FO = FO^3 for Linear Orders with Monotone Binary Relations

Marie Fortin
LSV, CNRS & ENS Paris-Saclay, Université Paris-Saclay, France
fortin@lsv.fr

Abstract

We show that over the class of linear orders with additional binary relations satisfying some monotonicity conditions, monadic first-order logic has the three-variable property. This generalizes (and gives a new proof of) several known results, including the fact that monadic first-order logic has the three-variable property over linear orders, as well as over $(\mathbb{R}, <, +1)$, and answers some open questions mentioned in a paper from Antonopoulos, Hunter, Raza and Worrell [FoSSaCS 2015]. Our proof is based on a translation of monadic first-order logic formulas into formulas of a star-free variant of Propositional Dynamic Logic, which are in turn easily expressible in monadic first-order logic with three variables.

2012 ACM Subject Classification Theory of computation → Logic

Keywords and phrases first-order logic, three-variable property, propositional dynamic logic

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.116

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version `https://arxiv.org/abs/1904.00189`

1 Introduction

Logics with a bounded number of variables have been extensively studied, in particular in the context of descriptive complexity [17, 18, 10, 21] and temporal logics [20, 7, 14, 16]. One recurring question of interest [7, 25, 19, 4, 26, 1] is to determine, in a given class $\mathcal{C}$ of structures, whether all properties expressible in monadic first-order logic (FO) can be defined in the fragment FO^k consisting of formulas which use at most k variables. (A same variable may be quantified over several times in a formula.) In fact, several non-equivalent versions of this question appear in the literature, many of which are compared in [15]. We say that $\mathcal{C}$ has the *k-variable property* if every formula of FO with at most k free variables is equivalent over $\mathcal{C}$ to a formula of FO^k. Note that this is strictly stronger than requiring that all *sentences* (without free variables) of FO are equivalent to some FO^k formulas. Indeed, Hodkinson and Simon gave an example of a class of structures where no sentence requires more than 3 variables, but which does not have the k-variable property for any k [15].

The problem of whether a given class of structures has the k-variable property is closely related to the question of the existence of an expressively complete temporal logic (with a finite set of FO-definable modalities). A temporal logic is called expressively complete if any first-order formula with a single free variable can be expressed in it. For instance, it is well-known that linear temporal logic (LTL) over Dedekind-complete time flows, or its extension with Stavi connectives over all time flows, are expressively complete for first-order logic [20, 8]. More recently, it was shown that over the real numbers equipped with binary relations $+q$ for all $q \in \mathbb{Q}$, metric temporal logic (MTL) is expressively complete [16]. However, the questions of having the k-variable property for some k or admitting an expressively complete temporal logic are incomparable in general: there exists a class of structures which admits a finite expressively complete set of temporal connectives but which does not have the k-variable

46th International Colloquium on Automata, Languages, and Programming (ICALP 2019).
Editors: Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi;
Article No. 116; pp. 116:1–116:13
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

property for any k [15], and there exists one which has the 3-variable property but for which no temporal logic is expressively complete [14]. However, Gabbay established that having the k-variable property implies the existence of a *multi-dimensional* expressively complete temporal logic, with multiple reference points [7].

Another classical approach to proving or disproving that a class of structures has the k-variable property is through Ehrenfeucht-Fraïssé games, with a bounded number of pebbles [12, 25, 19, 1]. Immerman and Kozen applied this to linear orders and bounded-degree trees [19], and Antonopoulos et al. to real-time signals [1].

Natural candidates for classes $\mathcal{C}$ which might have the k-variable property are classes of linearly ordered structures. Indeed, a typical counter-example to unrestricted structures having the k-variable property is a formula such as "there exists $k+1$ distinct elements which satisfy some predicate P". It is in general not expressible in FO^k, but it is easily expressible in FO^2 if all models are equipped with a linear order $<$. For instance for $k = 2$, we take the formula $\exists x.\ P(x) \wedge \exists y.(x < y \wedge P(y) \wedge \exists x.(y < x \wedge P(x)))$. As mentioned before, Immerman and Kozen showed that the class of linear orders has the 3-variable property [19]. However, adding a single binary relation suffices to obtain a class of linearly ordered structures which does not have the k-variable property for any k. Venema gave an example of a dense linear order with a single equivalence relation which does not have the k-variable property for any k [29]; this was adapted in [1] to give another example where the equivalence relation is replaced with a bijection. In fact, even for finite linear orders, Rossman [26] proved that the class of linearly ordered graphs does not have the k-variable property for any k, resolving a problem which had been open for more than 25 years [17]. Therefore, adding binary relations to linear orders while keeping the k-variable property requires some restrictions on the interpretation of the relation symbols.

On the positive side, Antonopoulos et al. proved that the class of structures over $(\mathbb{R}, <, +1)$ (or signals) has the 3-variable property [1]. Such structures have been studied in the context of real-time verification. As a corollary, they also showed that $(\mathbb{R}, <, f)$ has the 3-variable property for any linear function $f : x \mapsto ax + b$.

Contribution. We consider the class of linearly ordered structures with an additional (finite or infinite) number of binary *interval-preserving* relations. These are binary relations $\mathcal{R}$ such that, for all intervals I, any point which is in between two points of $\mathcal{R}(I)$ and has a preimage by $\mathcal{R}$ must have one in I. (We also require a symmetric condition of the converse relation $\mathcal{R}^{-1}$.) We show that FO over this class of structures also has the 3-variable property.

This generalizes results from [19] and [1] described above. Moreover, this answers some open questions mentioned in the conclusion of [1], which asked if the result could be extended from linear functions to polynomials over the reals, or other linear orders and families of monotone functions. In fact, all increasing or decreasing partial functions (over arbitrary linear orders) are special cases of interval-preserving relations, and thus covered by our result.

Our proof relies on different techniques than [19, 1], which were based on Ehrenfeucht-Fraïssé games. We give an effective translation from FO to FO^3 which goes through a star-free variant of Propositional Dynamic Logic (PDL) with converse. Propositional dynamic logic was introduced by Fischer and Ladner [6] to reason about program schemes, and has now found a large range of applications in artificial intelligence and verification [11, 5, 23, 22, 9]. It combines local formulas containing modal operators, and path formulas using the concatenation, union and Kleene star operations. Several extensions have been studied, including PDL with converse [27], intersection [3], or negation of atomic programs

[24]. The particular *star-free* variant of PDL we use here is in fact very similar to Tarski's relation algebras [28], which was used as a basis for formalizing set theory. It also corresponds to a *two-dimensional* temporal logic in the sense of Gabbay [7].

We applied similar proof techniques in [2], where we introduced a star-free variant of PDL and proved that it is equivalent to FO over *message sequence charts* (MSCs) (and thus obtained a 3-variable property result for MSCs as a corollary). MSCs are discrete partial orders which represent behaviors of concurrent message passing systems. They consist of a fixed, finite number of linear orders called process orders (one for each process in the system), together with FIFO binary message relations connecting matching send and receive actions. Having a (fixed) finite number of total orders instead of a single one is not an important difference, as we could always put them one after the other to extend them into a single linear order. FIFO relations are a special case of interval-preserving relations, thus the result of the present paper can in fact be seen as a strict generalization of our previous result in [2]. More importantly, a major difference between MSCs studied in [2] and the setting we consider here is that MSCs are *discrete* structures, whereas here we allow arbitrary linear orders. In fact, [2] relied on the definition of formulas describing the minimum or the maximum of some binary relations. As such, it is interesting to see that the same kind of techniques can still be applied to a priori very different linear orders.

Outline. In Section 2, we introduce interval-preserving relations and monadic first-order logic. In Section 3, we define star-free PDL, and prove some properties of its formulas. In Section 4, we give an effective translation from FO to star-free PDL, and explain its consequences. We conclude in Section 5.

2 Interval-preserving relations and first-order logic

In this section, we define the class of structures covered by our results, and recall the syntax of first-order logic.

Interval-preserving binary relations. Let $\mathcal{R} \subseteq A \times B$ be a binary relation between sets A and B. We write $a \mathrel{\mathcal{R}} b$ if $(a,b) \in \mathcal{R}$, and $\mathcal{R}(a) = \{b \in B \mid a \mathrel{\mathcal{R}} b\}$. For a subset $A' \subseteq A$, we also write $\mathcal{R}(A') = \bigcup_{a \in A'} \mathcal{R}(a)$. We define the *converse* of a relation $\mathcal{R}$ as $\mathcal{R}^{-1} = \{(b,a) \in B \times A \mid (a,b) \in \mathcal{R}\}$, and the *composition* of two binary relations $\mathcal{R}_1 \subseteq A \times B$ and $\mathcal{R}_2 \subseteq B \times C$ as $\mathcal{R}_1 \cdot \mathcal{R}_2 = \{(a,c) \in A \times C \mid \exists b \in B.\ (a,b) \in \mathcal{R}_1 \wedge (b,c) \in \mathcal{R}_2\}$. Finally, we write $\mathcal{R}^{\mathsf{c}} = (A \times B) \setminus \mathcal{R}$ for the *complement* of $\mathcal{R}$. Note that we have the following identities:

$$(\mathcal{R}_1 \cdot \mathcal{R}_2)^{-1} = {\mathcal{R}_2}^{-1} \cdot {\mathcal{R}_1}^{-1} \qquad (\mathcal{R}^{\mathsf{c}})^{-1} = \left(\mathcal{R}^{-1}\right)^{\mathsf{c}} \qquad (\mathcal{R}_1 \cap \mathcal{R}_2)^{-1} = \mathcal{R}_1^{-1} \cap \mathcal{R}_2^{-1}\,.$$

A linear order $\leq$ over a set A is a reflexive, transitive and antisymmetric relation $\leq \subseteq A \times A$ such that for all $a, b \in A$, we have $a \leq b$ or $b \leq a$. Let $(A, \leq)$ be a linearly ordered set. For $A' \subseteq A$, we also denote by $\leq$ the restriction of $\leq$ to A', so that $(A', \leq)$ is still a linearly ordered set. Moreover, for $a \in A$, we write $a < A'$ if for all $a' \in A'$, $a < a'$, and $A' < a$ if for all $a' \in A'$, $a' < a$.

An *interval* of $(A, \leq)$ is a set $I \subseteq A$ such that for all $a \leq b \leq c$ with $a, c \in I$, we have $b \in I$. For $a, b \in A$, we denote by $[a, b)$ the interval $\{c \in A \mid a \leq c < b\}$, and similarly for the intervals $[a,b]$, $(a,b]$, (a,b). We call a relation $\mathcal{R} \subseteq A \times B$ between two linearly ordered sets $(A, \leq_A)$ and $(B, \leq_B)$ *interval-preserving* if:

- For all intervals I of $(A, \leq_A)$, $\mathcal{R}(I)$ is an interval of $(\mathcal{R}(A), \leq_B)$.
- For all intervals J of $(B, \leq_B)$, $\mathcal{R}^{-1}(J)$ is an interval of $(\mathcal{R}^{-1}(B), \leq_A)$.

Figure 1 Definition of interval-preserving relations.

In other terms, for all $a_1 \mathrel{\mathcal{R}} b_1$ and $a_2 \mathrel{\mathcal{R}} b_2$ with $a_1, a_2 \in I$, for all $b_1 \leq_B b \leq_B b_2$, if there exists some $a \in A$ such that $a \mathrel{\mathcal{R}} b$, then there exists one in I (cf. Figure 1). Note that we do not require that *all* elements between b_1 and b_2 are in $\mathcal{R}(I)$, but only those which are in the image of $\mathcal{R}$. The second condition is symmetric: for all $a_1 \mathrel{\mathcal{R}} b_1$ and $a_2 \mathrel{\mathcal{R}} b_2$ with $b_1, b_2 \in J$, for all $a_1 \leq_A a \leq_A a_2$, if there exists some $b \in B$ such that $a \mathrel{\mathcal{R}} b$, then there exists one in J.

▶ **Example 1.** For any linear order $(A, \leq)$ and partial function $f : A \to A$, if f is increasing or decreasing then the relation $\{(a, f(a)) \mid a \in \mathrm{dom}(f)\}$ is interval-preserving.

As another example, consider a temporal structure $(A, \leq, \lambda)$ over a set of atomic propositions AP, where $\lambda : A \to 2^{\mathrm{AP}}$ indicates the set of propositions which are true at a given point. For $P, Q \in \mathrm{AP}$, we let $\mathsf{until}_{P,Q} = \{(a, b) \in A \times A \mid a < b \wedge Q \in \lambda(b) \wedge \forall a < c < b, P \in \lambda(c)\}$. Then $\mathsf{until}_{P,Q}$ is interval-preserving.

The following lemma states some simple closure properties of interval-preserving relations.

▶ **Lemma 2.** *Let $(A, \leq_A)$, $(B, \leq_B)$, $(C, \leq_C)$ be linearly ordered sets.*
1. *For all interval-preserving relation $\mathcal{R} \subseteq A \times B$, $\mathcal{R}^{-1}$ is interval-preserving.*
2. *For all interval-preserving relations $\mathcal{R}_1, \mathcal{R}_2 \subseteq A \times B$, $\mathcal{R}_1 \cap \mathcal{R}_2$ is interval-preserving.*
3. *For all interval-preserving relations $\mathcal{R}_1 \subseteq A \times B$ and $\mathcal{R}_2 \subseteq B \times C$, $\mathcal{R}_1 \cdot \mathcal{R}_2$ is interval-preserving.*

Proof. Part 1 follows from the fact that $(\mathcal{R}^{-1})^{-1} = \mathcal{R}$.

Let us prove 2. Since $(\mathcal{R}_1 \cap \mathcal{R}_2)^{-1} = \mathcal{R}_1^{-1} \cap \mathcal{R}_2^{-1}$, by symmetry, it suffices to prove that for all interval I of $(A, \leq)$, $(\mathcal{R}_1 \cap \mathcal{R}_2)(I)$ is an interval of $((\mathcal{R}_1 \cap \mathcal{R}_2)(A), \leq)$. Let $a_1, a_2 \in I$ and $b_1 \leq b \leq b_2$ such that $(a_1, b_1), (a_2, b_2) \in (\mathcal{R}_1 \cap \mathcal{R}_2)$ and $(a, b) \in (\mathcal{R}_1 \cap \mathcal{R}_2)$ for some $a \in A$. If $a \in I$, then we are done. Otherwise, suppose for instance that $a < a_1 \leq a_2$ (the other cases are similar). Since $\mathcal{R}_1$ is interval-preserving, there exists $a_1 \leq a' \leq a_2$ such that $a' \mathrel{\mathcal{R}_1} b$. Then, since $a < a_1 \leq a'$ and $\mathcal{R}_1^{-1}(b)$ is an interval of $(\mathcal{R}_1^{-1}(B), \leq_A)$, we obtain $a_1 \mathrel{\mathcal{R}_1} b$. Similarly, $a_1 \mathrel{\mathcal{R}_2} b$. Hence $a_1 \mathrel{(\mathcal{R}_1 \cap \mathcal{R}_2)} b$.

Let us show that 2 implies 3. Again, by symmetry, it suffices to prove that for all interval I of $(A, \leq_A)$, $(\mathcal{R}_1 \cdot \mathcal{R}_2)(I)$ is an interval of $((\mathcal{R}_1 \cdot \mathcal{R}_2)(A), \leq_C)$. Let $\mathcal{R}_3 \subseteq B \times C$ denote the relation $\mathcal{R}_1(A) \times C$. It is an interval-preserving relation between $(B, \leq_B)$ and $(C, \leq_C)$. Moreover, we have $(\mathcal{R}_1 \cdot \mathcal{R}_2)(A) = (\mathcal{R}_2 \cap \mathcal{R}_3)(B)$. Now, let I be some interval of $(A, \leq_A)$, and $J = \{b \in B \mid \exists b_1, b_2 \in \mathcal{R}_1(I), b_1 \leq b \leq b_2\}$. Then J is an interval of $(B, \leq_B)$. Moreover, since $\mathcal{R}_1$ is interval-preserving, we have $\mathcal{R}_1(I) = J \cap \mathcal{R}_1(A)$, hence

$$(\mathcal{R}_1 \cdot \mathcal{R}_2)(I) = \mathcal{R}_2(\mathcal{R}_1(I)) = \mathcal{R}_2(J \cap \mathcal{R}_1(A)) = (\mathcal{R}_2 \cap \mathcal{R}_3)(J)\,.$$

Then, according to 2, $(\mathcal{R}_1 \cdot \mathcal{R}_2)(I)$ is an interval of $((\mathcal{R}_2 \cap \mathcal{R}_3)(B), \leq_C)$, i.e., an interval of $((\mathcal{R}_1 \cdot \mathcal{R}_2)(A), \leq_C)$. ◀

Models. Let $\mathcal{P} = \{P, Q, \ldots\}$ be an infinite set of monadic predicates, and $\Gamma = \{\alpha, \beta, \ldots\}$ be a finite or infinite set of binary relation symbols. Throughout the paper, $\mathcal{M}$ will denote a structure $\mathcal{M} = (A, \leq, (\alpha^{\mathcal{M}})_{\alpha \in \Gamma}, (P^{\mathcal{M}})_{P \in \mathcal{P}})$ where $\leq$ is a linear order over A, $\alpha^{\mathcal{M}} \subseteq A \times A$ is an interval-preserving relation for all $\alpha \in \Gamma$, and $P^{\mathcal{M}} \subseteq A$ for all $P \in \mathcal{P}$.

Monadic first-order logic. We assume an infinite supply of variables $\mathcal{X} = \{x, y, \ldots\}$. The set $\mathrm{FO}[\Gamma, \leq]$ of monadic first-order logic formulas over Γ is defined as follows:

$$\Phi ::= P(x) \mid x \leq y \mid x = y \mid \alpha(x, y) \mid \Phi \vee \Phi \mid \neg\Phi \mid \exists x.\Phi\,, \qquad \text{where } x, y \in \mathcal{X}, P \in \mathcal{P}, \alpha \in \Gamma\,.$$

We assume that all formulas are interpreted over structures $\mathcal{M}$ defined as above. Given an $\mathrm{FO}[\Gamma, \leq]$ formula Φ, we denote by $\mathsf{Free}(\Phi)$ its set of free variables. We define the satisfaction relation $\mathcal{M}, \nu \models \Phi$ as usual, where $\mathcal{M} = (A, \leq, (\alpha^{\mathcal{M}})_{\alpha\in\Gamma}, (P^{\mathcal{M}})_{P\in\mathcal{P}})$ and $\nu : \mathsf{Free}(\Phi) \to A$ is an interpretation of the free variables of Φ. We say that two formulas $\Phi, \Psi \in \mathrm{FO}[\Gamma, \leq]$ are *equivalent*, written $\Phi \equiv \Psi$, if for all $\mathcal{M} = (A, \leq, (\alpha^{\mathcal{M}})_{\alpha\in\Gamma}, (P^{\mathcal{M}})_{P\in\mathcal{P}})$ and $\nu : \mathsf{Free}(\Phi) \cup \mathsf{Free}(\Psi) \to A$, we have $\mathcal{M}, \nu|_{\mathsf{Free}(\Phi)} \models \Phi$ if and only if $\mathcal{M}, \nu|_{\mathsf{Free}(\Psi)} \models \Psi$.

For $k \in \mathbb{N}$, we denote by $\mathrm{FO}^k[\Gamma, \leq]$ the set of first-order formulas with at most k variables. Note that a same variable may be quantified over several times in the formula.

▶ **Example 3.** Let $p : \mathbb{R} \to \mathbb{R}$ be a polynomial function, and $m_1 < \cdots < m_n$ its local extrema (we suppose that $n \geq 1$). Fix $\Gamma = \{p\}$. For convenience, we write $p(x) = y$ instead of $p(x, y)$ in $\mathrm{FO}[\Gamma, \leq]$ formulas. We focus on models of the form $\mathcal{M} = (\mathbb{R}, \leq, p^{\mathcal{M}}, (P^{\mathcal{M}})_{P\in\mathcal{P}})$ where $\leq$ is the usual ordering of the reals, and $p^{\mathcal{M}} = \{(x, p(x)) \mid x \in \mathbb{R}\}$. Let us describe an $\mathrm{FO}^3[\Gamma, \leq]$ formula $m_i \leq x$ such that for all $\mathcal{M}$ and $r \in \mathbb{R}$, we have $\mathcal{M}, [x \mapsto r] \models m_i \leq x$ if and only if $m_i \leq r$. First, we write $p(x) \leq p(y)$ for the $\mathrm{FO}^3[\Gamma, \leq]$ formula

$$\exists z.\, p(x) = z \wedge \exists x.\, (p(y) = x \wedge z \leq x)\,.$$

We can then define formulas $\min(x) \in \mathrm{FO}^3[\Gamma, \leq]$ and $\max(x) \in \mathrm{FO}^3[\Gamma, \leq]$ which state that x is a local minimum (resp. maximum) of p, for instance:

$$\begin{aligned}\min(x) = \; & (\exists z.\, z < x \wedge \forall y.\, (z < y \leq x \implies p(x) \leq p(y))) \wedge \\ & (\exists z.\, x < z \wedge \forall y.\, (x \leq y < z \implies p(x) \leq p(y)))\;.\end{aligned}$$

The formula $m_i \leq x$ then states that there exist at least i local extrema before x, alternating existential quantifications over y and z to identify them; for instance, $m_3 \leq x$ is the formula

$$\exists y.\, y \leq x \wedge (\min(y) \vee \max(y)) \wedge \exists z.\, z < y \wedge (\min(z) \vee \max(z)) \wedge \exists y.\, y < z \wedge (\min(y) \vee \max(y))\,.$$

3 Star-free Propositional Dynamic Logic

Star-free Propositional Dynamic Logic. Propositional dynamic logic (PDL) [6] consists of two sorts of formulas: state formulas which are evaluated at single elements, and path formulas which are evaluated at pairs of elements and allow to navigate inside the model. Here we consider a star-free variant of PDL (with converse). The syntax of *star-free propositional dynamic logic* over Γ, written $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$, is given below:

$$\begin{aligned} \varphi &::= P \mid \varphi \vee \varphi \mid \neg\varphi \mid \langle\pi\rangle\,\varphi && \text{(state formulas)} \\ \pi &::= \alpha \mid \leq \mid \{\varphi\}? \mid \pi^{-1} \mid \pi \cdot \pi \mid \pi \cup \pi \mid \pi \cap \pi \mid \pi^{\mathsf{c}} && \text{(path formulas)} \end{aligned}$$

where $P \in \mathcal{P}$ and $\alpha \in \Gamma$.

Compared to classical PDL, star-free PDL uses the operators $(\cdot, \cup, \cap, \mathsf{c})$ of star-free expressions, instead of the rational operators $(\cdot, \cup, *)$.

Let $\mathcal{M} = (A, \leq, (\alpha^{\mathcal{M}})_{\alpha\in\Gamma}, (P^{\mathcal{M}})_{P\in\mathcal{P}})$. The semantics $[\![\varphi]\!]^{\mathcal{M}} \subseteq A$ or $[\![\pi]\!]^{\mathcal{M}} \subseteq A \times A$ of a state or path formula in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ is defined below. The state formula $\langle\pi\rangle\,\varphi$ is true at a point $a \in A$ in $\mathcal{M}$ (that is, $a \in [\![\langle\pi\rangle\,\varphi]\!]^{\mathcal{M}}$) if there exists some $b \in A$ such that (a, b) satisfies

π and φ is true at b. The path formula $\{\varphi\}?$ is stationary and tests if the state formula φ is true. The semantics of other formulas is straightforward:

$$
\begin{aligned}
[\![P]\!]^{\mathcal{M}} &:= P^{\mathcal{M}} & [\![\varphi_1 \vee \varphi_2]\!]^{\mathcal{M}} &:= [\![\varphi_1]\!]^{\mathcal{M}} \cup [\![\varphi_2]\!]^{\mathcal{M}} \\
[\![\neg\varphi]\!]^{\mathcal{M}} &:= A \setminus [\![\varphi]\!]^{\mathcal{M}} & [\![\langle\pi\rangle\, \varphi]\!]^{\mathcal{M}} &:= \{a \in A \mid \exists b \in [\![\varphi]\!]^{\mathcal{M}},\ (a,b) \in [\![\pi]\!]^{\mathcal{M}}\} \\
[\![\alpha]\!]^{\mathcal{M}} &:= \alpha^{\mathcal{M}} & [\![\{\varphi\}?]\!]^{\mathcal{M}} &:= \{(a,a) \mid a \in [\![\varphi]\!]^{\mathcal{M}}\} \\
[\![\leq]\!]^{\mathcal{M}} &:= \leq & [\![\pi^{-1}]\!]^{\mathcal{M}} &:= \left([\![\pi]\!]^{\mathcal{M}}\right)^{-1} \\
[\![\pi_1 \cup \pi_2]\!]^{\mathcal{M}} &:= [\![\pi_1]\!]^{\mathcal{M}} \cup [\![\pi_2]\!]^{\mathcal{M}} & [\![\pi_1 \cap \pi_2]\!]^{\mathcal{M}} &:= [\![\pi_1]\!]^{\mathcal{M}} \cap [\![\pi_2]\!]^{\mathcal{M}} \\
[\![\pi^{\mathsf{c}}]\!]^{\mathcal{M}} &:= (A \times A) \setminus [\![\pi]\!]^{\mathcal{M}} & [\![\pi_1 \cdot \pi_2]\!]^{\mathcal{M}} &:= [\![\pi_1]\!]^{\mathcal{M}} \cdot [\![\pi_2]\!]^{\mathcal{M}} .
\end{aligned}
$$

For simplicity, we often write $[\![\varphi]\!]$ or $[\![\pi]\!]$ instead of $[\![\varphi]\!]^{\mathcal{M}}$ and $[\![\pi]\!]^{\mathcal{M}}$. We also write $\mathcal{M}, a \models \varphi$ if $a \in [\![\varphi]\!]^{\mathcal{M}}$, and $\mathcal{M}, a, b \models \pi$ if $(a,b) \in [\![\pi]\!]^{\mathcal{M}}$.

We use the abbreviations $\mathit{true} := P \vee \neg P$, $\mathit{false} := \neg\mathit{true}$, $\geq := (\leq)^{-1}$, $< := \geq^{\mathsf{c}}$, $> := \leq^{\mathsf{c}}$ and $\langle\pi\rangle := \langle\pi\rangle\, \mathit{true}$. For all $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ formulas π, we also define a state formula $\mathsf{loop}(\pi) := \langle \pi \cap \{\mathit{true}\}? \rangle$ which holds at a if and only if $(a,a) \in [\![\pi]\!]$.

► **Example 4.** Suppose that $\Gamma = \{+q \mid q \in \mathbb{Q}\}$, and that we consider only models over $\mathbb{R}$ and with $[\![+q]\!] = \{(r, r+q) \mid r \in \mathbb{R}\}$. Let $q, r \in \mathbb{Q}_{\geq 0}$ and $P, Q \in \mathcal{P}$. The formula $P \,\mathsf{U}_{(q,r)}\, Q$ of metric temporal logic, which holds at time $t \in \mathbb{R}$ if there exists $t + q < t' < t + r$ such that $t' \in [\![Q]\!]$ and for all $t < t'' < t'$, $t'' \in [\![P]\!]$, can be expressed in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ as follows:

$$P \,\mathsf{U}_{(q,r)}\, Q \equiv \left\langle (+q \cdot <) \cap (+r \cdot >) \cap (< \cdot \{\neg P\}? \cdot <)^{\mathsf{c}} \right\rangle Q\,.$$

An interval-preserving fragment of star-free PDL. We say that a path formula $\pi \in \mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ is *interval-preserving* if for all $\mathcal{M}$, $[\![\pi]\!]^{\mathcal{M}}$ is interval-preserving. Notice that $\leq$ and $\{\varphi\}?$ (for all φ) are interval-preserving. By Lemma 2 (and assumption on $[\![\alpha]\!]$), all $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ formulas constructed without the boolean operators $\cup$ and c are interval-preserving. However, the complement or the union of interval-preserving relations are not in general interval-preserving. We define below a fragment of $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ where all path formulas are interval-preserving, and which will turn out to be as expressive as $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ (and in fact, $\mathrm{FO}[\Gamma, \leq]$) when it comes to *state* formulas. To do so, we introduce several restrictions of π^{c} which are interval-preserving, and which suffice to characterize π^{c}.

Let us first look at the different reasons for which we may have $(a,b) \in [\![\pi^{\mathsf{c}}]\!]$, assuming that π is interval-preserving. To begin with, we focus on a. One first sufficient condition for having $b \notin [\![\pi]\!](a)$ is that $[\![\pi]\!](a) = \emptyset$. Now, suppose $[\![\pi]\!](a) \neq \emptyset$. If π is interval-preserving, there are only three possible cases in which $b \notin [\![\pi]\!](a)$: $b < [\![\pi]\!](a)$, or $[\![\pi]\!](a) < b$, or $[\![\pi^{-1}]\!](b) = \emptyset$. We define formulas $\mathsf{left}\ \pi$ and $\mathsf{right}\ \pi$ corresponding respectively to the first two cases. We let

$$
\begin{aligned}
\mathsf{left}\ \pi &= \{\langle\pi\rangle\}? \cdot (\pi \cdot \leq)^{\mathsf{c}}, \quad \text{i.e.} & (a,b) &\in [\![\mathsf{left}\ \pi]\!] \quad \text{iff} \quad b < [\![\pi]\!](a) \neq \emptyset \\
\mathsf{right}\ \pi &= \{\langle\pi\rangle\}? \cdot (\pi \cdot \geq)^{\mathsf{c}}, \quad \text{i.e.} & (a,b) &\in [\![\mathsf{right}\ \pi]\!] \quad \text{iff} \quad b > [\![\pi]\!](a) \neq \emptyset\,.
\end{aligned}
$$

Now, if we look at $[\![\pi^{-1}]\!](b)$ instead of $[\![\pi]\!](a)$, we can make the same observations, by symmetry: we have $(a,b) \in [\![\pi^{\mathsf{c}}]\!]$ if and only if $a \notin [\![\pi^{-1}]\!](b)$, and if π is interval-preserving, there are again only four possible cases: $[\![\pi^{-1}]\!](b) = \emptyset$, or $a < [\![\pi^{-1}]\!](b)$, or $a > [\![\pi^{-1}]\!](b)$, or $[\![\pi]\!](a) = \emptyset$.

Unfortunately, the formulas $\mathsf{left}\ \pi$ and $\mathsf{right}\ \pi$ are still not interval-preserving in general. However, if we take a more symmetric restriction of π^{c}, where we look at all the possible positions of b and a relatively to $[\![\pi]\!](a)$ and $[\![\pi^{-1}]\!](b)$, we obtain four cases, illustrated in Figure 2, which we will later show correspond to interval-preserving restrictions of π^{c}.

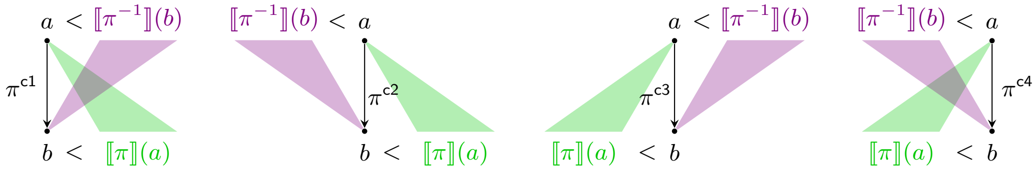


Figure 2 Definition of π^{c1}, π^{c2}, π^{c3} and π^{c4}, from left to right.

More precisely, let

$$\pi^{\mathsf{c1}} := \mathsf{left}\ \pi \cap \left(\mathsf{left}\ \left(\pi^{-1}\right)\right)^{-1}, \text{i.e.} \qquad (a,b) \in [\![\pi^{\mathsf{c1}}]\!] \quad \text{iff} \quad \begin{cases} b < [\![\pi]\!]\,(a) \neq \emptyset & \text{and} \\ a < [\![\pi^{-1}]\!]\,(b) \neq \emptyset \end{cases}$$

$$\pi^{\mathsf{c2}} := \mathsf{left}\ \pi \cap \left(\mathsf{right}\ \left(\pi^{-1}\right)\right)^{-1}, \text{i.e.} \qquad (a,b) \in [\![\pi^{\mathsf{c2}}]\!] \quad \text{iff} \quad \begin{cases} b < [\![\pi]\!]\,(a) \neq \emptyset & \text{and} \\ a > [\![\pi^{-1}]\!]\,(b) \neq \emptyset \end{cases}$$

$$\pi^{\mathsf{c3}} := \mathsf{right}\ \pi \cap \left(\mathsf{left}\ \left(\pi^{-1}\right)\right)^{-1}, \text{i.e.} \qquad (a,b) \in [\![\pi^{\mathsf{c3}}]\!] \quad \text{iff} \quad \begin{cases} b > [\![\pi]\!]\,(a) \neq \emptyset & \text{and} \\ a < [\![\pi^{-1}]\!]\,(b) \neq \emptyset \end{cases}$$

$$\pi^{\mathsf{c4}} := \mathsf{right}\ \pi \cap \left(\mathsf{right}\ \left(\pi^{-1}\right)\right)^{-1}, \text{i.e.} \qquad (a,b) \in [\![\pi^{\mathsf{c4}}]\!] \quad \text{iff} \quad \begin{cases} b > [\![\pi]\!]\,(a) \neq \emptyset & \text{and} \\ a > [\![\pi^{-1}]\!]\,(b) \neq \emptyset\,. \end{cases}$$

Notice that $\pi^{\mathsf{c3}} \equiv \left((\pi^{-1})^{\mathsf{c2}}\right)^{-1}$.

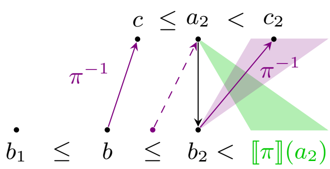
Let $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \cap, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ be the following restriction of $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$:

$$\varphi ::= P \mid \varphi \vee \varphi \mid \neg\varphi \mid \langle\pi\rangle\,\varphi$$
$$\pi ::= \alpha \mid \leq \mid \{\varphi\}? \mid \pi^{-1} \mid \pi \cdot \pi \mid \pi \cap \pi \mid \pi^{\mathsf{c1}} \mid \pi^{\mathsf{c2}} \mid \pi^{\mathsf{c3}} \mid \pi^{\mathsf{c4}}\,.$$

▶ **Lemma 5.** *All* $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \cap, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ *formulas are interval-preserving.*

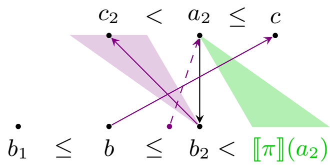
Proof. We proceed by induction on the formula. By assumption, α is interval-preserving for all $\alpha \in \Gamma$. Moreover, $\leq$ and $\{\varphi\}?$ are interval-preserving. For π^{-1}, $\pi_1 \cdot \pi_2$ and $\pi_1 \cap \pi_2$, we apply Lemma 2.

Suppose that π is interval-preserving. Let us show that π^{c1} is interval-preserving. Notice that $(\pi^{\mathsf{c1}})^{-1} \equiv (\pi^{-1})^{\mathsf{c1}}$. So we only need to show that for all intervals I, for all $b_1, b_2 \in [\![\pi^{\mathsf{c1}}]\!](I)$ and $b_1 \leq b \leq b_2$ such that $[\![(\pi^{\mathsf{c1}})^{-1}]\!](b) \neq \emptyset$, there exists $a \in I$ such that $(a,b) \in [\![\pi^{\mathsf{c1}}]\!]$. Let $a_2 \in I$ such that $(a_2, b_2) \in [\![\pi^{\mathsf{c1}}]\!]$. Let us show that we can in fact take $a = a_2$. The proof is illustrated in the picture below.



First, we have $b \leq b_2 < [\![\pi]\!](a_2) \neq \emptyset$. Moreover, $[\![\pi^{-1}]\!](b) \neq \emptyset$ (since $[\![(\pi^{\mathsf{c1}})^{-1}]\!](b) \neq \emptyset$). Now, suppose towards a contradiction that $a_2 \not< [\![\pi^{-1}]\!](b)$. Let $c \in [\![\pi^{-1}]\!](b)$ such that $c \leq a_2$. Since $(a_2, b_2) \in [\![\pi^{\mathsf{c1}}]\!]$, there exists $c_2 > a_2$ such that $(b_2, c_2) \in [\![\pi^{-1}]\!]$. We then have $c \leq a_2 < c_2$ and $[\![\pi]\!](a_2) \neq \emptyset$. Since π is interval-preserving, we obtain $a_2 \in [\![\pi^{-1}]\!]([b, b_2])$, a contradiction with the fact that $b_2 < [\![\pi]\!](a_2)$. Thus, $(a_2, b) \in [\![\pi^{\mathsf{c1}}]\!]$.

Let us show that π^{c2} is also interval-preserving. Similarly to the previous case, we show that for all $(a_2, b_2) \in [\![\pi^{\mathsf{c2}}]\!]$ and $b \leq b_2$ such that $[\![(\pi^{\mathsf{c2}})^{-1}]\!](b) \neq \emptyset$, we have $(a_2, b) \in [\![\pi^{\mathsf{c2}}]\!]$.



First, $b \leq b_2 < [\![\pi]\!](a_2) \neq \emptyset$, and $[\![\pi^{-1}]\!](b) \neq \emptyset$. Suppose towards a contradiction that $[\![\pi^{-1}]\!](b) \not< a_2$. Let $c \in [\![\pi^{-1}]\!](b)$ such that $a_2 \leq c$, and $c_2 \in [\![\pi^{-1}]\!](b_2)$. We have $c_2 < a_2 \leq c$, and $[\![\pi]\!](a_2) \neq \emptyset$. Since π is interval-preserving, we obtain $a_2 \in [\![\pi^{-1}]\!]([b, b_2])$, a contradiction with the fact that $b_2 < [\![\pi]\!](a_2)$. Symmetrically, let J be an interval, $a_1, a_2 \in [\![(\pi^{\mathsf{c2}})^{-1}]\!](J)$, and $a_1 \leq a \leq a_2$ such that $[\![\pi^{\mathsf{c2}}]\!](a) \neq \emptyset$. Then for any $b_1 \in J$ such that $(a_1, b_1) \in [\![\pi^{\mathsf{c2}}]\!]$, we also have $(a, b_1) \in [\![\pi^{\mathsf{c2}}]\!]$, hence $a \in [\![(\pi^{\mathsf{c2}})^{-1}]\!](J)$.

Since $\pi^{\mathsf{c3}} \equiv \left((\pi^{-1})^{\mathsf{c2}}\right)^{-1}$, this also implies that π^{c3} is interval-preserving.

Finally, the case of π^{c4} is symmetric to the case of π^{c1}: for all $(a_1, b_1) \in [\![\pi^{\mathsf{c4}}]\!]$ and $b_1 \leq b$ such that $[\![(\pi^{\mathsf{c4}})^{-1}]\!](b) \neq \emptyset$, we have $(a_1, b) \in [\![\pi^{\mathsf{c4}}]\!]$. ◀

4 Star-free PDL is expressively equivalent to FO

Let φ be a state formula in $\mathrm{PDL_{sf}}[\Gamma, \leq]$, and $\Phi(x)$ an $\mathrm{FO}[\Gamma, \leq]$ formula with a single free variable x. We say that φ and Φ are equivalent, written $\varphi \equiv \Phi(x)$, if for all $\mathcal{M}$ and elements a in $\mathcal{M}$, we have $\mathcal{M}, a \models \varphi$ if and only if $\mathcal{M}, [x \mapsto a] \models \Phi(x)$. Similarly, for a path formula $\pi \in \mathrm{PDL_{sf}}[\Gamma, \leq]$ and an $\mathrm{FO}[\Gamma, \leq]$ formula $\Phi(x, y)$ with exactly two free variables x and y, we write $\pi \equiv \Phi(x, y)$ if for all $\mathcal{M}$ and elements a, b in $\mathcal{M}$, we have $\mathcal{M}, a, b \models \pi$ if and only if $\mathcal{M}, [x \mapsto a, y \mapsto b] \models \Phi(x, y)$.

From $\mathrm{PDL_{sf}}[\Gamma, \leq]$ to $\mathrm{FO}^3[\Gamma, \leq]$. An easy induction shows that any formula in $\mathrm{PDL_{sf}}[\Gamma, \leq]$ can be translated into an $\mathrm{FO}[\Gamma, \leq]$ formula which uses at most three distinct variables:

▶ **Lemma 6.** *For every state formula $\varphi \in \mathrm{PDL_{sf}}[\Gamma, \leq]$, there exists a formula $\widetilde{\varphi}(x) \in \mathrm{FO}^3[\Gamma, \leq]$ such that $\varphi \equiv \widetilde{\varphi}(x)$. For every path formula $\pi \in \mathrm{PDL_{sf}}[\Gamma, \leq]$, there exists a formula $\widetilde{\pi}(x, y) \in \mathrm{FO}^3[\Gamma, \leq]$ such that $\pi \equiv \widetilde{\pi}(x, y)$.*

For the other direction, we will see that the fragment $\mathrm{PDL_{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ of $\mathrm{PDL_{sf}}[\Gamma, \leq]$ defined below is sufficient:

$$\varphi ::= P \mid \varphi \vee \varphi \mid \neg\varphi \mid \langle\pi\rangle\, \varphi \mid \mathsf{loop}(\pi)$$
$$\pi ::= \alpha \mid \leq \mid \{\varphi\}? \mid \pi^{-1} \mid \pi \cdot \pi \mid \pi^{\mathsf{c1}} \mid \pi^{\mathsf{c2}} \mid \pi^{\mathsf{c3}} \mid \pi^{\mathsf{c4}}\,.$$

This fragment is a restriction of $\mathrm{PDL_{sf}}[\Gamma, \leq, \cap, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$, where the intersection is only used for $\mathsf{loop}(\pi)$ formulas.

From $\mathrm{FO}[\Gamma, \leq]$ to $\mathrm{PDL_{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. The main result of the paper is an effective translation of $\mathrm{FO}[\Gamma, \leq]$ formulas into positive boolean combinations of formulas in $\mathrm{PDL_{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$:

▶ **Theorem 7.** *Every formula $\Phi \in \mathrm{FO}[\Gamma, \leq]$ with at least one free variable is equivalent to a positive boolean combination of formulas of the form $\widetilde{\pi}(x, y)$, where $x, y \in \mathsf{Free}(\Phi)$ and $\pi \in \mathrm{PDL_{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$.*

Note that the equivalent formula may also contain subformulas of the form $\widetilde{\pi}(x, x)$.

Before proving Theorem 7, we state some of its consequences.

▶ **Corollary 8.** *Every formula $\Phi \in \mathrm{FO}[\Gamma, \leq]$ with a single free variable is equivalent to some $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ state formula. Every formula $\Phi \in \mathrm{FO}[\Gamma, \leq]$ with two free variables is equivalent to some $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ path formula.*

Proof. If Φ has a single free variable x, it is equivalent to a positive boolean combination of formulas of the form $\widetilde{\pi}(x, x)$, which are themselves equivalent to the formulas $\mathsf{loop}(\pi)$. The combination of these $\mathsf{loop}(\pi)$ formulas is then a state formula of $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$.

If Φ has two free variables x and y, we obtain an equivalent positive boolean combination of formulas of the form $\widetilde{\pi}(x, y)$, $\widetilde{\pi}(y, x)$, $\widetilde{\pi}(x, x)$, or $\widetilde{\pi}(y, y)$. We can replace any subformula $\widetilde{\pi}(y, x)$ with $\widetilde{\pi^{-1}}(x, y)$, and any subformula $\widetilde{\pi}(x, x)$ with $\widetilde{\pi_1}(x, y) \vee \widetilde{\pi_2}(x, y)$, where $\pi_1 = (\{\mathsf{loop}(\pi)\}? \cdot \leq)$ and $\pi_2 = (\{\mathsf{loop}(\pi)\}? \cdot \geq)$, and similarly for formulas $\widetilde{\pi}(y, y)$. We obtain an equivalent positive boolean combination of formulas of the form $\widetilde{\pi}(x, y)$. Since $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ allows union and intersection of path formulas, this is equivalent to a $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ formula. ◀

Another consequence is that $\mathrm{FO}[\Gamma, \leq]$ over linear orders with interval-preserving relations has the three-variable property. More precisely:

▶ **Theorem 9.** *Any $\mathrm{FO}[\Gamma, \leq]$ formula is equivalent to a boolean combination of formulas in $\mathrm{FO}^3[\Gamma, \leq]$.*

This also allows us to answer an open question from [1], namely, whether structures over the real numbers with polynomial functions have the 3-variable property. Suppose that Γ is a set of polynomials $p : \mathbb{R} \to \mathbb{R}$. Let $\mathcal{M}_\Gamma = (\mathbb{R}, \leq, (p^{\mathcal{M}_\Gamma})_{p \in \Gamma})$, where $\leq$ is the usual ordering of the real numbers, and $p^{\mathcal{M}_\Gamma} = \{(x, p(x)) \mid x \in \mathbb{R}\}$ for all $p \in \Gamma$. Given an interpretation $h : \mathcal{P} \to 2^{\mathbb{R}}$ of the monadic predicates, we denote by $(\mathcal{M}_\Gamma, h)$ the structure $(\mathbb{R}, \leq, (p^{\mathcal{M}_\Gamma})_{p \in \Gamma}, (h(P))_{P \in \mathcal{P}})$. We say that two formulas $\Phi, \Psi \in \mathrm{FO}[\Gamma, \leq]$ are equivalent over $\mathcal{M}_\Gamma$, written $\Phi \equiv_{\mathcal{M}_\Gamma} \Psi$, if for all $h : \mathcal{P} \to 2^{\mathbb{R}}$ and $\nu : \mathsf{Free}(\Phi) \cup \mathsf{Free}(\Psi) \to \mathbb{R}$, we have $(\mathcal{M}_\Gamma, h), \nu|_{\mathsf{Free}(\Phi)} \models \Phi$ if and only if $(\mathcal{M}_\Gamma, h), \nu|_{\mathsf{Free}(\Psi)} \models \Psi$.

▶ **Theorem 10.** *For all $\Phi \in \mathrm{FO}[\Gamma, \leq]$, there exists a boolean combination Ψ of formulas in $\mathrm{FO}^3[\Gamma, \leq]$ such that $\Phi \equiv_{\mathcal{M}_\Gamma} \Psi$.*

Proof. Let $p \in \Gamma$, and $m_1 < \cdots < m_n$ its local extrema. We denote by $p_{(-\infty, m_1)}$, $p_{[m_1, m_2)}, \ldots, p_{[m_n, +\infty)}$ the (monotone) restrictions of p to intervals delimitated by its local extrema, and Δ_p the set of these partial functions. Let $\Delta = \bigcup_{p \in \Gamma} \Delta_p$. As above, we let $\mathcal{M}_\Delta = (\mathbb{R}, \leq, ({p_I}^{\mathcal{M}_\Delta})_{p_I \in \Delta})$, where $\leq$ is the usual ordering of the real numbers, and ${p_I}^{\mathcal{M}_\Delta} = \{(x, p(x)) \mid x \in I\}$. Note that ${p_I}^{\mathcal{M}_\Delta}$ is interval-preserving (cf. Example 1). We say that two formulas $\Phi \in \mathrm{FO}[\Gamma, \leq]$ and $\Psi \in \mathrm{FO}[\Delta, \leq]$ are equivalent, written $\Phi \equiv \Psi$, when for all $h : \mathcal{P} \to 2^{\mathbb{R}}$ and $\nu : \mathsf{Free}(\Phi) \cup \mathsf{Free}(\Psi) \to \mathbb{R}$, we have $(\mathcal{M}_\Gamma, h), \nu|_{\mathsf{Free}(\Phi)} \models \Phi$ if and only if $(\mathcal{M}_\Delta, h), \nu|_{\mathsf{Free}(\Psi)} \models \Psi$.

Let $\Phi \in \mathrm{FO}[\Gamma, \leq]$. The formula $\Psi \in \mathrm{FO}[\Delta, \leq]$ obtained by replacing each atomic formula $p(x, y)$ by $\bigvee_{p_I \in \Delta_p} p_I(x, y)$ is equivalent to Φ. Applying Theorem 9 to Ψ, we obtain another formula $\Psi' \in \mathrm{FO}[\Delta, \leq]$ such that $\Psi' \equiv \Psi$ and Ψ' is a boolean combination of formulas in $\mathrm{FO}^3[\Delta, \leq]$.

Following Example 3, one can construct for each $p_I \in \Delta$ a formula "$x \in I$" of $\mathrm{FO}^3[\Gamma, \leq]$ such that $(\mathcal{M}_\Gamma, h), \nu \models x \in I$ if and only if $\nu(x) \in I$. Consider now the formula $\Phi' \in \mathrm{FO}[\Gamma, \leq]$ obtained by replacing each atomic formula $p_I(x, y)$ in Ψ' by $x \in I \wedge p(x, y)$. Then $\Phi' \equiv \Psi'$, hence $\Phi \equiv_{\mathcal{M}_\Gamma} \Phi'$. Moreover, Φ' is a boolean combination of formulas in $\mathrm{FO}^3[\Gamma, \leq]$. ◀

The remainder of the section is devoted to the proof of Theorem 7.

Eliminating negations. The fact that all $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ path formulas are interval-preserving gives us a simple characterization of the complement of a path formula: we show below that an element b is in $[\![\pi^{\mathsf{c}}]\!](a)$ if it is to the left or to the right of all elements of $[\![\pi]\!](a)$, or if it does not satisfy $\langle \pi^{-1} \rangle$. We can then show that the complement of a path formula in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ is equivalent to a union of path formulas in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. This will allow us to deal with negation in the translation from $\mathrm{FO}[\Gamma, \leq]$ to $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$.

► **Lemma 11.** *For all path formulas $\pi \in \mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$, π^{c} is equivalent to a union of $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$ formulas.*

Proof. We show that

$$\begin{aligned}\pi^{\mathsf{c}} \equiv\ & (\{\neg \langle \pi \rangle\}? \cdot \leq) \cup (\{\neg \langle \pi \rangle\}? \cdot \geq) \cup \\ & (\leq \cdot \{\neg \langle \pi^{-1} \rangle\}?) \cup (\geq \cdot \{\neg \langle \pi^{-1} \rangle\}?) \cup \\ & (\pi^{\mathsf{c1}}) \cup (\pi^{\mathsf{c2}}) \cup (\pi^{\mathsf{c3}}) \cup (\pi^{\mathsf{c4}}) .\end{aligned}$$

We denote by π' the right-hand-side formula. First, for all a, b such that $[\![\pi]\!](a) = \emptyset$ or $[\![\pi^{-1}]\!](b) = \emptyset$, we have $(a, b) \in [\![\pi^{\mathsf{c}}]\!]$ and $(a, b) \in [\![\pi']\!]$. Now, suppose that $[\![\pi]\!](a) \neq \emptyset$ and $[\![\pi^{-1}]\!](b) \neq \emptyset$. We have $(a, b) \in [\![\pi']\!]$ if and only if $(a, b) \in [\![\pi^{\mathsf{c1}} \cup \pi^{\mathsf{c2}} \cup \pi^{\mathsf{c3}} \cup \pi^{\mathsf{c4}}]\!]$. Clearly, if $(a, b) \in [\![\pi^{\mathsf{c1}} \cup \pi^{\mathsf{c2}} \cup \pi^{\mathsf{c3}} \cup \pi^{\mathsf{c4}}]\!]$, then $(a, b) \in [\![\pi^{\mathsf{c}}]\!]$. Conversely, let us show that if $(a, b) \notin [\![\pi^{\mathsf{c1}} \cup \pi^{\mathsf{c2}} \cup \pi^{\mathsf{c3}} \cup \pi^{\mathsf{c4}}]\!]$ then $(a, b) \in [\![\pi]\!]$. In that case, we have either $a_1 \leq a \leq a_2$ for some $a_1, a_2 \in [\![\pi^{-1}]\!](b)$, or $b_1 \leq b \leq b_2$ for some $b_1, b_2 \in [\![\pi]\!](a)$. Since π is interval-preserving, we obtain $(a, b) \in [\![\pi]\!]$. ◄

Existential quantification. The elimination of existential quantifiers relies on the simple lemma below:

► **Lemma 12.** *Let $(A, \leq)$ be a linearly ordered set, and $I_1, \ldots, I_n$ intervals of $(A, \leq)$. Then $\bigcap_{1 \leq i \leq n} I_i \neq \emptyset$ if and only if for all $1 \leq i, j \leq n$, $I_i \cap I_j \neq \emptyset$.*

Proof. We show that there exists k and ℓ such that $\bigcap_{1 \leq i \leq n} I_i = I_k \cap I_\ell$, which implies the result. We define relations $\sqsubseteq_{\text{left}}$ and $\sqsubseteq_{\text{right}}$ over $\{I_1, \ldots, I_n\}$ which, intuitively, compare respectively the left and right bounds of the intervals:

$$\begin{aligned} I \sqsubseteq_{\text{left}} J \quad & \text{if} \quad \forall a \in J, \exists b \in I, b \leq a \\ I \sqsubseteq_{\text{right}} J \quad & \text{if} \quad \forall a \in I, \exists b \in J, a \leq b . \end{aligned}$$

It is easy to check that $\sqsubseteq_{\text{left}}$ and $\sqsubseteq_{\text{right}}$ are transitive, an that for all I and J, we have $I \sqsubseteq_{\text{left}} J$ or $J \sqsubseteq_{\text{left}} I$ (or both), and similarly for $\sqsubseteq_{\text{right}}$. Thus, there exists k such that $I_i \sqsubseteq_{\text{left}} I_k$ for all i, and ℓ such that $I_\ell \sqsubseteq_{\text{right}} I_i$ for all i. Then for all $a \in I_k \cap I_\ell$, for all i, there exists $b, b' \in I_i$ such that $b \leq a \leq b'$. Since I_i is an interval, we obtain $a \in I_i$. Hence $I_k \cap I_\ell = \bigcap_{1 \leq i \leq n} I_i$. ◄

The next lemma follows from an application of Lemma 12 to intervals of the form $[\![\pi_i]\!](a_i)$.

► **Lemma 13.** *Let $n \geq 1$. For all path formulas $\pi_1, \ldots, \pi_n$ and all state formulas φ in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$, the $\mathrm{FO}[\Gamma, \leq]$ formula*

$$\Phi(x_1, \ldots, x_n) = \exists x. \left(\widetilde{\varphi}(x) \wedge \bigwedge_{1 \leq i \leq n} \widetilde{\pi}_i(x_i, x) \right) \qquad (x_i \neq x \textit{ for all } i)$$

is equivalent to a positive boolean combination of formulas of the form $\widetilde{\pi}(x_j, x_k)$, with $1 \leq j, k \leq n$ and $\pi \in \mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$.

Proof. Let $\psi = \varphi \wedge \bigwedge_{1 \le i \le n} \langle \pi_i^{-1} \rangle$, and

$$\Psi(x_1, \dots, x_n) = \bigwedge_{1 \le i,j \le n} (\pi_i \cdot \widetilde{\{\psi\}?} \cdot \pi_j^{-1})(x_i, x_j)\,.$$

Let us show that $\Phi(x_1, \dots, x_n) \equiv \Psi(x_1, \dots, x_n)$. Let $\mathcal{M} = (A, \le, (\alpha^M)_{\alpha \in \Gamma}, (P^M)_{P \in \mathcal{P}})$, and $\nu : \{x_1, \dots, x_n\} \to A$. For all $1 \le i \le n$, let $I_i = [\![\pi_i]\!](\nu(x_i)) \cap [\![\psi]\!]$. Let us show that I_i is an interval of $([\![\psi]\!], \le)$. First, since π_i is interval-preserving, $[\![\pi_i]\!](\nu(x_i))$ is an interval of $([\![\langle \pi_i^{-1} \rangle]\!], \le)$. Thus, I_i is an interval of $([\![\langle \pi_i^{-1} \rangle]\!] \cap [\![\psi]\!], \le)$. But since $[\![\langle \pi_i^{-1} \rangle]\!] \subseteq [\![\psi]\!]$, this is simply $([\![\psi]\!], \le)$. Besides, it is easy to verify that

$$\mathcal{M}, \nu \models \Phi(x_1, \dots, x_n) \quad \iff \quad \bigcap_{1 \le i \le n} I_i \neq \emptyset\,.$$

Applying Lemma 12, we obtain

$$\begin{aligned} \mathcal{M}, \nu \models \Phi(x_1, \dots, x_n) \quad &\iff \quad \text{for all } 1 \le i,j \le n, I_i \cap I_j \neq \emptyset \\ &\iff \quad \text{for all } 1 \le i,j \le n, (\nu(x_i), \nu(x_j)) \in [\![\pi_i \cdot \{\psi\}? \cdot \pi_j^{-1}]\!] \\ &\iff \quad \mathcal{M}, \nu \models \Psi(x_1, \dots, x_n)\,. \end{aligned}$$ ◀

Translation from FO$[\Gamma, \le]$ to PDL$_{\mathsf{sf}}[\Gamma, \le, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. We are now ready to give the proof of Theorem 7.

Proof of Theorem 7. We assume that Φ is in prenex normal form, and prove the result by induction. The translation of atomic formulas $x \le y$ or $\alpha(x, y)$ is straightforward; moreover, $P(x) \equiv \widetilde{\{P\}?}(x, x)$, and $(x = y) \equiv \widetilde{\{true\}?}(x, y)$. Using Lemma 11 to eliminate negations, we obtain the result for all quantifier-free formulas.

The case $\Phi = \forall x.\Psi \equiv \neg \exists x. \neg \Psi$ reduces to the case of existential quantification, applying again Lemma 11 to eliminate negations.

We are left with the case $\Phi = \exists x.\Psi$. If x is not free in Ψ, then $\Phi \equiv \Psi$ (since Ψ has at least one free variable) and we are done by induction. Otherwise, assume that $\mathsf{Free}(\Psi) = \{x_1, \dots, x_n\}$ with $n > 1$ and $x = x_n$. By induction, Ψ is equivalent to a positive boolean combination of formulas of the form $\widetilde{\pi}(x_i, x_j)$ with $\pi \in \mathrm{PDL}_{\mathsf{sf}}[\Gamma, \le, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. We replace $\widetilde{\pi}(x_i, x_j)$ with $\widetilde{\pi^{-1}}(x_j, x_i)$ whenever $j < i$, and bring the resulting formula into disjunctive normal form. Each conjunct is then of the form $\Upsilon = \Upsilon_1 \wedge \Upsilon_2 \wedge \Upsilon_3$, where Υ_1 uses only variables from $\{x_1, \dots, x_{n-1}\}$, $\Upsilon_2 = \bigwedge_i \widetilde{\pi}_i(y_i, x)$ with $y_i = x_j$ for some $1 \le j < n$, and $\Upsilon_3 = \bigwedge_j \widetilde{\pi}_j(x, x)$. Note that $\Upsilon_3 \equiv \widetilde{\varphi}(x)$, where $\varphi = \bigwedge_j \mathsf{loop}(\pi_j)$. Then $\exists x.\Psi$ is equivalent to a finite disjunction of formulas

$$\exists x.\Upsilon \;\equiv\; \Upsilon_1 \wedge \exists x.\,(\Upsilon_2 \wedge \widetilde{\varphi}(x))$$

with Υ_1 and Υ_2 as above. If Υ_2 is empty, then we replace $\exists x.\widetilde{\varphi}(x)$ with the formula

$$(\le \cdot \{\varphi\}? \cdot \ge)(x_1, x_1) \vee (\ge \cdot \{\varphi\}? \cdot \le)(x_1, x_1)\,.$$

Otherwise, we apply Lemma 13 to $\exists x.\,(\Upsilon_2 \wedge \widetilde{\varphi}(x))$. In all cases, we obtain an equivalent formula which is a positive boolean combination of formulas $\widetilde{\pi}(x_i, x_j)$ with $1 \le i, j < n$ and $\pi \in \mathrm{PDL}_{\mathsf{sf}}[\Gamma, \le, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. ◀

▶ Remark 14. Without the assumption that all atomic binary relations are interval-preserving, $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ is still equivalent to $\mathrm{FO}^3[\Gamma, \leq]$. Indeed, in the proof of Theorem 7, the assumption that all atomic binary relations are interval-preserving is only used in Lemmas 11 and 13. But if Φ uses only three variables x, y and z, this assumption is not needed in the proof of Lemma 13. Indeed, we then have $\Phi(y, z) \equiv (\pi \cdot \widetilde{\{\varphi\}?} \cdot \pi'^{-1})(y, z)$, where π is the intersection of all π_i such that $x_i = y$, and π' is the intersection of all π_i such that $x_i = z$. Moreover, Lemma 11 is no longer needed if we translate an $\mathrm{FO}^3[\Gamma, \leq]$ formula into a positive boolean combination of $\mathrm{PDL}[\Gamma, \leq]$ formulas, since $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ allows to take the complement of a path formula. Note that the equivalence with $\mathrm{FO}^3[\Gamma, \leq]$ is already proven in [28] (for the calculus of relations).

5 Conclusion

We proved that every $\mathrm{FO}[\Gamma, \leq]$ formula over linear orders with interval-preserving binary relations can be translated into an equivalent positive boolean combination of path formulas in $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq, \mathsf{loop}, \mathsf{c1}, \mathsf{c2}, \mathsf{c3}, \mathsf{c4}]$. In particular, any $\mathrm{FO}[\Gamma, \leq]$ formula is equivalent to a boolean combination of formulas in $\mathrm{FO}^3[\Gamma, \leq]$, which shows that $\mathrm{FO}[\Gamma, \leq]$ has the three-variable property. This generalizes several known results.

It would be interesting to see if the equivalence between $\mathrm{FO}[\Gamma, \leq]$ and $\mathrm{PDL}_{\mathsf{sf}}[\Gamma, \leq]$ can be used as an intermediate step to prove that a temporal logic is expressively complete. It is not the case in general, since [13] provides an example of a class of structures which fits our assumptions but does not admit any expressively complete temporal logic. However, the equivalence could still be useful in more restricted settings.

References

1 Timos Antonopoulos, Paul Hunter, Shahab Raza, and James Worrell. Three Variables Suffice for Real-Time Logic. In *FoSSaCS 2015*, volume 9034 of *LNCS*, pages 361–374. Springer, 2015.

2 Benedikt Bollig, Marie Fortin, and Paul Gastin. It Is Easy to Be Wise After the Event: Communicating Finite-State Machines Capture First-Order Logic with “Happened Before”. In *CONCUR 2018*, volume 118 of *LIPIcs*, pages 7:1–7:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

3 Ryszard Danecki. Nondeterministic Propositional Dynamic Logic with intersection is decidable. In *Computation Theory*, pages 34–53. Springer Berlin Heidelberg, 1985.

4 Anuj Dawar. How Many First-order Variables are Needed on Finite Ordered Structures? In *We Will Show Them! (1)*, pages 489–520. College Publications, 2005.

5 G. De Giacomo and M. Lenzerini. Boosting the Correspondence between Description Logics and Propositional Dynamic Logics. In *Proceedings of the 12th National Conference on Artificial Intelligence, Seattle, WA, USA, July 31 - August 4, 1994, Volume 1.*, pages 205–212. AAAI Press / The MIT Press, 1994.

6 M. J. Fischer and R. E. Ladner. Propositional Dynamic Logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, 1979.

7 D. M. Gabbay. Expressive Functional Completeness in Tense Logic. In Uwe Mönnich, editor, *Aspects of Philosophical Logic: Some Logical Forays into Central Notions of Linguistics and Philosophy*, pages 91–117. Springer Netherlands, Dordrecht, 1981.

8 D. M. Gabbay, I. M. Hodkinson, and M. A. Reynolds. *Temporal expressive completeness in the presence of gaps*, volume Volume 2 of *Lecture Notes in Logic*, pages 89–121. Springer-Verlag, 1993.

9 S. Göller, M. Lohrey, and C. Lutz. PDL with intersection and converse: satisfiability and infinite-state model checking. *Journal of Symbolic Logic*, 74(1):279–314, 2009.

10 Martin Grohe. Finite variable logics in descriptive complexity theory. *Bulletin of Symbolic Logic*, 4(4):345–398, 1998.

11 J. Y. Halpern and Y. Moses. A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief. *Artif. Intell.*, 54(2):319–379, 1992.

12 L. Henkin. *Logical Systems Containing Only a Finite Number of Symbols*. Séminaire de Mathématiques Supérieures: Publications. Presses de l'Université de Montréal, 1967.

13 Yoram Hirshfeld and Alexander Rabinovich. Expressiveness of Metric modalities for continuous time. *Logical Methods in Computer Science*, 3(1), 2007.

14 Ian M. Hodkinson. Finite H-dimension does not imply expressive completeness. *J. Philosophical Logic*, 23(5):535–573, 1994.

15 Ian M. Hodkinson and András Simon. The k-variable property is stronger than H-dimension k. *J. Philosophical Logic*, 26(1):81–101, 1997.

16 Paul Hunter, Joël Ouaknine, and James Worrell. Expressive Completeness for Metric Temporal Logic. In *LICS*, pages 349–357. IEEE Computer Society, 2013.

17 Neil Immerman. Upper and Lower Bounds for First Order Expressibility. *J. Comput. Syst. Sci.*, 25(1):76–98, 1982.

18 Neil Immerman. DSPACE[n^{k}] = VAR[k+1]. In *Structure in Complexity Theory Conference*, pages 334–340. IEEE Computer Society, 1991.

19 Neil Immerman and Dexter Kozen. Definability with Bounded Number of Bound Variables. *Inf. Comput.*, 83(2):121–139, 1989.

20 H. Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, University of California, Los Angeles, 1968.

21 Michal Koucký, Clemens Lautemann, Sebastian Poloczek, and Denis Thérien. Circuit Lower Bounds via Ehrenfeucht-Fraisse Games. In *IEEE Conference on Computational Complexity*, pages 190–201. IEEE Computer Society, 2006.

22 M. Lange. Model checking propositional dynamic logic with all extras. *Journal of Applied Logic*, 4(1):39–49, 2006.

23 M. Lange and C. Lutz. 2-ExpTime lower bounds for Propositional Dynamic Logics with intersection. *Journal of Symbolic Logic*, 70(5):1072–1086, 2005.

24 Carsten Lutz and Dirk Walther. PDL with negation of atomic programs. *Journal of Applied Non-Classical Logics*, 15(2):189–213, 2005.

25 Bruno Poizat. Deux Ou Trois Choses Que Je Sais de Ln. *Journal of Symbolic Logic*, 47(3):641–658, 1982.

26 Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 721–730. ACM, 2008.

27 R. S. Streett. Propositional Dynamic Logic of Looping and Converse. In *Proceedings of STOC'81*, pages 375–383. ACM, 1981.

28 Alfred Tarski and Steven R. Givant. *A formalization of set theory without variables*. American Mathematical Society Providence, R.I, 1987.

29 Yde Venema. Expressiveness and Completeness of an Interval Tense Logic. *Notre Dame Journal of Formal Logic*, 31(4):529–547, 1990.