variable. For $k > 1$, the $(k - 1)$ -type for a given k -type τ is the set of $(k - 2)$ -types corresponding to the $(k - 1)$ -types that belong to τ .

3.3 Deciding formulas with types

Let $\mathcal{A}$ be a (n, m) -chain, with $n, m \geq 0$, and φ be a formula of $\text{MFO}(D, <)$ with n first-order and m second-order free variables. Let k be the quantifier depth of φ . From the k -type of $\mathcal{A}$, one can decide whether $\mathcal{A} \models \varphi$ holds or not.

Indeed, for the case $k = 0$, the 0-type of $\mathcal{A}$ corresponds to the set of all the atoms involving n first-order and m second-order variables that are satisfied by $\mathcal{A}$. Deciding the quantifier-free formula φ then amounts to evaluating the truth value of this formula when each such atom is replaced by $\top$, and all the others by $\perp$.

If $k > 0$, then the k -type τ of $\mathcal{A}$ corresponds to the set of $(k - 1)$ -types of the $(n + 1, m)$ -chains obtained by extending $\mathcal{A}$ with one additional first-order value $p_{n+1} \in D$, for all possible choices of p_{n+1} . If φ is of the form

$$\exists x_{n+1} \psi(x_1, x_2, \dots, x_{n+1}, X_1, X_2, \dots, X_m)$$

where the quantifier depth of ψ is $k - 1$, then deciding whether $\mathcal{A} \models \varphi$ holds reduces to checking whether ψ is satisfied by the chains that belong to at least one of the $(k - 1)$ -types that belong to τ . (Recall that all the chains in such a $(k - 1)$ -type satisfy the same formulas of quantifier depth $k - 1$.) There are finitely many such types, hence this decision procedure can be carried out by induction on k . Similarly, if φ is of the form

$$\forall x_{n+1} \psi(x_1, x_2, \dots, x_{n+1}, X_1, X_2, \dots, X_m)$$

then deciding $\mathcal{A} \models \varphi$ amounts to checking that ψ is satisfied by the chains that belong to every $(k - 1)$ -type that belongs to τ .

4 Constructible chains

We now discuss how to construct a particular class of chains by starting from basic building blocks and combining them with the help of composition operators. Before we introduce those elements and operators, we first show that the domain over which a chain is considered does not impact its k -type.

4.1 Domain equivalence

Let $D_1 = (a_1, b_1)$ and $D_2 = (a_2, b_2)$ be two domains, with $a_1, a_2 \in \mathbb{R} \cup \{-\infty\}$, $b_1, b_2 \in \mathbb{R} \cup \{+\infty\}$, $a_1 < b_1$ and $a_2 < b_2$. We denote by $\mathcal{A}_i \models_i \varphi$ the fact that a chain $\mathcal{A}_i$ over D_i satisfies a formula φ in $\text{MFO}(D_i, <)$.

Let f be a strictly increasing continuous total function from D_1 to D_2 , i.e., such that $f(D_1) = D_2$, and $f(x) < f(y)$ for every $x, y \in D_1$ such that $x < y$. This function transforms any (n, m) -chain $\mathcal{A}_1 = (p_{1,1}, p_{1,2}, \dots, p_{1,n}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ over D_1 into a (n, m) -chain $\mathcal{A}_2 = (p_{2,1}, p_{2,2}, \dots, p_{2,n}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$ over D_2 , where $p_{2,i} = f(p_{1,i})$ for every $i \in [1, n]$, and $P_{2,i} = \{f(p) \mid p \in P_{1,i}\}$ for every $i \in [1, m]$.

Theorem 2. *For any such f , $\mathcal{A}_1$, $\mathcal{A}_2$ and formula φ , we have*

$$\mathcal{A}_1 \models_1 \varphi \Leftrightarrow \mathcal{A}_2 \models_2 \varphi.$$

In other words, $\mathcal{A}_1$ and $\mathcal{A}_2$ share the same k -type for every $k \geq 0$.

Proof. Let k be the quantifier depth of φ . If $k = 0$, then $\mathcal{A}_1$ and $\mathcal{A}_2$ satisfy exactly the same subset of atoms, hence they share the same 0-type, and $\mathcal{A}_1 \models_1 \varphi \Leftrightarrow \mathcal{A}_2 \models_2 \varphi$. If $k > 0$, then one establishes by induction on k that the k -types of $\mathcal{A}_1$ and $\mathcal{A}_2$ are identical by construction, hence $\mathcal{A}_1 \models_1 \varphi \Leftrightarrow \mathcal{A}_2 \models_2 \varphi$ as well. $\square$

The following corollary shows that the choice of domain does not influence the satisfiability of formulas in $\text{MFO}(D, <)$, where D is any non-empty open interval of $\mathbb{R}$.

Corollary 3. *A formula φ is satisfiable in $\text{MFO}(D_1, <)$ iff it is satisfiable in $\text{MFO}(D_2, <)$.*

Proof. For every pair of domains D_1 and D_2 , there exists a strictly increasing continuous total function f from D_1 to D_2 . The result then follows from Theorem 2. $\square$

4.2 Elementary chains

We now describe the basic building blocks that will be used for constructing chains, and introduce a symbolic notation for them. The idea is that each instance of this notation describes a set of chains that share the same k -type (in other words, that are k -elementarily equivalent) for every $k \geq 0$. As explained previously, the domain D over which chains are defined can be chosen equal to any non-empty open interval of $\mathbb{R}$.

4.2.1 Constant chains

Let $m \geq 0$. The notation $\llbracket U \rrbracket$, with $U \subseteq \{1, 2, \dots, m\}$, describes a set containing a single $(0, m)$ -chain $(P_1, P_2, \dots, P_m)$ such that for every $i \in [1, m]$, $P_i = D$ if $i \in U$ and $P_i = \emptyset$ if $i \notin U$. In other words, the values i in U characterize the predicates P_i that are true over the whole domain D , while the other predicates are false everywhere.

4.2.2 Shuffle chains

Let $m \geq 0$ and $\ell > 0$. The notation $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$, with $U_i \subseteq \{1, 2, \dots, m\}$ for every $i \in [1, \ell]$, describes the set of all the $(0, m)$ -chains $(P_1, P_2, \dots, P_m)$ that satisfy the two following properties:

- For each $p \in D$, there exists $i \in [1, \ell]$ such that $U_i = \{j \in [1, m] \mid p \in P_j\}$. This means that the components U_i represent all the combinations of predicate values that occur over the elements of D .
- For all $i \in [1, \ell]$ and $p_1, p_2 \in D$ such that $p_1 < p_2$, there exists p such that $p_1 < p < p_2$ and $U_i = \{j \in [1, m] \mid p \in P_j\}$. In other words, for every bounded open interval in D and component U_i of the expression, there must exist a point p in this interval at which the combination of predicate values is represented by U_i .

Note that in the particular case where $\ell = 1$, this definition corresponds to the constant chains introduced in Section 4.2.1. This is the reason why the two notations coincide in this case.

Theorem 4. *Let $k \geq 0$. All the chains contained in the set $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$ share the same k -type.*

Proof. We actually prove a stronger property: Consider two (n, m) -chains $\mathcal{A}_1 = (p_{1,1}, p_{1,2}, \dots, p_{1,n}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ and $\mathcal{A}_2 = (p_{2,1}, p_{2,2}, \dots, p_{2,n}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$, with $n \geq 0$, such that

- $(P_{1,1}, P_{1,2}, \dots, P_{1,m}) \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$,
- $(P_{2,1}, P_{2,2}, \dots, P_{2,m}) \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$,
- $p_{1,i} \in P_{1,j} \Leftrightarrow p_{2,i} \in P_{2,j}$ for all $i \in [1, n]$ and $j \in [1, m]$, and
- $p_{1,i} \leq p_{1,j} \Leftrightarrow p_{2,i} \leq p_{2,j}$ for all $i, j \in [1, n]$.

We will establish that, under these hypotheses, $\mathcal{A}_1$ and $\mathcal{A}_2$ share the same k -type for every $k \geq 0$. The theorem then corresponds to the special case $n = 0$ of this property.

The proof is by induction on k . For the base case $k = 0$, the (n, m) -chains $\mathcal{A}_1$ and $\mathcal{A}_2$ satisfy exactly the same atoms, thus they have the same 0-type. For the inductive case $k > 0$, we know that the k -type of $\mathcal{A}_1$ corresponds to the set of $(k-1)$ -types of all the $(n+1, m)$ -chains $(p_{1,1}, p_{1,2}, \dots, p_{1,n+1}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ that can be obtained by extending $\mathcal{A}_1$ with one additional first-order value $p_{1,n+1} \in D$. Likewise, the k -type of $\mathcal{A}_2$ corresponds to the set of $(k-1)$ -types of all the $(n+1, m)$ -chains $(p_{2,1}, p_{2,2}, \dots, p_{2,n+1}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$ obtained by extending $\mathcal{A}_2$ with $p_{2,n+1} \in D$.

It is sufficient to show that for each choice of $p_{1,n+1} \in D$, there exists $p_{2,n+1} \in D$ such that the $(k-1)$ -types of $(p_{1,1}, p_{1,2}, \dots, p_{1,n+1}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ and $(p_{2,1}, p_{2,2}, \dots, p_{2,n+1}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$ are identical. (By symmetry of the problem, this implies that for each choice of $p_{2,n+1} \in D$, there exists a matching $p_{1,n+1} \in D$ as well.)

Consider a particular value of $p_{1,n+1} \in D$. Either this value is equal to some other $p_{1,i}$, with $i \in [1, n]$, or it is not. In the former case, picking $p_{2,n+1} = p_{2,i}$ entails that the chains $(p_{1,1}, p_{1,2}, \dots, p_{1,n+1}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ and $(p_{2,1}, p_{2,2}, \dots, p_{2,n+1}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$ have the same $(k-1)$ -type by inductive hypothesis.

If, on the other hand, $p_{1,n+1} \neq p_{1,i}$ for every $i \in [1, n]$, then there exists an open interval (α_1, β_1) of D , with $\alpha_1 \in \{-\infty, p_{1,1}, p_{1,2}, \dots, p_{1,n}\}$, $\beta_1 \in \{p_{1,1}, p_{1,2}, \dots, p_{1,n}, +\infty\}$ and $\alpha_1 < \beta_1$, such that this interval contains $p_{1,n+1}$ and does not contain any $p_{1,i}$ with $i \in [1, n]$. We then construct the matching interval (α_2, β_2) by setting $\alpha_2 \in \{-\infty, p_{2,1}, p_{2,2}, \dots, p_{2,n}\}$ and $\beta_2 \in \{p_{2,1}, p_{2,2}, \dots, p_{2,n}, +\infty\}$ so that

$\alpha_2 = p_{2,i}$ iff $\alpha_1 = p_{1,i}$, and $\beta_2 = p_{2,i}$ iff $\beta_1 = p_{1,i}$, for every $i \in [1, n]$. We then pick $p_{2,n+1}$ in the interval (α_2, β_2) , such that the value of $(P_{2,1}, P_{2,2}, \dots, P_{2,m})$ at $p_{2,n+1}$ is identical to the value of $(P_{1,1}, P_{1,2}, \dots, P_{1,m})$ at $p_{1,n+1}$. This is always possible by definition of shuffle chains, since both $(P_{1,1}, P_{1,2}, \dots, P_{1,m})$ and $(P_{2,1}, P_{2,2}, \dots, P_{2,m})$ belong to $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$; this implies that the combination of predicate values at each point corresponds to some set U_i , and that each U_i occurs at least once in every bounded open interval of D .

With such a choice of value for $p_{2,n+1}$, the chains $(p_{1,1}, p_{1,2}, \dots, p_{1,n+1}, P_{1,1}, P_{1,2}, \dots, P_{1,m})$ and $(p_{2,1}, p_{2,2}, \dots, p_{2,n+1}, P_{2,1}, P_{2,2}, \dots, P_{2,m})$ have the same $(k-1)$ -type by inductive hypothesis, which concludes the proof. $\square$

It may not be obvious that the set of $(0, m)$ -chains described by $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$ is not empty. We now show how to explicitly build such a chain $(P_1, P_2, \dots, P_m)$, over the domain $D = (0, 1)$, by a Cantor-set-like construction. First, we split evenly the interval $(0, 1)$ into ℓ parts by setting $p_i = i/\ell$ for each $i \in [1, \ell-1]$. We then fix, for each $i \in [1, \ell-1]$ and $j \in [1, m]$, $P_j(p_i) = \top$, i.e., $p_i \in P_j$, if $j \in U_i$, and $P_j(p_i) = \perp$ if $j \notin U_i$. Then, in every real interval $(0, p_1), (p_1, p_2), \dots, (p_{\ell-2}, p_{\ell-1}), (p_{\ell-1}, 1)$ delimited by the points p_i , we carry out recursively the same operation. Finally, for each $p \in (0, 1)$ for which the value of the predicates P_j is still undefined, we set for each $j \in [1, m]$, $P_j(p) = \top$ if $j \in U_\ell$ and $P_j(p) = \perp$ if $j \notin U_\ell$. As a side remark, notice that the particular chain built by this process has countably many elements corresponding to each U_i for each $i \in [1, \ell-1]$ and an uncountable number of elements corresponding to U_ℓ . This is of no consequence in our context.

4.2.3 Type computation

Let $m \geq 0, \ell > 0$ and $U_i \subseteq \{1, 2, \dots, m\}$ for every $i \in [1, \ell]$. We now show that for every $k \geq 0$, the k -type of the $(0, m)$ -chains in $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$ can be effectively computed.

Like in the proof of Theorem 4, we consider the more general case of (n, m) -chains $(p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$, with $n \geq 0$, such that

- $(P_1, P_2, \dots, P_m) \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$, and
- for each $i \in [1, n]$ and $j \in [1, m]$, the Boolean value $t_{i,j}$ such that $t_{i,j} = \top$ iff $p_i \in P_j$ is fixed and known.

We show how to compute the k -type of such chains, given $U_1, U_2, \dots, U_\ell, p_1, p_2, \dots, p_n$, and $t_{i,j}$ for all $i \in [1, n]$ and $j \in [1, m]$. The original problem then corresponds to the particular case where $n = 0$ of this more general setting.

For the case $k = 0$, we have to compute the set of all the atoms involving n first-order and m second-order variables that are satisfied by the considered chains. This is immediate: the truth values of atoms of the form $x_i = x_j$ and $x_i < x_j$ are obtained from the values of $p_1, p_2, \dots, p_n$, and those of the atoms $x_i \in X_j$ are provided by the truth values $t_{i,j}$.

For the case $k > 0$, we have to compute the set of $(k-1)$ -types of all the $(n+1, m)$ -chains that can be obtained by adding one new first-order value $p_{n+1} \in D$, for all possible choices of this value. The possible choices of p_{n+1} are

- to make p_{n+1} equal to some p_i with $i \in [1, n]$, if $n > 0$. In this case, we have $t_{n+1,j} = t_{i,j}$ for all $j \in [1, m]$.
- to make p_{n+1} different from all the p_i with $i \in [1, n]$ (this case allows $n = 0$). In this case,
 - there are only finitely many classes of values for p_{n+1} that can be distinguished by atoms of the form $x_i = x_j$ and $x_i \leq x_j$ (evaluated over all the p_i for $i \in [1, n+1]$).
 - the values of $t_{n+1,j}$, for $j \in [1, m]$ can correspond to any of the U_i with $i \in [1, \ell]$. More precisely, for every choice of $i \in [1, \ell]$, one can set the $t_{n+1,j}$, for each $j \in [1, m]$, so that $t_{n+1,j} = \top$ iff $j \in U_i$.

We have established that there are only finitely many choices of p_{n+1} and $t_{n+1,j}$ for every $j \in [1, m]$ that can lead to different $(k-1)$ -types for the $(n+1, m)$ -chains obtained by adding the new first-order value p_{n+1} . One can then explicitly enumerate those choices, and compute the corresponding $(k-1)$ -types by induction on k .

4.3 Composition operators

We now introduce operators that allow to transform and combine the sets of chains that can be described by our notation. For each of them, we define how the composite chains are constructed from their underlying components, and show that chains sharing identical notations always have the same k -type for all $k \geq 0$. Most proofs in this section are concrete applications of the composition method [6–10].

4.3.1 First-order variables reordering

Let S be a set of (n, m) -chains, with $n, m \geq 0$, and $V = [i_1; i_2; \dots; i_n]$ be a permutation of $[1; 2; \dots; n]$. The notation

$$S/V$$

describes the set of all the chains contained in S in which the first-order values are reordered according to V . Formally, S/V is the set of all the chains $(p_{i_1}, p_{i_2}, \dots, p_{i_n}, P_1, P_2, \dots, P_m)$ such that $(p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m) \in S$. In other words, each first-order variable x_j involved in the expression of S , with $j \in [1, n]$, is renamed x_{i_j} .

Theorem 5. *Let $k \geq 0$. All the chains contained in S/V share the same k -type, assuming that the same property holds for S .*

Proof. Consider two (n, m) -chains $\mathcal{A}$ and $\mathcal{B}$ belonging to the set described by S/V . Let us show that these two chains share the same k -type for any $k \geq 0$. By definition of the reordering operator, the chains $\mathcal{A}$ and $\mathcal{B}$ correspond respectively to (n, m) -chains $\mathcal{A}_0$ and $\mathcal{B}_0$ of S , in which the first-order values have been reordered according to V .

The proof is by induction on k . For the base case $k = 0$, the sets of atoms satisfied by $\mathcal{A}$ and $\mathcal{B}$ correspond respectively to the atoms satisfied by the (n, m) -chains $\mathcal{A}_0$ and $\mathcal{B}_0$ in which the first-order variables are renamed according to V . Since those sets of atoms are identical for $\mathcal{A}_0$ and $\mathcal{B}_0$ by hypothesis, they are identical for $\mathcal{A}$ and $\mathcal{B}$ as well after the renaming operation.

For the inductive case $k > 0$, the k -type τ of $\mathcal{A}$ corresponds to the set of $(k-1)$ -types of the (n, m) -chains obtained by extending $\mathcal{A}$ with one additional first-order value p_{n+1} , for all possible choices of this value. Such chains take the form $(p_{i_1}, p_{i_2}, \dots, p_{i_n}, p_{n+1}, P_1, P_2, \dots, P_m)$, where $(p_1, p_2, \dots, p_n, p_{n+1}, P_1, P_2, \dots, P_m)$ is an element of the $(k-1)$ -type of $\mathcal{A}_0$. In other words, the elements of the k -type of the elements of S/V can be obtained by applying the variable reordering induced by the permutation $[i_1; i_2; \dots; i_n; n+1]$ to the elements of the (supposed unique) k -type of the elements of S . By inductive hypothesis, the result of this operation is identical for all elements of S , in particular for $\mathcal{A}$ and $\mathcal{B}$, which therefore admit the same k -type. $\square$

4.3.2 Concatenation

Let S_1 and S_2 be sets of (respectively) (n_1, m) and (n_2, m) -chains, with $n_1, n_2, m \geq 0$. Thanks to Theorem 2, we can freely assign the domains of the chains in S_1 and S_2 to be any non-empty open interval of $\mathbb{R}$ without affecting their type. We fix the domain $(-\infty, 0)$ for S_1 and the domain $(0, +\infty)$ for S_2 . The notation

$$S_1 +_U^\ell S_2,$$

with $\ell \geq 0$ and $U \subseteq \{1, 2, \dots, m\}$, then describes the set of all the $(n_1 + \ell + n_2, m)$ -chains $(p_1, p_2, \dots, p_{n_1 + \ell + n_2}, P_1, P_2, \dots, P_m)$ over the domain $\mathbb{R}$ such that

- the chain $(p_1, p_2, \dots, p_{n_1}, P'_1, P'_2, \dots, P'_m)$ belongs to S_1 , where for each $i \in [1, m]$, P'_i is the restriction of P_i to the domain $(-\infty, 0)$, i.e., $P'_i = P_i \cap (-\infty, 0)$.
- $p_{n_1 + i} = 0$ for each $i \in [1, \ell]$.
- the chain $(p_{n_1 + \ell + 1}, \dots, p_{n_1 + \ell + n_2}, P''_1, P''_2, \dots, P''_m)$ belongs to S_2 , where for each $i \in [1, m]$, P''_i is the restriction of P_i to the domain $(0, +\infty)$, i.e., $P''_i = P_i \cap (0, +\infty)$.
- for each $i \in [1, m]$, $0 \in P_i$ iff $i \in U$.

In other words, the domain is split into two parts. The values of the predicates in the first part are described by S_1 together with those of the first n_1 first-order variables, and in the second part by S_2 with the n_2 last first-order variables. At the junction point between the two parts, there are ℓ additional first-order variables, and the values of the predicates at this point are specified by U . In this definition, the junction point is fixed at 0 for the sake of simplicity, but any other real value could indifferently be used, since by Theorem 2 the domains can be transformed at will. Note that the additional first-order variables are inserted after those from S_1 and before those from S_2 , in order to make the notation associative: the chains described by $(S_1 +_{U_1}^{\ell_1} S_2) +_{U_2}^{\ell_2} S_3$ are identical to those described by $S_1 +_{U_1}^{\ell_1} (S_2 +_{U_2}^{\ell_2} S_3)$, and one can simply write $S_1 +_{U_1}^{\ell_1} S_2 +_{U_2}^{\ell_2} S_3$.

In the particular case where no additional first-order variable is inserted, we write $S_1 +_U S_2$ as shorthand for $S_1 +_U^0 S_2$.

Theorem 6. *Let $k \geq 0$. All the chains contained in $S_1 +_U^\ell S_2$ share the same k -type, assuming that the same property holds for S_1 and S_2 .*

Proof. W.l.o.g., we assume that the domains assigned to the chains belonging to S_1 , S_2 and $S_1 +_U^\ell S_2$ are respectively $(-\infty, 0)$, $(0, +\infty)$ and $\mathbb{R}$.

Consider two $(n_1 + \ell + n_2, m)$ -chains $\mathcal{A}$ and $\mathcal{B}$ belonging to the set described by $S_1 +_U^\ell S_2$. Let us show that these two chains share the same k -type for any $k \geq 0$. The proof is by induction on k . For the base case $k = 0$, we inspect the atoms satisfied by $\mathcal{A}$ or $\mathcal{B}$. The set of these atoms is composed of

- the atom $\top$.
- the atoms satisfied by any (n_1, m) -chain $(p_1, p_2, \dots, p_{n_1}, P'_1, P'_2, \dots, P'_m)$ belonging to S_1 . Recall that by hypothesis, all such chains share the same 0-type, thus the set of such atoms does not depend on the particular chain being chosen.
- the atoms satisfied by any (n_2, m) -chain $(p_1, p_2, \dots, p_{n_2}, P''_1, P''_2, \dots, P''_m)$ belonging to S_2 , in which the first-order variables $x_1, x_2, \dots, x_{n_2}$ are (respectively) renamed $x_{n_1+\ell+1}, x_{n_1+\ell+2}, \dots, x_{n_1+\ell+n_2}$. (The same remark applies.)
- the atoms $x_{i_1} < x_{i_2}$, $x_{i_1} < x_j$ and $x_j < x_{i_2}$ for every $i_1 \in [1, n_1]$, $i_2 \in [n_1 + \ell + 1, n_1 + \ell + n_2]$ and $j \in [n_1 + 1, n_1 + \ell]$.
- the atoms $x_i \in X_j$ for every $i \in [n_1 + 1, n_1 + \ell]$ and $j \in U$.

This set of atoms is thus identical for $\mathcal{A}$ and $\mathcal{B}$, hence these two chains share the same 0-type.

For the inductive case $k > 0$, we have to characterize the set T of $(k - 1)$ -types of the $(n_1 + \ell + n_2 + 1, m)$ -chains obtained by extending $\mathcal{A}$ and $\mathcal{B}$ with an additional first-order value $p_{n_1+\ell+n_2+1} \in \mathbb{R}$, for all possible choices of this value. There are several possibilities:

- $p_{n_1+\ell+n_2+1} < 0$. Then, each element τ of the k -type of the chains in S_1 (supposed to be unique by hypothesis) provides the $(k - 1)$ -type of some set S_1^τ of $(n_1 + 1, m)$ -chains obtained by adding such a first-order value $p_{n_1+\ell+n_2+1}$ to a (n_1, m) -chain in S_1 . For each S_1^τ obtained in this way, the $(n_1 + \ell + n_2 + 1)$ -chains described by $S_1^\tau +_U^\ell S_2$ share the same $(k - 1)$ -type by inductive hypothesis. By Theorem 5, the $(n_1 + \ell + n_2 + 1)$ -chains described by $(S_1^\tau +_U^\ell S_2)/V$, with $V = [1; 2; \dots; n_1; n_1 + 2; n_1 + 3; \dots; n_1 + \ell + n_2 + 1; n_1 + 1]$ share the same $(k - 1)$ -type as well. This $(k - 1)$ -type is an element of T .
- $p_{n_1+\ell+n_2+1} > 0$. Similarly, each element τ of the k -type of the chains in S_2 provides the $(k - 1)$ -type of some set S_2^τ of $(n_2 + 1, m)$ -chains obtained by adding a first-order value to a (n_2, m) -chain in S_2 . For each S_2^τ obtained in this way, the $(n_1 + \ell + n_2 + 1)$ -chains described by $S_1 +_U^\ell S_2^\tau$ share the same $(k - 1)$ -type by inductive hypothesis. This $(k - 1)$ -type is an element of T .
- $p_{n_1+\ell+n_2+1} = 0$. In this case, the $(n_1 + \ell + n_2 + 1, m)$ -chains obtained by extending $\mathcal{A}$ and $\mathcal{B}$ with $p_{n_1+\ell+n_2+1}$ belong to the set described by $(S_1 +_U^{(\ell+1)} S_2)/V$, where $V = [1; 2; \dots; n_1 + \ell; n_1 + \ell + 2; n_1 + \ell + 3; \dots; n_1 + \ell + n_2 + 1; n_1 + \ell + 1]$. By inductive hypothesis, the $(n_1 + \ell + n_2 + 1, m)$ -chains described by $S_1 +_U^{(\ell+1)} S_2$ share the same $(k - 1)$ -type. By Theorem 5, the $(n_1 + \ell + n_2 + 1, m)$ -chains described by $(S_1 +_U^{(\ell+1)} S_2)/V$ share the same $(k - 1)$ -type as well. This $(k - 1)$ -type is an element of T .

The set T of all the $(k - 1)$ -types obtained in this way is finite, and corresponds to the k -types of $\mathcal{A}$ and $\mathcal{B}$. These k -types are therefore identical. $\square$

Note that the proof of Theorem 6 provides an algorithm for computing the k -type of the chains described by $S_1 +_U^\ell S_2$ from the k -types of those described by S_1 and S_2 , for any $k \geq 0$.

4.3.3 Omega repetition

Let S be a set of $(0, m)$ -chains, with $m \geq 0$. The notation

$$S_U^{+\omega},$$

with $U \subseteq \{1, 2, \dots, m\}$ describes the set of all the $(0, m)$ -chains that can be transformed by Theorem 2 into chains $(P_1, P_2, \dots, P_m)$ over the domain $(0, +\infty)$ such that

- for each $j \in \mathbb{N}$, the chain $(P'_1, P'_2, \dots, P'_m)$ belongs to S , where for each $i \in [1, m]$, P'_i is the restriction of P_i to the open interval $(j, j+1)$, i.e., $P'_i = P_i \cap (j, j+1)$. Note that this definition assigns the domain $(j, j+1)$ to the chain $(P'_1, P'_2, \dots, P'_m)$, but thanks to Theorem 2, we can w.l.o.g. transform this domain so that it matches the domain of the chains in S .
- for each $j \in \mathbb{N}_{>0}$ and $i \in \{1, 2, \dots, m\}$, we have $j \in P_i$ iff $i \in U$.

In other words, the domain is split into countably many successive open intervals, with an initial one. Inside every such interval, the values of the predicates are described by S . At the junction points between the intervals, the values of the predicates are specified by U .

Theorem 7. *Let $k \geq 0$. All the chains contained in $S_U^{+\omega}$ share the same k -type, assuming that the same property holds for S .*

Proof. We will prove a stronger property. Let S_1 be a set of (n, m) -chains that all share the same k -type for every $k \geq 0$, with $n, m \geq 0$, and let S_2 be a set of $(0, m)$ -chains that share the same k -type for every $k \geq 0$. (The types of the chains in S_1 and S_2 can be different.) Let $U \subseteq \{1, 2, \dots, m\}$. We show that all the (n, m) -chains belonging to the set described by

$$S_1 +_U (S_2)_U^{+\omega}$$

share the same k -type for every $k \geq 0$. The statement of the theorem then corresponds to the particular case where $S_1 = S_2 = S$.

We assume w.l.o.g. that the domains of S_1 and $(S_2)_U^{+\omega}$ are respectively $(-\infty, 0)$ and $(0, +\infty)$. Furthermore, we suppose that the $(0, m)$ -chains that belong to $(S_2)_U^{+\omega}$ match chains in S_2 over each open interval $(j, j+1)$ with $j \in \mathbb{N}$, i.e., that the junction points for the $\cdot_U^{+\omega}$ operator are fixed at the strictly positive integers.

Let $\mathcal{A}$ and $\mathcal{B}$ be two (n, m) -chains belonging to the set described by $S_1 +_U (S_2)_U^{+\omega}$. We show by induction on k that $\mathcal{A}$ and $\mathcal{B}$ have the same k -type.

For the base case $k = 0$, we need to characterize the set of atoms satisfied by $\mathcal{A}$ and $\mathcal{B}$. Since these atoms either correspond to $\top$ or involve a first-order variable, they necessarily belong to the 0-type of S_1 . They are thus identical for $\mathcal{A}$ and $\mathcal{B}$, hence these two (n, m) -chains have the same 0-type.

For the inductive case $k > 0$, we construct the set T of $(k-1)$ -types of the $(n+1, m)$ -chains obtained by extending $\mathcal{A}$ and $\mathcal{B}$ with an additional first-order value $p_{n+1} \in \mathbb{R}$, for all possible choices of this value. There are several possibilities:

- $p_{n+1} < 0$. Then, each element τ of the k -type of the chains in S_1 (supposed to be unique by hypothesis) provides the $(k-1)$ -type of some set S_1^τ of $(n+1, m)$ -chains obtained by adding such a first-order value p_{n+1} to a (n, m) -chain in S_1 . For each S_1^τ obtained in this way, the $(n+1, m)$ -chains described by $S_1^\tau +_U (S_2)_U^{+\omega}$ share the same $(k-1)$ -type by inductive hypothesis. This $(k-1)$ -type is an element of T .
- $j < p_{n+1} < j+1$ for some $j \in \mathbb{N}$. Let us assume (w.l.o.g. thanks to Theorem 2) that the $(0, m)$ -chains in S_2 are defined over the domain $(j, j+1)$. The k -type of these chains is unique by hypothesis. Each element τ of this k -type provides the $(k-1)$ -type of some set S_2^τ of $(1, m)$ -chains obtained by adding a first-order value p_{n+1} to a $(0, m)$ -chain in S_2 . For each S_2^τ obtained in this way, the $(n+1, m)$ -chains described by

$$S_3^{\tau, j} = S_1 +_U S_2 +_U S_2 +_U \cdots +_U S_2 +_U S_2^\tau,$$

where there are $j+1$ occurrences of the operator $+_U$, share the same $(k-1)$ -type by Theorem 6. Therefore, by inductive hypothesis, the $(n+1, m)$ -chains described by

$$S_3^{\tau, j} +_U (S_2)_U^{+\omega}$$

share the same $(k-1)$ -type. This $(k-1)$ -type is an element of T .

- $p_{n+1} = j$ for some $j \in \mathbb{N}$. In this case, the $(n+1, m)$ -chains obtained by extending $\mathcal{A}$ and $\mathcal{B}$ with p_{n+1} belong to the set described by

$$S_4^j +_U (S_2)_U^{+\omega}, \tag{2}$$

with

$$S_4^j = S_1 +_U S_2 +_U S_2 +_U \cdots +_U S_2 +_U^{1_U} S_2,$$

where there are $j+1$ occurrences of $+_U$ (counting $+_U^{1_U}$ as well). By Theorem 6, all the $(n+1, m)$ -chains described by S_4^j share the same $(k-1)$ -type, therefore by inductive hypothesis, the $(n+1, m)$ -chains described by (2) share the same $k-1$ type as well. This $(k-1)$ -type is an element of T .

The set T of the $(k-1)$ -types obtained in this way is finite thanks to Theorem 1, and corresponds to the k -types of $\mathcal{A}$ and $\mathcal{B}$. These k -types are therefore identical. $\square$

Unlike Theorem 6, the proof of Theorem 7 does not straightforwardly translate into an algorithm for computing the k -type of the $(0, m)$ -chains described by $S_U^{+\omega}$ from the k -type of the chains in S . The problem is that there are potentially infinitely many sets $S_3^{\tau, j}$ and S_4^j to consider.

Let us show that we can construct the set of $(k-1)$ -types of the chains contained in the sets $S_3^{\tau, j}$ and S_4^j , induced by all possible choices of $j \in \mathbb{N}$, with a finite computation. The main issue is to compute the $(k-1)$ -types of the $(0, m)$ -chains described by

$$S_2 +_U S_2 +_U \cdots +_U S_2, \tag{3}$$

where there can be any number of occurrences of $+_U$, from the (supposed unique) $(k-1)$ -type τ_2 of the $(0, m)$ -chains contained in S_2 .

Let $T^0 = \{\tau_2\}$. Then, for each $i \geq 0$, we define

$$T^{i+1} = T^i \cup \{\tau +_U \tau_2 \mid \tau \in T^i\},$$

where $\tau +_U \tau_2$ denotes the $(k-1)$ -type of the $(0, m)$ -chains obtained by applying the operator $+_U$ to a $(0, m)$ -chain of $(k-1)$ -type τ and another of $(k-1)$ -type τ_2 . By Theorem 6, the resulting $(k-1)$ -type is unique for given k , τ , τ_2 and U . This recurrence relation makes it possible to compute effectively T^i for successive values of i .

By Theorem 1, the number of possible $(k-1)$ -types for $(0, m)$ -chains is finite for given values of k and m . Since we have $T^i \subseteq T^{i+1}$ for every $i \geq 0$, the computation of T^i will eventually stabilize, i.e., we will have $T^{i+1} = T^i$ for some $i \geq 0$. Such a T^i then contains the $(k-1)$ -types of all the $(0, m)$ -chains of the form (3).

4.3.4 Reverse omega repetition

Let S be a set of $(0, m)$ -chains, with $m \geq 0$. The notation

$$S_U^{-\omega},$$

with $U \subseteq \{1, 2, \dots, m\}$ describes the set of all the $(0, m)$ -chains that can be transformed thanks to Theorem 2 into chains $(P_1, P_2, \dots, P_m)$ over the domain $(-\infty, 0)$ such that

- for each $j \in \mathbb{N}$, the chain $(P'_1, P'_2, \dots, P'_m)$ belongs to S , where for each $i \in [1, m]$, P'_i is the restriction of P_i to the open interval $(-j-1, -j)$, i.e., $P'_i = P_i \cap (-j-1, -j)$. Like for omega repetition, we transform w.l.o.g. the domain $(-j-1, -j)$ of $(P'_1, P'_2, \dots, P'_m)$ so that it matches the domain of the chains in S ;
- for each $j \in \mathbb{N}_{>0}$ and $i \in \{1, 2, \dots, m\}$, we have $-j \in P_i$ iff $i \in U$.

The operator $\cdot_U^{-\omega}$ produces sets of chains that are the mirror image of those generated by $\cdot_U^{+\omega}$. Here, the domain is split into countably many successive open intervals, with a final one. Inside every such interval, the values of the predicates are described by S . At the junction points between the intervals, the values of the predicates are specified by U .

Theorem 8. *Let $k \geq 0$. All the chains contained in $S_U^{-\omega}$ share the same k -type, assuming that the same property holds for S .*

Proof. The proof follows the same lines as the one of Theorem 7. It can be turned in a similar way into an algorithm for computing the k -type of the $(0, m)$ -chains described by $S_U^{-\omega}$ from the k -type of the $(0, m)$ -chains contained in S . $\square$

4.3.5 Shuffle composition

Let $S_1, S_2, \dots, S_\ell$ be sets of $(0, m)$ -chains, with $\ell > 0$ and $m \geq 0$. The notation

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \quad (4)$$

where $U_i, U'_i, U''_j \subseteq \{1, 2, \dots, m\}$ for all $i \in [1, \ell]$ and $j \in [1, h]$, with $h > 0$, describes a set S of $(0, m)$ -chains defined as follows. We partition the domain D of those chains into

- countably infinitely many closed intervals $[a_i, b_i]$, with $a_i, b_i \in D$ and $a_i < b_i$ for each $i \in \mathbb{N}$. These intervals must be such that any open interval (a, b) , with $a, b \in D$ and $a < b$, that is not included in any of them intersects infinitely many of them. This implies that, with respect to the natural ordering between the intervals $[a_i, b_i]$, for $i \in \mathbb{N}$, none of them admits an immediate successor or predecessor.
- the remaining points of D , that form a set C that does not include any non-empty open interval. This implies that C is nowhere dense, i.e., that its closure has an empty interior. (Indeed, this closure cannot contain any interior point of an interval $[a_i, b_i]$ of the partition.)

An explicit example of such a partition can be built by a construction similar to that of the Cantor set: Assume w.l.o.g. that D is the open interval $(0, 1)$. We set $[a_0, b_0] = [\frac{1}{3}, \frac{2}{3}]$, and then carry on the generation of $[a_i, b_i]$ for $i = 1, 2, \dots$ by recursively partitioning the subdomains $(0, \frac{1}{3})$ and $(\frac{2}{3}, 1)$, and so on, following the same procedure.

Once such a partition has been obtained, we assign an index from 1 to ℓ to every interval $[a_i, b_i]$, as well as an index from 1 to h to every element of C . This must be done in such a way that every open interval (a, b) , with $a, b \in D$ and $a < b$, for which $(a, b) \not\subseteq [a_i, b_i]$ for all $i \in \mathbb{N}$ intersects intervals and remaining points of every index. In other words, every index from 1 to ℓ must be assigned to some $[a_i, b_i]$ such that $[a_i, b_i] \cap (a, b) \neq \emptyset$. Likewise, every index from 1 to h must be assigned to some $c \in C$ such that $a < c < b$.

We are now ready to define the $(0, m)$ -chains that belong to the set S defined by (4). These chains are those that can be transformed thanks to Theorem 2 into chains $(P_1, P_2, \dots, P_m)$ such that:

- for each interval $[a_i, b_i]$ of index j , with $i \in \mathbb{N}$ and $j \in [1, \ell]$:
 - the $(0, m)$ -chain $(P'_1, P'_2, \dots, P'_m)$ belongs to S_j , where each P'_k , with $k \in [1, m]$, is the restriction of the corresponding P_k to the domain (a_i, b_i) , i.e., $P'_k = P_k \cap (a_i, b_i)$.
 - for each $k \in [1, m]$, $a_i \in P_k$ iff $k \in U_j$.
 - for each $k \in [1, m]$, $b_i \in P_k$ iff $k \in U'_j$.
- for each point $c \in C$ of index j , with $j \in [1, h]$, and for each $k \in [1, m]$, $c \in P_k$ iff $k \in U''_j$.

In other words, the $(0, m)$ -chains in S are obtained by densely interleaving closed intervals described by the triplets (U_i, S_i, U'_i) , for $i \in [1, \ell]$. In each such interval, the combinations of values of the predicates at the left and right boundaries are respectively specified by U_i and U'_i , and in the interior of this interval by S_i . At the remaining points of the domain, the predicates admit a dense combination of values specified by the sets U''_i , for $i \in [1, h]$.

Theorem 9. *Let $k \geq 0$. All the chains contained in*

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \quad (5)$$

share the same k -type, assuming that the same property holds for the sets $S_1, S_2, \dots, S_\ell$.

Proof. As for the previous theorems, we prove a stronger property. Let $S_1, S_2, \dots, S_\ell$, with $\ell > 0$, be sets of $(0, m)$ -chains, with $m \geq 0$ that share the same k -type (this type can be different for each S_i). Let S_0 be a set of (n, m) -chains, with $n \geq 0$, that share the same k -type. Let $U_i, U'_i, U''_j \subseteq \{1, 2, \dots, m\}$ for all $i \in [0, \ell]$ and $j \in [1, h]$, with $h > 0$. We will show that the (n, m) -chains that belong to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_0} S_0 +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \end{aligned} \quad (6)$$

share the same k -type. The statement of the theorem then corresponds to the particular case where $S_0 = S_i$, $U_0 = U_i$ and $U'_0 = U'_i$ for an arbitrary $i \in [1, \ell]$. Indeed, in this case, it follows from the definition of the shuffle composition operator that the sets of $(0, m)$ -chains described by (5) and (6) are identical. We will rely in the proof on another property of this operator, which is that for any $i \in [1, h]$, the sets of $(0, m)$ -chains described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U''_i} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \end{aligned}$$

and (5) are identical as well.

We assume w.l.o.g. that the domains of the chains described by the three successive terms of (6) are respectively $(-\infty, -1)$, $(-1, 1)$ and $(1, +\infty)$. In other words, we fix the junction points for the operators $+_{U_0}$ and $+_{U'_0}$ at respectively -1 and 1 .

Let $\mathcal{A}$ and $\mathcal{B}$ be two (n, m) -chains belonging to the set described by (6). Let us show by induction on k that $\mathcal{A}$ and $\mathcal{B}$ have the same k -type.

For the base case $k = 0$, we need to characterize the set of atoms satisfied by $\mathcal{A}$ and $\mathcal{B}$. Since these atoms either correspond to $\top$ or involve a first-order variable, they necessarily belong to the 0-type of S_0 . They are thus identical for $\mathcal{A}$ and $\mathcal{B}$, hence these two (n, m) -chains have the same 0-type.

For the inductive case $k > 0$, we construct the set T of $(k-1)$ -types of the $(n+1, m)$ -chains obtained by extending $\mathcal{A}$ and $\mathcal{B}$ with an additional first-order value $p_{n+1} \in \mathbb{R}$, for all possible choices of this value. The possibilities are the following:

- $p_{n+1} < -1$. Let $\mathcal{A} = (p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$. By definition of the shuffle composition operator, there exists a partition Π of $(-\infty, -1)$ into closed intervals and remaining points, over which the values of the predicates P_j within $(-\infty, -1)$ are characterized by the triplets $(U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell)$ and the

sets $U_1'', U_2'', \dots, U_h''$, for all $j \in [1, m]$. We denote those closed intervals by $[a_i, b_i]$, for $i \in \mathbb{N}$, and the set of remaining points by C . There are several possibilities:

- $a_i < p_{n+1} < b_i$ for some interval $[a_i, b_i]$ in Π , with $i \in \mathbb{N}$. Let $\eta \in [1, \ell]$ be the index assigned to this interval. This means that the restriction of $(P_1, P_2, \dots, P_m)$ to the interval (a_i, b_i) belongs to S_η , assuming (w.l.o.g. thanks to Theorem 2) that the domain of S_η is (a_i, b_i) . The k -type of the chains in S_η is unique by hypothesis. Each element τ of this k -type provides the $(k-1)$ -type of some set S_η^τ of $(1, m)$ -chains obtained by adding a first-order value p_{n+1} to a $(0, m)$ -chain in S_η . For each S_η^τ obtained in this way, the $(n+1, m)$ -chains described by

$$R_\eta^\tau = S_\eta^\tau +_{U'_\eta} \llbracket (U_1, S_1, U'_1), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, \dots, U''_h \rrbracket +_{U_0} S_0$$

share the same $(k-1)$ -type by inductive hypothesis and by Theorem 6. Therefore, by inductive hypothesis, the $(n+1, m)$ -chains described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_\eta} R_\eta^\tau +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \end{aligned}$$

share the same $(k-1)$ -type. After applying the reordering operator with respect to the permutation $[2; 3; \dots; n+1; 1]$, one obtains a $(k-1)$ -type that is an element of T .

- $p_{n+1} = a_i$ for some interval $[a_i, b_i]$ in Π , with $i \in \mathbb{N}$. Let $\eta \in [1, \ell]$ be the index assigned to this interval. Up to some reordering of the variables, the $(n+1, m)$ -chain obtained by extending $\mathcal{A}$ with p_{n+1} belongs to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U'_1} R_\eta +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \end{aligned} \tag{7}$$

where R_η is the set of $(n+1, m)$ -chains described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_\eta}^1 S_\eta +_{U'_\eta} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_0} S_0. \end{aligned}$$

By inductive hypothesis and Theorem 6, the $(n+1, m)$ -chains described by (7) share the same $(k-1)$ -type. After applying the reordering operator with respect to the permutation $[2; 3; \dots; n+1; 1]$, one obtains a $(k-1)$ -type that is an element of T .

- $p_{n+1} = b_i$ for some interval $[a_i, b_i]$ in Π , with $i \in \mathbb{N}$. Let $\eta \in [1, \ell]$ be the index assigned to this interval. Up to some reordering of the variables, the $(n+1, m)$ -chain obtained by extending $\mathcal{A}$ with p_{n+1} belongs to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_\eta} R'_\eta +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \end{aligned} \quad (8)$$

where R'_η is the set of $(n+1, m)$ -chains described by

$$S_\eta +_{U'_\eta}^1 \llbracket (U_1, S_1, U'_1), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, \dots, U''_h \rrbracket +_{U_0} S_0.$$

By inductive hypothesis and Theorem 6, the $(n+1, m)$ -chains described by (8) share the same $(k-1)$ -type. After applying the reordering operator with respect to the permutation $[2; 3; \dots; n+1; 1]$, one obtains a $(k-1)$ -type that is an element of T .

- $p_{n+1} \in C$. Let $\eta \in [1, h]$ be the index assigned to p_{n+1} . Up to some reordering of the variables, the $(n+1, m)$ -chain obtained by extending $\mathcal{A}$ with p_{n+1} belongs to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U''_1} R''_\eta +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \end{aligned} \quad (9)$$

where R''_η is the set of $(n+1, m)$ -chains described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U''_\eta}^1 \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_0} S_0. \end{aligned}$$

By inductive hypothesis and Theorem 6, the $(n+1, m)$ -chains described by (9) share the same $(k-1)$ -type. After applying the reordering operator with respect to the permutation $[2; 3; \dots; n+1; 1]$, one obtains a $(k-1)$ -type that is an element of T .

- $p_{n+1} > 1$. This is the symmetrical situation of the previous one. It can be handled in a similar way.
- $-1 < p_{n+1} < 1$. Then, each element τ of the k -type of the chains in S_0 (supposed to be unique by hypothesis) provides the $(k-1)$ -type of some set S_0^τ of $(n+1, m)$ -chains obtained by adding such a first-order value p_{n+1} to a (n, m) -chain in S_0 . For each S_0^τ obtained in this way, the $(n+1, m)$ -chains described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_0} S_0^\tau +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \end{aligned}$$

share the same $(k-1)$ -type by inductive hypothesis. This $(k-1)$ -type is an element of T .

- $p_{n+1} = -1$. Then, up to some reordering of the variables, the $(n+1, m)$ -chain obtained by extending $\mathcal{A}$ with p_{n+1} belongs to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U''_1} R +_{U'_0} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \end{aligned} \quad (10)$$

where R is the set of $(n+1, m)$ -chains described by

$$\llbracket (U_1, S_1, U'_1), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, \dots, U''_h \rrbracket +_{U_0}^1 S_0.$$

By inductive hypothesis and Theorem 6, the $(n+1, m)$ -chains described by (10) share the same $(k-1)$ -type. After applying the reordering operator with respect to the permutation $[2; 3; \dots; n+1; 1]$, one obtains a $(k-1)$ -type that is an element of T .

- $p_{n+1} = 1$. Then, the $(n+1, m)$ -chain obtained by extending $\mathcal{A}$ with p_{n+1} belongs to the set described by

$$\begin{aligned} & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket \\ & \quad +_{U_0} R' +_{U''_1} \\ & \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket, \end{aligned} \quad (11)$$

where R' is the set of $(n+1, m)$ -chains described by

$$S_0 +_{U_0}^1 \llbracket (U_1, S_1, U'_1), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, \dots, U''_h \rrbracket.$$

By inductive hypothesis and Theorem 6, the $(n+1, m)$ -chains described by (11) share the same $(k-1)$ -type. This $(k-1)$ -type is an element of T .

The set T of the $(k-1)$ -types obtained in this way does only depend on the choice of the sets S_i and U_i, U'_i, U''_j , for $i \in [0, \ell]$ and $j \in [0, h]$, and not on the particular values of p_i and P_j , for $p_i \in [1, n]$ and $P_j \in [1, m]$. This set T is thus identical for $\mathcal{A}$ and $\mathcal{B}$, which therefore have the same k -type. $\square$

Note that the proof of Theorem 9 provides an algorithm for computing the k -type of the chains described by

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket$$

from the k -types of the sets $S_1, S_2, \dots, S_\ell$, for any $k \geq 0$.

4.4 Synthesis

The following definition characterizes the class of chains that can be constructed by starting from the elementary chains of Section 4.2, and applying finitely many times the operators of Section 4.3.

Definition 10. A (n, m) -chain, with $n, m \geq 0$, is said to be *constructible* if it belongs to a set of chains described by

- $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$ for some $\ell > 0$ and $U_i \subseteq \{1, 2, \dots, m\}$ for all $i \in [1, \ell]$, with $n = 0$,
- S/V , where V is a permutation of $[1; 2; \dots; n]$ and S is a set of constructible (n, m) -chains,
- $S_1 +_U^\ell S_2$, where $0 \leq \ell \leq n$, $U \subseteq \{1, 2, \dots, m\}$, S_1 is a set of constructible (n_1, m) -chains, and S_2 is a set of constructible (n_2, m) -chains, with $n_1 + \ell + n_2 = n$,
- $S_U^{+\omega}$ or $S_U^{-\omega}$, where $n = 0$, $U \subseteq \{1, 2, \dots, m\}$ and S is a set of constructible $(0, m)$ -chains, or
- $\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket$, where $n = 0$, $\ell, h > 0$, $U_i, U'_i, U''_j \subseteq \{1, 2, \dots, m\}$ for all $i \in [1, \ell]$ and $j \in [1, h]$, and $S_1, S_2, \dots, S_\ell$ are sets of constructible $(0, m)$ -chains.

Definition 10 provides a symbolic notation for sets of chains: we call a *chain term* an expression obtained by applying finitely many times reordering, concatenation, repetition or shuffle composition operators to elementary sets of chains of the form $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$. In Sections 4.2 and 4.3, we have established that for any $n, m, k \geq 0$, the (n, m) -chains described by a chain term share the same k -type. Moreover, we have developed algorithms for computing the k -type of the chains in every elementary set, as well as in every set obtained by applying reordering, concatenation, repetition or shuffle composition operators to sets of chains with a known k -type. This leads to the following result.

Theorem 11. *There exists an algorithm for computing, for any $k, n, m \geq 0$, the k -type of the constructible (n, m) -chains described by a given chain term.*

5 Deciding satisfiability

By applying the results of Section 3.3, one obtains the following corollary to Theorem 11.

Corollary 12. *For every constructible (n, m) -chain $\mathcal{A}$, with $n, m \geq 0$ described by a chain term and formula φ of $\text{MFO}(\mathbb{R}, <)$, it is decidable whether $\mathcal{A} \models \varphi$.*

Our goal now consists in establishing that for every (n, m) -chain $\mathcal{A}$, with $n, m \geq 0$, and $k \geq 0$, there exists a constructible (n, m) -chain $\mathcal{B}$ such that $\mathcal{A} \equiv_k \mathcal{B}$. An important consequence of this property will be that every formula of $\text{MFO}(\mathbb{R}, <)$ that is satisfiable admits a model that takes the form of a constructible chain.

The results of this section follow the same approach as in [5], working in $\mathbb{R}$ instead of in an arbitrary totally ordered set.

5.1 Good chains

A (n, m) -chain $\mathcal{A}$, with $n, m \geq 0$, is said to be *k -good*, with $k \geq 0$, if there exists a constructible (n, m) -chain $\mathcal{B}$ such that $\mathcal{A} \equiv_k \mathcal{B}$. From Definition 10 and the results of Section 4.3, we have the following lemma (corresponding to Lemma 7 of [5]), stating that the class of k -good chains is closed under reordering, concatenation, omega and reverse omega repetition, and shuffle composition operators.

Lemma 13. *The following properties hold:*

- If S is a set of k -good (n, m) -chains for $k, n, m \geq 0$, then S/V is a set of k -good (n, m) -chains, for any permutation V of $[1; 2; \dots; n]$.
- If S_1 and S_2 are sets of k -good (n_1, m) and (n_2, m) -chains respectively, for $k, n_1, n_2, m \geq 0$, then $S_1 +_U^j S_2$ is a set of k -good $(n_1 + j + n_2, m)$ -chains, for any $j \geq 0$ and $U \subseteq \{1, 2, \dots, m\}$.
- If $S_1, S_2, \dots, S_\ell$ are sets of k -good $(0, m)$ -chains, for $k, m \geq 0$ and $\ell \geq 1$, then the $(0, m)$ -chains in $\llbracket (U_1, S_1, U_1'), (U_2, S_2, U_2'), \dots, (U_\ell, S_\ell, U_\ell'), U_1'', U_2'', \dots, U_h'' \rrbracket$ are k -good, for any $U_i, U_i', U_j'' \subseteq \{1, 2, \dots, m\}$, for all $i \in [1, \ell]$, $j \in [1, h]$, with $h > 0$.
- If S is a set of k -good $(0, m)$ -chains for $k, m \geq 0$, then $(S)_U^{+\omega}$ and $(S)_U^{-\omega}$ are sets of k -good $(0, m)$ -chains, for any $U \subseteq \{1, 2, \dots, m\}$.

5.2 Segments

In the following definition, we abuse the notation by writing $\mathcal{A}$ for the set $\{\mathcal{A}\}$ containing only the chain $\mathcal{A}$, for the sake of simplicity.

Definition 14. A (n_1, m) -chain $\mathcal{A}_1$ is a segment of a (n_2, m) -chain $\mathcal{A}_2$, with $0 \leq n_1 \leq n_2$ and $m \geq 0$, if at least one of the following conditions is satisfied:

- There exists a permutation V of $[1; 2; \dots; n_2]$ such that

$$\mathcal{A}_2 = \mathcal{A}_1/V. \quad (12)$$

- There exist $n_3 \geq 0$, a (n_3, m) -chain $\mathcal{A}_3$, a set $U \subseteq \{1, 2, \dots, m\}$, $\ell \geq 0$, and a permutation V of $[1; 2; \dots; n_2]$ such that $n_2 = n_3 + \ell + n_1$ and

$$\mathcal{A}_2 = (\mathcal{A}_3 +_U^\ell \mathcal{A}_1)/V. \quad (13)$$

- There exist $n_4 \geq 0$, a (n_4, m) -chain $\mathcal{A}_4$, a set $U \subseteq \{1, 2, \dots, m\}$, $\ell \geq 0$, and a permutation V of $[1; 2; \dots; n_2]$ such that $n_2 = n_1 + \ell + n_4$ and

$$\mathcal{A}_2 = (\mathcal{A}_1 +_U^\ell \mathcal{A}_4)/V. \quad (14)$$

- There exist $n_3, n_4 \geq 0$, a (n_3, m) -chain $\mathcal{A}_3$, a (n_4, m) -chain $\mathcal{A}_4$, sets $U_1, U_2 \subseteq \{1, 2, \dots, m\}$, $\ell_1, \ell_2 \geq 0$, and a permutation V of $[1; 2; \dots; n_2]$ such that $n_2 = n_3 + \ell_1 + n_1 + \ell_2 + n_4$ and

$$\mathcal{A}_2 = (\mathcal{A}_3 +_{U_1}^{\ell_1} \mathcal{A}_1 +_{U_2}^{\ell_2} \mathcal{A}_4)/V. \quad (15)$$

In this last case, the segment $\mathcal{A}_1$ is said to be bounded.

The bounded segment $\mathcal{A}_1$ describes the content of the chain $\mathcal{A}_2$ between two points a and b in $\mathbb{R}$, which correspond to the junction points of the two concatenation operators in (15). We will write

$$\mathcal{A}_1 = [\mathcal{A}_2]_{(a,b)}$$

when we want to make this property explicit, in the particular case that the first-order variables reordering is irrelevant, that is, (15) holds with either $n_1 = 0$, or $n_1 = n_2 > 0$ and $V = [1; 2; \dots; n_2]$. Similarly, for unbounded segments, we write

$\mathcal{A}_1 = [\mathcal{A}_2]_{(-\infty, +\infty)}$, $\mathcal{A}_1 = [\mathcal{A}_2]_{(-\infty, b)}$ and $\mathcal{A}_1 = [\mathcal{A}_2]_{(a, +\infty)}$ when (respectively) (12), (14) and (13) are satisfied, under the same assumption on n_1 , n_2 and V .

The following result corresponds to Lemma 8 of [5].

Lemma 15. *If every bounded segment of a (n, m) -chain $\mathcal{A}$ is k -good, for $n, m, k \geq 0$, then $\mathcal{A}$ is k -good as well.*

Proof. It is always possible to decompose $\mathcal{A}$ into

$$\mathcal{A} = (\mathcal{A}_1 +_U \mathcal{A}_2)/V \quad (16)$$

where $\mathcal{A}_1$ and $\mathcal{A}_2$ are respectively (n_1, m) and (n_2, m) -chains, with $n_1 + n_2 = n$, $U \subseteq \{1, 2, \dots, m\}$, and V is a permutation of $[1; 2; \dots; n_1 + n_2]$. Indeed, it suffices to fix the junction point of the concatenation operator at any point $p \in \mathbb{R}$ that does not coincide with a first-order value in $\mathcal{A}$. Let us consider (w.l.o.g. thanks to Theorem 2) that we have $p = 0$. The second-order values that are true at 0 then determine the set U .

By Lemma 13, to prove that $\mathcal{A}$ is k -good, it is sufficient to establish that $\mathcal{A}_1$ and $\mathcal{A}_2$ are both k -good. We only give the proof for $\mathcal{A}_2$; the case of $\mathcal{A}_1$ is symmetric and can be handled in a similar way.

Let $\mathcal{A} = (p_1, p_2, \dots, p_n, P_1, P_2, \dots, P_m)$. We pick an arbitrary infinite sequence $d_0, d_1, d_2, \dots \in \mathbb{R}_{>0}$ of points in the domain $(0, +\infty)$ of $\mathcal{A}_2$, such that

- $d_{i+1} > d_i$ for all $i \in \mathbb{N}$. In other words, the sequence is strictly increasing.
- $\forall d > 0 : \exists i \in \mathbb{N} : d_i > d$. In other words, the sequence does not admit an upper bound in $\mathbb{R}$.
- $p_i < d_0$ for all $i \in [1, n]$. In other words, the sequence begins after the greatest first-order value in $\mathcal{A}$.

Every unordered pair of values (d_i, d_j) with $i, j \in \mathbb{N}$ and $i \neq j$ determines a bounded segment of $\mathcal{A}$, which corresponds to $[\mathcal{A}]_{(d_i, d_j)}$ if $d_i < d_j$ and to $[\mathcal{A}]_{(d_j, d_i)}$ if $d_i > d_j$. We can partition the set of all pairs (d_i, d_j) according to

- the k -type of the bounded segment that they determine, together with
- the combination of values taken by the predicates $P_1, P_2, \dots, P_m$ at $d_{i,j} = \min(d_i, d_j)$. Formally, this combination is represented by the set $U_{i,j} = \{\ell \in [1, m] \mid d_{i,j} \in P_\ell\}$.

The bounded segments involved in this partitioning scheme have the dimension $(0, m)$, hence by Theorem 1, their k -type can only take finitely many values. The possible values of the sets $U_{i,j}$ are also finite. It follows that our partition is finite.

By the infinite version of Ramsey's theorem [24], there exists an infinite subsequence $\delta_0, \delta_1, \delta_2, \dots$ of $d_0, d_1, d_2, \dots$ such that every pair (δ_i, δ_j) with $i, j \in \mathbb{N}$ and $i \neq j$ belongs to the same class of the partition. Note that this sequence $\delta_0, \delta_1, \delta_2, \dots$ does not admit an upper bound in $\mathbb{R}$. It is such that the k -type of the bounded interval $[\mathcal{A}]_{(\delta_i, \delta_{i+1})}$ is identical for all $i \in \mathbb{N}$, and the combination of values taken by the predicates $P_1, P_2, \dots, P_m$ at all the δ_i is identical as well. We therefore have

$$\mathcal{A}_2 \equiv_k [\mathcal{A}]_{(0, \delta_0)} +_{U'} ([\mathcal{A}]_{(\delta_0, \delta_1)})_{U'}^{+\omega},$$

for some $U' \subseteq \{1, 2, \dots, m\}$.

By hypothesis, the bounded segments $[\mathcal{A}]_{(0, \delta_0)}$ and $[\mathcal{A}]_{(\delta_0, \delta_1)}$ involved in this expression are k -good. It follows by Lemma 13 that $\mathcal{A}_2$ is k -good as well. $\square$

The reciprocal of the property expressed by Lemma 15 also holds, for bounded and as well as unbounded segments:

Lemma 16. *Every segment of a k -good (n, m) -chain $\mathcal{A}$ is k -good as well, for any $n, m, k \geq 0$.*

Proof. The proof is by structural induction on the description of a set of chains that are k -elementarily equivalent to $\mathcal{A}$. For the base case where $\mathcal{A} \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$ for some $U_1, U_2, \dots, U_\ell \subseteq \{1, 2, \dots, m\}$, with $\ell > 0$, we have that every segment of $\mathcal{A}$ also belongs to $\llbracket U_1, U_2, \dots, U_\ell \rrbracket$ by definition of shuffle chains. Every such segment is thus k -good.

For the inductive case, we consider separately each operator.

- Every segment $\mathcal{A}$ of a (n, m) -chain in S/V , where S is a set of k -good chains and V is a permutation of $[1; 2; \dots; n]$, is such that

$$\mathcal{A} = \mathcal{A}_1/V_1,$$

where the (n_1, m) -chain $\mathcal{A}_1$ is a segment of a chain in S , with $0 \leq n_1 \leq n$, and V_1 is a permutation of $[1; 2; \dots; n_1]$. By inductive hypothesis and Lemma 13, $\mathcal{A}$ is k -good.

- Every segment $\mathcal{A}$ of a (n, m) -chain in $S_1 +_U^\ell S_2$, where S_1, S_2 are sets of k -good chains, $\ell \geq 0$ and $U \subseteq \{1, 2, \dots, m\}$, is such that either
 - $\mathcal{A}$ is a segment of a chain belonging to S_1 or S_2 . In this case, $\mathcal{A}$ is k -good by inductive hypothesis,
 - or $\mathcal{A} = (\mathcal{A}_1 +_U^\ell \mathcal{A}_2)/V$, where $\mathcal{A}_1$ is a segment of dimension (n_1, m) of a chain in S_1 , with $n_1 \geq 0$, $\mathcal{A}_2$ is a segment of dimension (n_2, m) of a chain in S_2 , with $n_2 \geq 0$, and V is a permutation of $[1; 2; \dots; n_1 + n_2 + \ell]$. This implies, by inductive hypothesis and Lemma 13, that $\mathcal{A}$ is k -good.
- Every segment $\mathcal{A}$ of a $(0, m)$ -chain in $S_U^{+\omega}$ or $S_U^{-\omega}$, where S is a set of k -good $(0, m)$ -chains and $U \subseteq \{1, 2, \dots, m\}$, is such that
 - either $\mathcal{A}$ is a segment of a chain belonging to S . In this case, $\mathcal{A}$ is k -good by inductive hypothesis.
 - or $\mathcal{A} = \mathcal{A}_1 +_U \mathcal{A}_2 +_U \mathcal{A}_2 +_U \dots +_U \mathcal{A}_2 +_U \mathcal{A}_3$, where there are any (nonzero) number of occurrences of the operator $+_U$, $\mathcal{A}_1$ and $\mathcal{A}_3$ are segments of $(0, m)$ -chains in S , and $\mathcal{A}_2$ is a $(0, m)$ -chain in S . By inductive hypothesis and Lemma 13, $\mathcal{A}$ is then k -good.
- Every segment $\mathcal{A}$ of a $(0, m)$ -chain in

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket,$$

where $\ell, h \geq 0$, $S_1, S_2, \dots, S_\ell$ are sets of k -good $(0, m)$ -chains and $U_i, U'_i, U''_j \subseteq \{1, 2, \dots, m\}$ for all $i \in [0, \ell]$ and $j \in [0, h]$, is such that

- either $\mathcal{A}$ is a segment of a chain belonging to $S_1, S_2, \dots$ or S_ℓ . In this case, $\mathcal{A}$ is k -good by inductive hypothesis.
- $\mathcal{A} = \mathcal{A}_1 +_{U'_i} \mathcal{A}_2 +_{U_j} \mathcal{A}_3$, where $\mathcal{A}_1$ is a segment of a $(0, m)$ -chain in some set S_i , with $i \in [1, \ell]$, $\mathcal{A}_3$ is a segment of a $(0, m)$ -chain in some set S_j , with $j \in [1, \ell]$, and $\mathcal{A}_2$ is a $(0, m)$ -chain in

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket.$$

By inductive hypothesis and Lemma 13, this implies that $\mathcal{A}$ is k -good.

- $\mathcal{A} = \mathcal{A}_1 +_{U'_i} \mathcal{A}_2$, where $\mathcal{A}_1$ is a segment of a $(0, m)$ -chain in some set S_i , with $i \in [1, \ell]$ and $\mathcal{A}_2$ is a $(0, m)$ -chain in

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket.$$

By inductive hypothesis and Lemma 13, this implies that $\mathcal{A}$ is k -good.

- or $\mathcal{A} = \mathcal{A}_2 +_{U_j} \mathcal{A}_3$, where $\mathcal{A}_3$ is a segment of a $(0, m)$ -chain in some set S_j , with $j \in [1, \ell]$, and $\mathcal{A}_2$ is a $(0, m)$ -chain in

$$\llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket.$$

By inductive hypothesis and Lemma 13, this implies that $\mathcal{A}$ is k -good.

□

Lemmas 15 and 16 reduce the problem of establishing that a chain is k -good with respect to its full domain to the same problem over bounded subsets of this domain. This result is quite powerful. Consider for instance a $(0, m)$ -chain $\mathcal{A}$ defined over the domain $(0, +\infty)$, such that for every $i \in \mathbb{N}$, the bounded segment $[\mathcal{A}]_{(i, i+1)}$ belongs to the set described by $(\dots((S_U^{+\omega})_U^{+\omega})_U^{+\omega} \dots)^{+\omega}$, where there are i occurrences of the repetition operator $\cdot_U^{+\omega}$, S is a set of k -good $(0, m)$ -chains, and $U \subseteq \{1, 2, \dots, m\}$. In other words, we have (using improperly the notation):

$$\mathcal{A} = \mathcal{A}_0 +_U (\mathcal{A}_1)^{+\omega} +_U ((\mathcal{A}_2)^{+\omega})^{+\omega} +_U (((\mathcal{A}_3)^{+\omega})^{+\omega})^{+\omega} +_U \dots,$$

where $\mathcal{A}_i$ is k -good for every $i \in \mathbb{N}$. Even though the previous expression is not a valid chain term — since it involves infinitely many occurrences of the concatenation operator $+_U$ — it follows from Lemmas 15 and 16 that the chain $\mathcal{A}$ is k -good, and therefore that it belongs to some set that can be described by a chain term.

5.3 Goodness is everywhere

The following result corresponds to Lemma 9 of [5], adapted to the domain $\mathbb{R}$.

Lemma 17. *For every $k \geq 0$ and (n, m) -chain $\mathcal{A}$, with $n, m \geq 0$, $\mathcal{A}$ is k -good.*

Proof. We first observe that if $n > 0$, then $\mathcal{A}$ can be decomposed into

$$\mathcal{A} = (\mathcal{A}_1 +_{U_1}^{\ell_1} \mathcal{A}_2 +_{U_2}^{\ell_2} \dots +_{U_j}^{\ell_j} \mathcal{A}_{j+1})/V,$$

where $0 < j \leq n$, $\ell_1, \ell_2, \dots, \ell_j \geq 0$ with $\ell_1 + \ell_2 + \dots + \ell_j = n$, $U_i \subseteq \{1, 2, \dots, m\}$ for all $i \in [1, j]$, V is a permutation of $[1, 2, \dots, n]$, and $\mathcal{A}_i$ is a $(0, m)$ -chain for each $i \in [1, j+1]$. By Lemma 13, it is then sufficient to establish the lemma for each $\mathcal{A}_i$, i.e., we can assume w.l.o.g. that $n = 0$.

Let therefore $\mathcal{A} = (P_1, P_2, \dots, P_m)$ be a $(0, m)$ -chain. Recall that the domain is assumed to be $\mathbb{R}$. We define the set $R \subseteq \mathbb{R}$ as the union of the intervals (a, b) , with $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{+\infty\}$ and $a < b$, for which there exist $\ell > 0$ and $U_1, U_2, \dots, U_\ell \subseteq \{1, 2, \dots, m\}$ such that $[\mathcal{A}]_{(a,b)} \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$.

Let us show that the set R is dense in $\mathbb{R}$, i.e., that every $a, b \in \mathbb{R}$ with $a < b$ are such that $R \cap (a, b) \neq \emptyset$. By contradiction, assume that there exists a non-empty open interval (a, b) of $\mathbb{R}$ inside which no bounded segment of $\mathcal{A}$ is a constant or shuffle chain. We define

$$\mathcal{U}_{(a,b)} = \{\{j \in [1, m] \mid p \in P_j\} \mid p \in \mathbb{R} \wedge a < p < b\}.$$

In other words, $\mathcal{U}_{(a,b)}$ describes the combinations of predicates values that occur inside the interval (a, b) .

We then perform the following operation. If there exist $\alpha, \beta \in \mathbb{R}$ such that $a < \alpha < \beta < b$, i.e., (α, β) is a proper sub-interval of (a, b) , and $\mathcal{U}_{(\alpha,\beta)} \subsetneq \mathcal{U}_{(a,b)}$, then we replace (a, b) with (α, β) and repeat the same procedure. Note that, since m is fixed, the number of possible distinct sets $\mathcal{U}_{(a,b)}$ is finite, hence this procedure will terminate after a finite number of steps. The final interval (a, b) is then such that $[\mathcal{A}]_{(a,b)} \in \llbracket U_1, U_2, \dots, U_\ell \rrbracket$, where $U_1, U_2, \dots, U_\ell$ are the elements of $\mathcal{U}_{(a,b)}$. This contradicts the hypothesis that within our original interval (a, b) , the chain $\mathcal{A}$ does not admit a constant or shuffle bounded segment.

At this stage, we have established that there exists a partition Π_1 of the domain $\mathbb{R}$ into

- open intervals (a_i, b_i) , with $a_i \in \mathbb{R} \cup \{-\infty\}$, $b_i \in \mathbb{R} \cup \{+\infty\}$ and $a_i < b_i$ for either all $i \in \mathbb{N}$ or all $i \in [1, q]$ with $q > 0$. For each such interval, there exist $\ell_i > 0$ and $U_{i,1}, U_{i,2}, \dots, U_{i,\ell_i} \subseteq \{1, 2, \dots, m\}$ such that $[\mathcal{A}]_{(a_i,b_i)} \in \llbracket U_{i,1}, U_{i,2}, \dots, U_{i,\ell_i} \rrbracket$. Note that the number of such intervals is necessarily finite or countable, since every of them contains a rational value that does not belong to the others. The set of all those intervals is denoted by I_1 .
- remaining points $c \in \mathbb{R}$. The set C_1 of those remaining points does not include any non-empty open interval, i.e., C_1 is nowhere dense. (Note that C_1 is the complement of a union of open sets, and is therefore closed.)

We now define a relation $\sim$ on the elements of I_1 : Two intervals $(a_1, b_1), (a_2, b_2) \in I_1$ are such that $(a_1, b_1) \sim (a_2, b_2)$ iff either

- $(a_1, b_1) = (a_2, b_2)$,
- $a_1 < b_2$ and the segment $[\mathcal{A}]_{(a_1,b_2)}$ is k -good, or
- $a_2 < b_1$ and the segment $[\mathcal{A}]_{(a_2,b_1)}$ is k -good.

Let us show that $\sim$ is an equivalence relation. Indeed, if $(a_1, b_1) \sim (a_2, b_2)$ and $(a_2, b_2) \sim (a_3, b_3)$, then one has $(a_1, b_1) \sim (a_3, b_3)$ as a consequence of the definition of $\sim$ together with Lemmas 13 and 16.

Each equivalence class E of $\sim$ is a set of disjoint open intervals (a_i, b_i) . For each such class, we define

$$\alpha_E = \inf\{a_i \mid (a_i, b_i) \in E\},$$

and

$$\beta_E = \sup\{b_i \mid (a_i, b_i) \in E\}.$$

We thus have $\alpha_E \in \mathbb{R} \cup \{-\infty\}$, $\beta_E \in \mathbb{R} \cup \{+\infty\}$ and $\alpha_E < \beta_E$.

Each open interval (α_E, β_E) is such that the segment $[\mathcal{A}]_{(\alpha_E, \beta_E)}$ is k -good. Indeed, for every bounded sub-interval (a, b) of (α_E, β_E) , i.e., such that $\alpha_E < a < b < \beta_E$, there exist $(a_1, b_1) \in E$ and $(a_2, b_2) \in E$ such that $a_1 < a$ and $b < b_2$. By definition of the equivalence relation $\sim$, this implies that $[\mathcal{A}]_{(a_1, b_2)}$ is k -good, from which Lemma 16 entails that $[\mathcal{A}]_{(a, b)}$ is k -good. Since this property holds for every bounded segment of $[\mathcal{A}]_{(\alpha_E, \beta_E)}$, it follows from Lemma 15 that $[\mathcal{A}]_{(\alpha_E, \beta_E)}$ is k -good.

By merging together equivalent intervals belonging to I_1 , i.e., by computing the new set of intervals $I_2 = I_1 / \sim$ where $/$ denotes partition refinement, and removing from C_1 the points that are covered by intervals in I_2 , i.e., by setting $C_2 = C_1 \setminus \cup_{(a_i, b_i) \in I_2} (a_i, b_i)$, we obtain a partition Π_2 of $\mathbb{R}$ into

- a set I_2 of open intervals (a_i, b_i) , with $a_i \in \mathbb{R} \cup \{-\infty\}$, $b_i \in \mathbb{R} \cup \{+\infty\}$ and $a_i < b_i$ for either all $i \in \mathbb{N}$ or all $i \in [1, q]$ with $q > 0$. For each such interval, the segment $[\mathcal{A}]_{(a_i, b_i)}$ is k -good.
- a closed set C_2 of remaining points. This set does not include any non-empty open interval, i.e., C_2 is nowhere dense.

The intervals in I_2 cannot be adjacent, i.e., I_2 does not contain intervals (a_1, b_1) and (a_2, b_2) such that $b_1 = a_2$. Indeed, in such a case, the segment $[\mathcal{A}]_{(a_1, b_2)}$ would be k -good thanks to Lemma 13, and the interval (a_1, b_2) would belong to I_2 . (Recall that the intervals in I_2 are elements of a partition of $\mathbb{R}$, hence they are disjoint.)

We conclude the proof by showing that I_2 can only contain a single interval $(-\infty, +\infty)$, which implies that $\mathcal{A}$ is k -good. Assume, by contradiction, that I_2 contains at least two distinct intervals (a_1, b_1) and (a_2, b_2) . We can impose w.l.o.g. $b_1 < a_2$. For every such pair of intervals, we define

- the set $\mathcal{T}_{(a_1, b_1), (a_2, b_2)}$ of k -types of all the intervals $(\alpha, \beta) \in I_2$ such that $b_1 < \alpha < \beta < a_2$. This set cannot be empty, otherwise C_2 would include a non-empty open interval within (b_1, a_2) .
- the set $\mathcal{U}_{(a_1, b_1), (a_2, b_2)} = \{\{j \in [1, m] \mid p \in P_j\} \mid p \in C_2 \wedge b_1 < p < a_2\}$, that characterizes the combinations of values of the predicates that occur between b_1 and a_2 . This set cannot be empty, since $C_2 \cap (b_1, a_2) \neq \emptyset$. (Otherwise, (b_1, a_2) would belong to I_2 , which is impossible since this interval is adjacent to both (a_1, b_1) and (a_2, b_2) .)

We then perform the following operation. If there exist $(a_3, b_3) \in I_2$ and $(a_4, b_4) \in I_2$ such that $b_1 < a_3 < b_3 < a_4 < b_4 < a_2$, and

$$\mathcal{T}_{(a_3, b_3), (a_4, b_4)} \subsetneq \mathcal{T}_{(a_1, b_1), (a_2, b_2)} \text{ or } \mathcal{U}_{(a_3, b_3), (a_4, b_4)} \subsetneq \mathcal{U}_{(a_1, b_1), (a_2, b_2)},$$

we replace (a_1, b_1) by (a_3, b_3) and (a_2, b_2) by (a_4, b_4) , and repeat the same procedure. Note that, by Theorem 1, the set of possible k -types for the segments $[\mathcal{A}]_{(a_i, b_i)}$ with $(a_i, b_i) \in I_2$ is finite, hence the number of possible sets $\mathcal{T}_{(a_1, b_1), (a_2, b_2)}$ is finite. Furthermore, the number of possible sets $\mathcal{U}_{(a_1, b_1), (a_2, b_2)}$ is finite as well since m is fixed. This procedure will thus terminate in a finite number of steps. The final set $\mathcal{T}_{(a_1, b_1), (a_2, b_2)}$ will be non-empty, and will contain the k -types of some number $\ell > 0$ of sets $S_1, S_2, \dots, S_\ell$ of k -good $(0, m)$ -chains, corresponding to bounded segments $[\mathcal{A}]_{(\alpha, \beta)}$ for some intervals $(\alpha, \beta) \in I_2$. The final set $\mathcal{U}_{(a_1, b_1), (a_2, b_2)}$ will also be non-empty, and will contain some number $h > 0$ of sets $U_1, U_2, \dots, U_h \subseteq [1, m]$. One then has

$$[\mathcal{A}]_{(b_1, a_2)} \in \llbracket (U_1, S_1, U'_1), (U_2, S_2, U'_2), \dots, (U_\ell, S_\ell, U'_\ell), U''_1, U''_2, \dots, U''_h \rrbracket.$$

This implies that $[\mathcal{A}]_{(b_1, a_2)}$ is k -good, hence by Lemma 13 that $[\mathcal{A}]_{(a_1, b_2)}$ is k -good as well. We therefore have $(a_1, b_1) \sim (a_2, b_2)$, which contradicts the fact that (a_1, b_1) and (a_2, b_2) belong to different equivalence classes of $\sim$. $\square$

5.4 Wrap up

We have just established the following result.

Theorem 18. *For every (n, m) -chain $\mathcal{A}$, with $n, m \geq 0$, and $k \geq 0$, there exists a constructible (n, m) -chain $\mathcal{B}$ such that $\mathcal{A} \equiv_k \mathcal{B}$.*

This theorem admits the following important corollary.

Corollary 19. *A formula φ of $\text{MFO}(\mathbb{R}, <)$ with n free first-order and m free second-order variables, with $n, m \geq 0$, is satisfiable iff there exists a constructible (n, m) -chain $\mathcal{A}$ such that $\mathcal{A} \models \varphi$.*

Proof. One can apply Theorem 18 with k equal to the quantifier depth of the formula φ . $\square$

We are now ready to prove the main result of this paper.

Theorem 20. *Satisfiability is decidable for formulas in $\text{MFO}(\mathbb{R}, <)$.*

Proof. By Corollary 19, a semi-decision procedure can be obtained by enumerating all possible constructible chains, or more precisely the chain terms that we have introduced for describing sets of such chains, and checking for each of them whether the formula is satisfied.

To get full satisfiability, a criterion is needed for stopping this procedure at some point when the formula is not satisfiable. Let n and m be respectively the numbers of free first-order and second-order variables in the formula, and k be its quantifier depth. Using the results developed in Sections 4.2 and 4.3 together with Theorem 11, one can compute the set T of all possible k -types of constructible (n, m) -chains. This computation can be carried out by starting with the set of possible k -types of all constant and shuffle (n, m) -chains, and then recursively adding to this set the k -types of the chains that are obtained by applying reordering, concatenation, repetition or shuffle composition operators to chains contained in the set. Since the number of possible k -types of (n, m) -chains is finite by Theorem 1, this computation will eventually terminate. One can then check explicitly, for each $\tau \in T$, whether (n, m) -chains of k -type τ satisfy the formula, using the results of Section 4. $\square$

Conclusion

This paper compiles, revisits, recasts, and joins former results that build up a decision procedure for the monadic first-order theory of order over real numbers $\text{MFO}(\mathbb{R}, <)$. We believe the proof is self-contained enough to be a good basis for a formal theory mechanically verified within a proof assistant.

Essentially, decidability comes from the fact that a formula with k nested quantifiers, n variables, and m monadic predicates admits interpretations that split up in a finite number of classes called k -types for (n, m) -chains. These classes can be represented and generated explicitly thanks to a symbolic notation system for sets of interpretations. It is then possible to check each of these classes one by one for a model of the formula.

Our core motivation to revisit decision procedures for logic fragments incorporating arithmetic is that more insight into them may become useful to designing precise instantiation techniques and heuristics, notably for Satisfiability Modulo Theory (SMT) solvers. In the case of $\text{MFO}(\mathbb{R}, <)$, the k -types somehow embed the various ways to instantiate the k variables in the nested quantifiers. In the long term, this knowledge might lead to SMT instantiation techniques that complement the generic ones, and are more aware of interpreted symbols, notably the order predicate on $\mathbb{R}$.

The theoretical decision procedure presented in this paper is not realistically suited for a practical implementation. The results in [19] provide some inspiration for a more effective procedure following a similar approach, but the constraints on the set of considered formulas prevent a direct transposition of the algorithms. Our current research is focused on developing a reasonably practical decision procedure for $\text{MFO}(\mathbb{R}, <)$, based on advanced forms of automata [25]. An emptiness test over $\mathbb{R}$ for these automata is introduced in [26], and a first step towards universal first-order quantification of these automata is described in [27].

Acknowledgments

This research was partially supported in part by an Amazon Research Award, Fall 2022 CFP. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of Amazon. This work was also partially supported by the Fonds de la Recherche Scientifique - FNRS WEAVE under Grant number T.W028.23. Finally, the authors wish to thank the anonymous reviewers for their helpful and constructive comments.

ORCID

Bernard Boigelot: <https://orcid.org/0009-0009-4721-3824>

Pascal Fontaine: <https://orcid.org/0000-0003-4700-6031>

Baptiste Vergain: <https://orcid.org/0009-0003-5545-4579>

References

- [1] Büchi, J.R.: On a decision method in restricted second order arithmetic. In: Proc.

- Intl. Congr. on Logic, Methodology and Philosophy of Science, pp. 1–12 (1962)
- [2] Büchi, J.R.: The Monadic Second-order Theory of All Countable Ordinals. Decidable theories, vol. 2. Springer, Berlin, Heidelberg (1973)
 - [3] Gurevich, Y., Magidor, M., Shelah, S.: The monadic theory of ω_2 . The Journal of Symbolic Logic **48**(2), 387–398 (1983)
 - [4] Ehrenfeucht, A.: Decidability of the theory of linear ordering relation. Notices of the American mathematical society **6**(3), 268–269 (1959)
 - [5] Läuchli, H., Leonard, J.: On the elementary theory of linear order. Fundamenta Mathematicae **59**(1), 109–116 (1966)
 - [6] Fraïssé, R.: Sur l’extension aux relations de quelques propriétés des ordres. Annales scientifiques de l’École Normale Supérieure **71**(4), 363–388 (1954)
 - [7] Ehrenfeucht, A.: An application of games to the completeness problem for formalized theories. Fundamenta Mathematicae **49**(2), 129–141 (1961)
 - [8] Läuchli, H.: A decision procedure for the weak second-order theory of linear order. In: Contributions to Mathematical Logic. Studies in Logic and the Foundations of Mathematics, vol. 50, pp. 189–197. Elsevier, Amsterdam (1968)
 - [9] Shelah, S.: The monadic theory of order. Annals of Mathematics **102**(3), 379–419 (1975)
 - [10] Thomas, W.: Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In: Structures in Logic and Computer Science: A Selection of Essays in Honor of A. Ehrenfeucht. Lecture Notes in Computer Science, pp. 118–143 (1997)
 - [11] Rabin, M.O.: Decidability of second-order theories and automata on infinite trees. Transactions of the American mathematical society **141**, 1–35 (1969)
 - [12] Kamp, H.: Tense logic and the theory of linear order. PhD thesis, University of California at Los Angeles (1968)
 - [13] Rabinovich, A.: A proof of Kamp’s theorem. Logical Methods in Computer Science **10**(1) (2014)
 - [14] Gabbay, D., Pnueli, A., Shelah, S., Stavi, J.: On the temporal analysis of fairness. In: Proc. 7th ACM Symp. on Principles of Programming Languages (POPL), pp. 163–170. ACM Press, New-York (1980)
 - [15] Burgess, J.P., Gurevich, Y.: The decision problem for linear temporal logic. Notre Dame Journal of Formal Logic, 115–128 (1985)

- [16] Reynolds, M.: The complexity of the temporal logic with “until” over general linear time. *Journal of Computer and System Sciences* **66**(2), 393–426 (2003)
- [17] Reynolds, M.: The complexity of decision problems for linear temporal logics. *Journal of Studies in Logic* **3**(1), 19–50 (2010)
- [18] Reynolds, M.: The complexity of temporal logic over the reals. *Annals of Pure and Applied Logic* **161**(8), 1063–1096 (2010)
- [19] Rabinovich, A.: Temporal logics over linear time domains are in PSPACE. *Information and Computation* **210**, 40–67 (2012)
- [20] Barrett, C., Kroening, D., Melham, T.: Problem solving for the 21st century: Efficient solvers for satisfiability modulo theories. Technical Report 3, London Mathematical Society and Smith Institute for Industrial Mathematics and System Engineering (2014)
- [21] Barrett, C., Sebastiani, R., Seshia, S.A., Tinelli, C.: Satisfiability modulo theories. In: *Handbook of Satisfiability. Frontiers in Artificial Intelligence and Applications*, vol. 185, pp. 825–885. IOS Press, Amsterdam (2009). Chap. 26
- [22] Habermehl, P., Hecko, M., Havlena, V., Holík, L., Lengál, O.: Algebraic reasoning meets automata in solving linear integer arithmetic. In: *Proc. Intl. Conf. on Computer Aided Verification (CAV). Lecture Notes in Computer Science*, vol. 14681, pp. 42–67. Springer, Berlin, Heidelberg (2024)
- [23] Ebbinghaus, H.-D., Flum, J.: *Finite Model Theory*. Springer, Berlin, Heidelberg (1995)
- [24] Ramsey, F.P.: On a problem of formal logic. *Proceedings of the London Mathematical Society* **30**, 264–286 (1930)
- [25] Bruyère, V., Carton, O.: Automata on linear orderings. *Journal of Computer and System Sciences* **73**(1), 1–24 (2007)
- [26] Boigelot, B., Fontaine, P., Vergain, B.: Non-emptiness test for automata over words indexed by the reals and rationals. In: Fazekas, S.Z. (ed.) *Proc. Intl. Conf. on Implementation and Application of Automata (CIAA). Lecture Notes in Computer Science*, vol. 15015, pp. 94–108. Springer, Cham (2024)
- [27] Boigelot, B., Fontaine, P., Vergain, B.: Universal first-order quantification over automata. In: Nagy, B. (ed.) *Proc. Intl. Conf. on Implementation and Application of Automata (CIAA). Lecture Notes in Computer Science*, vol. 14151, pp. 91–102. Springer, Cham (2023)