# Definability over Linear Constraints

**Michael Benedikt**[1] **H. Jerome Keisler**[2]

[1] Bell Laboratories, 263 Shuman Blvd., Naperville, IL 60566, USA, Email: benedikt@research.bell-labs.com
[2] University of Wisconsin, Madison Wisconsin 53706, USA, Email: keisler@math.wisc.edu

**Abstract.** We settle a number of questions concerning definability in first order logics with an extra predicate symbol ranging over semi-linear sets. These questions are motivated by the constraint database model for representing spatial data. We give new results both on the positive and negative side: we show that in first-order logic one cannot query a semi-linear set as to whether or not it contains a line, or whether or not it contains the line segment between two given points. However, we show that some of these queries become definable if one makes small restrictions on the semi-linear sets considered. We also define query languages that allow limited second-order quantification, and prove bounds on the expressive power of these richer languages.

## 1 Introduction

Much recent work in the foundations of spatial databases concerns the modeling of spatial information by *constraint sets*: Boolean combinations of linear or polynomial inequalities. Constraint sets can be effectively queried using variants of first-order logic; this is the basic idea behind *constraint query languages* ([9], [7]). The most well-studied languages in this family are the *first-order linear constraint language* $FO_{LIN}$ and the *first-order polynomial constraint language* $FO_{POLY}$. By a *semi-linear set* we mean a subset of a Euclidean space $R^n$ which is definable by a linear constraint, that is, a quantifier-free first order formula in the real ordered group $\langle R, +, -, < \rangle$. (Here and throughout this paper, formulas may have parameters from $R$). By quantifier elimination, each first order formula is equivalent to a linear constraint in the real ordered group. $FO_{LIN}$ is the first-order language with the vocabulary of the real ordered group plus an extra predicate symbol $S$ which ranges over semi-linear sets. Every $FO_{LIN}$ sentence defines a collection of semi-linear sets. In $FO_{POLY}$, the product symbol $\times$ is added to the vocabulary, and the extra predicate symbol $S$ ranges over the *semi-algebraic sets*— the subsets of $R^n$ which are definable by polynomial constraints, i.e. quantifier-free (or first order) formulas in the ordered field of reals.

A basic question, then, concerns the expressive power of these languages. Which families of definable sets (semi-linear sets for $FO_{LIN}$, semi-algebraic sets for $FO_{POLY}$) can be defined by a sentence in the language of the real ordered group (respectively field) with an extra predicate symbol ranging over the definable sets? More generally, given a family of sets $F$ in Euclidean space, one can ask: Which subfamilies of $F$ can be defined by a sentence in $FO_{LIN}$ or $FO_{POLY}$ with an extra predicate symbol ranging over $F$. Recent work has clarified many questions about the expressive power of $FO_{LIN}$ and $FO_{POLY}$ with an extra predicate symbol ranging over the *finite* subsets of Euclidean space ([3], [9]). There are also a number of recent results about the expressiveness of $FO_{POLY}$ with an extra predicate symbol ranging over the semi-algebraic sets ([5], [8]). However, the expressiveness of $FO_{LIN}$ with an extra predicate symbol ranging over the semi-linear sets is much less understood. Let's consider the following examples in the Euclidean plane:

$$Colinear = \{A \subset R^2 : \text{ all points in } A \text{ are colinear } \}$$

$$Is.Line = \{A \subset R^2 : A \text{ is the graph of some line } \}$$

$$Cont.Line = \{A \subset R^2 : A \text{ contains the graph of some line } \}$$

$$Lin.Reach = \{(A, \mathrm{a}, \mathrm{b}) : A \subset R^2, \mathrm{a}, \mathrm{b} \in R^2,\ A \text{ contains the line segment from a to b}\}$$

In the last example, $FO_{LIN}$ has extra constant symbols ranging over $R$, in addition to the extra predicate symbol $S$. Each of the four examples is easily seen to be definable by a sentence in $FO_{POLY}$, since there one can quantify over lines. Are they also definable in $FO_{LIN}$? It is fairly straightforward to show that *Colinear* is *not* definable by a sentence in $FO_{LIN}$ (see [1] and remark 1 below). However *Is.Line* is definable in $FO_{LIN}$ [2]. A semi-linear set $A$ belongs to the collection *Is.Line* iff it is either a vertical line or is the graph of a function and has the property that if $\mathrm{x}, \mathrm{y}, \mathrm{z} \in A$ then $\mathrm{x} + (\mathrm{y} - \mathrm{z}) \in A$.

It was asked in [1] whether *Cont.Line* is definable in $FO_{LIN}$. This and related questions were also considered in [2]. The general question is: under what circumstances can we ask questions about the existence of lines or line-segments in $FO_{LIN}$? If it appears that a query cannot be expressed in $FO_{LIN}$, how can we prove this? There is a need for techniques to show that a family of semi-linear sets is or is not definable in $FO_{LIN}$. In this paper we introduce such techniques, particularly methods from nonstandard analysis, and use them to resolve numerous questions about the definability of sets such as *Cont.Line* and *Lin.Reach*. We also show some positive results, giving that over certain classes of sets queries with quantification over lines are expressible. For example, we show that *Cont.Line* is undefinable over arbitrary semi-linear sets, but we also show that *Cont.Line* is definable if the extra predicate symbol ranges over an interesting subclass, the thin semi-linear sets. We also investigate several natural languages between $FO_{LIN}$ and $FO_{POLY}$. We give game characterizations of definability in these languages, and extend several of the undefinability results.

**Organization:** Section 2 introduces the notation and basic definitions. Section 3 gives the basic negative results showing that a number of queries are undefinable in $FO_{LIN}$. Section 4 considers fragments of second-order logic over the language $FO_{LIN}$. Section 5 gives positive results showing that certain queries are definable when the extra predicate $S$ ranges over restricted classes of semi-linear sets. Section 6 is about the definability of the property that any two points are $n$-linked (connected by a polygonal path with at most $n$ segments). Conclusions are given in Section 7.

## 2 Notation

### 2.1 The language $FO_{LIN}$

We start with a signature $\mathcal{S}$ consisting of predicate and constant symbols. For simplicity we will confine our attention to the case where $\mathcal{S} = \langle S, \mathrm{c} \rangle$ has only one binary predicate symbol $S$, and a sequence c of constant symbols of length $l$. The sequence c of constant symbols may be empty, that is, $l = 0$.

We let $FO_{LIN}(\mathcal{S})$ be the first-order language over the vocabulary $\mathcal{S} \cup \{+, -, <\}$. When $\mathcal{S}$ is clear from context we refer simply to $FO_{LIN}$. The first order structures for this vocabulary have the form $\langle R, +, -, <, A, \mathrm{a} \rangle$ where $A \subset R^2$ interprets $S$ and $\mathrm{a} \in R^l$ interprets c. Since all the structures under consideration have the same $\langle R, +, -, < \rangle$ part, we concentrate on the other part and define an $\mathcal{S}$-structure to be an object $\mathcal{A} = \langle A, \mathrm{a} \rangle$ where $A \subset R^2$ interprets $S$ and $\mathrm{a} \in R^l$ interprets c.

An $\mathcal{S}$-structure $\mathcal{A} = \langle A, \mathrm{a} \rangle$ *satisfies* a sentence $\phi \in FO_{LIN}(\mathcal{S})$, in symbols $\mathcal{A} \models \phi$, exactly when the corresponding first order structure $\langle R, +, -, <, \mathcal{A} \rangle = \langle R, +, -, <, A, \mathrm{a} \rangle$ satisfies $\phi$. If the relation $A$ is semi-linear, we say that $\mathcal{A}$ is a *semi-linear structure*, or *semi-linear instance*. Semi-algebraic and semi-analytic structures are defined similarly. Any collection of $\mathcal{S}$-structures is called a (Euclidean) *query*. By a *semi-linear query*, we will mean a collection of semi-linear structures. We say that a query $X$ is $FO_{LIN}$*-definable* if there is an $FO_{LIN}$ sentence $\phi$ such that:

$$\text{for every semi-linear structure } \mathcal{A}, \quad \mathcal{A} \models \phi \text{ if and only if } \mathcal{A} \in X.$$

We will sometimes consider the following generalization where the family of all semi-linear structures is replaced by another family of structures. Given a "base" query $F$ (not necessarily semi-linear), we say that a query $X$ is *$FO_{LIN}$-definable over $F$* if there is a sentence $\phi$ of $FO_{LIN}$ such that:

$$\text{for every } \mathcal{A} \in F, \quad \mathcal{A} \models \phi \text{ if and only if } \mathcal{A} \in X.$$

Thus when a base $F$ is not mentioned, it is understood to be the family of all semi-linear structures. Some examples of base queries $F$ that will arise in this paper are the families of semi-linear structures $\mathcal{A} = \langle A, \mathrm{a} \rangle$ such that the relation $A$ is:

- finite;
- *thin*, that is, has an empty interior;
- *1-bounded*, that is, is contained in a unit square $[0,1]^2$;
- has at most $n$ *singular points* (a singular point is a vertex of the boundary of $A$).

We note that if a query is $FO_{LIN}$ definable over $F$ then it is $FO_{LIN}$ definable over any subcollection $E \subseteq F$. Thus definability results are stronger when $F$ is larger, while undefinability results are stronger when $F$ is smaller. We will study the definability of some natural queries in $FO_{LIN}$ and its extensions. As a starting point, we recall a known fact, which is a consequence of the classical result that the product function cannot be defined in the first order theory of $\langle R, +, -, < \rangle$.

*Remark 1.* The query *Colinear* is not $FO_{LIN}$ definable over the collection of subsets of the plane of cardinality three.

### 2.2 Nonstandard Analysis and Undefinability

We will use notions from nonstandard analysis as a tool in many of our proofs. However, all of our results are statements about the standard reals.

The main place where nonstandard notions will be used is in characterizations of definability in query languages (as in, e.g., [4]). We assume familiarity with basic notions of nonstandard analysis (see [6]), but give a briefer-than-brief review here.

$N$ denotes the set of positive integers. For any set $U$, the *superstructure* $V(U)$ with base set $U$ is defined as $V(U) = \bigcup_{n \in N} V_n(U)$ where $V_1(U) = U$, and $V_{n+1}(U) = V_n(U) \cup \{X : X \subset V_n(U)\}$. Note in particular that $U \in V(U)$. We will work with the superstructure $\langle V(U), \in \rangle$ considered as a structure for the first-order language with the binary relation $\in$. A *bounded quantifier formula* in this language is a formula built up from atomic formulas by the logical connectives and the bounded quantifiers: $\forall X \in Y$, $\exists X \in Y$, where $X$ and $Y$ are variables. Almost all of "classical" mathematics can be done within the superstructure $V(R)$ based on the set $R$ of reals.

A *nonstandard universe* (based on $R$) consists of a pair of superstructures $V(R)$ and $V({}^*R)$ and a mapping $* : V(R) \to V({}^*R)$ such that:

1. ${}^*R$ is a proper extension of $R$
2. For each $r \in R$, ${}^*r = r$
3. **(Transfer Principle)** For any bounded quantifier formula $\phi(v_1, \ldots, v_n)$ and any list $a_1, \ldots, a_n$ of elements from $V(R)$, $\phi(a_1, \ldots, a_n)$ is true in $V(R)$ if and only if $\phi({}^*a_1, \ldots, {}^*a_n)$ is true in $V({}^*R)$.

We will fix a nonstandard universe once and for all.

Note that $^*R$ is the image of the element $R \in V(R)$, and $^*R \in V(^*R)$. An element $B \in V(^*R)$ is *standard* if it is in the image of the $*$-map, that is, $B = {^*A}$ for some $A \in V(R)$, and *internal* if it is an element of a standard set, that is, $B \in {^*A}$ for some $A \in V(R)$.

Some examples of standard sets are $^*R$, the usual order relation and arithmetic operations on $^*R$, and the sets $^*Z$, $^*Q$, and $^*N$. To improve readability, we ordinarily drop the $*$ from the order relation and arithmetic operations of $^*R$.

All standard sets are internal, all elements of internal sets internal, and any finite subset of an internal set is internal. Other examples of internal sets are the closed intervals $^*[a, b]$ where $a, b \in {^*R}$, and more generally the sets and relations which are first-order definable in the structure $\langle {^*R}, {^*N}, <, +, -, \times \rangle$.

An element $r \in {^*R}$ is *finite* if $|r| < n$ for some $n \in N$, and *infinitesimal* if $|r| < 1/n$ for all $n \in N$. For $r, s \in {^*R}$, we write $r \approx s$ if $|r - s|$ is infinitesimal. For each finite $r \in {^*R}$, there is a unique standard real number ${^o r} \in R$, called the *standard part* of $r$, such that ${^o r} \approx r$.

Three important consequences of the definition are:

- $N$ is a proper initial segment of $^*N$ in the natural ordering.
- Every nonempty *internal* subset of $^*R$ which has an upper bound has a least upper bound.
- Every infinite internal set is uncountable (and has cardinality at least the continuum).

It follows that infinite and positive infinitesimal elements of $^*R$ exist. In fact, there are uncountably many infinite $K \in {^*N}$, and uncountably many infinitesimals in $^*Q$.

Some examples of sets in $V(^*R)$ which are not internal are: any nonempty subset of $^*R$ which has an upper bound but no least upper bound (such as $R$, the set of finite elements, or the set of infinitesimals), any countably infinite set, the set of all finite subsets of $^*R$, and the standard part function $^o$.

By the Transfer Principle, the mapping $*$ is an elementary embedding of the ordered ring $\langle Z, +, -, \times, < \rangle$ into $\langle {^*Z}, +, -, \times, < \rangle$, and similarly for $R$ and $^*R$. Many of the facts we need from nonstandard analysis can be derived from these elementary embedding results.

When a set $A \in V(R)$ has a name, say the set of *widgets*, the elements of $^*A$ are called **widgets*, or *hyperwidgets*. For example, $^*R$ is the set of *hyperreal numbers*, and the image of the collection of semi-linear sets is the collection of *hypersemi-linear sets*. Thus every hypersemi-linear set is internal. Hypersemi-linear sets will appear in many of our proofs. When discussing properties of a hypersemi-linear set, we will often drop the "hyper" prefix; for example, we will usually write "line" rather than "hyperline", and "connected" rather than "hyperconnected".

By the Transfer Principle, any set which is definable by a first order formula in $\langle {^*R}, +, -, < \rangle$ is hypersemi-linear. In fact, any set which is definable by a first order hyperformula with parameters in $\langle {^*R}, +, -, < \rangle$ (or, equivalently, by a hyperfinite Boolean combination of linear constraints) is still hypersemi-linear.

The following proposition will be useful in proving undefinability results. In this proposition, $\equiv$ stands for the elementary equivalence relation between first order structures.

**Proposition 1.** *Suppose $X$ and $F$ are Euclidean queries. Then the following are equivalent:*

1. *$X$ is not $FO_{LIN}$-definable over $F$.*
2. *There are hyperstructures $\mathcal{A}, \mathcal{B} \in {^*F}$ such that $\mathcal{A} \in {^*X}$ and $\mathcal{B} \notin {^*X}$, but*

$$\langle {^*R}, +, -, <, \mathcal{A} \rangle \equiv \langle {^*R}, +, -, <, \mathcal{B} \rangle.$$

Proof: We will only use the direction from 2 to 1 in this paper, so we prove that direction here and leave the converse as an exercise. Assume that 1 fails but 2 holds. Let $\phi$ be an $FO_{LIN}$ sentence which defines $X$ over $F$. By the Transfer Principle, since $\mathcal{A} \in {}^*X$, we have $\mathcal{A} \models \phi$ and $\langle {}^*R, +, -, <, \mathcal{A} \rangle \models \phi$. Similarly, since $\mathcal{B} \notin {}^*X$, we have $\mathcal{B} \models \neg\phi$ and $\langle {}^*R, +, -, <, \mathcal{B} \rangle \models \neg\phi$. This contradicts 2. $\square$

## 3 Undefinability in first-order logic

Our first result is that *Lin.Reach* is not first-order.

**Theorem 1.** *The query Lin.Reach is not definable in $FO_{LIN}$.*

Proof: To do this we construct a hypersemi-linear set $A$ as follows. Let $\delta$ be a positive infinitesimal and $m$ be an element of ${}^*(0,1)$ which satisfy requirements to be given below. Let $B$ be the set of parallel lines $y = K\delta + mx$, with $K \in {}^*Z$, and let $A$ be the intersection of $B$ with the hyperreal unit square ${}^*[0,1]^2$. Thus $\delta$ will be the vertical distance between line segments, and $m$ will be the slope of any of the lines. See Figure 1.

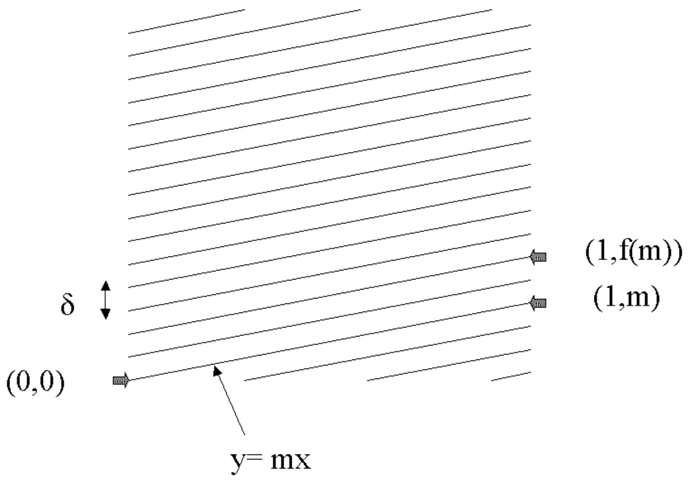


**Fig. 1.** The Basic Construction

Let $D = \{K : K/n \in {}^*Z \text{ for all } n \in N\}$. $D$ is the largest divisible subgroup of $\langle {}^*Z, +, - \rangle$. $D$ is nontrivial, since $K! \in D$ for any infinite $K \in {}^*N$. Choose hyperrational numbers $m, \delta \in {}^*Q \cap {}^*[0,1]$ so that:

- $m = J/L \in {}^*Q$ where $J, L \in {}^*N$ and $J \leq L$
- The standard part ${}^\circ m$ (which belongs to $[0,1]$) is irrational
- $H = 1/\delta$ is in $D$ and is such that $H/L \in D$ (i.e. $H$ is $L$ times something in $D$)

Our aim is to find a function $f$ satisfying the following requirements. Let $\Delta(x) = f(x) - x$. We say that a function $f : {}^*R \to {}^*R$ is *good* if for all $x, y \in {}^*R$ :

- $f$ is a bijection

- $f$ is order preserving
- $f$ is linear, i.e. $f(x+y) = f(x) + f(y)$
- $f(\epsilon) = \epsilon$ for all infinitesimal $\epsilon$
- $f(z) = z$ for all $z \in {}^*Z$
- $f(x) \approx x$ (hence $\Delta(x) \approx 0$)
- $\Delta(x)/\delta \in D$ and $\Delta(x)m/\delta \in D$
- $f(m) \neq m$

For any function $f$ on the hyperreal line, we let $f^2$ be the map on the plane defined by $f^2((x,y)) = (f(x), f(y))$.

**Claim 1** *For any good function $f$, $f^2$ maps any point lying on a line of the form $y = K\delta + mx$, with $K \in {}^*Z$, to a point on another line of this form.*

Proof: Given $y = K\delta + mx$, we show that $(f(x), f(y))$ is of the required form: i.e. $(f(y) - mf(x))/\delta \in {}^*Z$. We have

$$f(y) - mf(x) = f(K\delta) + f(mx) - mf(x) = K\delta + \Delta(K\delta) + mx + \Delta(mx) - m(x + \Delta(x)).$$

This is $K\delta + \Delta(K\delta) + \Delta(mx) - m\Delta(x)$. But each of these four terms is a multiple of $\delta$, because for every $z$, $\Delta(z)/\delta \in D \subset {}^*Z$, and similarly, $m\Delta(x)/\delta \in {}^*Z$. □

**Lemma 1.** *There is a good function $f$.*

The proof of this lemma will take some time. Let $R_0$ be the smallest divisible subgroup of $\langle {}^*R, +, - \rangle$ that contains both ${}^*Z$ and the set of all infinitesimals. Note that $R_0$ is just the set of all $x \in {}^*R$ such that the standard part of the fractional part of $x$ is rational.

**Claim 2** *$R_0$ is equal to the set of all $x \in {}^*R$ such that $x = p + n + \epsilon$ for some $p \in Q \cap [0,1)$, $n \in {}^*Z$, and $\epsilon \approx 0$. Moreover, for each $x \in R_0$, the decomposition $x = p + n + \epsilon$ is unique.*

Proof: The first statement is clear. To prove uniqueness, suppose $p + n + \epsilon = p' + n' + \epsilon'$. Then $n - n' = (p' - p) + (\epsilon' - \epsilon)$. The left side of this equation belongs to ${}^*Z$, and the right side is finite and has standard part in $(-1,1)$. Therefore both sides of the equation are equal to 0. Therefore $n = n'$ and $p' \approx p$. It follows that $p' = p$ and hence $\epsilon' = \epsilon$. □

Now let $\mathbf{c}$ be the cardinality of the continuum. As usual, we identify $\mathbf{c}$ with the set of all ordinals of cardinality less than $\mathbf{c}$. Let $\{r_\alpha : \alpha < \mathbf{c}\}$ be an enumeration of the set of all reals in $[0,1)$. Starting with $R_0$ defined above, we build an increasing chain of sets $R_\alpha$, $\alpha < \mathbf{c}$ by the following transfinite recursion. For limit ordinals $\gamma < \mathbf{c}$, put $R_\gamma = \bigcup_{\beta<\gamma} R_\beta$. For successor ordinals $\alpha + 1$, let $R_{\alpha+1}$ be the smallest divisible subgroup of $\langle {}^*R, +, - \rangle$ that contains both $R_\alpha$ and $r_\alpha$. We then have ${}^*R = \bigcup_{\alpha<\mathbf{c}} R_\alpha$. Note that in the case that $r_\alpha \in R_\alpha$, $R_{\alpha+1}$ is just $R_\alpha$.

**Claim 3** *$R_{\alpha+1}$ is equal to the set of all $x \in {}^*R$ such that $x = pr_\alpha + n + y$ for some $p \in Q$, $n \in {}^*Z$, and $y \in {}^*[0,1) \cap R_\alpha$. Moreover, if $r_\alpha \notin R_\alpha$, then for each $x \in R_{\alpha+1}$ the decomposition $x = pr_\alpha + n + y$ is unique.*

Proof: To prove uniqueness, suppose $pr_\alpha + n + y = p'r_\alpha + n' + y'$. If $p \neq p'$, then $r_\alpha = ((n - n') + (y - y'))/(p' - p) \in R_\alpha$. On the other hand, if $p = p'$, then $n + y = n' + y'$, so $n - n' = y' - y$. The left side of this equation belongs to ${}^*Z$ and the right side belongs to ${}^*(-1,1)$. Therefore both sides are equal to 0, so $n = n'$ and $y = y'$. □

**Claim 4** *For each $\alpha < \mathbf{c}$, $R_\alpha$ contains fewer than $\mathbf{c}$ real numbers.*

Proof: We will prove by transfinite induction that for each $\alpha < \mathbf{c}$, $|R_\alpha \cap R| \le \aleph_0 + |\alpha| < \mathbf{c}$. By Claim 2, $R_0 \cap R = Q$, so $|R_0 \cap R| = \aleph_0$. If $\alpha$ is a limit ordinal, then by inductive hypothesis, $|R_\alpha \cap R| \le \sum_{\beta<\alpha}(\aleph_0 + |\beta|) = \aleph_0 + |\alpha|$.

Now assume the result for $\alpha$. By Claim 3 we have $|R_{\alpha+1} \cap R| \le |Q| \times |R_\alpha \cap R| \le \aleph_0 + |\alpha|$. Thus the result holds for $\alpha + 1$, and the induction is complete. □

**Claim 5** *There exist infinitely many positive infinitesimal $\epsilon \in {}^*Q$ such that $\epsilon/\delta \in D$ and $\epsilon m/\delta \in D$.*

Proof: By hypothesis, $1/\delta = H$ where $H \in D$ and $H/L \in D$, so $mH = J(H/L) \in D$. Thus for all $n \in N$, $H/n \in {}^*Z$ and $mH/n \in {}^*Z$. Using the Transfer Principle, there is a least $n \in {}^*N \setminus N$ such that $H/n \notin {}^*Z$ or $mH/n \notin {}^*Z$. Therefore for all sufficiently small infinite $K \in {}^*N$ we have $H/K \in D$ and $mH/K \in D$. Thus $\epsilon = 1/K$ has the required property. □

With another transfinite recursion, we build an increasing chain of functions $f \ {}^{-}\ R_\alpha : R_\alpha \to {}^*R$, $\alpha < \mathbf{c}$. Let $f \ {}^{-}\ R_0$ be the identity function on $R_0$, take unions at limit ordinals, and define $f \ {}^{-}\ R_{\alpha+1}$ as follows when $r_\alpha \notin R_\alpha$: By Claim 5 we may choose a positive infinitesimal $\epsilon_\alpha \in {}^*Q$ such that $\epsilon_\alpha H \in D$ and $\epsilon_\alpha mH \in D$. For $x \in R_{\alpha+1}$, put $x = pr_\alpha + n + y$ as in Claim 3, and define $f(x) = p(r_\alpha + \epsilon_\alpha) + f(n + y)$. When $x \in R_\alpha$, we have $p = 0$, so the new value of $f(x)$ agrees with the old. Taking the union, we have a function $f : {}^*R \to {}^*R$.

We now add one more requirement in the construction which will insure that $f(m) \ne m$. Consider the first $\beta$ such that $m \in R_\beta$. Since ${}^o m$ is irrational, $m \notin R_0$, so $\beta > 0$. Then $\beta$ must be a successor ordinal, $\beta = \alpha + 1$. Since $m \notin R_\alpha$ we have $R_{\alpha+1} \ne R_\alpha$ and thus $r_\alpha \notin R_\alpha$. We then have a unique decomposition $m = pr_\alpha + n + y$ with $p \ne 0$. Any two different choices of the infinitesimal $\epsilon_\alpha$ will result in different values for $f(m)$, so we can choose $\epsilon_\alpha$ in such a way that $f(m) \ne m$.

**Claim 6** *The function $f$ is good.*

Proof: We verify the requirement that $\Delta(x)/\delta \in D$ and $\Delta(x)m/\delta \in D$ for all $x \in {}^*R$. Let $\beta = \alpha + 1$ be the first ordinal such that $x \in R_\beta$, and put $x = pr_\alpha + n + y$. We argue by induction on $\alpha$. We have $\Delta(x) = f(x) - x = p\epsilon_\alpha + \Delta(n + y)$, where $n + y \in R_\alpha$ and $p \in Q$. By definition, $\epsilon_\alpha/\delta \in D$, and by inductive hypothesis, $\Delta(n + y)/\delta \in D$. Since $D$ is a divisible group, it follows that $\Delta(x)/\delta \in D$. The proof that $\Delta(x)m/\delta \in D$ is similar. The other requirements on $f$ are easily proved by induction on $\alpha$. □

This completes the proof of Lemma 1.

We now use the good function $f$ to complete the proof of Theorem 1. Consider the hypersemi-linear structures $\mathcal{A} = \langle A, 0, 0, 1, m \rangle$ and $\mathcal{B} = \langle A, 0, 0, 1, f(m) \rangle$. The two points $(0, 0), (1, m)$ are on the same line segment in $A$, so *Lin.Reach* holds in $\mathcal{A}$. But $f(m) \ne m$, so *Lin.Reach* fails in $\mathcal{B}$. Since $f$ is good, it is an automorphism of the hyperreal ordered group that maps the relation and constants in $\mathcal{A}$ to those in $\mathcal{B}$. Hence $\langle {}^*R, +, -, <, \mathcal{A} \rangle \equiv \langle {}^*R, +, -, <, \mathcal{B} \rangle$. By Proposition 1, *Lin.Reach* is not $FO_{LIN}$ definable. □

The above proof shows more. Let *Conn* be the query

$$Conn = \{(A, \mathrm{a}, \mathrm{b}) : A \subset R^2, \mathrm{a}, \mathrm{b} \in R^2, \text{ a is connected to b in } A\}.$$

**Corollary 1.** *Lin.Reach is not $FO_{LIN}$-definable over the collection $F$ of 1-bounded thin semi-linear sets. In fact, there is no sentence $\psi$ of $FO_{LIN}(\mathcal{S})$ such that $(Lin.Reach \to \psi) \wedge (\psi \to Conn)$ holds in all $\mathcal{A} \in F$.*

Proof: The graph $A$ is 1-bounded and thin. The points $(0, 0), (1, m)$ are on the same line segment in the structure $\mathcal{A}$, but the corresponding points $(0, 0), (1, f(m))$ are not even connected in the structure $\mathcal{B}$. Thus $\psi$ would have to hold in $\mathcal{A}$ but fail in $\mathcal{B}$, so $\psi$ cannot be a sentence of $FO_{LIN}$. □

The proof of Theorem 1 can be modified to show that *Lin.Reach* is $FO_{LIN}$ undefinable over other families. For example, the preceding proof uses hypersemi-linear sets with infinitely many parallel line segments, which corresponds to a family of semi-linear sets with unboundedly many parallel line segments, and thus unboundedly many singular points. The next result has a finite bound on the number of singular points but a nonempty interior.

**Theorem 2.** *Lin.Reach is not $FO_{LIN}$-definable over the collection of 1-bounded semi-linear sets with ten singular points.*

Proof: We modify the set $A$ to form two new sets $T$ and $U$ (see Figure 2). To form $T$, we first replace the unit square by the "unit right triangle" with vertices $(0,0),(1,0),(1,1)$. Then we add all of the interior of the unit right triangle except for two similar infinitesimal triangular windows around the points $(0,0)$ and $(1,m)$ with height $< \delta$. These windows are so small that they miss all of the original family of parallel lines of slope $m$ except for the line segment through $(0,0),(1,m)$. The set $T$ formed in this way is a 1-bounded hypersemi-linear set with ten singular points. The set $U$ is similar except that the second triangular window is around the point $(1, f(m))$. We use the same function $f$ as before. $f$ is an isomorphism between the structures $\langle T,0,0,1,m \rangle$ and $\langle U,0,0,1,f(m) \rangle$. The points $(0,0),(1,m)$ are on a line segment contained in $T$ but the points $(0,0),(1,f(m))$ are not on a line segment contained in $U$. $\square$

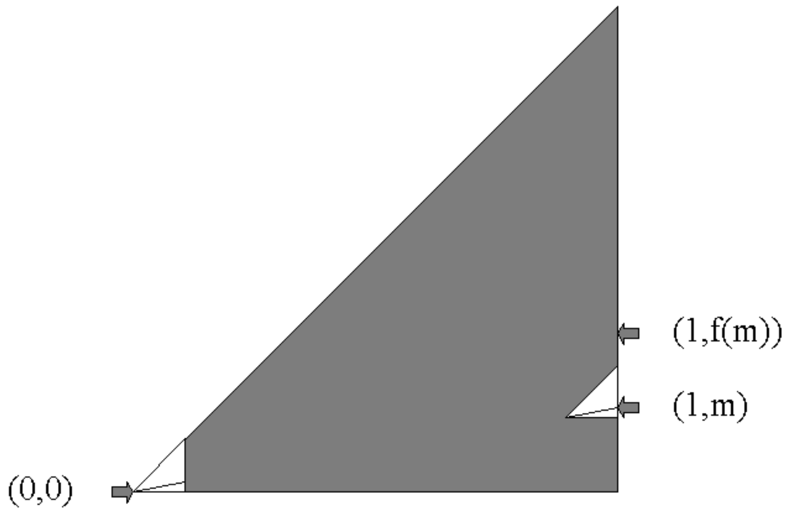


**Fig. 2.** Modified Construction

We will now show that the query *Cont.Line* is not first order definable. Here the signature has a binary predicate symbol but no constants.

**Theorem 3.** *The query Cont.Line is not definable in $FO_{LIN}$. In fact, it's not even $FO_{LIN}$-definable over the collection of semi-linear sets with ten singular points.*

Proof: We modify the construction in the previous theorem. Let $T, U$, and $f$ be as in the proof of Theorem 2. We will add two infinite cones to $T$ as follows. Let $C_1$ be the cone heading towards $-\infty$ with apex $(0,0)$ and bounded by two rays having standard rational slopes $r$ and $s$ where $0 < r < {}^{\circ}m < s < 1$. See Figure 3.

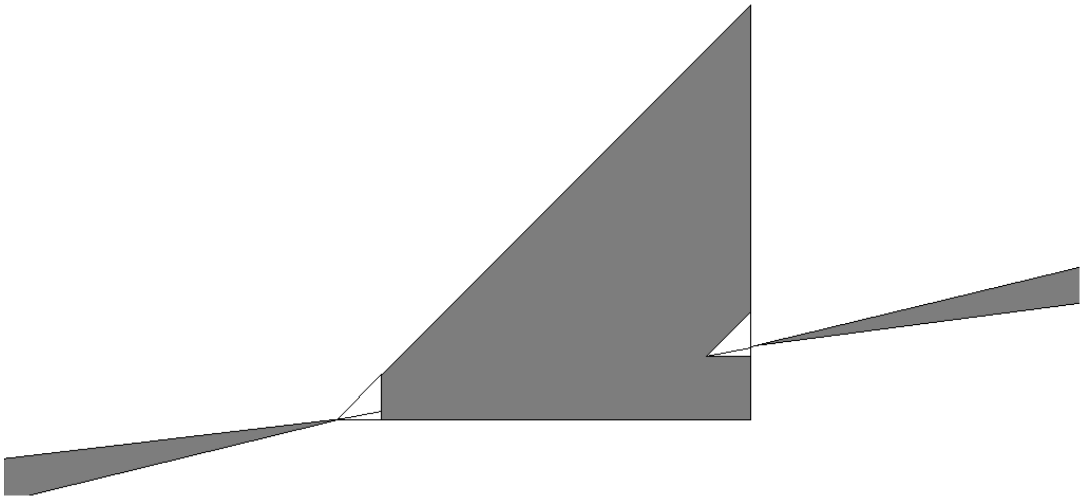

**Fig. 3.** Contains a line

$C_1$ is a standard cone defined by rational numbers, and hence will be fixed by $f$. Let $C_2$ be the cone with apex $(1, m)$ and bounded by two lines having rational slopes $r'$ and $s'$ with $0 < r' < \circ m < s' < 1$. The base of $C_2$ will be moved by $f$, and the rationality of the slopes will guarantee that the boundary lines map to boundary lines. Thus $C_2$ maps under $f$ to the cone $C_2'$ with apex $(1, f(m))$ and boundary rays having slopes $s'$ and $t'$. Let $T_1 = T \cup C_1 \cup C_2$ and $U_1 = U \cup C_1 \cup C_2'$. These sets again have ten singular points. $T_1$ contains a line and $T_2$ does not. It is clear that $f$ maps $T_1$ to $T_2$, which shows that the corresponding structures are equivalent in $FO_{LIN}$. □

## 4 Undefinability in restrictions of second-order logic

Since we've seen that $FO_{LIN}$ cannot express queries that we're interested in, we now investigate definability in extensions of $FO_{LIN}$. The queries we consider will be obviously definable in $FO_{POLY}$ (introduced in the Introduction). It is natural to look, then, for languages between $FO_{POLY}$ and $FO_{LIN}$.

The language $Lin - \Pi_1$ has the symbols of $FO_{LIN}$ plus a set $\mathcal{M}$ of unary function variables $M_i$. A $Lin - \Pi_1$ formula has the form $\forall M_1 \ldots \forall M_n \ \phi$, where $\phi$ is a first-order formula over the signature $\{+, -, <\} \cup \mathcal{S} \cup \mathcal{M}$. The semantics are as follows: A $Lin - \Pi_1$ sentence $\forall M_1 \ldots \forall M_n \ \phi$ holds in a structure $\mathcal{A} = \langle A, \mathrm{a} \rangle$ iff for all interpretations of the $M_i$ by unary scalar multiplication functions $\times_{m_i}$ (where $m_i \in R$ and $\times_{m_i}(x) = m_i x$), the sentence $\phi$ holds in the first-order structure $\langle R, +, -, <, \mathcal{A}, \times_{m_1}, \ldots, \times_{m_n} \rangle$. Equivalently, one can think of a sentence $\forall M_1 \ldots \forall M_n \phi$ as saying "for all lines $M_1 \ldots M_n$, $\phi$ holds".

A query $X$ is said to be in $Lin - \Pi_1$ iff there is some $Lin - \Pi_1$ sentence $\theta$ such that for each semi-linear structure $\mathcal{A}$, $\mathcal{A} \in X$ if and only if $\mathcal{A} \models \theta$. We define $Lin - \Sigma_1$ to be the class of queries whose complement is in $Lin - \Pi_1$, and define $Lin - \Delta_1$ to be $Lin - \Pi_1 \cap Lin - \Sigma_1$. Given a base query $F$, the classes of queries which are $Lin - \Pi_1$ over $F$, $Lin - \Sigma_1$ over $F$, and $Lin - \Delta_1$ over $F$ are defined similarly, with "for each $\mathcal{A} \in F$" in place of "for each semi-linear $\mathcal{A}$".

It is clear that over every base query $F$ we have

$$FO_{LIN} \subseteq Lin - \Delta_1, \quad Lin - \Pi_1 \subseteq FO_{POLY}, \quad Lin - \Sigma_1 \subseteq FO_{POLY}.$$

It is also easily seen that the query *Colinear* is in $Lin - \Delta_1$. Since this query is not $FO_{LIN}$ definable, we have

$$FO_{LIN} \neq Lin - \Delta_1,$$

even over the collection $F$ of planar sets of cardinality 3. Another example of a query which is in $Lin - \Delta_1$ but not in $FO_{LIN}$ is *Lin.Reach*. To see that *Lin.Reach* is in $Lin - \Delta_1$, note that the line segment from a to b is contained in a semi-linear set $A$ iff for some (every) $m$ such that b is on the line through a with slope $m$, every point on this line between a and b lies in $A$. But Theorem 1 shows that *Lin.Reach* is not $FO_{LIN}$ definable.

Our next proposition shows that $FO_{POLY}$ is exactly the closure of $Lin - \Pi_1$, and also of $Lin - \Sigma_1$, under first order quantification. Let us say that two sentences $\phi, \psi$ are *equivalent over* $\mathcal{S}$ if for every $\mathcal{S}$-structure $\mathcal{A}$, $\mathcal{A} \models \phi$ if and only if $\mathcal{A} \models \psi$. Say that two languages are equivalent if every sentence in one language is equivalent to a sentence in the other language.

**Proposition 2.** *(i) The set of universal sentences of $FO_{POLY}$ is equivalent to the set of universal sentences of $Lin - \Pi_1$.*

*(ii) The set of existential sentences of $FO_{POLY}$ is equivalent to the set of existential sentences of $Lin - \Sigma_1$.*

*(iii) $FO_{POLY}$ is equivalent to the closure of $Lin - \Delta_1$ under first order quantification, and also to the closure of $Lin - \Pi_1$ under all the first-order operations $\{\forall, \exists, \wedge, \vee, \neg\}$.*

Proof: It suffices to prove the non-trivial direction of (i). Then (ii) follows by applying (i) to the negation, and the non-trivial direction of (iii) follows by applying both (i) and (ii) to quantifier-free sentences of $FO_{POLY}$ and then using the prenex normal form theorem.

Consider a universal sentence $\forall \mathrm{y}\, \theta(\mathrm{c}, \mathrm{y})$ of $FO_{POLY}$. We need an equivalent universal sentence of $Lin - \Pi_1$. Let $\tau_1, \ldots, \tau_k$ be the set of all terms and subterms of terms which occur in $\theta$. Introduce second order variables $M_1, \ldots, M_k$. For each $j \leq k$, we define a term $\sigma_j$ of $Lin - \Pi_1$ by induction on the complexity of $\tau_j$. The only nontrivial case in this definition is the product case: If $\tau_j$ is $\tau_h \times \tau_i$, then $\sigma_j$ is $M_h(\sigma_i)$. Form $\theta'$ by replacing each term $\tau_j$ in $\theta$ by $\sigma_j$. The required sentence of $Lin - \Pi_1$ is

$$\forall M_1 \cdots \forall M_k \forall \mathrm{y} \left( \bigwedge_{j=1}^{k} M_j(1) = \sigma_j \ \rightarrow \theta'(M_1, \ldots, M_k, \mathrm{c}, \mathrm{y}) \right).$$

□

Our next goal is:

**Theorem 4.** *$Lin - \Delta_1$, $Lin - \Sigma_1$, $Lin - \Pi_1$, and $FO_{POLY}$ are all distinct.*

To prove this theorem, all we need is an example of a query that is in $Lin - \Sigma_1 \setminus Lin - \Pi_1$. The negation will then be in $Lin - \Pi_1 \setminus Lin - \Sigma_1$, and Theorem 4 will follow. We first note a sufficient condition for not being in $Lin - \Sigma_1$:

For a semi-linear query $X$, the game $G_n(X)$ is played as follows: The game begins with Duplicator choosing a hypersemi-linear structure $\mathcal{A} \in {}^*X$. Spoiler chooses $n$ hyperreals $L_1 \ldots L_n$. Duplicator responds by choosing a hypersemi-linear structure $\mathcal{B} \notin {}^*X$ and hyperreals $M_1 \ldots M_n$. Spoiler and Duplicator then play the standard Ehrenfeucht-Fraisse game on the first order structures $\langle {}^*R, +, -, <, \mathcal{A}, \times_{L_1} \ldots \times_{L_n} \rangle$ and $\langle {}^*R, +, -, <, \mathcal{B}, \times_{M_1}, \ldots \times_{M_n} \rangle$.

**Lemma 2.** *If Duplicator can win $G_n(X)$ for every $n \in N$, then $X$ is not in $Lin - \Sigma_1$.*

Proof: Suppose $X$ is definable by $\exists L_1 \ldots \exists L_n\ \phi$. Then Duplicator's first move must be a structure $\mathcal{A} \in {}^*X$, so $\mathcal{A} \models \exists L_1 \ldots \exists L_n\ \phi$. Spoiler can then choose hyperreal numbers $L_1 \ldots L_n$ which satisfy $\phi$ in $\mathcal{A}$. Duplicator must respond with a structure $\mathcal{B} \notin {}^*X$ and hyperreals $M_1 \ldots M_n$. Since $\mathcal{B}$ is not in ${}^*X$, the $M_i$ cannot satisfy $\phi$ in $\mathcal{B}$. Hence Duplicator cannot have a winning strategy for the game $G_n(X)$. □

We will now give an example which completes the proof of Theorem 4. We will work in the plane, and let $\mathbf{0} = (0,0)$. Consider the query:

$$SomeTriple = \{A \subseteq R^2 : (\exists \mathrm{x} \in A)(\exists \mathrm{y} \in A)(\exists \mathrm{z} \in A)[\mathrm{x}, \mathrm{y}, \mathrm{z} \text{ are distinct and colinear}]\}.$$

$SomeTriple$ is clearly $FO_{POLY}$ definable, and is even in $Lin - \Sigma_1$.

**Theorem 5.** *The query SomeTriple is not in* $Lin - \Pi_1$.

Proof idea: Let $X$ be the complement of $SomeTriple$. Fix a (standard) integer $n$. We construct a finite set $A$ that will be Duplicator's opening move in a winning strategy in the game $G_n(X)$. Let $\mathrm{m} = (m_1, \ldots, m_{n+1})$ be an $n+1$-tuple of hyperrational numbers in ${}^*(0,1)$ such that $0 < {}^o m_1 < \cdots < {}^o m_{n+1} < 1$, and the real tuple ${}^o\mathrm{m}$ is algebraically independent over the (standard) rationals (that is, no polynomial with rational coefficients has root ${}^o\mathrm{m}$ ). For each $k = 1, \ldots, n+1$, choose a hyperrational number $q_k$ such that $q_k \approx (m_k)^2$ but $q_k \neq (m_k)^2$. Let $A$ be the finite set consisting of $\mathbf{0}$, the points $\mathrm{x}_k = (m_k, q_k)$, and the points $\mathrm{y}_k = (k, k m_k)$, where $k = 1, \ldots, n+1$. One can show that $A$ belongs to ${}^*X$, and hence is a legal move for Duplicator. Spoiler must then respond by choosing an $n$-tuple of multipliers $r_1, \ldots, r_n \in {}^*R$. One can then construct a winning response by Duplicator by building an automorphism of the hyperreal ordered group, in a manner similar to the proof of Lemma 1. However, this time the construction requires a nonstandard universe which has an additional property, called $\omega_1$-*saturation*. The complete proof is given in the full paper. □

Theorem 4 now follows immediately from Theorem 5.

As usual, the proof of the theorem shows more. Let $SomeTriple0$ be the query

$$SomeTriple0 = \{A \subseteq R^2 : \mathbf{0} \in A \wedge (\exists \mathrm{y} \in A)(\exists \mathrm{z} \in A)[\mathbf{0}, \mathrm{y}, \mathrm{z} \text{ are distinct and colinear}]\}.$$

Clearly, $SomeTriple0$ implies $SomeTriple$. The proof of Theorem 5 shows that:

**Corollary 2.** *There is no* $Lin - \Pi_1$ *sentence* $\theta$ *such that* $(SomeTriple0 \rightarrow \theta) \wedge (\theta \rightarrow SomeTriple)$ *holds in every semi-linear structure.*

A query is $Lin - \Pi_1(n)$ if it can be defined by a $Lin - \Pi_1$ sentence with at most $n$ second order quantifiers. Similarly for $Lin - \Sigma_1(n)$. A second corollary to the proof is:

**Corollary 3.** *The query SomeTriple is not* $Lin - \Pi_1$ *over the collection of finite sets in the plane. In fact, for each* $n$, *SomeTriple is not* $Lin - \Pi_1(n)$ *over the collection of planar sets of cardinality* $2n+3$.

A natural variant of $SomeTriple0$ is the query:

$$AllTriples = \{A \subseteq R^2 : (\forall \mathrm{x} \in A)\ (\exists \mathrm{y} \in A)\ [\mathbf{0}, \mathrm{x}, \mathrm{y} \text{ are distinct and colinear}]\}.$$

Note that this query implies that $\mathbf{0} \notin A$. $AllTriples$ is clearly $FO_{POLY}$ definable, and is even in $Lin - \Pi_1$.

**Theorem 6.** *AllTriples is not in* $Lin - \Sigma_1$.

The proof is a variation of the proof for Theorem 5, and is given in the full paper.

## 5 Definability over thin sets

There is an intuition that queries such as *Lin.Reach* and *Cont.Line* are "almost" first-order definable. The previous section gave one formalization of this. Namely, we saw that *Lin.Reach* is in $Lin - \Delta_1$, and is thus "very close" to being first-order definable. What about *Cont.Line*? It is certainly in $Lin - \Sigma_1$, and we saw in Theorem 3 that it is not $FO_{LIN}$ definable. However, we do not know whether or not the query *Cont.Line* is in $Lin - \Pi_1$. We will give two positive results showing that *Cont.Line* and *Lin.Reach* are $FO_{LIN}$ definable over restricted classes of semi-linear structures.

We note that in Theorem 3 we made use of examples that had nonempty, and in fact unbounded, interiors. We show here that this is essential. We show that *Cont.Line* is definable over thin semi-linear instances: i.e. there is an $FO_{LIN}$ sentence $\phi$ such that for each thin semi-linear structure $\mathcal{A}$, $\mathcal{A} \models \phi \leftrightarrow \mathcal{A} \in Cont.Line$.

**Theorem 7.** *Cont.Line is $FO_{LIN}$ definable over the family of thin semi-linear sets.*

Proof: A point is called *regular* in $S$ if it is not singular in $S$. If $U(\mathrm{y},\mathrm{z})$ denotes the open rectangle with corners $\mathrm{y},\mathrm{z}$, then $\mathrm{x} \in U(\mathrm{y},\mathrm{z})$ can be expressed by the formula $y_1 < x_1 < z_1 \wedge y_2 < x_2 < z_2$. Thus if $S$ is thin one can say that x is regular in $S$ with the formula $Reg(\mathrm{x})$:

$$\exists \mathrm{y}\, \exists \mathrm{z}\, \exists \mathrm{u}\, [\mathrm{u} \neq \mathbf{0} \wedge (S \cap U(\mathrm{y},\mathrm{z})) \text{ contains } \mathrm{x}+\mathrm{u}, \mathrm{x}-\mathrm{u} \text{ and is closed under midpoints}].$$

Let $Cong(\mathrm{x},\mathrm{y})$ (for congruence) say that x and y are regular in $S$, and the line segments in $S$ containing x and y are parallel (that is, for all sufficiently small z, $\mathrm{x}+\mathrm{z} \in S \leftrightarrow \mathrm{y}+\mathrm{z} \in S$).

Let $MinSame(\mathrm{x},t,y)$ say that $y$ is minimal such that $Cong(\mathrm{x},t,y)$, that is,
$Cong(\mathrm{x},t,y) \wedge \forall u(Cong(\mathrm{x},t,u) \rightarrow y \leq u)$.

Let $Far(t)$ say that there are no singular points of $S$ with horizontal coordinate $\geq t$.

Let $Vert$ be the sentence asserting that $S$ contains a vertical line.

Finally, let $\phi$ be the sentence

$Vert \vee \exists \mathrm{x}\ (Reg(\mathrm{x}) \wedge \forall t\, \forall t'\ [Far(t) \wedge Far(t') \rightarrow$
$\exists y\ \exists y'[MinSame(\mathrm{x},t,y) \wedge MinSame(\mathrm{x},t',y') \wedge (\mathrm{x}+(t',y')-(t,y)) \in S]\,])$.

We claim that $\phi$ is the required sentence. Suppose $S$ is a thin semi-linear set in *Cont.Line*. If $S$ contains a vertical line then it satisfies $\phi$. Suppose $S$ contains a nonvertical line $L$, and let x be any regular point on $L$. Let $t$ and $t'$ be points beyond any singular point of $S$ with $0 < t < t'$. To the right of $t$, $S$ can only consist of finitely many rays that never intersect, including the part of $L$ beyond $t$. Among these rays is a lowest ray $L'$ which is parallel to $L$. There exist $y, y'$ such that $(t,y)$ and $(t',y')$ are on $L'$. Then $MinSame(\mathrm{x},t,y)$ and $MinSame(\mathrm{x},t',y')$. Since $L'$ is parallel to $L$, the sum of x and the vector between $(t,y)$ and $(t',y')$ is on $L$, and hence is in $S$.

Conversely, suppose that $S$ satisfies $\phi$. Suppose $Vert$ does not hold, and let x be any witness for the rest of $\phi$. Locally, $S$ looks like a line $L$ through x, since x is a regular point and $S$ is thin. We show that $L$ is actually contained in $S$. Let $m$ be the slope of $L$. For any $d$ we can find $t$ and $t'$ satisfying $Far(t) \wedge Far(t')$ with $t - t' = d$. Since $S$ satisfies $\phi$, there are minimal $y$ and $y'$ such that in neighborhoods of both $(t,y)$ and $(t',y')$, $S$ is a line with slope $m$. Since $t$ and $t'$ are far out, the only possibility is that $(t,y)$ and $(t',y')$ actually are on the same line of $S$, which means that the slope of the line segment between them must be $m$. Hence when we add the vector between $(t,y)$ and $(t',y')$ to x, we get the point on $L$ at horizontal distance $d$ away from x. By assumption, this point is in $S$. Since $d$ was arbitrary, this shows that $S$ contains $L$. □

The above argument also shows somewhat more.

**Corollary 4.** *Cont.Line is $FO_{LIN}$ definable over the family of semi-linear sets with 1-bounded interior.*

Turning back to *Lin.Reach*, recall that Corollary 1 showed that *Lin.Reach* is *not* $FO_{LIN}$ definable over the family of thin semi-linear structures. However, there we used structures with unboundedly many singular points. The following positive result shows that this is essential.

**Theorem 8.** *For each $n$, the query Lin.Reach is $FO_{LIN}$ definable over the collection $T_n$ of thin semi-linear structures with at most $n$ singular points.*

Proof: Fix $n$. Let $Vert(\mathrm{a},\mathrm{b})$ say that $S$ contains a vertical line segment from a to b. Let $\mathrm{x} \leq_1 \mathrm{y}$ mean that the horizontal coordinate of the point x is $\leq$ the horizontal coordinate of y, and define $=_1$ and $<_1$ similarly. Let $Mid(\mathrm{x},\mathrm{y})$ be the midpoint between x and y. The following sentence defines *Lin.Reach* over $T_n$.

$$Vert(\mathrm{a},\mathrm{b}) \vee \exists \mathrm{x}^1 \in S \cdots \exists \mathrm{x}^n \in S \; \{\mathrm{x}^1 = \mathrm{a} \wedge \mathrm{x}^n = \mathrm{b} \wedge (\mathrm{x}^1 <_1 \mathrm{x}^2 \leq_1 \cdots \leq_1 \mathrm{x}^n) \wedge$$

$$\forall \mathrm{y} \left( Reg(\mathrm{y}) \vee \mathrm{y} <_1 \mathrm{a} \vee \mathrm{y} =_1 \mathrm{x}^1 \vee \cdots \vee \mathrm{y} =_1 \mathrm{x}^n \vee \mathrm{b} <_1 \mathrm{y} \right) \wedge$$

$$\bigwedge_{k=1}^{n-1} \left[ \mathrm{x}^k <_1 \mathrm{x}^{k+1} \rightarrow \left( Mid(\mathrm{x}^k,\mathrm{x}^{k+1}) \in S \wedge Cong(Mid(\mathrm{x}^k,\mathrm{x}^{k+1}), Mid(\mathrm{x}^1,\mathrm{x}^2)) \right) \right] \}$$

It says that either $S$ contains a vertical segment from a to b, or there is a sequence of points $\mathrm{x}^1, \ldots \mathrm{x}^n$ in $S$ beginning at a, ending at b, and with increasing horizontal coordinates, such that for each pair $\mathrm{x}^k, \mathrm{x}^{k+1}$, there are no singular points with horizontal coordinate strictly between them, and all the midpoints $Mid(\mathrm{x}^k,\mathrm{x}^{k+1})$ belong to $S$ with the same slope.

This implies that the line segment from a to b passes through all the points $\mathrm{x}^1, \ldots, \mathrm{x}^n$ and is contained in $S$. Conversely, if $S$ contains the segment from a to b and the segment is not vertical, the above sentence is satisfied by taking $\mathrm{x}^1, \ldots, \mathrm{x}^n$ to include a, b, and each point on the line segment between them that agrees with some singular point in the horizontal coordinate. □

## 6 Definability of n-linked

We say a semi-linear set $A$ is *$n$-linked* if for any two points x and y in $A$, there is a polygonal path from x to y consisting of at most $n$ line segments. $n$-linkedness is a natural connectivity property of semi-linear sets, and in this section we will give a fairly complete description of the definability of $n$-linkedness for $n \in N$ (summarized at the end of the section). Somewhat surprisingly, the answer will be "yes" for some values of $n$ and "no" for others.

**Theorem 9.** 1-*linked is $FO_{LIN}$ definable.* 2-*linked is $FO_{LIN}$ definable over the family of thin semi-linear sets.*

Proof: 1-linked is definable by the statement: For any two points in $S$, their midpoint is in $S$.

For 2-linked, let $Parallel(\mathrm{x},\mathrm{y})$ say that y is regular and there is a line segment of $S$ containing x that has the same slope as the line segment through y. Since $S$ is thin, this can be expressed by the formula:

$$\mathrm{x} \in S \wedge Reg(\mathrm{y}) \wedge (\forall \mathrm{u} \text{ sufficiently close to } \mathbf{0})[\mathrm{y} + \mathrm{u} \in S \rightarrow (\mathrm{x} + \mathrm{u} \in S \vee \mathrm{x} - \mathrm{u} \in S)].$$

As before, we let $Mid(\mathrm{x},\mathrm{y})$ denote the midpoint between x and y. Now consider the sentence $\phi$:

$$\begin{aligned} \forall \mathrm{x} \forall \mathrm{y} \; ( Reg(\mathrm{x}) \wedge Reg(\mathrm{y}) \rightarrow [ Parallel(\mathrm{x},\mathrm{y}) \rightarrow Mid(\mathrm{x},\mathrm{y}) \in S ] \wedge \\ [ \neg Parallel(\mathrm{x},\mathrm{y}) \rightarrow \exists \mathrm{z} \; ( \neg Reg(\mathrm{z}) \wedge \\ Parallel(\mathrm{z},\mathrm{x}) \wedge Parallel(\mathrm{z},\mathrm{y}) \wedge Mid(\mathrm{z},\mathrm{y}) \in S \wedge Mid(\mathrm{z},\mathrm{x}) \in S ) ] ) \end{aligned}$$

This sentence says that any two regular points either have the same slope and their midpoint is in $S$, or they have different slopes and there is a singular point realizing both slopes whose midpoint with each of the original points is in $S$.

We assume $\phi$, and show that $S$ is 2-linked. Note that it suffices to show that every two regular points are connected by a path consisting of two line segments (since for two singular points $\mathrm{x}$ and $\mathrm{y}$ we can find regular points $\mathrm{x}'$ and $\mathrm{y}'$ nearby such that any path from $\mathrm{x}'$ to $\mathrm{y}'$ extends to a path from $\mathrm{x}$ to $\mathrm{y}$ with the same number of segments). Suppose $\mathrm{x}$ and $\mathrm{y}$ are in $S$ and are regular. Case 1: The slope of $S$ at $\mathrm{x}$ is the same as the slope at $\mathrm{y}$. In this case the line segment $L$ between $\mathrm{x}$ and $\mathrm{y}$ must be in $S$; if not, we can find regular points $\mathrm{p}$ and $\mathrm{q}$ on $L$ such that the midpoint of $\mathrm{p}$ and $\mathrm{q}$ is not in $S$, which contradicts $\phi$. Case 2: The slopes at $\mathrm{x}$ and $\mathrm{y}$ are different. Choose $\mathrm{z}$ as given by $\phi$. Reasoning as before, we see that the line segment between $\mathrm{z}$ and $\mathrm{y}$ must be in $S$, and the same for the segment between $\mathrm{z}$ and $\mathrm{x}$. But this shows that $S$ is 2-linked.

Conversely suppose that $S$ is two-linked. For any regular $\mathrm{x}$ and $\mathrm{y}$ in $S$, either they are one-linked — in which case they satisfy the $Parallel(\mathrm{x},\mathrm{y})$ clause — or they are (minimally) two-linked – in which case they satisfy the $\neg Parallel(\mathrm{x},\mathrm{y})$ clause. □

We now show that the assumption that $A$ is thin is necessary:

**Corollary 5.** *2-linked is not $FO_{LIN}$ definable, even over the family of 1-bounded semi-linear sets with ten singular points.*

Proof: This follows from the proof of Theorem 2. That proof uses a pair of 1-bounded hypersemi-linear structures with ten singular points which are elementarily equivalent in $FO_{LIN}$, but one structure is 2-linked and the other is not. □

**Corollary 6.** *For each $n > 1$, $n$-linked is not $FO_{LIN}$ definable over the collection of 1-bounded semi-linear sets with $n+8$ singular points.*

Proof: Start with the hypersemi-linear structures used in the preceding proof, and attach an infinitesimal small jagged line with $n-2$ segments to a regular point in the middle of one of the triangular windows. This gives a pair of elementarily equivalent structures, one of which is $n$-linked and the other is not. □

We now return to the thin semi-linear sets.

**Corollary 7.** *For each finite $n$ and $k$, $k$-linked is $FO_{LIN}$ definable over the collection $T_n$ of thin semi-linear sets with at most $n$ singular points.*

Proof: By Theorem 8, the query *Lin.Reach* is definable over $T_n$. Let $LR_n(\mathrm{x},\mathrm{y})$ be a $FO_{LIN}$ sentence defining it. Then $k$-linked is defined over $T_n$ by the following sentence:

$$\forall \mathrm{x} \forall \mathrm{y} \exists \mathrm{z}_1 \cdots \exists \mathrm{z}_k \, (\mathrm{z}_1 = \mathrm{x} \wedge \mathrm{z}_k = \mathrm{y} \wedge LR_n(\mathrm{z}_1,\mathrm{z}_2) \wedge \cdots \wedge LR_n(\mathrm{z}_{k-1},\mathrm{z}_k)) .$$

□

The following results can be proved by modifying these constructions in this paper. The proofs will be given in the full paper.

**Theorem 10.** *3-linked is not $FO_{LIN}$ definable over the family of thin semi-linear sets.*

**Theorem 11.** *There is no $FO_{LIN}$ sentence $\phi$ such that*

$$(4\text{-}linked \rightarrow \phi) \wedge (\phi \rightarrow Conn)$$

*holds in all thin semi-linear sets $A$. In particular, for each $k \geq 4$, $k$-linked is not $FO_{LIN}$ definable over the family of thin semi-linear sets.*

We do not know whether the above theorem can be improved by replacing 4-linked by 3-linked. Putting together all of the above results we have:

| **Summary of Definability for $n$-linked** | | | |
|---|---|---|---|
| $n$ | Semi-linear | ... and Thin | ... and Boundedly Many Singularities |
| 1 | Yes | Yes | Yes |
| 2 | No | Yes | Yes |
| $\geq 3$ | No | No | Yes |

Here 'Yes' in a box for integer $n$ and class $C$ means that $n$-linked is definable over $C$.

## 7 Conclusions and Future Work

Questions concerning definability with an extra predicate, even for a well-understood structure such as the real ordered group, turn out to be surprisingly complex. The answers are also a bit counterintuitive: the results here show that seemingly slight modifications of either the query definition or the class of definable sets can make or break definability. It would clearly be desirable to find general topological conditions on a family of sets that guarantee definability and include the interesting definable examples here. Our results, however, indicate that this will be a difficult (perhaps impossible) task.

Given the undefinability results, it seems natural to look for intermediate languages between the first-order linear and polynomial query languages, which can define the queries considered here. We have introduced the intermediate languages: $Lin - \Pi_1$, $Lin - \Sigma_1$, and $Lin - \Delta_1$ here. Although we have shown separation results for these classes, we have not investigated the separation question for the entire prefix hierarchy $Lin - \Pi_n$, $Lin - \Sigma_n$. Moreover, for several of the queries mentioned here, the question of membership in $Lin - \Pi_1$ is open.

## References


1. F. Afrati, S. Cosmadakis, S. Grumbach and G. Kuper. Linear vs. polynomial constraints in database query languages. In *Proceedings of the Second International Workshop on Principles and Practice of Constraint Programming* pp. 181-192. Springer LNCS 874, 1995.
2. F. Afrati, T. Andronikos,and T. Kavalieros. On the Expressiveness of Query Languages with Linear Constraints; Capturing Desirable Spatial Properties. In *Proceedings of the Second International Workshop on Constraint Database Systems, CDB '97*, pp. 105-115. Springer LNCS, Vol. 1191, 1997.
3. M. Benedikt, G. Dong, L. Libkin and L. Wong. Relational expressive power of constraint query languages. *J. ACM*, 45 (1998), pp. 1–34.
4. M. Benedikt and H. J. Keisler. Expressive Power of Unary Counters In *Structures in Logic and Computer Science*, Mycielski, Rozenberg, and Salomaa (eds.) Springer LNCS 1261 1997.
5. O. Chapuis and P. Koiran. Definability of Geometric properties in algebraically closed fields. To appear in *Mathematical Logic Quarterly.*
6. C.C. Chang and H.J. Keisler. *Model Theory.* North Holland, 1990.
7. P. Kanellakis, G. Kuper, and P. Revesz. Constraint query languages. *JCSS*, 51 (1995), 26–52.
8. B. Kuijpers, J. Paredaens and J. Van den Bussche. On topological elementary equivalence of spatial databases. In *Proceedings of the Sixth International Conference on Database Theory* , pp. 432–446. Springer LNCS 1186, 1997.
9. G. Kuper, L. Libkin and J. Paredaens, eds. *Constraint Databases.* Springer Verlag, 2000.