Definability with a Predicate for a Semi-Linear Set

Michael Benedikt H. Jerome Keisler

Abstract

We settle a number of questions concerning definability in first order logics with an extra predicate symbol ranging over semi-linear sets. 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.

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* ([12], [9]), first-order languages with extra free predicates for definable sets. 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 finite Boolean combination of linear inequalities with rational coefficients. A linear constraint is just a quantifier-free first order formula in the additive real ordered group $\mathcal{R} = \langle R, +, -, <, 0, 1 \rangle$. By quantifier elimination, each first order formula in the language of $\mathcal{R}$ (without parameters) defines a semi-linear set. Every $FO+LIN$ sentence defines a collection of semi-linear sets. $FO+LIN$ is the first-order language with an atomic formula for each linear inequality with rational coefficients and an extra predicate symbol S which ranges over semi-linear sets.

In $FO+POLY$, the binary 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 underlying structure (the additive real ordered group or the real 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 ([4], [12]). 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 ([7], [11]). 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:

$$Co.Linear = \{A \subset R^2 : \text{ all points in } A \text{ are collinear }\}$$

$$Is.Line = \{A \subset R^2 : A \text{ is a line }\}$$

$$Cont.Line = \{A \subset R^2 : A \text{ contains a line }\}$$

$$Lin.Reach = \{\langle A, \mathbf{a}, \mathbf{b}\rangle : A \subset R^2, \mathbf{a}, \mathbf{b} \in R^2,\ A \text{ contains the line segment } \mathbf{ab}\}$$

$$Lin.Meet = \{\langle A, \mathbf{a})\rangle : A \subset R^2, A \text{ contains two lines which intersect in } \mathbf{a}\}$$

In the last two examples, $FO+LIN$ has extra constant symbols ranging over R, in addition to the extra predicate symbol S. Each of the five examples is easily seen to be definable by a sentence in $FO+POLY$, since there one can quantify over lines. It is shown in [2] that $Is.Line$ is definable in $FO+LIN$. 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 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in A$ then $\mathbf{x} + (\mathbf{y} - \mathbf{z}) \in A$. The paper [2] proved more: There is a sentence of $FO+LIN$ which defines $Is.Line$ over all subsets of R^2, rather than just over the class of semi-linear sets. On the other hand, it is shown in [1] that $Co.Linear$ is *not* definable by a sentence in $FO + LIN$.

It was conjectured in [1] that $Cont.Line$ is not definable in $FO + LIN$, and also that $Lin.Meet$ is not definable in $FO+LIN$. 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? In this paper we introduce techniques for showing that a family of sets is not definable in $FO + LIN$. In the preliminary version of this paper [6], we used these techniques to show that $Cont.Line$ and $Lin.Reach$ are not definable in $FO + LIN$. Since then, we found much simpler proofs that $Cont.Line$ and $Lin.Reach$ are not definable in $FO + LIN$, which we give in Section 2. We will use the new techniques to show that several other queries are not definable in $FO+LIN$, including: "The boundary of A contains the line segment **ab**", and "A is 2-linked", (i.e. any two points of A can be connected by a path with at most two legs).

To complement these negative results, we will prove positive results, showing that queries which are very similar to $Cont.Line$ and $Lin.Reach$ are definable in $FO + LIN$. For example, we show that the set of semi-linear sets A such that the boundary of A contains a line is definable in $FO + LIN$.

We will also investigate definability in two languages between $FO+LIN$ and $FO + POLY$. The language $\Pi_1 + LIN$ is obtained by adding universal quantifiers over lines to $FO + LIN$, while $\Sigma_1 + LIN$ is obtained by adding existential quantifiers over lines to $FO+LIN$. While $Cont.Line$ is obviously definable in $\Sigma_1 + LIN$, we will show that $Cont.Line$ and related queries are not definable in $\Pi_1 + LIN$.

Organization: In Section 2 we introduce notation and basic definitions, including the language $FO + LIN$, and prove some elementary undefinability results. In Section 3 we give a brief review of the notions we need from nonstandard analysis, and prove some model-theoretic results about the additive ordered group of hyperreal numbers which are useful in undefinability proofs. Section 4 concerns the definability or undefinability of queries related to $Lin.Reach$ in $FO+LIN$. In Section 5 we introduce the universal and existential second-order extensions of $FO+LIN$ and show that these extensions are proper. Section 6 is about the definability problem for queries related to $Cont.Line$ in $FO+LIN$ and its second-order extensions. Section 7 concerns the $FO + LIN$-definability problem for the property that any two points are n-linked (connected by a polygonal path with at most n segments) for various n. Section 8 discusses extensions of our results to higher dimensions. Finally, conclusions are given in Section 9.

2 The Language $FO + LIN$

We will introduce some basic notation for the language $FO + LIN$. We will then prove some elementary undefinability results which are based on the well-known fact that multiplication is not definable in the additive real ordered group.

2.1 Notation

The sets of real numbers, rational numbers, and natural numbers are denoted by R, Q, and N respectively. We start with the additive real ordered group $\mathcal{R} = \langle R, +, -, <, 0, 1 \rangle$ which serves as a framework for all standard models, and a signature $\mathcal{S}$ consisting of predicate and constant symbols. For simplicity we will confine our attention to the case where $\mathcal{S} = \langle S, \mathbf{c} \rangle$ has only one binary predicate symbol S, and a tuple $\mathbf{c}$ of constant symbols of length l. The tuple $\mathbf{c}$ may be empty, that is, $l = 0$.

We let $FO + LIN$ be the first-order language with the vocabulary $\mathcal{S} \cup \{+, -, <, 0, 1\}$. The standard first order structures for this vocabulary have the form $(\mathcal{R}, A, \mathbf{a})$ where $A \subset R^2$ interprets S and $\mathbf{a} \in R^l$ interprets $\mathbf{c}$. Since all the structures under consideration have the same $\mathcal{R}$ part, we concentrate on the other part and define a real structure with **signature** $\mathcal{S}$ to be an object $\mathcal{A} = \langle A, \mathbf{a} \rangle$ where $A \subset R^2$ interprets S and $\mathbf{a} \in R^l$ interprets $\mathbf{c}$.

A structure $\mathcal{A} = \langle A, \mathbf{a} \rangle$ **satisfies** a sentence $\psi \in FO + LIN$, in symbols $\mathcal{A} \models \psi$, exactly when the corresponding first order structure $(\mathcal{R}, \mathcal{A}) = (\mathcal{R}, A, \mathbf{a})$ satisfies ψ. By a **semi-linear relation** we mean a relation which is definable by a first-order formula *without parameters* in $\mathcal{R}$. If the relation A is semi-linear and $\mathbf{a}$ is also definable in $\mathcal{R}$, 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 structures with signature $\mathcal{S}$ is called a (Boolean) **query**. We say that a query X is **$FO{+}LIN$-definable** if there is an $FO + LIN$ sentence ψ such that:

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

A query X with signature $\mathcal{S} = (S, \mathbf{c})$ where $|c| = l$ can also be viewed as a function from $P(R^2)$ to R^l, but this view does not match the syntax of $FO + LIN$ as well. 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, we say that a query X is **$FO+LIN$-definable over** F if there is a sentence ψ of $FO + LIN$ such that:

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

Thus when a base F is not mentioned, it is understood to be the family of all semi-linear structures.

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.

One particular query which will frequently be used as a base query in this paper is the class of thin semi-linear structures, defined by

$$Thin = \{\mathcal{A} : A \text{ is semi-linear and has empty interior}\}.$$

We will study the definability of some natural queries in $FO+LIN$, and in extensions of $FO + LIN$. We will need the following notation.

Definition 2.1 *The* **boundary of** *A is the set ∂A of all points $\mathbf{x}$ such that every open rectangle around $\mathbf{x}$ meets both A and $R^2 \setminus A$.*

A point $\mathbf{x}$ will be called **regular** *in a semi-linear set A if for some open rectangle U around $\mathbf{x}$, and $A \cap U$ is a line segment with $\mathbf{x}$ in $A \cap U$ and $\mathbf{x}$ not an endpoint of $A \cap U$.*

$\mathbf{x}$ is said to be **singular** *in A if $\mathbf{x}$ is a boundary point of A which is not regular in A.*

If $\mathbf{x}$ is singular in A, the **degree** *of $\mathbf{x}$ in A is the number of edges in the boundary of A which end in $\mathbf{x}$.*

Lemma 2.2 *Let n be a natural number. Each of the following queries is $FO + LIN$-definable.*

$$Thin.$$

$$Card(n) = \{\mathcal{A} : A \text{ has cardinality } n\}.$$

$$Bounded = \{\mathcal{A} : A \text{ is bounded}\}.$$

$$Singular(n) = \{\mathcal{A} : A \text{ has exactly } n \text{ singular points}\}.$$

$$Degree(n) = \{\mathcal{A} : \text{ each singular } \mathbf{x} \text{ in } A \text{ has degree } \leq n\}.$$

Proof: We prove that $Singular(n)$ is $FO + LIN$-definable. It suffices to show that there is an $FO + LIN$ formula which says that $\mathbf{x}$ is a singular point of S. Let $U(\mathbf{y}, \mathbf{z})$ denote the open rectangle with corners $\mathbf{y}, \mathbf{z}$. Then $\mathbf{x} \in U(\mathbf{y}, \mathbf{z})$ can be expressed by the formula

$$y_1 < x_1 < z_1 \wedge y_2 < x_2 < z_2.$$

Using this, it is an easy exercise to find an $FO + LIN$ formula $\partial S(\mathbf{x})$ which says that $\mathbf{x}$ is on the boundary of S. The following formula $Reg(\mathbf{x})$ says that $\mathbf{x}$ is regular in S:

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

Then the formula $\partial S(\mathbf{x}) \cap \neg Reg(\mathbf{x})$ says that $\mathbf{x}$ is a singular point of S.

■

2.2 Elementary undefinability results

We recall a known result from [1], which is a consequence of the classical result that the product function cannot be defined in the first order theory of $\mathcal{R}$.

Theorem 2.3 *The query Co.Linear is not $FO + LIN$-definable. In fact,*

$$Co.Linear \cap Card(3)$$

is not $FO + LIN$-definable.

One way to prove this is to note that if the query were definable, then there would be a relation in R^3 definable in $\mathcal{R}$ (with no extra relation symbols) that agrees with the graph of the multiplication function on the rationals. It is easy (e.g. by examining a cell decomposition for the relation, which would have to be defined using a finite set of linear functions) that no such relation can exist. The same argument shows that the set of collinear triples $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ in R^2 is not definable in the structure $\mathcal{R}$. Using this fact, we now show that the queries *Lin.Reach* and *Cont.Line* are undefinable in $FO + LIN$.

Theorem 2.4 *The query Lin.Reach is not $FO + LIN$-definable. In fact,*

$$Lin.Reach \cap Singular(1) \cap Degree(4)$$

is not $FO + LIN$-definable.

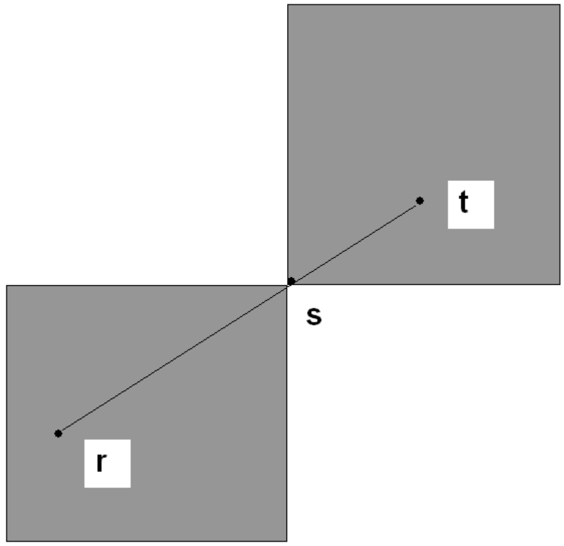

Figure 1: Theorem 2.4—The set $A_{\mathbf{s}}$.

Proof: Consider three rational points $\mathbf{r}, \mathbf{s}, \mathbf{t}$ in R^2. Without loss of generality, we restrict our attention to the case that $r_1 < s_1 < t_1$ and $r_2 < s_2 < t_2$. Let

$$A_{\mathbf{s}} = \{\mathbf{x} : x_1 \leq s_1 \text{ iff } x_2 \leq s_2\}.$$

Then the structure $\langle A_{\mathbf{s}}, \mathbf{r}, \mathbf{t} \rangle$ is semi-linear and has one singular point of degree 4. See Figure 1. $\langle A_{\mathbf{s}}, \mathbf{r}, \mathbf{t} \rangle$ satisfies *Lin.Reach* if and only if the triple $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is collinear, because the line segment $\mathbf{rt}$ is contained in $A_{\mathbf{s}}$ if and only if it passes through $\mathbf{s}$. It follows that $Lin.Reach \cap Singular(1)$ is not definable in $FO + LIN$. ∎

Theorem 2.5 *The query Cont.Line is not $FO + LIN$-definable. In fact,*

$$Cont.Line \cap Singular(3) \cap Degree(4)$$

is not $FO + LIN$-definable.

Proof: Consider triples of rational points $\mathbf{r}, \mathbf{s}, \mathbf{t}$ in R^2 with $r_1 < s_1 < t_1$ and $r_2 < s_2 < t_2$. Let A be the semi-linear set shown in Figure 2, with the understanding that all of the boundary lines have rational slope. (Ignore the dotted line for now). The set A has three singular points of degree 4, and contains a line if and only if the triple $(\mathbf{r}, \mathbf{s}, \mathbf{t})$ is collinear. ∎

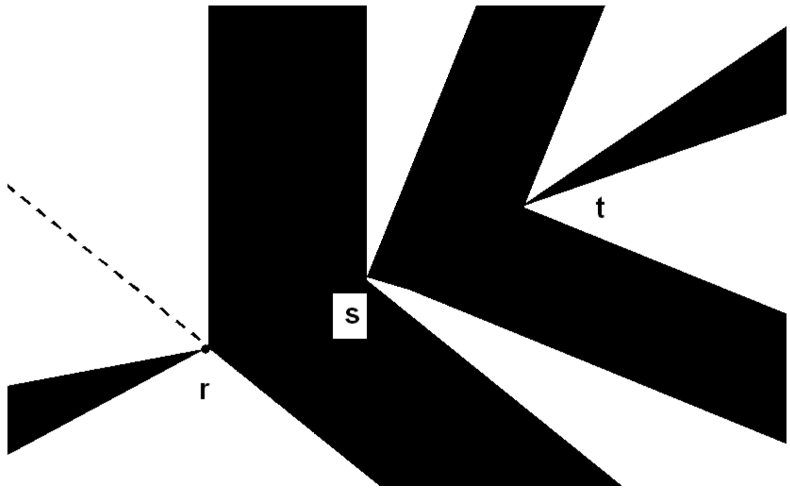

Figure 2: Theorem 2.5 and Corollary 2.6.

Corollary 2.6 *The query Lin.Meet is not $FO + LIN$-definable. In fact,*

$$Lin.Meet \cap Singular(3) \cap Degree(4)$$

is not $FO + LIN$-definable.

Proof: Use the proof of Theorem 2.5, but modify the set A by rotating the vertical ray through $\mathbf{r}$ to the dotted line. Now the dotted line is a boundary line of A through $\mathbf{r}$. ∎

By contrast, it was shown in [1] that $Lin.Meet$ is $FO + LIN$-definable over the class of semi-linear structures $\mathcal{A}$ such that A is exactly the union of two lines. We can improve this using Lemma 2.2.

Corollary 2.7 *The query Lin.Meet is $FO+LIN$-definable over the class F of semi-linear structures $\mathcal{A}$ such that A is exactly the union of finitely many lines.*

Proof: A structure $\langle A, \mathbf{a} \rangle \in F$ belongs to $Lin.Meet$ if and only if $\mathbf{a}$ is singular in A. ∎

3 Infinitesimals and Undefinability

We will use notions from nonstandard analysis as a tool in proving further undefinability results. However, the statements of these undefinability results are standard. The main place where nonstandard notions will be used is in characterizations of definability in query languages (as in, e.g., [5]). We assume familiarity with basic notions of nonstandard analysis (see [8]), but give a briefer-than-brief review here.

3.1 Some Nonstandard Analysis

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 equality and 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 $\psi(v_1, \ldots, v_n)$ and any list $a_1, \ldots, a_n$ of elements from $V(R)$, $\psi(a_1, \ldots, a_n)$ is true in $V(R)$ if and only if $\psi({}^*a_1, \ldots, {}^*a_n)$ is true in $V({}^*R)$.
4. **(Saturation Principle)** For any set of sets $A \in V(R)$, any countable decreasing chain of nonempty sets $B_n \in {}^*A$ has a nonempty intersection.

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 are 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 $^*\mathcal{R}$, or even in the structure $\langle {}^*R, {}^*N, <, +, -, \times \rangle$.

The Saturation Principle says that any decreasing chain of nonempty internal sets has a nonempty intersection. A consequence we will need is that the structure $^*\mathcal{R}$ is ω_1-saturated, that is, any countable decreasing chain of nonempty definable sets in $^*\mathcal{R}$ has a nonempty intersection.

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*, y is a *hypermultiple* of x if $y = zx$ for some $z \in {}^*Z$, 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 ${}^*\mathcal{R}$ is hypersemi-linear. In fact, any set which is definable by a first order hyperformula in ${}^*\mathcal{R}$ (or, equivalently, by a hyperfinite Boolean combination of linear constraints with hyperrational coefficients) is still hypersemi-linear.

3.2 Elementary equivalence over the hyperreals

We will present some model-theoretic results about the hyperreal ordered additive group ${}^*\mathcal{R}$ which are useful for proving undefinability in $FO + LIN$. $\mathbf{a}, \mathbf{b}$ will denote finite sequences of hyperreal numbers. $\equiv$ is the elementary equivalence relation for first order logic.

Here is a useful "nonstandard" sufficient condition for a query to be $FO+LIN$-undefinable.

Proposition 3.1 *Let X and F be queries with signature $\mathcal{S}$. Suppose that there are hyperstructures $\mathcal{A}, \mathcal{B} \in {}^*F$ such that $\mathcal{A} \in {}^*X$ and $\mathcal{B} \notin {}^*X$, but*

$$({}^*\mathcal{R}, \mathcal{A}) \equiv ({}^*\mathcal{R}, \mathcal{B}).$$

Then X is not $FO + LIN$-definable over F.

Proof: Assume that some $FO + LIN$ sentence ψ defines X over F. By the Transfer Principle, since $\mathcal{A} \in {}^*X$, we have $({}^*\mathcal{R}, \mathcal{A}) \models \psi$. Similarly, since $\mathcal{B} \notin {}^*X$, we have $({}^*\mathcal{R}, \mathcal{B}) \models \neg\psi$. This contradicts the hypotheses, so X cannot be $FO + LIN$-definable over F. ■

In order to use this sufficient condition, we need conditions which imply that two structures are elementarily equivalent. One such condition from the literature is partial isomorphism. A **partial isomorphism** from $\mathcal{A}$ to $\mathcal{B}$ is a binary relation $\simeq$ between tuples $\mathbf{a}$ in $\mathcal{A}$ and $\mathbf{b}$ in $\mathcal{B}$ of the same length such that:

(a) $\emptyset \simeq \emptyset$,

(b) If $\mathbf{a} \simeq \mathbf{b}$ then $\mathbf{a}$ and $\mathbf{b}$ satisfy the same atomic sentences,

(c) If $\mathbf{a} \simeq \mathbf{b}$ then for each $c \in \mathcal{A}$ there exists $d \in \mathcal{B}$ with $(\mathbf{a}, c) \simeq (\mathbf{b}, d)$, and vice versa.

Karp's Theorem [Ka65], states that two structures $\mathcal{A}$ and $\mathcal{B}$ are partially isomorphic if and only if they are elementarily equivalent in the infinitary language $L_{\infty,\omega}$ with finite quantifiers, negation, and infinite conjunctions and disjunctions.

Our next goal will be to develop a particular partial automorphism of ${}^*\mathcal{R}$, which will be denoted by $\sim$. As a preliminary step we will introduce the notion of a dispersed tuple.

Let ${}^*\mathcal{R}(\mathbf{a})$ be the substructure of ${}^*\mathcal{R}$ generated by a k-tuple $\mathbf{a} = \langle a_1, \cdots a_k \rangle$. Then the universe of ${}^*\mathcal{R}(\mathbf{a})$ is the set of all hyperreal numbers of the form

$$\mathbf{p} \cdot \mathbf{a} + s = p_1 a_1 + \cdots + p_k a_k + s$$

where $\mathbf{p} = \langle p_1, \ldots, p_k \rangle \in Q^k$ and $s \in Q$.

Definition 3.2 *A tuple* $\mathbf{a}$ *is* **dispersed** *(in ${}^*\mathcal{R}$) iff ${}^*\mathcal{R}(\mathbf{a})$ does not contain any positive infinitesimals.*

A k-tuple $\mathbf{x} = (x_1, \ldots, x_k)$ in ${}^*R$ is said to be **infinitesimal** *iff $x_i \approx 0$ for each $i \leq k$. Two k-tuples $\mathbf{x}$ and $\mathbf{y}$ are said to be* **infinitesimally close***, in symbols $\mathbf{x} \approx \mathbf{y}$, iff $x_i \approx y_i$ for each $i \leq k$.*

We say that $\mathbf{a}$ *and* $\mathbf{b}$ *are* **equivalent over the infinitesimals***, in symbols $\mathbf{a} \sim \mathbf{b}$, iff $({}^*\mathcal{R}, \mathbf{a}, \mathbf{x}) \equiv ({}^*\mathcal{R}, \mathbf{b}, \mathbf{x})$ for all infinitesimal tuples $\mathbf{x}$.*

In other words, $\mathbf{a} \sim \mathbf{b}$ if and only if the expansions of $({}^*\mathcal{R}, \mathbf{a})$ and $({}^*\mathcal{R}, \mathbf{b})$ formed by adding constant symbols for every infinitesimal are elementarily equivalent, that is,

$$({}^*\mathcal{R}, \mathbf{a}, \delta)_{\delta \approx 0} \equiv ({}^*\mathcal{R}, \mathbf{b}, \delta)_{\delta \approx 0}.$$

We collect some easy facts about dispersed tuples and the $\sim$ relation in the next two lemmas.

Lemma 3.3 *(i) Any tuple* $\mathbf{a}$ *of real numbers is dispersed.*

(ii) Suppose $\mathbf{a}$ *is dispersed and $({}^*\mathcal{R}, \mathbf{a}) \equiv ({}^*\mathcal{R}, \mathbf{b})$. Then* $\mathbf{b}$ *is dispersed.*

(iii) Suppose $\mathbf{a}$ *is dispersed and $\mathbf{b} \in ({}^*\mathcal{R}(\mathbf{a}))^k$ for some k. Then* $\mathbf{b}$ *is dispersed.*

Proof: (i) If $\mathbf{a} \in R^n$ then ${}^*\mathcal{R}(\mathbf{a})$ is a subset of R, so it contains no positive infinitesimals.

(ii) Suppose $\mathbf{b}$ is not dispersed. Then $\mathbf{p} \cdot \mathbf{b} + s \in {}^*\mathcal{R}(\mathbf{b})$ is positive infinitesimal for some $\mathbf{p} \in Q^k$ and $s \in Q$. Then

$$0 < \mathbf{p} \cdot \mathbf{b} + s < 1/n$$

for each positive integer n. Since $({}^*\mathcal{R}, \mathbf{a}) \equiv ({}^*\mathcal{R}, \mathbf{b})$,

$$0 < \mathbf{p} \cdot \mathbf{a} + s < 1/n$$

for each positive integer n, contradicting the hypothesis that $\mathbf{a}$ is dispersed.

(iii) We have ${}^*\mathcal{R}(\mathbf{b}) \subseteq {}^*\mathcal{R}(\mathbf{a})$, so ${}^*\mathcal{R}(\mathbf{b})$ has no positive infinitesimals and thus $\mathbf{b}$ is dispersed. ■

Lemma 3.4 *(i) $\sim$ is an equivalence relation.*

(ii) $\mathbf{a} \sim \mathbf{b}$ implies $({}^*\mathcal{R}, \mathbf{a}) \equiv ({}^*\mathcal{R}, \mathbf{b})$.

(iii) Suppose $\mathbf{a} \sim \mathbf{b}$ and $\mathbf{x}$ is infinitesimal. Then

$$(\mathbf{a}, \mathbf{x}) \sim (\mathbf{b}, \mathbf{x}).$$

(iv) Suppose $\mathbf{a} \sim \mathbf{b}$, $\mathbf{x}$ is infinitesimal, and $|\mathbf{x}| = |\mathbf{a}|$. Then

$$\mathbf{a} + \mathbf{x} \sim \mathbf{b} + \mathbf{x}.$$

Proof: Parts (i) and (ii) are trivial.

(iii) Since $\mathbf{a} \sim \mathbf{b}$, we have

$$({}^*\mathcal{R}, \mathbf{a}, \mathbf{x}, \mathbf{y}) \equiv ({}^*\mathcal{R}, \mathbf{b}, \mathbf{x}, \mathbf{y}). \tag{1}$$

for any infinitesimal tuple $\mathbf{y}$. This shows that $(\mathbf{a}, \mathbf{x}) \sim (\mathbf{b}, \mathbf{x})$.

(iv) It follows from (1) that

$$({}^*\mathcal{R}, \mathbf{a} + \mathbf{x}, \mathbf{y}) \equiv ({}^*\mathcal{R}, \mathbf{b} + \mathbf{x}, \mathbf{y}),$$

for any infinitesimal $\mathbf{y}$, and hence $\mathbf{a} + \mathbf{x} \sim \mathbf{b} + \mathbf{x}$. ■

Lemma 3.5 *Suppose $\mathbf{a}$ is dispersed. Then for any n-tuple $\mathbf{b}$ in ${}^*R$ there is an n-tuple $\mathbf{c} \approx \mathbf{b}$ such that $(\mathbf{a}, \mathbf{c})$ is dispersed.*

Proof: The result is trivial for $n = 0$. For the case $n = 1$, let $b \in {}^*R$.

If $(\mathbf{a}, b)$ is dispersed we simply take $c = b$. If $b \approx 0$ we may take $c = 0$, since $(\mathbf{a}, 0)$ is dispersed by Lemma 3.3 (iii).

We now consider the general case where $(\mathbf{a}, b)$ is not dispersed. Then there exist rational $\mathbf{p}$, q, and s such that $\mathbf{p} \cdot \mathbf{a} + qb + s$ is a positive infinitesimal. Since $\mathbf{a}$ is dispersed, $\mathbf{p} \cdot \mathbf{a} + s$ cannot be a positive infinitesimal. Therefore $q \neq 0$. Then $(1/q)(\mathbf{p} \cdot \mathbf{a} + s) + b \approx 0$, and $b \approx c$ where $c = -(1/q)(\mathbf{p} \cdot \mathbf{a} + s)$.

We have $c \in {}^*\mathcal{R}(\mathbf{a})$, so $(\mathbf{a}, c)$ is dispersed by Lemma 3.3 (iii). This proves the lemma for $n = 1$.

We now argue by induction. Assume that the result holds for n where $n > 0$, and let $(\mathbf{b}, d) \in ({}^*R)^{n+1}$. By inductive hypothesis there exists $\mathbf{c} \approx \mathbf{b}$ such that $(\mathbf{a}, \mathbf{c})$ is dispersed. Using the result for $n = 1$, there exists $e \approx d$ such that $(\mathbf{a}, \mathbf{c}, e)$ is dispersed. Since $(\mathbf{c}, e) \approx (\mathbf{b}, d)$, the proof is complete. ∎

Proposition 3.6 *If* $({}^*\mathcal{R}, \mathbf{a}) \equiv ({}^*\mathcal{R}, \mathbf{b})$ *and* $\mathbf{a}$ *is dispersed, then* $\mathbf{a} \sim \mathbf{b}$.

Proof: By Lemma 3.3 (ii), $\mathbf{b}$ is dispersed. Since the theory of $\mathcal{R}$ admits quantifier elimination, it suffices to prove that for each infinitesimal n-tuple $\mathbf{x}$, $({}^*\mathcal{R}, \mathbf{a}, \mathbf{x})$ and $({}^*\mathcal{R}, \mathbf{b}, \mathbf{x})$ satisfy the same quantifier-free formulas. For this it suffices to show that for each rational $\mathbf{p}$, s, and $\mathbf{q}$, if $\mathbf{p} \cdot \mathbf{a} + s + \mathbf{q} \cdot \mathbf{x} \geq 0$ then $\mathbf{p} \cdot \mathbf{b} + s + \mathbf{q} \cdot \mathbf{x} \geq 0$. Suppose $\mathbf{p} \cdot \mathbf{a} + s + \mathbf{q} \cdot \mathbf{x} \geq 0$. Since $\mathbf{x}$ is infinitesimal, $\mathbf{q} \cdot \mathbf{x} \approx 0$. We distinguish three cases.

Case 1: $\mathbf{p} \cdot \mathbf{a} + s > 0$. Then $\mathbf{p} \cdot \mathbf{b} + s > 0$. Since $\mathbf{b}$ is dispersed in ${}^*\mathcal{R}$, $\mathbf{p} \cdot \mathbf{b} + s > t$ for some real $t > 0$. Therefore $\mathbf{p} \cdot \mathbf{b} + s + \mathbf{q} \cdot \mathbf{x} > t/2 > 0$.

Case 2: $\mathbf{p} \cdot \mathbf{a} + s = 0$. Then $\mathbf{p} \cdot \mathbf{b} + s = 0$, so

$$\mathbf{p} \cdot \mathbf{b} + s + \mathbf{q} \cdot \mathbf{x} = \mathbf{q} \cdot \mathbf{x} = \mathbf{p} \cdot \mathbf{a} + s + \mathbf{q} \cdot \mathbf{x} \geq 0.$$

Case 3: $\mathbf{p} \cdot \mathbf{a} + s < 0$. Then $\mathbf{p} \cdot \mathbf{a} + s < t$ for some real $t < 0$, and $\mathbf{p} \cdot \mathbf{a} + s + \mathbf{q} \cdot \mathbf{x} < t/2 < 0$, contradiction. ∎

We call a k-tuple $\mathbf{a}$ of hyperreal numbers **linearly independent** if $\mathbf{p} \cdot \mathbf{a} + s \neq 0$ for every non-zero $(k+1)$-tuple of rational numbers $(\mathbf{p}, s)$, and **algebraically independent** if $f(\mathbf{a}) \neq 0$ for every non-zero polynomial $f(\mathbf{u})$ with rational coefficients.

Corollary 3.7 *If* $\mathbf{a}$ *is a linearly independent tuple of (standard) real numbers, and* $\mathbf{b} \approx \mathbf{a}$, *then* $({}^*\mathcal{R}, \mathbf{a}) \sim ({}^*\mathcal{R}, \mathbf{b})$.

Proof: We first note that $({}^*\mathcal{R}, \mathbf{a}) \equiv ({}^*\mathcal{R}, \mathbf{b})$, because for any non-zero rational $(\mathbf{p}, s)$, we have

$$0 \neq \mathbf{p} \cdot \mathbf{a} + s = \mathbf{p} \cdot {}^\circ\mathbf{b} + s = {}^\circ(\mathbf{p} \cdot \mathbf{b} + s),$$

and hence

$$\mathbf{p} \cdot \mathbf{a} + s \geq 0 \text{ iff } \mathbf{p} \cdot \mathbf{b} + s \geq 0.$$

The result now follows from Proposition 3.6. ∎

Proposition 3.8 *The relation $\sim$ is a partial automorphism of ${}^*\mathcal{R}$.*

Proof: Suppose $\mathbf{a} \sim \mathbf{b}$ and $c \in {}^*R$. By Lemma 3.4 (ii), $\mathbf{a}$ and $\mathbf{b}$ satisfy the same atomic formulas in ${}^*\mathcal{R}$. By Lemma 3.5 there is a tuple $\mathbf{a}' \approx \mathbf{a}$ and an element $c' \approx c$ such that $(\mathbf{a}', c')$ is dispersed. Then $\mathbf{a}'$ is dispersed by Lemma 3.3 (iii).

Let $\mathbf{x} = \mathbf{a}' - \mathbf{a}$ and $y = c' - c$. $\mathbf{x}$ and y are infinitesimal. By Lemma 3.4 (iv) we have $\mathbf{a} + \mathbf{x} \sim \mathbf{b} + \mathbf{x}$, or in other words, $\mathbf{a}' \sim \mathbf{b}'$.

By elementary equivalence, each first order formula satisfied by c' in $({}^*\mathcal{R}, \mathbf{a}')$ is satisfiable in $({}^*\mathcal{R}, \mathbf{b}')$. Since ${}^*\mathcal{R}$ is ω_1-saturated, it follows that there exists $d' \in {}^*R$ such that $({}^*\mathcal{R}, \mathbf{a}', c') \equiv ({}^*\mathcal{R}, \mathbf{b}', d')$. Since $(\mathbf{a}', c')$ is dispersed, we have $(\mathbf{a}', c') \sim (\mathbf{b}', d')$ by Proposition 3.6. Using Lemma 3.4 (iv) again, we have $(\mathbf{a}' - \mathbf{x}, c' - y) \sim (\mathbf{b}' - \mathbf{x}, d' - y)$. Taking $d = d' - y$, we may write this as $(\mathbf{a}, c) \sim (\mathbf{b}, d)$. ■

We now apply the preceding results to get a sufficient condition for two $FO + LIN$ structures to be elementarily equivalent. In fact, we will get the even stronger conclusion that the structures are partially isomorphic, and hence elementarily equivalent in the infinitary language $L_{\infty,\omega}$.

Theorem 3.9 *Let A and B be subsets of ${}^*R^2$. Suppose that:*

(a) $\mathbf{a} \sim \mathbf{b}$*,*

(b) For all $c, d \in {}^*R^2$ such that $(\mathbf{a}, c) \sim (\mathbf{b}, d)$, we have $c \in A$ if and only if $d \in B$.

Then the relation

$$\mathbf{u} \simeq \mathbf{v} \text{ iff } (\mathbf{a}, \mathbf{u}) \sim (\mathbf{b}, \mathbf{v})$$

is a partial isomorphism from $({}^*\mathcal{R}, A, \mathbf{a})$ to $({}^*\mathcal{R}, B, \mathbf{b})$, and hence $({}^*\mathcal{R}, A, \mathbf{a})$, $({}^*\mathcal{R}, B, \mathbf{b})$ are elementarily equivalent in $L_{\infty,\omega}$.

Proof: By (a) and Proposition 3.8, $\simeq$ is a partial isomorphism from $({}^*\mathcal{R}, \mathbf{a})$ to $({}^*\mathcal{R}, \mathbf{b})$. By (b), the relation $\mathbf{u} \simeq \mathbf{v}$ implies that $\mathbf{u}$ and $\mathbf{v}$ satisfy the same atomic formulas in the structures $({}^*\mathcal{R}, A, \mathbf{a})$ and $({}^*\mathcal{R}, B, \mathbf{b})$. The result now follows from Karp's Theorem. ■

3.3 Good Automorphisms of the Hyperreal Numbers

Our next theorem shows that the structure ${}^*\mathcal{R}$ has automorphisms which are good in the following sense.

Definition 3.10 *Let δ be a positive infinitesimal hyperrational number, let m be a hyperrational number between 0 and 1, and let f be a function from $^*R$ to $^*R$. We say that the triple (δ, m, f) is* **good** *if:*

(a) $f(\varepsilon) = \varepsilon$ for all $\varepsilon \approx 0$.
(b) $f(n) = n$ for all $n \in {}^*N$.
(c) $f(m) \neq m$.
(d) f is an automorphism of $({}^*\mathcal{R}, L)$ where

$$L = \{(x, mx + K\delta) : x \in {}^*R, K \in {}^*Z\}.$$

The key condition is (d), which implies that f is an automorphism of $^*\mathcal{R}$. The extra binary relation L is internal, and is the union of a set of parallel lines of slope m with the vertical distance δ between neighboring lines. We will use good triples to build pairs of elementarily equivalent hypersemi-linear structures $({}^*\mathcal{R}, \mathcal{A})$ and $({}^*\mathcal{R}, f(\mathcal{A}))$, which will be used to show that certain queries are not $FO + LIN$-definable.

Theorem 3.11 *There exists a good triple (δ, m, f).*

Proof: Let

$$D = \{K : K/n \in {}^*Z \text{ for all } n \in N\}.$$

D is the largest divisible subgroup of $\langle {}^*Z, +, -, 0 \rangle$. D is nontrivial, since $K! \in D$ for any infinite $K \in {}^*N$.

We first choose δ. Pick a positive hyperinteger $H \in D$, and let $\delta = H^{-2}$.

We next choose $m \in {}^*Q$ such that:

- $m = J/H$ where $J \in {}^*N$ and $0 < J < H$.
- The standard part $^\circ m$ is irrational.

To get such an m, just choose any irrational $r \in (0, 1)$ and take J so that $J/H \approx r$. We note that m has standard part $^\circ m \in (0, 1)$.

Given a function f on the hyperreal line, let $\Delta(x) = f(x) - x$ and let f^2 be the map on the hyperreal plane defined by $f^2((x, y)) = (f(x), f(y))$.

We first obtain a sufficient condition for f to preserve the binary relation L in part (d).

Claim 1 *Suppose f has the property that $H\Delta(u) \in D$ for every $u \in {}^*R$. Then for all $(x, y) \in R^2$ we have $(x, y) \in L$ if and only if $f^2(x, y) \in L$.*

Proof: Consider any point $(x, y) \in {}^*R^2$. By definition, $(x, y) \in L$ if and only if $y - mx$ is a hypermultiple of δ. We must show that $y - mx$ is a hypermultiple of δ if and only if $f(y) - mf(x)$ is a hypermultiple of δ. It suffices to prove that the difference

$$(f(y) - mf(x)) - (y - mx) = \Delta(y) - m\Delta(x)$$

is a hypermultiple of δ. By hypothesis, $\Delta(y) = Y/H = YH\delta$ and $\Delta(x) = X/H = XH\delta$ for some hyperintegers $Y, X \in D$. Therefore

$$\Delta(y) - m\Delta(x) = YH\delta - (J/H)XH\delta = (YH - JX)\delta.$$

■

We will now define a function f by transfinite recursion such that (δ, m, f) is good. We will build f in such a way that $H\Delta(x) \in D$ for all $x \in {}^*R$. In view of Claim 1, this will insure that f preserves the relation L.

Let R_0 be the smallest divisible subgroup of $\langle {}^*R, +, -, 0, 1 \rangle$ that contains ${}^*Z$ 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 $x = p + n + \varepsilon$ for some $p \in Q \cap [0, 1)$, $n \in {}^*Z$, and $\varepsilon \approx 0$. Moreover, for each $x \in R_0$, the decomposition $x = p + n + \varepsilon$ is unique.*

Proof: The first statement is clear. To prove uniqueness, suppose $p + n + \varepsilon = p' + n' + \varepsilon'$. Then $n - n' = (p' - p) + (\varepsilon' - \varepsilon)$. 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 $\varepsilon' = \varepsilon$. ■

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_α, $\alpha < \mathbf{c}$ 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 $\langle {}^*R, +, -, 0, 1 \rangle$ that contains both R_α and r_α. 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_α.

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_α contains fewer than $\mathbf{c}$ real numbers.*

Proof: We will prove by transfinite induction that for each $\alpha < \mathbf{c}$,

$$|R_\alpha \cap R| \leq \aleph_0 + |\alpha| < \mathbf{c}.$$

By Claim 2, $R_0 \cap R = Q$, so $|R_0 \cap R| = \aleph_0$. If α is a limit ordinal, then by inductive hypothesis,

$$|R_\alpha \cap R| \leq \sum_{\beta < \alpha} (\aleph_0 + |\beta|) = \aleph_0 + |\alpha|.$$

Now assume the result for α. By Claim 3 we have

$$|R_{\alpha+1} \cap R| \leq |Q| \times |R_\alpha \cap R| \leq \aleph_0 + |\alpha|.$$

Thus the result holds for $\alpha + 1$, and the induction is complete. ■

Claim 5 *There exist infinitely many positive infinitesimal $\varepsilon \in {}^*Q$ such that $H\varepsilon \in D$.*

Proof: By hypothesis, $H \in D$. Using the Transfer Principle, there is a least $K \in {}^*N$ such that $H/K \notin {}^*Z$. Since $H \in D$, $H/n \in {}^*Z$ for all $n \in N$, and hence K must be infinite. Therefore for all infinite $L < K$ in ${}^*N$ we have $H/L \in D$, so $\varepsilon = 1/L$ has the required property. ■

With another transfinite recursion, we build an increasing chain of functions $f \restriction R_\alpha : R_\alpha \to {}^*R$, $\alpha < \mathbf{c}$. Let $f \restriction R_0$ be the identity function on R_0, take unions at limit ordinals, and define $f \restriction R_{\alpha+1}$ as follows when $r_\alpha \notin R_\alpha$: By Claim 5 we may choose a positive infinitesimal $\varepsilon_\alpha \in {}^*Q$ such that $H\varepsilon_\alpha \in D$. For $x \in R_{\alpha+1}$, put $x = pr_\alpha + n + y$ as in Claim 3, and define $f(x) = p(r_\alpha + \varepsilon_\alpha) + f(n + y)$. 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 $f(m) \neq m$. Consider the first β such that $m \in R_\beta$. Since ${}^o m \in [0,1] \setminus Q$, $m \notin R_0$, so $\beta > 0$. Then β must be a successor ordinal, $\beta = \alpha + 1$. Since $m \notin R_\alpha$ we have $R_{\alpha+1} \neq R_\alpha$ and thus $r_\alpha \notin R_\alpha$. We then have a unique decomposition $m = pr_\alpha + n + y$ with $p \neq 0$. Any two different choices of the infinitesimal ε_α will result in different values for $f(m)$, so we can choose ε_α in such a way that $f(m) \neq m$.

Claim 6 *The function f is good.*

Proof: We first verify that $H\Delta(x) \in D$ 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 α. We have $\Delta(x) = f(x) - x = p\varepsilon_\alpha + \Delta(n+y)$, where $n+y \in R_\alpha$ and $p \in Q$. By definition, $H\varepsilon_\alpha \in D$, and by inductive hypothesis, $H\Delta(n+y) \in D$. Since D is a divisible group, it follows that $H\Delta(x) \in D$. By Claim 1, f preserves the relation L. The other requirements on (δ, m, f) are now easily proved by induction on α. ■

This completes the proof of Theorem 3.11. ■

Corollary 3.12 *If (δ, m, f) is a good triple, then $({}^*\mathcal{R}, \mathcal{A}) \equiv ({}^*\mathcal{R}, f(\mathcal{A}))$ for any $FO + LIN$ hyperstructure $\mathcal{A}$.*

Proof: f is an isomorphism between these two structures. ■

The main point of good triples is that in many cases they give us enough control to insure that $f(\mathcal{A})$ is internal even though f is external, and the inequality $f(m) \neq m$ can be used to get $\mathcal{A} \in {}^*X$ and $f(\mathcal{A}) \notin {}^*X$ for a query X.

4 The Query *Lin.Reach*

In this section we investigate the $FO+LIN$-definability of queries which are related to *Lin.Reach*.

By truncating the semi-linear set $A_{\mathbf{s}}$ in the proof of Theorem 2.4, one can show that the query

$$Lin.Reach \cap Bounded \cap Singular(5) \cap Degree(4)$$

is not $FO+LIN$-definable. This can be improved by applying Theorem 3.9.

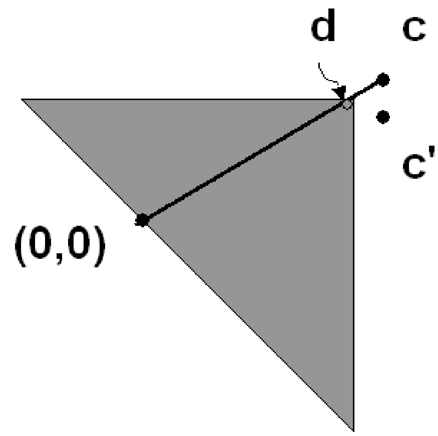

Figure 3: Theorem 4.1—The set A_1 and the point $\mathbf{c}'$.

Theorem 4.1 *The query*

$$Lin.Reach \cap Bounded \cap Singular(4) \cap Degree(3)$$

is not $FO + LIN$-definable.

Proof: We will construct two hypersemi-linear sets A_1 and B_1 as follows. (See Figure 3). Choose hyperrational numbers δ and m such that δ is positive infinitesimal and m has standard part $^\circ m \in (0,1) \setminus Q$. Let

$$m' = m - \delta, \quad \mathbf{c} = (1, m), \quad \mathbf{c}' = (1, m'),$$

$$\mathbf{d} = (1 - \delta, m - m\delta), \quad \mathbf{d}' = \mathbf{d} - (0, \delta) = (1 - \delta, m' - m\delta).$$

Define A_1 to be union of the right isosceles triangle with horizontal and vertical edges at $\mathbf{d}$ and hypotenuse through $(0,0)$, and the infinitesimal line segment (or "twig") from $\mathbf{d}$ to $\mathbf{c}$. Let B_1 be the union of the right isosceles triangle with horizontal and vertical edges at $\mathbf{d}'$ and hypotenuse through $(0,0)$, and the infinitesimal line segment from $\mathbf{d}'$ to $\mathbf{c}'$.

A_1 and B_1 are hypersemi-linear sets because δ and m are hyperrational. Moreover, they are bounded and have 4 singular points, each of degree at most 3.

We will use Theorem 3.9 to show that $({}^*\mathcal{R}, A_1, m) \equiv ({}^*\mathcal{R}, B_1, m')$. We must verify the following two hypotheses:

(a) $m \sim m'$, that is, m and m' are equivalent over the infinitesimals.

(b) Whenever $(m,x,y) \sim (m',x',y')$, we have $(x,y) \in A_1$ iff $(x',y') \in B_1$.

Hypothesis (a) follows from Corollary 3.7. To prove (b), assume that $(m,x,y) \sim (m',x',y')$. Suppose first that $(x,y) \in A_1$.

Case 1: (x,y) is on the infinitesimal twig. Then $\mathbf{z} = (x,y) - (1,m) \approx (0,0)$, so $({}^*\mathcal{R}, m, x, y, \mathbf{z}) \equiv ({}^*\mathcal{R}, m', x', y', \mathbf{z})$. Therefore $\mathbf{z} = (x',y') - (1,m')$, hence $(x',y') \in B_1$ (again on the infinitesimal twig).

Case 2: (x,y) is on the triangle. Then

$$x \leq 1 - \delta,\ y \leq m - m\delta,\ x + y \geq 0.$$

We have $({}^*\mathcal{R}, m, x, y, \delta, m\delta) \equiv ({}^*\mathcal{R}, m', x', y', \delta, m\delta)$. Therefore

$$x' \leq 1 - \delta,\ y' \leq m' - m\delta,\ x' + y' \geq 0,$$

so $(x',y') \in B_1$ (on the triangle).

The same argument shows that if $(x',y') \in B_1$ then $(x,y) \in A_1$. This proves hypothesis (b). It follows that $({}^*\mathcal{R}, A_1, m) \equiv ({}^*\mathcal{R}, B_1, m')$, and hence $({}^*\mathcal{R}, A_1, \mathbf{c}) \equiv ({}^*\mathcal{R}, B_1, \mathbf{c}')$.

A_1 contains the line segment from $(0,0)$ to $\mathbf{c}$, which passes through the singular point $\mathbf{d}$. However, B_1 does not contain the line segment from $(0,0)$ to $\mathbf{c}'$, because the segment from $\mathbf{c}'$ to the singular point $\mathbf{d}'$ has slope m, but the segment from $\mathbf{c}'$ to $(0,0)$ has slope m' and thus misses $\mathbf{d}'$.

The theorem now follows from Proposition 3.1. ■

We remark that by adding a superfluous polygonal line with rational vertices, one can show that the preceding theorem still holds if $Singular(4)$ is replaced by $Singular(m)$ where $m \geq 4$. It can also be shown that $Singular(4)$ is best possible.

We now ask whether *Lin.Reach* is definable over the family of thin semi-linear sets. This can also be formulated as a definability problem with respect to the family of all semi-linear sets by considering the boundary ∂A of a set A.

Definition 4.2 *For a structure* $\mathcal{A} = \langle A, \mathbf{a} \rangle$*, let* $\partial\mathcal{A} = \langle \partial A, \mathbf{a} \rangle$*. For each query* X*, let* $X(\partial S)$ *denote the query* $\{\mathcal{A} : \partial\mathcal{A} \in X\}$.

The following lemma shows that any undefinability result for $X(\partial S)$ is stronger than the corresponding undefinability result for X.

Lemma 4.3 *(i) For any query X, if $X(\partial S)$ is $FO+LIN$-undefinable then X is $FO+LIN$-undefinable.*

(ii) X is $FO+LIN$-definable over $Thin$ if and only if $X(\partial S)$ is $FO+LIN$-definable.

Proof: (i) For any $FO+LIN$ sentence $\psi(S)$, the definition of the boundary can be used directly to obtain an $FO+LIN$ sentence $\psi(\partial S)$ such that for each $\mathcal{A}$, $\mathcal{A} \models \psi(S)$ if and only if $\partial\mathcal{A} \models \psi(\partial S)$. Therefore if $\psi(S)$ defines the query X then $\psi(\partial S)$ defines the query $X(\partial S)$.

(ii) For any semi-linear set A, the boundary ∂A is a thin closed semi-linear set. Moreover, any thin closed set is equal to its boundary. It follows that if an $FO+LIN$ sentence $\psi(S)$ defines a query X over the thin closed semi-linear sets then $\psi(\partial S)$ defines $X(\partial S)$. And if an $FO+LIN$ sentence θ defines $X(\partial S)$, then θ defines X over the thin closed semi-linear sets. ∎

We will now use Theorem 3.11 to show that *Lin.Reach* is not definable over $Thin$.

Theorem 4.4 *(i) The query Lin.Reach is not $FO+LIN$-definable over $Thin$.*

(ii) The query Lin.Reach(∂S) is not $FO+LIN$-definable.

Remark 4.5 *Parts (i) and (ii) of Theorem 4.4 are equivalent by Lemma 4.3. Part (i) is a statement of the form "X is not $FO+LIN$-definable over $Thin$", and part (ii) is of the form "$X(\partial S)$ is not $FO+LIN$-definable". In the remainder of this paper we will prove other definability results which have two versions analogous to (i) and (ii). To avoid repetition, from now on we will only state version (i).*

Proof: We prove (i). By Theorem 3.11, there exists a good triple (δ, m, f). Let L be the set defined in Definition 3.10. Each line in L has slope m, and δ is the vertical distance between adjacent lines. The set L is hyperthin, but is not hypersemi-linear (a predicate for the integers is needed to define L).

Now define A_2 to be the intersection of L with the hyperreal unit square ${}^*[0,1]^2$. See Figure 4. Since δ and m are hyperrational numbers, A_2 is a thin hypersemi-linear set. Since the function f is good, it maps the set A_2 onto itself.

Consider the hypersemi-linear structures $\mathcal{A}_2 = \langle A_2, 0, 0, 1, m \rangle$ and $\mathcal{B}_2 = \langle A_2, 0, 0, 1, f(m) \rangle$. The two points $(0,0), (1,m)$ are on the same line segment in A_2. Then *Lin.Reach* holds in $\mathcal{A}_2$. But $f(m) \neq m$, so *Lin.Reach*

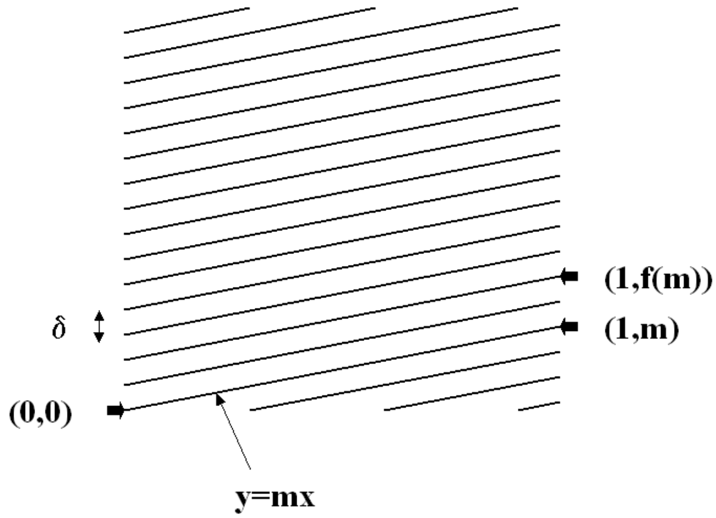

Figure 4: Theorem 4.4—The set A_2.

fails in $\mathcal{B}_2$. Since f is good, it is an automorphism of ${}^*\mathcal{R}$ which maps the relation and constants in $\mathcal{A}_2$ to those in $\mathcal{B}_2$. Hence $({}^*\mathcal{R}, \mathcal{A}_2) \equiv ({}^*\mathcal{R}, \mathcal{B}_2)$. By Proposition 3.1, *Lin.Reach* is not $FO + LIN$-definable over $Thin$. ■

Note that the above proof would not work if we took A_2 to be just the single line segment from $(0,0)$ to $(1,m)$, because then the image $B = f(A_2)$ would be an external set. The infinite set of parallel lines was needed in order to "hide" the fact that the image of each particular line segment is external by arranging things so that $f(A_2) = A_2$. Later on we will use a different set A such that $f(A) \neq A$, but we will still need the image $f(A)$ to be an internal set.

The proof of Theorem 4.4 has several additional consequences.

Definition 4.6 *Let Conn be the query*

$$Conn = \{\langle A, \mathbf{a}, \mathbf{b} \rangle : A \subset R^2, \mathbf{a}, \mathbf{b} \in R^2, \mathbf{a} \textit{ is connected to } \mathbf{b} \textit{ in } A\}.$$

Note that *Lin.Reach* $\subset$ *Conn*.

Corollary 4.7 *There is no query Y such that Lin.Reach $\subseteq Y \subseteq$ Conn and Y is $FO + LIN$-definable over $Thin$.*

In particular, Conn is not $FO + LIN$-definable over $Thin$.

Proof: Viewed from the nonstandard world, the set A_2 is thin, and the points $(0,0),(1,m)$ are on the same line segment in $A_2 3$, but the corresponding points $(0,0),(1,f(m))$ are not even connected in the set A_2. Thus $\mathcal{A}_2 \in {}^*Y$ but $\mathcal{B}_2 \notin {}^*Y$. ■

The proof of Theorem 4.4 used models with infinitely many singular points. The following positive result shows that this is essential.

Theorem 4.8 *For each n, the query*

$$Lin.Reach \cap Singular(n)$$

is $FO + LIN$-definable over $Thin$.

Proof: Fix n. Since $Singular(n)$ is $FO+LIN$-definable by Lemma 2.2, it suffices to prove that $Lin.Reach$ is $FO+LIN$-definable over the family T_n of thin semi-linear sets with n singular points. Let $Cong(\mathbf{x},\mathbf{y})$ (for congruence) say that $Reg(\mathbf{x})$, $Reg(\mathbf{y})$, and the line segments in S containing $\mathbf{x}$ and $\mathbf{y}$ are parallel (that is, for all sufficiently small $\mathbf{z}$, $\mathbf{x}+\mathbf{z} \in S \leftrightarrow \mathbf{y}+\mathbf{z} \in S$). Then $Cong(\mathbf{x},\mathbf{y})$ is an equivalence relation on the set of regular points of S, and each equivalence class is definable with a parameter from the class, hence for every $FO + LIN$ query one can ask whether it holds on some or all classes. Moreover, a semi-linear structure $(S,\mathbf{a},\mathbf{b})$ satisfies $Lin.Reach$ if and only $(S_0,\mathbf{a},\mathbf{b})$ satisfies $Lin.Reach$ where S_0 is the closure in S of one of these equivalence classes. Therefore, it suffices to prove that $Lin.Reach$ is $FO + LIN$-definable over the family U_n of all semi-linear sets in T_n such that all regular points are congruent.

Now let $(S,\mathbf{a},\mathbf{b}) \in U_n$ and let m be the slope of the lines in S (possibly vertical). Let $Vert(\mathbf{a},\mathbf{b})$ say that S contains a vertical line segment from $\mathbf{a}$ to $\mathbf{b}$. Let $Proj$ say that for every x such that $a_1 \le x \le b_1$ there exists y with $(x,y) \in S$. Let Env be the lower envelope of S, that is, the closure of the set

$$\{(x,y) : a_1 < x < b_1 \wedge y = \inf\{z : (x,z) \in S\}\}.$$

Suppose for the moment that $Proj$ holds and the slope m is not vertical. Let Seg be the line segment with slope m from $\mathbf{a}$ to a point with horizontal coordinate b_1. We claim that the set Seg is first order definable in $(\mathcal{R},S,\mathbf{a},\mathbf{b})$.

To see this, we first observe that the lower envelope Env is the union of at most n line segments $L_0,\dots,L_k$ of slope m, where $k \le n$, L_i has endpoints

$(x_i, y_i), (x_{i+1}, z_{i+1})$, and $a_1 = x_0 < x_1 < \ldots < x_k < x_{k+1} = b_1$. Thus $y_{i+1} - z_{i+1}$ is the vertical jump between the end of L_i and the start of L_{i+1}.

The set *Env* is definable in $(\mathcal{R}, S, \mathbf{a}, \mathbf{b})$, and so are the numbers x_i, y_i, z_{i+1}, $i \leq k$. *Seg* is the closure of the set of all points (x, y) such that for some $j \leq k$ and some t, $x_j < x < x_{j+1}$, $(x, t) \in Env$, and

$$t - y_0 = (y - a_2) + \sum_{1 \leq i \leq j} (z_i - y_i).$$

This proves the claim that *Seg* is definable in $(\mathcal{R}, S, \mathbf{a}, \mathbf{b})$.

Finally, *Lin.Reach* holds in $(S, \mathbf{a}, \mathbf{b})$ if and only if the sentence

$$Vert(\mathbf{a}, \mathbf{b}) \vee [Proj \wedge Seg(\mathbf{b}) \wedge \forall \mathbf{x}\, (Seg(\mathbf{x}) \rightarrow \mathbf{x} \in S)]$$

holds. ■

In the above result, and in all the results where we prove definability in $FO + LIN$ in this paper, the defining formula does not use the constant symbol 1. The proof shows definability not only in the language of $\langle R, <, +, -, 0, 1, \mathcal{A} \rangle$ (which is $FO + LIN$), but also in the language of $\langle R, <, +, -, 0, \mathcal{A} \rangle$.

In the next section we will consider second order extensions of $FO + LIN$. Before taking up these second order extensions, let us take a brief look at two first order extensions of $FO + LIN$.

Let $FO + LIN_R$ be the language obtained from $FO + LIN$ by adding a parameter (i.e. a new constant symbol) for each $r \in R$. All of the undefinability results in this paper hold for $FO + LIN_R$ as well as for $FO + LIN$. Since each formula of $FO + LIN_R$ contains only finitely many parameters, it suffices to show that the proofs go through when symbols for any particular finite collection of real numbers are added to the language.

A larger extension of $FO + LIN$ is the language $FO + LIN(R)$ formed by adding a new unary function symbol $r(\cdot)$ for each $r \in R$, which stands for the scalar product function $x \mapsto rx$. All of the elementary undefinability results in Section 2 still hold for $FO + LIN(R)$. Moreover, Theorem 3.9 still holds for the language $FO + LIN(R)$. The proof is the same except that for a given tuple $\mathbf{r} = \langle r_1, \cdots, r_n \rangle$ of real numbers, the structure ${}^*\mathcal{R}$ is everywhere replaced by the structure $({}^*\mathcal{R}, r_1(\cdot), \ldots, r_n(\cdot))$, and "rational" is replaced by "in the field generated by $\mathbf{r}$".

It follows that Theorem 4.1 holds for $FO + LIN(R)$, that is, the query

$$Lin.Reach \cap Bounded \cap Singular(4) \cap Degree(3)$$

is not $FO + LIN(R)$-definable. The same remark will apply to several other applications of Theorem 3.9 which will be given later on in this paper. However, the proof of Theorem 3.11, the existence of good triples, does not carry over to the language $FO + LIN(R)$. For this reason, the question whether *Lin.Reach* is definable in $FO + LIN(R)$ over the family of thin semi-linear sets remains open.

5 Restrictions of Second-order Logic

Since we have seen that $FO + LIN$ cannot express several queries that we're interested in, we now investigate definability in extensions of $FO + LIN$. The queries we consider will be obviously definable in $FO + POLY$, the expansion of $\langle R, +, -, \times, < \rangle$ by a binary predicate S. It is natural to look, then, for languages between $FO + LIN$ and $FO + POLY$.

We begin by adding second order unary function variables M_i to the language $FO + LIN$. Hereafter it will be understood that $FO + LIN$ has been enhanced by adding the unary function variables M_i. The function variable M_i is to be interpreted by a scalar product function $m_i(\cdot)$ for some real number m_i.

A $\Pi_1 + LIN$-formula is an expression of the form $\forall M_1 \ldots \forall M_n\ \psi$, where ψ is a formula of $FO + LIN$ with function variables. A $\Sigma_1 + LIN$-formula is an expression of the form $\exists M_1 \ldots \exists M_n\ \psi$, where ψ is a formula of $FO + LIN$ with function variables.

$\Pi_1 + LIN$-formulas and $\Sigma_1 + LIN$-formulas are semantically interpreted in the natural way. A query X is $\Pi_1 + LIN$-definable over F, or simply $\Pi_1 + LIN$ over F, if and only if there is a $\Pi_1 + LIN$ sentence θ such that for every semi-linear structure $\mathcal{A} \in F$, $\mathcal{A} \models \theta$ if and only if $\mathcal{A} \in X$. Similarly, a query is $\Sigma_1 + LIN$ over F iff its complement is $\Pi_1 + LIN$ over F. A query is said to be $\Delta_1 + LIN$ over F iff it is both $\Pi_1 + LIN$ and $\Sigma_1 + LIN$ over F.

Our convention when we omit F will be different for the languages $\Pi_1 + LIN$, $\Sigma_1 + LIN$, and $\Delta_1 + LIN$ than it was for $FO + LIN$—we will now allow parameters from R. A structure $\mathcal{A} = \langle A, \mathbf{a} \rangle$ with signature $\mathcal{S}$ is **semi-linear over** R if A is definable by a first order formula in $\mathcal{R}$ *with parameters from* R. By quantifier elimination, this is equivalent to the condition that A is definable by a finite Boolean combination of inequalities of the form $a_1x_1 + \cdots + a_kx_k \leq r$ where the coefficients a_i are rational and r is real. In particular, every finite relation $A \subseteq R^k$ is semi-linear over R. We say that a query X is $\Pi_1 + LIN$, $\Sigma_1 + LIN$, or $\Delta_1 + LIN$ if it is $\Pi_1 + LIN$, $\Sigma_1 + LIN$, or $\Delta_1 + LIN$ respectively over the family of structures which are semi-linear over R. Thus

$$\Delta_1 + LIN = \Pi_1 + LIN \cap \Sigma_1 + LIN.$$

An equivalent way to form the languages $\Pi_1 + LIN$ and $\Sigma_1 + LIN$ is to add second order variables which range over lines, so that the second order quantifiers would say "for all lines," or "there exists a line." The languages formed in this way would have slightly different syntax but the same expressive power as $\Pi_1 + LIN$ and $\Sigma_1 + LIN$.

It is easily seen that over every collection F of subsets of R^2 we have

$$FO{+}LIN \subseteq \Delta_1{+}LIN, \quad \Pi_1{+}LIN \subseteq FO{+}POLY, \quad \Sigma_1{+}LIN \subseteq FO{+}POLY.$$

It is also easily seen that the query *Co.Linear* is in $\Delta_1 + LIN$. Since this query is not $FO + LIN$-definable, we see that

$$FO + LIN \neq \Delta_1 + LIN$$

even over the collection of sets of cardinality 3. Another example of a query which is in $\Delta_1 + LIN$ but not in $FO + LIN$ is the query *Lin.Reach*. It is easily seen that *Lin.Reach* is in $\Delta_1 + LIN$, since $\mathbf{x}$ reaches $\mathbf{y}$ by a line segment in S iff for some (every) m such that $\mathbf{y}$ is on the line through $\mathbf{x}$ with slope m, every point on this line between $\mathbf{x}$ and $\mathbf{y}$ lies in S. But Theorem 2.4 shows that *Lin.Reach* is not in $FO + LIN$.

Our next proposition shows that $FO + POLY$ is exactly the closure of $\Pi_1 + LIN$, and also of $\Sigma_1 + LIN$, under first order quantification. Let us say that two sentences are **equivalent (over $\mathcal{S}$)** iff they define the same query over the collection of all structures $(\mathcal{R}, \mathcal{A})$ with signature $\mathcal{S}$. Say that two languages are **equivalent** iff every sentence in one language is equivalent to a sentence in the other language.

Proposition 5.1 *(i) The set of universal sentences of $FO{+}POLY$ is equivalent to the set of universal sentences of $\Pi_1 + LIN$.*

(ii) The set of existential sentences of $FO + POLY$ is equivalent to the set of existential sentences of $\Sigma_1 + LIN$.

(iii) $FO + POLY$ is equivalent to the closure of $\Delta_1 + LIN$ under first order quantification, and also to the closure of $\Pi_1 + LIN$ 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 $\forall \mathbf{y}\, \theta(\mathbf{c}, \mathbf{y})$ of $FO + POLY$. We need an equivalent universal sentence of $\Pi_1 + LIN$. Let $\tau_1, \ldots, \tau_k$ be the set of all terms and subterms of terms which occur in θ. Introduce second order variables $M_1, \ldots, M_k$. For each $j \leq k$, we define a term σ_j of $\Pi_1 + LIN$ by induction on the complexity of τ_j. The only nontrivial case in this definition is the product case: If τ_j is $\tau_h \cdot \tau_i$, then σ_j is $M_h(\sigma_i)$. Form θ' by replacing each term τ_j in θ by σ_j. The required sentence of $\Pi_1 + LIN$ is

$$\forall \mathbf{M} \forall \mathbf{y} \left(\bigwedge_{j=1}^{k} M_j(1) = \sigma_j \ \rightarrow \theta'(\mathbf{M}, \mathbf{c}, \mathbf{y}) \right).$$

■

We will sometimes add second order constants, that is, extra unary function symbols, to the language of $\mathcal{R}$. For an n-tuple $\mathbf{t} = \langle t_1, \ldots, t_n \rangle$ of real numbers, let $\mathcal{R}(\mathbf{t})$ be the structure obtained by adding second order constants $t_1(\cdot), \ldots, t_n(\cdot)$ to $\mathcal{R}$, where $t_i(\cdot)$ stands for the scalar product function.

We will also consider hyperreal second order constants. For an n-tuple $\mathbf{t} = \langle t_1, \ldots, t_n \rangle$ of hyperreal numbers, let ${}^*\mathcal{R}(\mathbf{t})$ be the structure obtained by adding second order constants $t_1(\cdot), \ldots, t_n(\cdot)$ to ${}^*\mathcal{R}$.

Our next goal is:

Theorem 5.2 *The classes $\Delta_1 + LIN$, $\Sigma_1 + LIN$, $\Pi_1 + LIN$, and $FO + POLY$ are all distinct.*

To prove this theorem, all we need is an example of a query with signature $\{S\}$ that is $\Sigma_1 + LIN$ but not $\Pi_1 + LIN$-definable. The negation will then be $\Pi_1 + LIN$ but not $\Sigma_1 + LIN$-definable, and Theorem 5.2 will follow. We first give a sufficient condition for not being $\Sigma_1 + LIN$-definable.

Definition 5.3 *For a collection X of semi-linear sets over R and a natural number n, the game $G_n(X)$ is played as follows: Duplicator chooses a set $A \in {}^*X$. Spoiler chooses an n-tuple of hyperreal numbers $\mathbf{t} = \langle t_1 \ldots t_n \rangle$. Duplicator chooses a hypersemi-linear set B over ${}^*R$ that is not in ${}^*X$. Duplicator wins the game if the structures $({}^*\mathcal{R}(\mathbf{t}), A)$ and $({}^*\mathcal{R}(\mathbf{t}), B)$ are first-order elementarily equivalent, and otherwise Spoiler wins.*

The game $G_n(X)$ could, of course, be refined further, but the above definition is sufficient for our purposes here.

Lemma 5.4 *If Duplicator has a winning strategy for $G_n(X)$ for every $n \in N$, then X is not $\Sigma_1 + LIN$-definable.*

Proof: Suppose X is definable by a $\Sigma_1 + LIN$-sentence $\exists T_1 \ldots \exists T_n \psi$. Duplicator's first move in the winning strategy for $G_n(X)$ must be a set $A \in {}^*X$. Then $({}^*\mathcal{R}, A) \models \exists T_1 \ldots \exists T_n \psi$. Spoiler can then choose hyperreal numbers $t_1 \ldots t_n$ such that ψ holds in $({}^*\mathcal{R}(\mathbf{t}), A)$. Duplicator must respond with a hypersemi-linear set B over ${}^*R$ such that $B \notin {}^*X$. Then the formula ψ cannot hold in $({}^*\mathcal{R}(\mathbf{t}), B)$. Thus $({}^*\mathcal{R}(\mathbf{t}), A)$ and $({}^*\mathcal{R}(\mathbf{t}), B)$ are not elementarily equivalent. Therefore Duplicator cannot have a winning strategy for $G_n(X)$. ■

We will now give an example which completes the proof of Theorem 5.2. Consider the query *SomeTriple*:

$\{S : (\exists \mathbf{x} \in S)(\exists \mathbf{y} \in S)(\exists \mathbf{z} \in S)[(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are distinct and collinear$]\}$.

SomeTriple is clearly $FO + POLY$-definable, and is even $\Sigma_1 + LIN$-definable.

Theorem 5.5 *The query SomeTriple is not $\Pi_1 + LIN$-definable.*

Proof: Let X be the set of all semi-linear sets over R in which *SomeTriple* fails. For each (standard) integer n our goal is to produce a winning strategy for Duplicator in the game $G_n(X)$. Duplicator's opening move is the set A_3 which is constructed as follows: Let $\mathbf{r} = (r_1, \ldots, r_{n+1})$ be an $n+1$-tuple of hyperreal numbers in ${}^*(0,1)$ such that $0 < {}^o r_1 < \cdots < {}^o r_{n+1} < 1$, and the real tuple ${}^o\mathbf{r}$ is algebraically independent, (that is, no polynomial with rational coefficients has root ${}^o\mathbf{r}$). Then any $(n+1)$-tuple $\mathbf{u}$ of hyperreal numbers which is infinitely close to $\mathbf{r}$ is also algebraically independent, because for each rational polynomial $p(\mathbf{x})$, if $p(\mathbf{u}) = 0$ then $p({}^o\mathbf{r}) = {}^o p(\mathbf{u}) = 0$.

For each i, choose a hyperreal number q_i such that $q_i \approx (r_i)^2$ but $q_i \neq (r_i)^2$. Let A_3 be the finite set consisting of the point $\mathbf{0} = (0,0)$, the points $\mathbf{x}_i = (r_i, q_i)$, and the points $\mathbf{y}_i = (i, ir_i)$, where $i = 1, \ldots, n+1$. This is the fan-shaped set shown in Figure 5.

For each i, the points $(\mathbf{0}, \mathbf{x}_i, \mathbf{y}_i)$ are not collinear because $q_i \neq (r_i)^2$. It follows from the algebraic independence of ${}^o\mathbf{r}$ that for any other distinct triple of points of A_3, the standard parts are not collinear and hence the triple itself is not collinear. Therefore A_3 belongs to ${}^*X$, and is a legal move for Duplicator.

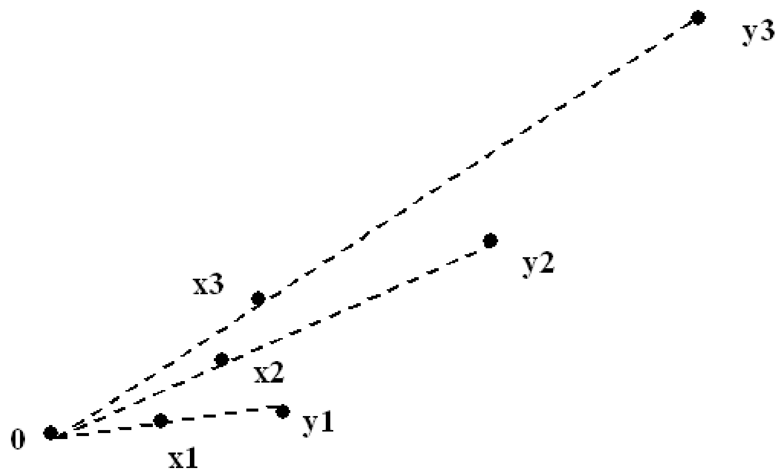

Figure 5: Theorem 5.5. The set A_3; the triples are almost collinear.

In the game $G_n(X)$, Spoiler must respond to Duplicator's opening move A_3 by choosing n scalar multipliers $t_1 \ldots t_n$ in ${}^*R$. Let P be the algebraic closure of $\{t_1 \ldots t_n\} \cup Q$ within ${}^*R$. Then P has transcendence rank n over Q, so P cannot contain an algebraically independent $(n+1)$-tuple. Since any $(n+1)$-tuple $\mathbf{u} \approx \mathbf{x}$ is algebraically independent, there is a $j \in \{1, \ldots, n+1\}$ such that P is disjoint from the monad of r_j. To simplify notation, we may assume that $j = 1$.

We must now construct Duplicator's response, which will be a finite set $B_3 \notin {}^*X$. First, form the structure ${}^*\mathcal{R}(P)$ by adding a new second order constant $p(\cdot)$ for scalar multiplication for each $p \in P$. Since P and R are real closed, it follows from Tarski's theorem that ${}^*\mathcal{R}(P)$ admits quantifier elimination. Next, let D be the substructure of ${}^*\mathcal{R}(P)$ generated by r_1. Thus $P \subset D$ and $r_1 \in D$. We now prove two claims.

Claim 7 *For each element $d \in D$ there are unique $s, p \in P$ such that $d = s + p(r_1)$.*

Proof: The set of all elements of the form $s+p(r_1)$ is clearly a substructure of ${}^*\mathcal{R}(P)$, so s, p exist. They are unique because if $s + p(r_1) = s' + p'(r_1)$ and

$p \neq p'$, then $(s - s') + (p - p')(r_1) = 0$ and hence $r_1 = -(s - s')/(p - p') \in P$, contrary to hypothesis. ■

Claim 8 *D is disjoint from the monad of r_1^2.*

Proof: Suppose to the contrary that $r_1^2 \approx d$ for some $d \in D$. Let $d = s + p(r_1)$ as in Claim 7. Consider the parabola $g(x) = s + p(x) - x^2$. We have $g(r_1) \approx 0$. The slope at r_1 is $g'(r_1) = p - 2r_1$. Since P is disjoint from the monad of r_1, we cannot have $p - 2r_1 \approx 0$, so the slope $g'(r_1)$ is not infinitesimal. Therefore the parabola crosses the x axis at some point $x_0 \approx r_1$, and $x_0 \in P$ because P is algebraically closed, contradiction. ■

Since q_1 belongs to the monad of r_1^2 which is disjoint from D, q_1 and r_1^2 are in the same cut in the ordering of D. It follows by quantifier elimination that

$$({}^*\mathcal{R}(P), r_1, q_1) \equiv ({}^*\mathcal{R}(P), r_1, r_1^2).$$

By the Saturation Principle, the relation $({}^*\mathcal{R}(P), \mathbf{x}) \equiv ({}^*\mathcal{R}(P), \mathbf{y})$ has the back and forth property. Since the set A is finite and contains the points $(0,0)$, (r_1, q_1), and $(1, r_1)$, it follows that there is a finite set B_3 which contains the points $(0,0)$, (r_1, r_1^2) and $(1, r_1)$ such that $({}^*\mathcal{R}(P), A_3) \equiv ({}^*\mathcal{R}(P), B_3)$. Then $({}^*\mathcal{R}(\mathbf{t}), A_3) \equiv ({}^*\mathcal{R}(\mathbf{t}), B_3)$. But $B_3 \notin {}^*X$ because the triple $((0,0), (r_1, r_1^2), (1, r_1))$ is collinear. Therefore Duplicator wins the game $G_n(X)$. This completes the proof of Theorem 5.5. ■

Theorem 5.2 now follows immediately from Theorem 5.5

As usual, the proof of the theorem shows more. We give three corollaries of the construction.

Corollary 5.6 *SomeTriple is not $\Pi_1 + LIN$-definable over the collection of finite sets.* ■

A query is said to be $\Pi_1 + LIN(n)$ if it can be defined by a $\Pi_1 + LIN$ sentence with at most n second order quantifiers. Similarly for $\Sigma_1 + LIN(n)$.

Corollary 5.7 *For each n, SomeTriple is not $\Pi_1 + LIN(n)$-definable over the collection of sets of cardinality $2n + 3$.* ■

Let $SomeTriple0$ be the query

$\{S : \mathbf{0} \in S \wedge (\exists \mathbf{y} \in S)(\exists \mathbf{z} \in S)[(\mathbf{0}, \mathbf{y}, \mathbf{z})$ are distinct and collinear$]\}$.

Clearly, $SomeTriple0$ implies $SomeTriple$. The proof of Theorem 5.5 works for $SomeTriple0$ as well. Namely:

Corollary 5.8 *There is no* $\Pi_1 + LIN$*-definable query* Y *such that*

$$SomeTriple0 \subseteq Y \subseteq SomeTriple.$$

■

A natural variant of *SomeTriple*0 is the query

$$AllTriples = \{A : (\forall \mathbf{x} \in A)(\exists \mathbf{y} \in A)(\mathbf{0}, \mathbf{x}, \mathbf{y}) \text{ are distinct and collinear}\}.$$

Note that this query implies that $\mathbf{0} \notin A$. *AllTriples* is clearly $FO + POLY$-definable, and is even $\Pi_1 + LIN$-definable. Using a modification of the proof of Theorem 5.5 we can show:

Theorem 5.9 *AllTriples is not* $\Sigma_1 + LIN$*-definable over the family of finite sets.*

Proof:

We will use Lemma 5.4 and a modification of the proof of Theorem 5.5. See Figure 6.

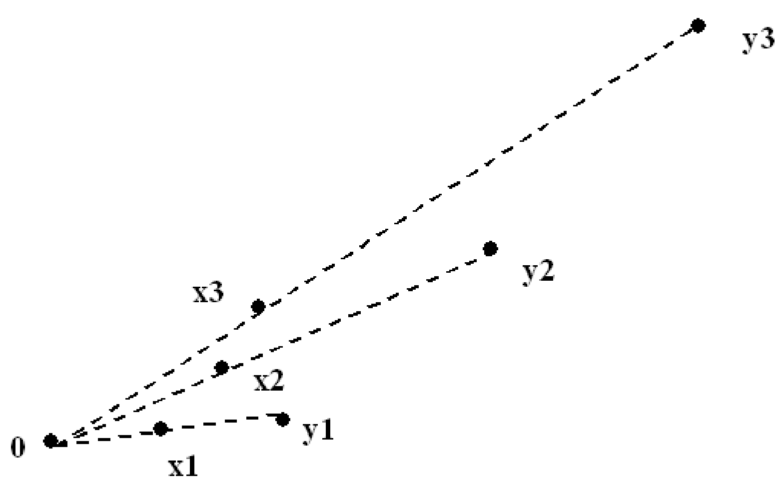

Figure 6: AllTriples. The set A_4; $(0, x_i, y_i)$ are collinear. Move one x_i vertically.

Take a standard n. This time Duplicator's first move must be a set A_4 for which *AllTriples* is true. Choose $\mathbf{r} = \langle r_1, \ldots, r_{n+1} \rangle$ as in the proof of Theorem 5.5. Let A_4 be the finite set consisting of the points $\mathbf{x}_i = (r_i, r_i^2)$ and $\mathbf{y}_i = (i, ir_i)$ for $i = 1, \ldots, n+1$. *AllTriples* is true in A_4. Again, Spoiler chooses an n-tuple of hyperreal numbers $t_1, \ldots, t_n$, and there is a $j \leq n+1$ such that the algebraic closure P of $Q \cup \{t_1, \ldots, t_n\}$ within $^*R$ is disjoint from the monad of r_j. For simplicity we assume $j = 1$.

This time Duplicator's move is a finite set B_4 such that $(^*\mathcal{R}(\mathbf{t}), A_4) \equiv (^*\mathcal{R}(\mathbf{t}), B_4)$ and B_4 contains points $(1, r_1)$ and (r_1, q_1) where $q_1 \approx r_1^2$ but $q_1 \neq r_1^2$. Then *AllTriples* is false in B_4, because the points $\mathbf{0}, (r_1, q_1), (1, r_1)$ are not collinear. Therefore Duplicator wins the game $G_n(AllTriples)$. ∎

Corollary 5.10 *For each n, AllTriples is not $\Sigma_1 + LIN(n)$-definable over the collection of sets of cardinality $2n+2$.*

6 The Queries *Cont.Line* and *Lin.Meet*

We have seen in Section 2 that *Cont.Line* and *Lin.Meet* are not $FO+LIN$-definable. We will now show that they are $FO+LIN$-definable over the thin semi-linear sets.

It is easily seen that *Cont.Line* and *Lin.Meet* are $\Sigma_1 + LIN$-definable. We will complement this by showing that they are not $\Pi_1 + LIN$-definable.

Intuitively, the query *Cont.Line* is "almost" first-order definable. For example, the query "S contains a ray" is easily definable in $FO + LIN$, because a semi-linear set contains a ray if and only if it is unbounded. The proof of Theorem 2.5 used semi-linear sets which are not thin. Our next theorem shows that this is essential, and reinforces the idea that *Cont.Line* is almost first-order definable.

Theorem 6.1 *Cont.Line is $FO+LIN$-definable over Thin.*

Proof: We restrict our attention to thin semi-linear sets in the following argument. We will freely use the particular $FO+LIN$ formulas introduced in the proofs of Lemma 2.2 and Theorem 4.8.

Let $MinSame(\mathbf{x}, t, y)$ say that y is minimal such that $Cong(\mathbf{x}, t, y)$, that is,

$$Cong(\mathbf{x}, t, y) \wedge \forall u(Cong(\mathbf{x}, t, u) \rightarrow y \leq u).$$

Let $Vert$ be the sentence asserting that S contains a vertical line. Let $OnLine(\mathbf{x})$ be the formula

$\exists s \forall t \, \forall t' \, [(s < t \wedge s < t') \rightarrow \exists y \exists y' [MinSame(\mathbf{x}, t, y) \wedge MinSame(\mathbf{x}, t', y') \wedge (\mathbf{x} + (t', y') - (t, y)) \in S\,]\,]$.

Finally, let ψ be the sentence

$$Vert \vee \exists \mathbf{x}[Reg(\mathbf{x}) \wedge OnLine(\mathbf{x})].$$

We claim that for any regular point $\mathbf{x}$ in a thin semi-linear set A, $OnLine(\mathbf{x})$ holds if and only if A contains a non-vertical line through $\mathbf{x}$. From this it follows easily that ψ is the required sentence.

Suppose A is a thin semi-linear set and $\mathbf{x}$ is a regular point of A which is not on a vertical line segment in A. We make a preliminary observation. If the set $J = \{\mathbf{y} : MinSame(\mathbf{x}, \mathbf{y})\}$ is unbounded on the right, then J is contained in the union of finitely many parallel lines, and it follows that there exists s and a line K such that J contains the part of K to the right of s.

Assume that A contains a line L through $\mathbf{x}$. Then L is not vertical. The set J is unbounded on the right, and there exist s and K as in the above paragraph. Moreover, K is parallel to L. Thus whenever $t, t' > s$, $(t, y) \in K$, and $(t', y') \in K$, the sum of $\mathbf{x}$ and the vector between (t, y) and (t', y') belongs to L, which is a subset of A. Thus $OnLine(\mathbf{x})$ holds.

For the converse, suppose that $OnLine(\mathbf{x})$ holds. Since $\mathbf{x}$ is regular, there is a line L through $\mathbf{x}$ such that $A = L$ in some open rectangle around $\mathbf{x}$. We show that L is actually contained in A. Let m be the slope of L. It follows from $OnLine(\mathbf{x})$ that the set J is unbounded on the right. Therefore there is a line K of slope m and an s such that J contains the part of K to the right of s. For any d we can find $t, t' > s$ with $t - t' = d$. Then there are y, y' such that $(t, y), (t', y') \in K$. When we add the vector between (t, y) and (t', y') to $\mathbf{x}$, we get the point on L at horizontal distance d away from $\mathbf{x}$. By $OnLine(\mathbf{x})$, this point is in A. Since d was arbitrary, this shows that A contains L, and proves the claim. ■

The above argument shows somewhat more.

Corollary 6.2 *(i) Cont.Line is FO+LIN-definable over the family of semi-linear sets A such that the interior of A is bounded on the right.*

(ii) Cont.Line is $FO + LIN$-definable over the family of thin semi-algebraic sets.

(iii) For each o-minimal expansion $\mathcal{R}'$ of $\mathcal{R}$, Cont.Line is $FO + LIN$-definable over the family of thin $\mathcal{R}'$-definable sets.

Proof: In each case, the $FO+LIN$ sentence ψ in the proof of Theorem 6.1 defines $Cont.Line$. ■

Corollary 6.3 *$Lin.Meet$ is $FO+LIN$-definable over $Thin$.*

Proof: Using the formula $OnLine(\mathbf{x})$ from the proof of Theorem 6.1, one can build an $FO+LIN$ sentence $\theta(\mathbf{a})$ which says that every open rectangle containing $\mathbf{a}$ contains two regular points $\mathbf{x}, \mathbf{y}$ such that $\neg Cong(\mathbf{x}, \mathbf{y})$, and A contains a line through $\mathbf{x}$ and a line through $\mathbf{y}$. This sentence defines $Lin.Meet$ over $Thin$. ■

We now prove the undefinability results over $\Pi_1 + LIN$.

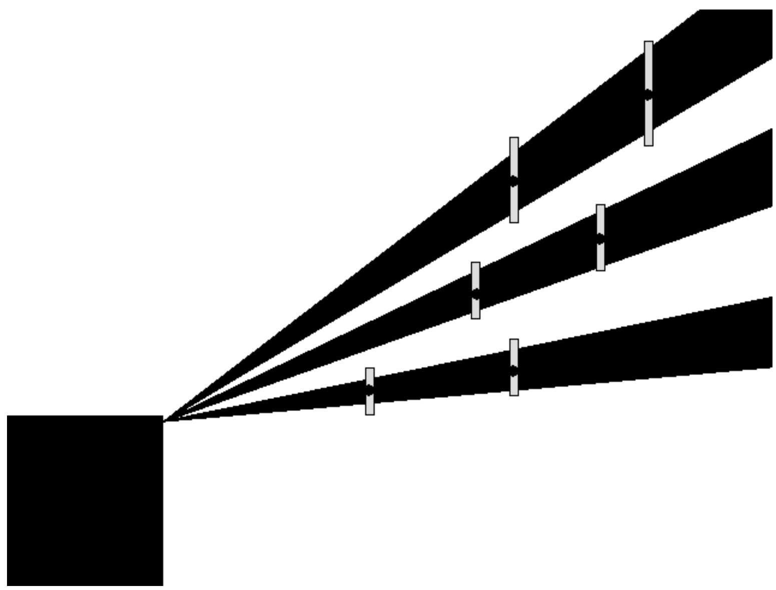

Figure 7: Theorem 6.4. The set A_5.

Theorem 6.4 *$Cont.Line$ is not $\Pi_1 + LIN$-definable.*

Proof: We build upon the construction used in Theorem 5.5, where it was shown that $SomeTriple$ is not in $\Pi_1 + LIN$. Let X be the family of semi-linear sets over R in which $Cont.Line$ fails. Given n, let

$$A_3 = \{\mathbf{0}, \mathbf{x}_1, \mathbf{y}_1, \ldots, \mathbf{x}_{n+1}, \mathbf{y}_{n+1}\}$$

be the finite set constructed in the proof of Theorem 5.5. Now take $n+1$ wedges W_i such that each W_i has vertex $(0,0)$, has rational slopes between 0 and 1, and contains the points $\mathbf{x}_i$ and $\mathbf{y}_i$, and such that for $i \neq j$, W_i and W_j are disjoint except for the origin $(0,0)$. For each i, form the set U_i by removing from W_i the vertical lines through $\mathbf{x}_i$ and $\mathbf{y}_i$ and then adding back the points $\mathbf{x}_i$ and $\mathbf{y}_i$. The vertical lines which are removed are "barriers" which are crossed by "bridges" $\mathbf{x}_i$ and $\mathbf{y}_i$. Finally, let

$$A_5 = U_1 \cup \cdots \cup U_{n+1} \cup \{(u,v) : u \leq 0 \wedge v \leq 0\}.$$

Let B_3 be the finite set used in Duplicator's strategy in the proof of Theorem 5.5 and let B_5 be the set formed in the same way as A_5 (with the same rational slopes). The sets A_5 and B_5 are hypersemi-linear over $^*R$. The proof of Theorem 5.5 shows that $({}^*\mathcal{R}, A_5) \equiv ({}^*\mathcal{R}, B_5)$. One can see from the figure that *Cont.Line* fails in A_5 because none of the triples $(\mathbf{0}, \mathbf{x}_i, \mathbf{y}_i)$ are collinear, but *Cont.Line* holds in B_5 because one of the triples $(\mathbf{0}, \mathbf{u}_j, \mathbf{v}_j)$ is collinear. Therefore Duplicator has a winning strategy for the game $G_n(X)$, and the theorem is proved. ■

Corollary 6.5 *The query Lin.Meet is not $\Pi_1 + LIN$-definable.*

Proof: Argue as in the proof of Theorem 6.4, but add a vertical line through the origin to the sets A_5 and B_5. ■

The following positive result shows that the use of models with unboundedly many singular points or unbounded degrees is essential in the proof of Theorem 6.4. This result should be contrasted with Theorem 2.5.

Theorem 6.6 *For each $k, m \in N$, the query*

$$Cont.Line \cap Singular(k) \cap Degree(m)$$

is $\Pi_1 + LIN$-definable.

Proof: Let $h = k(k+m)$. The defining $\Pi_1 + LIN$-sentence says that there are k singular points, each singular point has degree at most m, and either S contains a vertical line, or for all $M_1, \ldots, M_h$, if the slope of each regular point is equal to M_i for some $i \leq h$, and the slope of the segment between each pair of singular points is equal to M_i for some $i \leq h$, then S contains a line of slope M_i for some $i \leq h$. It is clear that this sentence implies that S contains a line.

For the converse, note that if A contains a line L, then L can be moved within A to a line which is either vertical, parallel to the segment through a regular point of A, or parallel to the segment between two singular points of A. ■

7 The Queries *n-Linked*, $n \in N$

Definition 7.1 *For each n, let n-Linked be the class of semi-linear sets A such that for any two points* **x** *and* **y** *in A, A contains a polygonal path from* **x** *to* **y** *consisting of at most n line segments.*

Note that a semi-linear set is connected if and only if it is n-linked for some n. It is easily seen that for each n the query *n-Linked* is definable in $FO + POLY$. We will give a fairly complete description of the $FO + LIN$-definability of *n-Linked* for $n \in N$, summarized in a table at the end of the section. Like $Cont.Line$, the query *n-Linked* will behave differently over the family of thin semi-linear sets than over the family of all semi-linear sets.

Proposition 7.2 *1-Linked is $FO + LIN$-definable.*

Proof: 1-*Linked* is defined by the $FO + LIN$ sentence

$$\forall \mathbf{x} \in S \forall \mathbf{y} \in S Mid(\mathbf{x}, \mathbf{y}) \in S.$$

■

Note that a semi-linear set A is in 1-*Linked* if and only if it is convex.

Theorem 7.3 *2-Linked is $FO + LIN$-definable over $Thin$.*

Proof: Let $Par(\mathbf{x}, \mathbf{y})$ say that each rectangle around **x** contains a point **z** such that $Cong(\mathbf{y}, \mathbf{z})$. As before, we let $Mid(\mathbf{x}, \mathbf{y})$ denote the midpoint between **x** and **y**. Now consider the following $FO + LIN$ sentence ψ:

$$\forall \mathbf{x} \forall \mathbf{y} (\, Reg(\mathbf{x}) \wedge Reg(\mathbf{y}) \rightarrow [\, Par(\mathbf{x}, \mathbf{y}) \rightarrow Mid(\mathbf{x}, \mathbf{y}) \in S \,] \wedge$$

$$[\, \neg Par(\mathbf{x}, \mathbf{y}) \rightarrow \exists \mathbf{z} (\, Par(\mathbf{z}, \mathbf{x}) \wedge Par(\mathbf{z}, \mathbf{y}) \,) \,] \,).$$

ψ 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 in S which realizes both slopes.

We claim that a thin semi-linear set A is 2-linked if and only if it satisfies ψ. It is easily seen that if A is 2-linked then A satisfies ψ.

For the converse, assume that A satisfies ψ. To show that A is 2-linked, it suffices to prove that any two regular points $\mathbf{x}, \mathbf{y}$ in A are 2-linked. Call a pair of points $\mathbf{x}, \mathbf{y}$ **bad** if $Reg(\mathbf{x}), Reg(\mathbf{y})$, and $Par(\mathbf{x}, \mathbf{y})$, but $\mathbf{x}, \mathbf{y}$ are not 1-linked in A. By ψ, if $\mathbf{x}, \mathbf{y}$ is a bad pair then every rectangle around $Mid(\mathbf{x}, \mathbf{y})$ must contain a point $\mathbf{z}$ such that then either $\mathbf{x}, \mathbf{z}$ or $\mathbf{z}, \mathbf{y}$ is another bad pair of points. It follows that there are bad pairs of points which are arbitrarily close to each other. This cannot happen in a thin semi-linear set. Therefore there are no bad pairs of points.

Now suppose $Reg(\mathbf{x}), Reg(\mathbf{y})$ and not $Par(\mathbf{x}, \mathbf{y})$. By ψ there is a point $\mathbf{z} \in A$ such that $Par(\mathbf{z}, \mathbf{x})$ and $Par(\mathbf{z}, \mathbf{y})$. Then each rectangle around $\mathbf{z}$ contains regular points $\mathbf{u}, \mathbf{v}$ such that $Par(\mathbf{u}, \mathbf{x})$ and $Par(\mathbf{v}, \mathbf{y})$. Since there are no bad pairs, A must contain the line segments from $\mathbf{x}$ to $\mathbf{z}$ and from $\mathbf{z}$ to $\mathbf{y}$, so $\mathbf{x}$ and $\mathbf{y}$ are 2-linked in A. ■

We obtain several undefinability results by modifying the proofs of our earlier theorems.

Theorem 7.4 *2-Linked is not $FO + LIN$-definable. In fact, each of the queries*

$$2\text{-}Linked \cap Bounded \cap Singular(6) \cap Degree(3),$$

$$2\text{-}Linked \cap Singular(4) \cap Degree(3)$$

is not $FO + LIN$-definable.

Proof: For the first query, start with the hypersemi-linear sets A_1 and B_1 introduced in the proof of Theorem 4.1. Form new hypersemi-linear sets A_6 and B_6 by attaching to $(0,0)$ an "infinitesimal twig" of slope m from $(0,0)$ to $(-\delta, -m\delta)$, as in the left part of Figure 8. A_6 and B_6 have six singular points with degree ≤ 3. As before, we have $({}^*\mathcal{R}, A_6) \equiv ({}^*\mathcal{R}, B_6)$ by Theorem 3.9. However, A_6 is 2-linked but B_6 is not.

For the second query, let A_7 be the union of the infinite vertical strip $[\delta, 1-\delta] \times {}^*R$ and the line segment from $(0,0)$ to $(1,m)$, and let B_7 be the same as A_7 except that the infinitesimal twig on the right is shifted down by δ. (See the right part of Figure 8). Again, A_7 and B_7 are hypersemi-linear

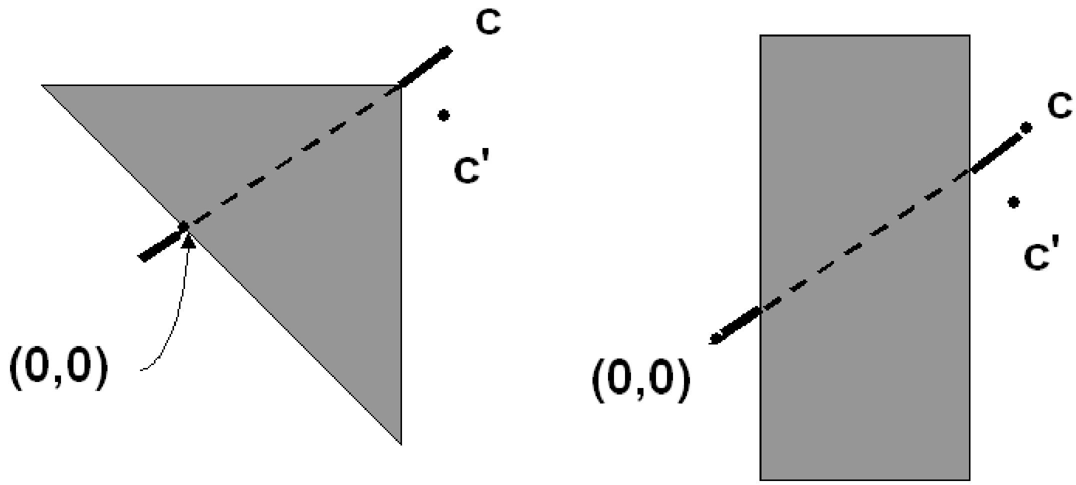

Figure 8: Theorem 7.4—The sets A_6, B_6 and A_7, B_7

sets such that $({}^*\mathcal{R}, A_7) \equiv ({}^*\mathcal{R}, B_7)$, and each set has four singular points, all of degree at most 3. However, A_7 is 2-linked but B_7 is not. ■

The next result shows that the queries in Theorem 7.4 are $FO + LIN$-definable over $Thin$.

Proposition 7.5 *For each finite n and k, the query*

$$k\text{-}Linked \cap Singular(n)$$

is $FO + LIN$-definable over $Thin$.

Proof: By Lemma 2.2 and Theorem 4.8, the query $Lin.Reach$ is $FO + LIN$-definable over the class $T_n = Thin \cap Singular(n)$. The defining sentence must be of the form $\psi_n(\mathbf{a}, \mathbf{b})$ for some $FO + LIN$ formula $\psi_n(\mathbf{x}, \mathbf{y})$. Then k-*Linked* is defined over T_n by the following $FO + LIN$ sentence:

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

■

We next obtain a negative interpolation result between 3-*Linked* and k-*Linked*.

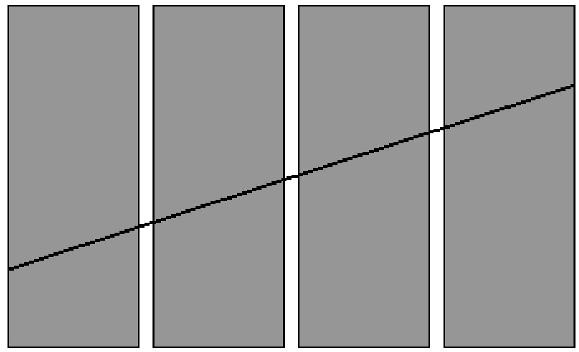

Figure 9: Theorem 7.6—The sets A_8, B_8

Theorem 7.6 *There does not exist an $FO + LIN$-definable query Y and natural number $k \geq 3$ such that*

$$3\text{-}Linked \subseteq Y \subseteq k\text{-}Linked.$$

Thus for each $k \geq 3$, k-Linked is not $FO + LIN$-definable.

Proof: Fix a natural number $k \geq 3$. Let Y be a query which is definable by an $FO+LIN$ sentence ψ. Choose $k+1$ linearly independent real numbers $\hat{r}_1 < \cdots < \hat{r}_k < \hat{m}$ in the interval $(0, 1)$ and choose hyperrational numbers r_i, m with standard part $\hat{r}_i, \hat{m}$. Form the set A_8 by taking the vertical strip $[0, 1] \times {}^*R$, removing the k vertical lines $\{r_i\} \times {}^*R$ for $i = 1, \ldots, k$, and adding the line segment L from $(0, 0)$ to $(1, m)$. The removed vertical lines are "barriers" and the segment from $(0, 0)$ to $(1, m)$ has a bridge across each barrier.

Let δ be a positive infinitesimal hyperrational number. Let B_8 be the same as A_8 except that for each even $i \leq k$, the bridge on the line segment L with horizontal coordinate r_i is shifted vertically upward by δ. See Figure 9. The sets A_8 and B_8 are hypersemi-linear sets. The set A_8 is 3-linked, but B_8 is not k-linked, because it takes two segments to get to the first and last bridge, and $(k-1)$ segments to cross all k bridges. By Corollary 3.7 and Theorem 3.9, the structures $({}^*\mathcal{R}, A_8)$ and $({}^*\mathcal{R}, B_8)$ are elementarily equivalent. If 3-*Linked* $\subseteq Y$, then $A_8 \models \psi$, so $B_8 \models \psi$ and thus Y is not contained in k-*Linked*. ■

The preceding results leave two problems open.

Problem 7.7 *(i) Can Theorem 7.6 be improved by replacing 3-Linked by 2-Linked?*

(ii) Is there an $FO + LIN$-definable query Y such that

$$3\text{-}\mathit{Linked} \subseteq Y \subseteq \mathit{Connected}?$$

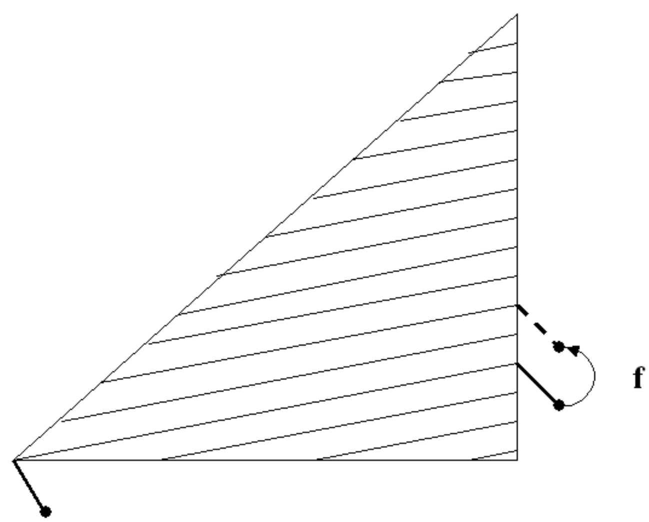

Figure 10: Theorem 7.8—The sets A_9, B_9

We now return to the question of definability over the family of thin semi-linear sets.

Theorem 7.8 *The query 3-Linked is not $FO + LIN$-definable over Thin.*

Proof: Let C be the restriction of the hypersemi-linear set A_2 from the proof of Theorem 4.4 to the right triangle with vertices $(0,0), (1,0), (1,1)$. Now form the hyper-semilinear set A_9 by adding to C the boundary of the right triangle, and two extra infinitesimal twigs, which are line segments with rational slopes attached to the points $(0,0)$ and $(1,m)$. See Figure 10. The set A_9 is thin and 3-linked.

By Theorem 3.11, there is a good triple (δ, m, f). The map f leaves the left twig fixed, but shifts the right twig vertically upward. f sends A_9 to a set B_9 with the same bounding triangle, but the two infinitesimal twigs are no longer on the same line of slope m. The set B_9 is not 3-linked, because it takes four segments to connect the ends of the two infinitesimal twigs. ■

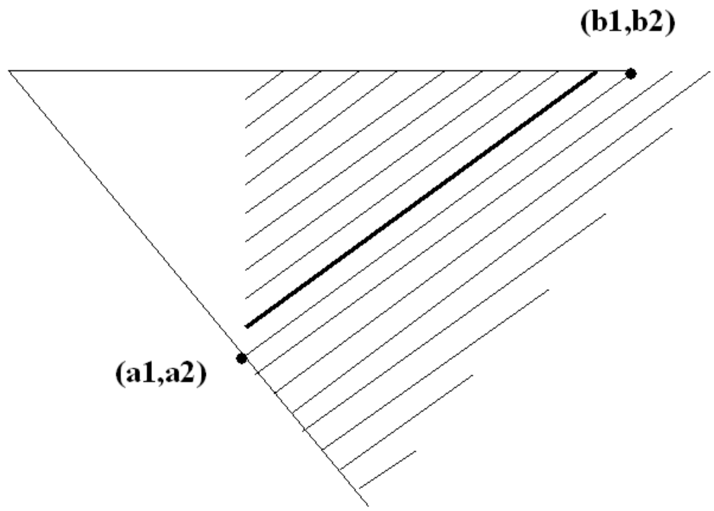

Figure 11: Theorem 7.9—The set A_{10}

A similar argument can be used to show that the query 4-*Linked* is not $FO + LIN$-definable over the thin semi-linear sets. By yet another modification, we can map a 4-linked set to a set that is not connected at all, and obtain a stronger interpolation result.

Theorem 7.9 *There is no class Y of semi-linear sets such that Y is $FO + LIN$-definable over Thin and*

$$4\text{-}Linked \subseteq Y \subseteq Connected.$$

In particular, this shows that Connected is not $FO + LIN$-definable over Thin, and also that for each $k \geq 4$, k-Linked is not $FO + LIN$-definable over Thin.

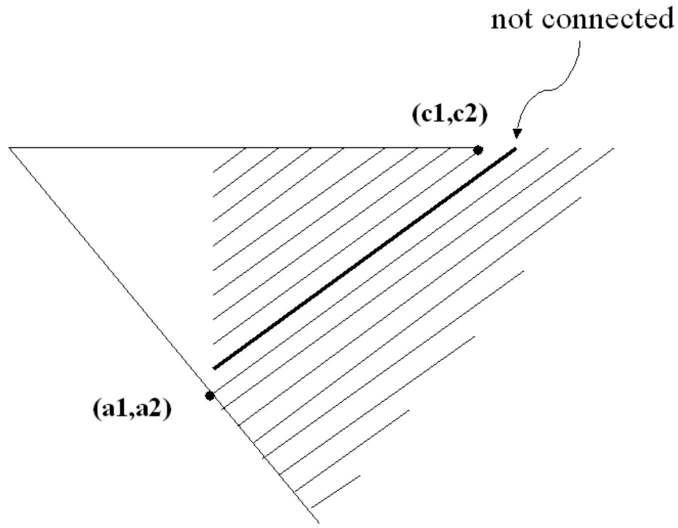

Figure 12: Theorem 7.9—The set B_{10}

Proof: Again let (δ, m, f) be a good triple. We take the set A_2 from the proof of Theorem 4.4, but modify its boundary so that the opposite sides are not parallel but have rational slopes, as shown in Figure 11. We choose one of the line segments from a point **a** on the left boundary to a point **b** on the top boundary. We let A_{10} be the hypersemi-linear set that includes the parallel lines within the boundary, plus the left boundary and the part of the top boundary to the left of the point **b**. Finally, we extend these two boundary lines until they intersect. As one can see from Figure 11, the set A_{10} is 4-linked and contains the line segment from **a** to **b**.

Since the boundaries of the triangle in A_{10} have rational slope, the map f takes A_{10} to a hypersemi-linear set B_{10} whose boundaries have the same slopes. f fixes the point **a**, but moves the point **b** horizontally to the left. $f(\mathbf{b})$ is the point where the upper boundary line of B_{10} ends. One can see from Figure 12 that, viewed from the nonstandard world, the set B_{10} is not connected. ■

Problem 7.10 *Can Theorem 7.9 can be improved by replacing 4-Linked by 3-Linked?*

We will also leave open the question of whether the queries *n-Linked* are definable in the second order languages $\Pi_1 + LIN$ and $\Sigma_1 + LIN$.

We summarize our results in this section with a table.

Summary of $FO + LIN$-definability for k-*Linked*			
k	Arbitary	Thin	Thin with n singular points
1	Yes	Yes	Yes
2	No	Yes	Yes
≥ 3	No	No	Yes

Here 'Yes' in a box for integer k and class C means that k-*Linked* is definable over C.

8 Extensions to Higher Dimensions

In this section we consider the $FO + LIN$-definability of higher dimensional analogs of the query $Cont.Line$. We first consider the three-dimensional case.

Theorem 8.1 *Let S be a ternary relation symbol.*
(i) The query "S contains a plane" is not $\Pi_1 + LIN$-definable.
(ii) The query "S contains a line" is not $\Pi_1 + LIN$-definable over the family of thin semi-linear sets.

Proof: (i) Note that a set $A \subseteq R^2$ contains a line if and only if the product $A \times R$ contains a (vertical) plane. Since $Cont.Line$ is not $\Pi_1 + LIN$-definable by Theorem 6.4, it follows that "S contains a plane" is not $\Pi_1 + LIN$-definable.
(ii) A set $A \subseteq R^2$ contains a line if and only if the product $A \times \{0\}$ contains a line. The product $A \times \{0\}$ is thin. It follows that "S contains a line" is not $\Pi_1 + LIN$-definable over the thin semi-linear sets. ∎

Theorem 8.2 *Let S be a ternary relation symbol. The query "S contains a plane" is $FO + LIN$-definable over the family of thin semi-linear sets.*

Proof: Let A be a thin semi-linear subset of R^3. A contains a plane if and only if either
(1) A contains a vertical plane, or
(2) A contains a plane which is the graph of a function of the form $z = ax + by + c$ where $a, b, c \in R$.

We first give an $FO + LIN$ sentence saying that (1) holds. Let $Vert(\mathbf{t})$ be the $FO + LIN$ formula $\forall z S(\mathbf{t}, z)$. Then $A \models Vert(\mathbf{t})$ iff the vertical line $\{(\mathbf{t}, z) : z \in R\}$ is contained in A. For each thin A, the set $V_A = \{\mathbf{t} \in R^2 : A \models Vert(\mathbf{t})\}$ must be a thin semi-linear set in R^2. By Theorem 6.1, there is a $FO + LIN$ sentence $Vert$ such that for every thin semi-linear set A, $A \models Vert$ iff V_A contains a line, and it is clear that V_A contains a line iff A contains a vertical plane.

We now find an $FO + LIN$ sentence which expresses (2). Let $Reg(\mathbf{x})$ say that $\mathbf{x}$ is a regular point of A, which now means that in some open box containing $\mathbf{x}$ the boundary of A coincides with a plane. We call two points $\mathbf{x}$ and $\mathbf{y}$ congruent, in symbols $Cong(\mathbf{x}, \mathbf{y})$, if they are both in A, and if for some U containing $\mathbf{0}$ we have

$$(\forall \mathbf{v} \in U)\mathbf{x} + \mathbf{v} \in A \leftrightarrow \mathbf{y} + \mathbf{v} \in A.$$

For $\mathbf{x} \in R^3, \mathbf{t} \in R^2$, let $MinSame(\mathbf{x}, \mathbf{t}, y)$ say that y is the least real u such that $Cong(\mathbf{x}, (\mathbf{t}, u))$. Let $OnPlane(\mathbf{x})$ be the analog of $OnLine(\mathbf{x})$ from the proof of Theorem 6.1. It says that for any two points $\mathbf{t}, \mathbf{t}'$ with sufficiently large coordinates in the horizontal plane,

$$MinSame(\mathbf{x}, \mathbf{t}, y) \wedge MinSame(\mathbf{x}, \mathbf{t}', y') \rightarrow (\mathbf{x} + (\mathbf{t}, y) - (\mathbf{t}', y')) \in S.$$

Let ψ be the $FO + LIN$ sentence

$$\exists \mathbf{x}[Reg(\mathbf{x}) \wedge OnPlane(\mathbf{x})].$$

Then ψ expresses (2), and thus $Vert \vee \psi$ says that A contains a plane. ■

The theory of o-minimal structures gives an appropriate notion of dimension for semi-linear subsets of R^n. Equivalently, one can define the dimension of a semi-linear set A to be the largest m such that A has a subset which is topologically equivalent to a ball in R^m.

Note that in n-dimensional Euclidean space, a semi-linear set is thin if and only if it has dimension at most $n - 1$. Our results can be generalized to the following theorem in n variables.

Theorem 8.3 *Let $n > k > 0$, and let $X(n, k)$ be the query saying that an n-ary semi-linear relation contains a k-dimensional hyperplane. Then:*

(i) $X(n, k)$ is $\Sigma_1 + LIN$-definable.

(ii) $X(n,k)$ is $FO+LIN$-definable over the collection of semi-linear sets of dimension $\leq k$.

(iii) $X(n,k)$ is not $\Pi_1 + LIN$-definable over the collection of semi-linear sets of dimension $\leq k+1$.

Proof: Part (i) is clear. Part (ii) proved by a direct generalization of the proof of Theorem 8.2. For (iii), we observe that for any set A in the plane, the product $C = A \times R^{k-1}$ has dimension $\leq k+1$, and A contains a line if and only if C contains a k-dimensional hyperplane. By Theorem 6.4, $Cont.Line$ is not $\Pi_1 + LIN$-definable, and (iii) follows. ■

9 Conclusions and Future Work

Questions concerning definability with an extra predicate, even for a well-understood structure such as the real ordered additive 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 $\Pi_1 + LIN$, $\Sigma_1 + LIN$, and $\Delta_1 + LIN$ and have found natural examples of queries which are $\Sigma_1 + LIN$-definable but not $\Pi_1 + LIN$-definable. We have left to the future the analogous separation question for the higher prefix classes $\Pi_n + LIN$, $\Sigma_n + LIN$.

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 First-Order Constraint Languages. In First CONTESSA Workshop on Constraint Databases, pp. 22-39.

[3] 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.

[4] M. Benedikt, G. Dong, L. Libkin and L. Wong. Relational expressive power of constraint query languages. *J. ACM*, 45 (1998), pp. 1–34.

[5] 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.

[6] M. Benedikt and H. J. Keisler. Definability over Linear Constraints To appear.

[7] O. Chapuis and P. Koiran. Definability of geometric properties in algebraically closed fields. To appear in *Mathematical Logic Quarterly*.

[8] C.C. Chang and H.J. Keisler. *Model Theory*. North Holland, 1990.

[9] P. Kanellakis, G. Kuper, and P. Revesz. Constraint query languages. *JCSS*, 51 (1995), 26–52.

[10] C. Karp. Finite quantifier equivalence. In *The Theory of Models*, pages 407-412, North-Holland, 1965.

[11] 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.

[12] G. Kuper, L. Libkin and J. Paredaens, eds. *Constraint Databases*. Springer Verlag, 2000.

Addresses

Bell Laboratories
263 Shuman Blvd.
Naperville, IL 60566, USA
Email: benedikt@research.bell-labs.com

University of Wisconsin-Madison
Department of Mathematics Van Vleck Hall
480 Lincoln Drive
Madison, Wisconsin 53706-1388, USA
Email: keisler@math.wisc.edu
home page: http://www.math.wisc.edu/~keisler