EXTENDED ORDER-GENERIC QUERIES

OLEG V. BELEGRADEK, ALEXEI P. STOLBOUSHKIN,
AND MICHAEL A. TAITSLIN

Abstract. We consider relational databases organized over an ordered domain with some additional relations—a typical example is the ordered domain of rational numbers together with the operation of addition. In the focus of our study are the first-order (FO) queries that are invariant under order-preserving "permutations"—such queries are called order-generic. It has recently been discovered that for some domains order-generic FO queries fail to express more than pure order queries. For example, every order-generic FO query over rational numbers with $+$ can be rewritten without $+$. For some other domains, however, this is not the case.

We provide very general conditions on the FO theory of the domain that ensure the collapse of order-generic extended FO queries to pure order queries over this domain: the *Pseudo-finite Homogeneity Property* and a stronger *Isolation Property*. We further distinguish one broad class of domains satisfying the Isolation Property, the so-called *quasi-o-minimal* domains. This class includes all the o-minimal domains, but also the ordered group of integer numbers and the ordered semigroup of natural numbers, and some other domains.

An important difference of this paper from the recent series of related papers is that we generalize all the notions to the case of finitely representable database states—as opposed to finite states—and develop a general lifting technique that, essentially, allows us to extend any result of the kind we are interested in, from finite to finitely-representable states. We show, however, that these results cannot be transferred to arbitrary infinite states.

Date: May 30, 2001.

1991 *Mathematics Subject Classification.* Primary 68P15; Secondary 03C40, 03C52.

Key words and phrases. Pseudo-finite homogeneity property, isolation property, quasi-o-minimal structure, local genericity, finite database state, finitely representable database state.

A part of this work was done while Oleg Belegradek was visiting the Fields Institute for Research in Mathematical Sciences in Toronto (January–March, 1997).

This work of Alexei P. Stolboushkin was partially supported by NSF Grant CCR 9403809.

A part of this research was carried out while M. A. Taitslin was visiting UCLA (partially supported by NSF Grant CCR 9403809), DIMACS and Princeton (partially supported by a grant from Princeton University). The work of the author was partially supported by the Russian Foundation of Basic Research (project code: 96-01-00086).

M. A. Taitslin is the corresponding author. The address for correspondence: 62 Mozhaiskogo St., apt.265, Tver, Russia 170043.

Contents

1. Introduction

1.1. **Infinite domains.** In the relational model of databases introduced by E.F. Codd a database state is thought of as a finite collection of relations between elements. For example, a father-son relation can be represented in the form of one binary relation (or a two-column table). The names of the relations and their arities are fixed and are called a database scheme. Particular information stored in the relations of a given scheme is called a database state. As we acquire more and more information about fathers and sons, the database states change, but the scheme (one binary relation) does not.

Database relations (tables) are always going to be finite. However, it is often convenient to assume that there is an infinite *domain*—for example, the integer or rational numbers or the strings—such that the data elements are chosen from this domain. Functions and relations defined over the entire domain, like $<$ and $+$, may also be used in querying, for example, if the language of first-order logic FO is used as the query language, its formulas may use database relations as well as the domain relations, while variables range over the entire domain. A study of databases over domains equipped with an additional structure ('constraint databases') and the expressive power of the corresponding query languages ('constraint query languages') was started by Kannelakis et al. [KKR90, KKR95].

1.2. **Finitely representable relations.** The database relations are finite, but answers yielded by relational queries may or may not be finite. This makes the traditional relational model not closed, in the sense that the output of queries is of a different nature than the input.

Kanellakis et al. [KKR90, KKR95] considered the real ordered fields and groups as domains and observed that, since the first-order theories of these admit elimination of quantifiers, the answers to first-order queries can be represented as quantifier-free first-order formulas, and then, if we allow database relations to be arbitrary relations representable by quantifier-free first-order formulas to begin with, the so modified relational model becomes closed in the above sense. The quantifier-free definable relations are called *finitely representable* (for short, f.r.); the term is due to [GS94].

Finitely representable databases are a logical choice, because finitely representable relations appear as results of queries dealing with finite relations anyway, and it is also a natural choice in many applications, say, in geographical databases (cf. [KKR90, KKR95]) or spatial databases ([PVV94]).

1.3. **Ordered domains and generic queries.** The original notion of generic query [CH80][1] referred to the =-generic queries over finite database states, that is, the queries (over finite states) which are preserved under arbitrary permutations of the domain. Some practically interesting queries, say, graph properties, are indeed =-generic.

The expressive power of the pure FO with respect to generic queries is, however, severely limited—a classical example is the inexpressibility of the parity query asserting that the cardinality of a finite relation in the database scheme is even. One of the ways to try to enhance the expressive power of the query language is to allow *domain functions/relations,* or *givens* to be used in the queries, as we mentioned above. The simplest example is the relation $<$ of linear order. Throwing in such givens obviously increases the expressive power of FO, but what is often not obvious is whether any new *generic* queries become expressible.

Yu. Gurevich [Gur90] showed that there are =-generic queries that are FO expressible over finite states with $<$, but not without. Here is a version of his example.

Let K be the class of all finite Boolean algebras with an even number of atoms. This class cannot be axiomatized within the class of finite structures by a first-order sentence because if it were then, by compactness, there would exist infinite atomic Boolean algebras B and B' such that the sentence would hold in B but fail in B', in contradiction to the completeness of the theory of infinite atomic Boolean algebras (see e.g. [CK90]). However, $K_<$, the class of expansions of the algebras in K by linear orders, is axiomatizable in the class of finite structures by a first-order sentence ψ which is the conjunction of the axioms for Boolean algebras, the linear order axioms, and a sentence expressing that there is an element containing exactly the atoms at even

[1] [CH80] used a different term for this notion.

positions (in the ordering induced on the atoms) and containing the last atom. It follows that, over any infinite ordered universe U, the FO query corresponding to ψ is not equivalent to a pure FO query for finite states, even though it is $=$-generic.

Now consider the ordered set of rational numbers. Since its first-order theory admits quantifier elimination, any FO query defines a mapping that maps every finitely representable database state into a new finitely representable relation. Since any finite partial $<$-isomorphism of the rationals can be extended to an $<$-automorphism, this mapping is invariant under finite partial order isomorphisms. In this sense, all the queries over finitely represented states over the rationals that can be expressed in the first-order language FO are *locally $<$-generic.*

There are some rather simple locally $<$-generic queries, however, that are not FO-expressible. For example, the Boolean query that says that the cardinality of a finite set—a unary relation—is even, is not expressible in FO. More examples can be found in [KKR90, KKR95, KG94]. The problem we are interested in is to try to increase the expressive power of FO, while preserving local $<$-genericity.

Note that although the language FO($<$) of first-order logic with a relation of linear order does indeed express more $=$-generic queries than the pure FO, parity continues to be inexpressible. Naturally, we may ask whether, over a certain ordered domain, it is possible to express even more $=$-generic queries using extended signatures. We observe that, with each $=$-generic query being locally $<$-generic, the collapse results like the ones established in this paper are automatically transferred to the case of $=$-genericity.

Over rational numbers, the FO queries that only use $<$ were shown to have the uniform data complexity AC^0 [KG94]. Attempts to distinguish the resulting extended queries from order queries in this domain using specific combinatorial or spatial queries not in AC^0 —like parity, Eulerian traversal, or region connectivity—have been unsuccessful, and, finally, [GST95, GS95] proved the AC^0 uniform data complexity for the extended queries over finitely representable inputs *with integer constants only.* However, the question of whether or not extended queries are more expressive than order queries, has remained open, and as Grumbach and Su [GS94] pointed out, " ... there is a serious lack of proof techniques ... " in this area.

In Section 3 (Theorem 3.2) we show that, if all possible states were considered, no translation would be possible even in such a simple example as the additive group of rational numbers.

But of course the really interesting cases are those of finite and of finitely representable database states. In Section 4 we show that these two cases can be treated uniformly. One of the main results of the present paper, Theorem 4.9, is that, over every ordered domain, finitely representable states can be uniformly represented as finite states of another database scheme, with the additional property that these finite database states are FO expressible (in the restricted language) in the old database scheme, and vice versa. This technique, in effect, allows us to lift any result on translatability of extended queries into restricted queries over *finite database states*, to the *finitely representable* states. The recursiveness of translation is preserved as well. This result first appeared in [ST96].

This technique can also be used to expand applicability of several other results for finite database states to the case of finitely representable states.

1.4. **Collapse results.** Paradaens et al. [PVV95] considered real numbers with $+$, and showed that, over finite database states, locally $<$-generic extended queries can be recursively[2] translated into restricted (pure order) queries. Due to our lifting result, the same is automatically true for all finitely representable states.

In [ST96], Stolboushkin and Taitslin proved a more general result on recursive translation of locally $<$-generic extended into restricted FO queries over an arbitrary ordered divisible Abelian group, thus answering, for example, the question from [GST95, GS95] of the role of addition in databases over rational numbers. At around the same time, in their breakthrough paper [BDLW96], Benedikt et al. established the collapse of locally $<$-generic extended queries to restricted queries for a broad class of domains called *o-minimal* domains.[3] Their proof uses non-standard analysis. The notion of o-minimality was introduced in [PS86, KPS86, PS88]. Examples of o-minimal structures include real numbers with $+$, $\times$, exponentiation and $<$, as well as many other structures. Since every ordered divisible Abelian group is o-minimal, this, in one sense, covers the results of [PVV95, ST96].

Notice, however, that, in another sense, the results are of a different nature. First, the proof in [BDLW96] is not constructive and does not give an algorithm for translation. Further, this proof cannot be made constructive. Indeed, take an o-minimal structure whose first-order theory is undecidable, while the first-order theory of $<$ alone is decidable, for example, the structure $(\mathbb{R}, +, \times, <, c)$, where c is a non-computable real number. If a recursive translation existed, this would lead to a contradiction.

In this paper, we suggest an approach that gives even stronger non-effective collapse results. The approach is based on the observation that expressibility of a locally $<$-generic extended query over finite states over a universe as a restricted query is a property of the complete first-order theory of the universe rather than the universe itself. Therefore we can use the well-known model-theoretic technique of saturated models to study this property of the theory of the universe.

First, in Theorem 5.1, we give a necessary and sufficient condition for an extended query to be equivalent to a restricted query. In essence, this is a new definability theorem.

[2]Although the algorithm is not explicit in their paper.

[3]Again, they considered finite states only, but this can be lifted to finitely representable states using the technique in this paper.

Secondly, in Theorem 5.2, this technique is further refined for locally $<$-generic queries.

Thirdly, we formulate a very general condition on the domain—the so-called *Pseudo-finite Homogeneity Property*—that ensures collapse of locally $<$-generic FO queries over this domain to pure order queries (see Theorem 5.4). However, proving the Pseudo-finite Homogeneity Property for a specific domain may be a bit technical. We introduce a condition on the domain, the Isolation Property, which ensures the Pseudo-finite Homogeneity Property (Theorem 5.8).

Fourthly, we identify a broad class of domains—the so-called *quasi-o-minimal domains*—which all satisfy the Isolation Property (Theorem 5.12).

Examples of quasi o-minimal domains include the following:

- all o-minimal domains;
- the integer or natural numbers with $+, <$;
- the ordered set of real numbers with the distinguished subset of rational numbers; and
- ordered unions of o-minimal domains.

The Isolation Property is broader than quasi-o-minimality: for example, every structure of the form $(A, <, E)$, where $<$ is a dense linear ordering on the set A, and E is an equivalence relation on A with two dense classes, satisfies the Isolation Property but is not quasi-o-minimal.

The Pseudo-finite Homogeneity Property is broader than the Isolation Property: we prove that for the structure $(\mathbb{R}, +, <, \mathbb{Q})$ the Pseudo-finite Homogeneity Property holds but the Isolation Property fails. Still, over this structure, every locally $<$-generic extended query over finite states is equivalent to a restricted query. This immediately implies the analogous collapse result for any structure of the form $(A, <, E)$, where $(A, <)$ is a dense linearly ordered set without endpoints, and E is an equivalence relation on A with infinitely many classes all of which are dense.

Further, due to the Pseudo-finite Homogeneity Property, the collapse result holds for any structure of the form $(\mathbb{R}, +, <, F, f_\alpha)_{\alpha \in F}$, where F is a subfield of $\mathbb{R}$, and f_α is a name for the unary operation of multiplication by the scalar α. However, it is easy to see that for the structure $(\mathbb{R}, +, \times, <, \mathbb{Q})$ the collapse result fails. This indicates that our collapse results are edging toward the limit.

The starting point of our work on generic collapse for finite states was an attempt to 'standardize' the proof in [BDLW96], in other words, to find a proof which does not use a non-standard analysis. It turns out, however, that some of the ideas of that paper work in a much broader context. Thus, our results cover the previously known, and several new, cases. Yet the proofs are relatively compact and straightforward. These results are presented in Section 5.

Note that some of the notions and the technique we introduce turned out to be interesting in its own right, from the model-theoretic point of view (see [BPW97], [FZ97]).

Recently, after acceptance of our paper for publication, a very interesting paper of Baldwin and Benedikt [BB97] appeared, where they showed that the expressive power of the extended query language over finite states over an infinite domain is determined by stability-theoretic properties of the domain. In particular, they proved the collapse result for locally generic queries over ordered domains without the independence property. This covers our collapse result for quasi-o-minimal domains. However, it is not clear whether every theory with the Pseudo-finite Homogeneity Property does not have the independence property.

A draft version of this paper was published as [BST96]. A short abstract covering some of the results in this paper was presented in [BST97].

2. Preliminaries

A structure of a relational signature L is a non-empty set with a mapping that assigns to every relational symbol in L a relation of the same arity over the set. Let U be an infinite structure over the signature L. This structure is called the *universe.* In this paper, we always consider ordered universes. This means that L includes a binary relational symbol $<$ whose interpretation in U satisfies the axioms of linear order. Let us denote $L_0 = \{<\}$. A *database scheme* SC is a finite collection of relational symbols of fixed arities. A *database state* (over U) is an assignment to these relational symbols of concrete relations of corresponding arities over U. These relations are called *database relations.* A database state is called a *finite database state* if all the relations are finite. We fix a database scheme SC and denote $L_0^+ = L_0 \bigcup SC$, and $L^+ = L \bigcup SC$.

Let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ be the sets of all natural, integer, rational, and real numbers, respectively. Practically, the most interesting cases of universes are:

- $(\mathbb{Q}, <, +)$ and $(\mathbb{R}, <, +)$
- $(\mathbb{Z}, <, +)$ and $(\mathbb{N}, <, +)$
- $(\mathbb{R}, <, +, \times)$

The set of all elements of the universe that occur in some tuple in some relation of a database state s is called the active domain of s; we denote it by $AD(s)$. We denote $AD(x)$ a first-order formula of signature SC which says that x is an element of the active domain. For a subset X of the universe U, a database state s over U is called a database state over X iff $AD(s)$ is a subset of X.

For convenience, we will consider database schemes that contain not only relation symbols, but also finitely many constant symbols. A database state over a universe U for such a scheme is a mapping that assigns to any relation symbol in the scheme a relation on U of the corresponding arity, and to any constant symbol in the scheme an element in U. In this case the active domain of a database state is defined to be the union of the active domain of the relational part of the state and the set of values of all constants of the scheme. For a relational database scheme SC, denote by SC_k the scheme $SC \cup \{c_1, \ldots, c_k\}$, where the c_i are new constant symbols.

A *database query* can formally be defined as a mapping that takes in a database state (of a fixed database scheme), and produces a new relation, of a fixed arity, over U. Thus, every query has an arity. Specifically, queries of arity 0 are called *Boolean queries.* A Boolean query defines a mapping from database states to $\{0, 1\}$, or, in other words, a set of database states of a given database scheme.

Queries can be formulated using *query languages,* the simplest being the language of first-order logic FO. Formulas (queries) of this language use $=$, as well as the relational symbols of the signature and of the database scheme. Thus, a database state essentially defines a structure of a larger signature with U as the domain; then a formula with n free variables defines an n-ary relation over U; sentences define Boolean queries. We say that two L^+-formulas with n free variables are equivalent over finite states over U if they define the same n-ary query.

Clearly, two L^+-formulas $\phi(x_1, \ldots, x_k)$ and $\psi(x_1, \ldots, x_k)$ are equivalent over finite states over a universe U if and only if the sentences $\phi(c_1, \ldots, c_k)$ and $\psi(c_1, \ldots, c_k)$ of signature $L \cup SC_k$ are equivalent over finite states over U.

A relation on U is said to be *finitely representable* if it can be defined by a quantifier-free formula using $=, <$, and constants for the elements of U. Such a formula is called a *finite representation* of the relation; if each element whose name occurs in the formula belongs to a set X the formula is called a *finite representation* over X. A database state is said to be *finitely representable* if every relation corresponding to a relation name from SC is finitely representable.

We consider two languages for querying. Queries of the first one are FO formulas of the signature L_0^+—we call them *restricted.* Queries of the second language are FO formulas of the signature L^+—we call them *extended.*

It is easy to see that the restricted queries are *generic,* in the sense that they are preserved under order-preserving permutations of U. In other words, if $f : U \to U$ is an automorphism of $\langle U, < \rangle$, and a restricted query Q maps a database state s to a relation R, then Q maps $f(s)$ to $f(R)$; in other words, $Q(f(s)) = f(Q(s))$. Extended queries may be not generic. Here for any state s over U, we define $f(s)$ as a state over U such that for any R of SC and any $\bar{a} \in U$,

$$((f(s))(R))(f(\bar{a})) \text{ iff } (s(R))(\bar{a})$$

where $f(a_1, \ldots, a_m) = (f(a_1), \ldots, f(a_m))$.

More generally, let M be an L-structure, and K be a class of states over M. We call an L-formula $\phi(\bar{x})$ generic over the states from K if, for any state s from K, any tuple $\bar{a}$ in M, and any L_0-automorphism f of the structure

M, $\phi(\bar{a})$ holds in (M, s) iff $\phi(f(\bar{a}))$ holds in $(M, f(s))$, in symbols,

$$(M, s) \models \phi(\bar{a}) \quad \text{iff} \quad (M, f(s)) \models \phi(f(\bar{a})).$$

We will also use a stronger notion of *locally generic* query. It was proposed in [BDLW96]. A k-ary query Q is said to be *locally generic over finite states* if $\bar{a} \in Q(s)$ iff $f(\bar{a}) \in Q(f(s))$, for any partial $<$-isomorphism $f : X \to U$ with $X \subseteq U$, for any finite state s over X, and for any k-tuple $\bar{a}$ in X.

For any finite representation σ over a subset X of U of a relation and for any partial $<$-isomorphism $f : X \to U$, a finite representation $f(\sigma)$ of the relation can be naturally defined, by replacing any parameter a that occur in σ with the parameter $f(a)$. So, for finitely representable states, the notion of local genericity can be defined as follows. A k-ary query Q is said to be *locally generic over finitely representable states* if $\bar{a} \in Q(\sigma)$ iff $f(\bar{a}) \in Q(f(\sigma))$, for any partial $<$-isomorphism $f : X \to U$ with $X \subseteq U$, for any finite representation σ over X, and for any k-tuple $\bar{a}$ in X. Here we denote by $Q(\sigma)$ the state into which the query Q transforms the state finitely represented by σ.

Since a finite n-ary relation $\{\bar{a}_1, \ldots, \bar{a}_n\}$, where $\bar{a}_i = (a_{i1}, \ldots, a_{in})$, can be finitely represented by the formula $\bigvee_{i=1}^{m} \bigwedge_{j=1}^{n} x_j = a_{ij}$, every query which is locally generic over finitely representable states is locally generic over finite states, too. On the other hand, a query which is locally generic even over *all* states can be not locally generic over finitely representable states: an example is the Boolean query '$P \neq \emptyset$'; it is obviously locally generic over all states over $\mathbb{Z}$, but $0 < x < 1$ defines in $\mathbb{Z}$ the empty set, even though the set defined by $0 < x < 2$ in $\mathbb{Z}$ is not empty.

The notions of genericity and local genericity for SC_k-queries are defined exactly the same way as for SC-queries. Clearly, a k-ary SC-query $\phi(x_1, \ldots, x_k)$ is generic (locally generic) over U iff the Boolean SC_k-query $\phi(c_1, \ldots, c_k)$ is generic (locally generic) over U.

Of course, every locally generic query is generic. Conversely, for some domains, every generic query is locally generic. It was noted in [BDLW96] that a sufficient condition for this is the so called *double transitivity* of the domain: a domain is called doubly transitive if for any $a_1 < b_1$ and $a_2 < b_2$ in the domain, there exists an $<$-automorphism of the domain mapping a_1 to a_2, and b_1 to b_2. (A proof: if U is doubly transitive any partial $<$-isomorphism $f : X \to U$ with finite $X \subseteq U$ can be extended to an $<$-automorphism of U.) For instance, the real and rational numbers are doubly transitive, while the integer numbers are not. The Boolean query *'there are even and odd numbers in P'* is an example of a query which is generic but not locally generic over finite states over $\mathbb{Z}$. Moreover, even a restricted query can be not locally generic: for example, the restricted Boolean query *'P is convex'* is not locally generic over finite states over $\mathbb{Z}$.

3. Impossibility of translation over arbitrary states

The goal of this section is to compare restricted and generic extended queries from the viewpoint of their expressive power over *all possible states,* whether finitely representable or not. We show that, in general, extended generic querying is more expressive than restricted, even in very simple situations. In fact, the results in this section can be proved, with a little more trouble, for all recursive states.

Theorem 3.1. *There is an extended Boolean query Q over $(\mathbb{Z}, +, <)$, the ordered group of integer numbers, such that*

(1) *Q is generic over all database states; in particular, Q is generic over all finite states;*
(2) *Q is not equivalent, over finite database states, to a restricted query; in particular, Q is not equivalent, over all database states, to a restricted query.*

Proof. Let Q be the query *'there are even and odd numbers in P'.* The query Q is generic over all possible database states. Indeed, the order automorphisms of the integers are exactly the maps of the form $x \mapsto x + n$; clearly, the query Q agrees on P and $P + n$, for any integer n.

Clearly, there is an extended FO query that expresses Q, for all possible database states.

We show that Q cannot be expressed as a FO restricted query, for finite database states. Towards a contradiction, suppose there exists a first-order sentence φ of the signature $L = \{<, P\}$ such that an L-structure $(\mathbb{Z}, <, X)$ with a finite X satisfies φ iff there are even and odd numbers in X. Let Γ be the infinite set of L-sentences that assert:

- $<$ is a discrete linear order without endpoints
- P has endpoints
- every element of P, which is not maximal in P, has a successor in P
- every element of P, which is not minimal in P, has a predecessor in P
- between any two elements of P there are infinitely many points
- P has infinitely many elements

Clearly, for every finite subset Δ of Γ, both of the sets Δ, φ and $\Delta, \neg\varphi$ have models of the form $(\mathbb{Z}, <, X)$ with a finite X. Then, by compactness, Γ, φ and $\Gamma, \neg\varphi$ are both consistent. Using standard model-theoretic arguments (for example, a Fraïssé-Ehrenfeucht game) one can show that the theory Γ is complete. A contradiction. ∎

Note that Q constructed is obviously not locally generic. Moreover, it will be shown in Section 5 that *every* FO extended query, which is locally generic over finite states over $(\mathbb{Z}, <, +)$, is equivalent, for finite database states, to an FO restricted query (Theorem 5.12). A similar result will be proved for $(\mathbb{Q}, <, +)$, the ordered group of rational numbers.

By the way, the mentioned result from Section 5 concerning $\mathbb{Z}$ has a curious corollary: *the query '$|P|$ is even' cannot be expressed as an extended FO query for finite database states over* $(\mathbb{Z}, <, +)$, *as opposed to the query 'there are even and odd numbers in P'*. Indeed, the query '$|P|$ *is even*' is obviously locally generic, even over all database states, and essentially the same arguments, as in the proof of Theorem 3.1, show that the query is not equivalent, for finite database states, to an FO restricted query over $(\mathbb{Z}, <, +)$. Note, for contrast, that the query '$|P|$ *is finite*' can be expressed as a restricted FO query over $(\mathbb{Z}, <, +)$, because a set of integers is finite iff it is bounded.

It is natural to ask whether Theorem 3.1 holds for $\mathbb{Q}$ instead of $\mathbb{Z}$. In this situation, in contrast to the case of $(\mathbb{Z}, <, +)$, the notions of genericity and local genericity coincide. However, for $(\mathbb{Q}, <, +)$, we will give an example of an extended query which is generic over all database states, but not equivalent, over all database states, to a restricted one. That example draws a line between finite and finitely representable database states, on one side, and essentially infinite states, on the other.

In fact, we will prove a more general result:

Theorem 3.2. *The extended querying is more expressive over all the database states than the restricted one with respect to generic Boolean queries over any divisible Archimedean ordered Abelian group not isomorphic to the ordered group of reals.*

Classical examples of divisible Archimedean ordered Abelian groups are the ordered groups of rational and real numbers. It is known that, up to isomorphism, Archimedean ordered groups are exactly the subgroups of the ordered group of reals. The following is an example of a uncountable divisible Archimedean ordered Abelian group not isomorphic to the ordered group of reals.

Let b be an irrational number. Consider a basis B of $\mathbb{R}$ over $\mathbb{Q}$, containing b; clearly B is of power of continuum. Let G be a $\mathbb{Q}$-subspace of $\mathbb{R}$ generated by $B \setminus \{b\}$. Then G is a required group. Indeed, clearly it is Archimedean and divisible. The orders on G and $\mathbb{R}$ are not isomorphic: the order on G is not complete as b is not in G.

We don't know whether the result of Theorem 3.2 holds for the ordered group of reals.

To prove Theorem 3.2, it suffices to prove the following two results.

Theorem 3.3. *Let $(A, +, <)$ be an Archimedean ordered Abelian group. Then the finiteness of database states over A is expressible by an extended FO query.*

Theorem 3.4. *Let A be a set of reals containing $\mathbb{Q}$. Then the finiteness of database states over A is expressible by a restricted FO query iff $A = \mathbb{R}$.*

Note, for contrast, that in $(\mathbb{Z}, <)$ the finiteness is expressible by a restricted FO query because a set of integers is finite iff it is bounded.

Proof of Theorem 3.3. It is easy to see that a database state over A is finite iff its active domain is bounded and there is a positive $a \in A$ such that $a \leqslant |x - y|$, for any distinct x and y in the active domain. The latter can be obviously expressed in the first-order language of ordered groups. ∎

Proof of Theorem 3.4. If $A = \mathbb{R}$, a database state over A is finite iff its active domain is bounded and has no limit points; the latter is obviously expressible by a restricted FO query.

Suppose $A \neq \mathbb{R}$. Then in $\mathbb{Q}$ there are $a_0 < a_1 < \dots$ and $a'_0 > a'_1 \dots$ with $\sup a_n = \inf a'_n \notin A$.

Towards a contradiction, assume that there is a sentence φ of the signature $L = \{<, R\}$, where R is a unary relation symbol, such that, for any $X \subseteq A$, the sentence φ holds in $(A, <, X)$ iff X is finite.

Let T be the first-order L-theory that says:

- $<$ is a dense order without endpoints
- R has a minimal element and a maximal one provided $R \neq \emptyset$
- each element has a successor in R provided the element is not an upper bound for R
- each element has a predecessor in R provided the element is not a lower bound for R

Let T' be the theory of all structures of the form $(B, <, F)$, where $(B, <)$ is a dense ordered set without endpoints and F is a finite subset of B. As the theory of dense ordered sets without endpoints is complete, T' is the theory of all structures of the form $(A, <, F)$, where F is a finite subset of A. Therefore $\varphi \in T'$. Clearly, $T \subseteq T'$; we will show that T is an axiom system for T'. Then a contradiction follows because, for $X = \{a_i, a'_i : i < \omega\}$, the structure $(A, <, X)$ is a model of $T, \neg\varphi$.

We need to show that every model of T with infinite R is a model of T'. As, by compactness, T' has a model with infinite R, it suffices to show that any two models of T with infinite R are elementarily equivalent. The latter is an easy corollary of the well-known results on the completeness of the following two theories:

- T_0, the theory of infinite orderings with maximal and minimal elements in which every non-maximal element has a successor and every non-minimal element has a predecessor;
- T_1, the theory of dense orderings without endpoints.

Indeed, it suffices to prove that any two special model of T of the same power λ with infinite R are isomorphic. (For the definition and the properties of special models we need, see Section 5). For every such a model M, the ordering $(R^M, <)$ is a special model of power λ of T_0; so its isomorphism type is uniquely determined, due to the completeness of T_0. For any subsequent elements a and b in R^M, the interval (a, b) in M is a special model of T_1 of power λ; so its isomorphism type is uniquely determined, due to the completeness of T_1. The same is true for the intervals $(-\infty, \min(R^M))$ and

$(\max(R^M), \infty)$ in M. Since every element of $M \setminus R^M$ belongs to one of the intervals of the forms above, the isomorphism type of M is uniquely determined, and we are done. ∎

In the special case when A is countable, Theorem 3.4 admits an especially simple proof. If φ expressed the finiteness of R in $(A, <, R)$ then, by compactness and the Löwenheim-Skolem theorem, there would be an infinite subset in a countable dense ordered set without endpoints, for which φ holds. As every countable dense ordered set without endpoints is isomorphic to $(A, <)$, we would have a contradiction with the choice of φ.

We conclude the section with some observations concerning translations over all states. There is a connection between the problem of equivalence of generic extended queries over arbitrary states and classical definability results.

Theorem 3.5. *Let T be an L-theory. Suppose an L^+-formula $\phi(\bar{x})$ is L_0-generic over arbitrary states over any model of T. Then there is a finite set of L_0^+-formulas $\phi_1(\bar{x}), \dots, \phi_n(\bar{x})$ such that $T \vdash \bigvee_{i=1}^n \forall \bar{x}\, (\phi(\bar{x}) \leftrightarrow \phi_i(\bar{x}))$. In particular, over any model of T, the formula ϕ is equivalent over arbitrary states to an L_0^+-formula.*

Here the symbol $\vdash$ denotes, as usual, the provability relation for the first-order logic.

The following proof works not only for $L_0 = \{<\}$ but for an arbitrary L_0.

We deduce the result from the following Svenonius' definability theorem proven in 1959 (see [Hod93], Corollary 10.5.3).

Svenonius' Definability Theorem. *Let L_0, L be first-order languages with $L_0 \subseteq L$, and R a relation symbol of L. Let T be an L-theory such that for every model M of T the relation R^M is invariant under all L_0-automorphisms of M. Then there is a finite set of L_0-formulas $\psi_1(\bar{x}), \dots, \psi_n(\bar{x})$ such that $T \vdash \bigvee_{i=1}^n \forall \bar{x}\, (R(\bar{x}) \leftrightarrow \psi_i(\bar{x}))$.*

Proof of the Theorem 3.5. Let R be a new relation name of arity $n+1$, where n is the length of $\bar{x}$. Consider the $L^+(R)$-theory

$$T' = T \cup \{\forall \bar{x} y\, (R(\bar{x}, y) \leftrightarrow \phi(\bar{x}))\}.$$

(If ϕ is not a sentence, we may let R to be an n-ary relation name and omit the y; however, as in Svenonius' theorem the arity of R is supposed to be nonzero, we introduce the y when ϕ is a sentence.)

Due to the L_0-genericity of ϕ over all states over any model of T, in any model of T' the interpretation of R is invariant under all L_0^+-automorphisms. Then, by Svenonius' theorem, there are L_0^+-formulas $\psi_1(\bar{x}, y), \dots, \psi_n(\bar{x}, y)$ such that

$$T' \vdash \bigvee_{i=1}^{n} \forall \bar{x} y\, (R(\bar{x}, y) \leftrightarrow \psi_i(\bar{x}, y)).$$

It follows that

$$T \vdash \bigvee_{i=1}^{n} \forall \bar{x} \, (\phi(\bar{x}) \leftrightarrow \exists y \, \psi_i(\bar{x}, y)),$$

and we are done. ∎

In the Theorem 3.5, it is essential that we assume the L_0-genericity of ϕ over arbitrary states over *all* models of T: even in the case of complete T it can happen that ϕ is L_0-generic and equivalent to an L_0^+-formula over one model and is L_0-generic and not equivalent to an L_0^+-formula over another model. Such an example—for the theory of ordered nonzero divisible Abelian groups—is contained in our paper (see Theorems 3.3 and 3.4).

4. Canonical representation for finitely representable relations

The goal of this section is to show that finite and finitely representable states can be treated uniformly. To achieve this goal, we show in Theorem 4.9 that finitely representable relations can be represented uniformly by finite relations, of a different signature, such that these f.r. states and their finite "codes" can be mapped to each other by restricted queries.

We begin with a simple example. Consider the following finitely represented unary relation P_0 on $\mathbb{Q}$:

$$(1 < x \leq 2) \vee (3 \leq x < 4) \vee (5 < x).$$

To reconstruct P_0, it suffices to know:

- the set of constants B_0 used in the representation—in our case it is $\{1, 2, 3, 4, 5\}$, and, moreover,
- which of the singletons and the open intervals

  $$(-\infty, 1), \quad \{1\}, \quad (1, 2), \quad \{2\}, \quad (2, 3), \quad \{3\}, \quad (3, 4), \quad \{4\}, \quad (4, 5), \quad (5, \infty)$$

  are contained in P_0.

The latter sets can be characterized as minimal open intervals and singletons which can be defined using constants from B_0; we will call them B_0-minimal 1-cells. Clearly, P_0 is the union of all B_0-minimal 1-cells which are contained in P_0.

The set of constants over which P_0 can be defined is not uniquely determined by P_0; for example, P_0 can be represented over the set $\{0, 1, 2, 3, 4, 5\}$ by the formula

$$((0 < x \leq 2) \vee (3 \leq x < 4) \vee (5 < x)) \wedge (1 < x).$$

However, the constant 0 is obviously irrelevant here as it is shadowed by the second conjunct. In fact, the constants which are really relevant are just the boundary points of P_0. Clearly, there is a unary restricted FO query δ which, for any subset P of $\mathbb{Q}$ as an input, yields its boundary as an answer.

We show that the information which is contained in the second item can be obtained from P_0 by means of several FO restricted queries (which can be uniformly applied to any finitely represented subset of $\mathbb{Q}$).

Let B be a boundary set of such a P; it is a finite set. There are 5 types of B-minimal 1-cells:

$$(0)\ \mathbb{Q}; \qquad (1)\ \{b\}; \qquad (2)\ (-\infty, b); \qquad (3)\ (b, \infty); \qquad (4)\ (b, b');$$

here $b, b' \in B$, and in all the cases except (1) the 1-cells do not contain any point from B.

Consider the following 5 relations S_i on B ($i < 5$). The relation S_0 is 0-ary; the relations S_1, S_2 and S_3 are unary, and the relation S_4 is binary. We put

- $S_0 = \mathbf{true}$ iff $\mathbb{Q} = P$
- $S_1(b)$ holds iff $b \in B$
- $S_2(b)$ holds iff b is the least element of B and P contains $(-\infty, b)$
- $S_3(b)$ holds iff b is the greatest element of B and P contains (b, ∞)
- $S_4(b, b')$ holds iff b, b' are subsequent elements of B, and P contains the interval (b, b')

In other words, S_i contains the information which of the B-minimal 1-cells of type (i) are contained in P.

It is easy to write down a restricted FO query σ_i which, for every finitely represented subset P of $\mathbb{Q}$ as an input, yield the finite relation S_i as the answer. Clearly, P can be recovered from the finite relations B and S_i's by means of a FO query π, because P is the union of all B-minimal 1-cells which are contained in P.

So, we have shown that, for any finitely representable subset P of Q, we can find, uniformly in P by means of FO queries, a finite collection of finite relations on $\mathbb{Q}$, from which P can be uniformly recovered by means of a FO query.

Our goal is to prove an analogous result for finitely represented relations of arbitrary arity. Here the idea is essentially the same, but some new important points appears. We illustrate this for the case of arity 2.

Consider a binary finitely represented relation P on $\mathbb{Q}$. We may assume that P is contained in one of the three relations

$$\{(a,b) : a < b\}, \qquad \{(a,b) : a > b\}, \qquad \{(a,b) : a = b\},$$

because P is the disjoint union of the intersections of P with these three relations, and the intersections are finitely representable. Assume, for example, that $P \subseteq \{(a,b) : a < b\} = D$.

It can be proven (it is not obvious!) that among the finite subsets of $\mathbb{Q}$, over which P can be finitely represented, there is the least one; we call it the set of boundary points for P and denote it by ∂P. For example, if P is the relation

$$(0 < x < y < 1) \vee (1 < x < y < 2)$$

then $\partial P = \{0, 1, 2\}$. Moreover, it turns out that ∂P can be obtained from P by means of a FO query, uniformly in P.

For a finite subset B of $\mathbb{Q}$, we define a simple binary relation on $\mathbb{Q}$ over B to be a finite union of "rectangulars" defined over B; that is, the *simple* binary relations over B are those which can be finitely represented by disjunctions of conjunctions of formulas of the forms

$$x = b, \quad x < b, \quad x > b, \quad y = b, \quad y < b, \quad y > b,$$

where the parameters b are taken from B. It is easy to show that simple binary relations on $\mathbb{Q}$ can be characterized as the binary relations that are invariant under all order automorphisms of $\mathbb{Q}$ which stabilize each element of B.

We will show that there is the least simple binary relation on $\mathbb{Q}$ over ∂P containing P; we denote it by $\mathrm{Inv}(P)$. For the P from the example above, $\mathrm{Inv}(P)$ is the union of two squares, $(0, 1) \times (0, 1)$ and $(1, 2) \times (1, 2)$. It is easy to see that $\mathrm{Inv}(P)$ can be uniformly obtained from P by means of a binary first-order query. It can be shown—it is the crucial point—that it always the case that $P = \mathrm{Inv}(P) \cap D$.

We call a set of the form $I \times J$, where I and J are B-minimal 1-cells, a B-minimal 2-cell. It is easy to see that any simple binary relation over B can be uniquely decomposed into a disjoint union of B-minimal 2-cells. For instance, in the example above $\mathrm{Inv}(P)$ is the disjoint union of two $\{0, 1, 2\}$-cells $(0, 1) \times (0, 1)$ and $(1, 2) \times (1, 2)$. Clearly, there are finitely many B-minimal 2-cells, for any finite B.

If we know ∂P, to reconstruct $\mathrm{Inv}(P)$, we need only to know which of the ∂P-minimal 2-cells are contained in $\mathrm{Inv}(P)$.

As there are 5 types of ∂P-minimal 1-cells, there are 25 types of ∂P-minimal 2-cells. With every type $\tau \in 5^2$ of ∂P-minimal 2-cells we associate a relation S_τ on the finite set ∂P with the information which of the ∂P-minimal 2-cells are contained in $\mathrm{Inv}(P)$. The arity of S_τ depends on τ and is equal to the number of the b's involved in representations of 2-cells of type τ.

Say, with the 2-cells $I \times J$, where I is of type (1) and J is of type (4), we associate the following ternary relation S_{14} on ∂P. For $a, b, c \in \partial P$, we consider $S_{14}(a, b, c)$ to be true iff $\{a\} \times (b, c)$ is a ∂P-minimal 2-cell and is contained in $\mathrm{Inv}(P)$.

For P in the example above, $S_{44} = \{(0, 1, 0, 1), (1, 2, 1, 2)\}$, $S_{00} = \mathbf{false}$, and $S_\tau = \emptyset$ for all remaining $\tau \in 5^2$.

For any $\tau \in 5^2$, we can uniformly obtain the finite relation S_τ on ∂P from $\mathrm{Inv}(P)$ and ∂P (and so from P) by means of a FO query.

On the other hand, if we know the finite relations ∂P and all S_τ's, we can uniformly recover $\mathrm{Inv}(P)$ from them, by means of a FO query. As $P = \mathrm{Inv}(P) \cap D$, the same is true for P. So in the binary case we have the result we need.

Actually, the arguments above work not only for $\mathbb{Q}$ but for an arbitrary dense ordered set. However, we will prove the result not only in the dense case, but for an arbitrary linear ordered set. In that case some extra technical problems with the definition of ∂P arise, because in general situation for a finitely representable relation the least set over which the relation can be defined does not exist: for example, in $\mathbb{Z}$ the relation $x > 0$ can be represented not only over $\{0\}$ but also over $\{1\}$, because $x > 0$ iff $x \geq 1$. Nevertheless, even in general case it turns out to be possible to define for any finitely represented relation P a certain canonical finite set of parameters ∂P, over which P can be defined and which can be uniformly obtained from P by means of a FO query.

Now we pass to the general case. We work over an arbitrary (but fixed) linearly ordered set U.

Our aim is to find for finitely representable relations on U, in a sense, a canonical finite representation. We begin with a special case of the so called simple relations.

A relation R on U is said to be *simple* if it can be finitely represented by a disjunction of conjunctions of formulas of the forms $x < c$, $x > c$, or $x = c$, where x is a variable and c is a name of an element of U.

For $k \geqslant 1$, we call sets of the form $I_1 \times \cdots \times I_k$, where each I_j is a singleton or an open interval in U *k-cells* in U. Here an open interval in U is a set defined by a formula of one of the following forms: $x = x$; $a < x < b$; $x < a$; $a < x$. Geometrically, simple k-ary relations over a set B can be described as unions of finitely many k-cells such that all parameters involved in their representations belongs to B. Of course, neither the k-cells nor the set of parameters B are uniquely determined by the simple relation. Simple relations can be characterized as follows.

Lemma 4.1. *For a finite set B, a relation P on U is a simple relation over B iff P is B-invariant.*

Here P is said to be *B-invariant* if $\bar{a} \in P$ iff $\bar{b} \in P$, for any tuples $\bar{a}$ and $\bar{b}$ such that a_i and b_i are positioned the same way with respect to the elements of B, for all i. The proof of the lemma is obvious.

Consider the following binary relation $E_k(\bar{x}, \bar{y})$ on U^k: $x_i < x_j$ iff $y_i < y_j$, for $1 \leqslant i, j \leqslant k$. It is an equivalence relation with finitely many classes. *Fix such a class D.* The relation $\{(i,j) : a_i = a_j, \text{ for } \bar{a} \in D\}$ is an equivalence relation on $\{1, \ldots, k\}$; let $J_1, \ldots, J_s$ be its classes.

For $P \subseteq D$ and $u, v \in U$, we say that u and v are *P-inseparable* if for any $\bar{a}, \bar{b}$ in D and any $i \in \{1, \ldots, s\}$

$$(\bigwedge_{j \notin J_i} a_j = b_j \wedge \bigwedge_{j \in J_i} (a_j = u \wedge b_j = v)) \rightarrow (P(\bar{a}) \equiv P(\bar{b})).$$

Hence, this relation of inseparability is expressible as a restricted query.

For $P \subseteq D$, an element $x \in U$ is said to be a *boundary point* of P if either there is y such that $y < x$ and for any $y < x$ there is a pair of P-separable elements in $[y, x]$, or there is y such that $y > x$ and for any $y > x$ there is a pair of P-separable elements in $[x, y]$. Here $[u, v] = \{z : u \leqslant z \leqslant v\}$. Denote by ∂P the set of boundary points of P. It will be observed in Lemma 4.3 that any boundary point of P is a constant in every definition of P or is adjacent to such a constant. So the set of the boundary points for each prime relation P is finite and can be expressed as an restricted query. A relation is called prime iff it is $S \cap D$ for a simple relation S.

In Lemmas 4.2–4.5 below, let S be a simple relation defined over a finite set B, and $P = S \cap D$.

Lemma 4.2. *If $x < y$ and $[x, y] \cap B = \emptyset$, there is no pair of P-separable elements in $[x, y]$.*

Proof. Let $u, v \in [x, y]$. Let $\bar{a}, \bar{b} \in D$, and $1 \leqslant i \leqslant s$. Suppose $a_j = b_j$ for $j \notin J_i$, and $a_j = u$, $b_j = v$ for $j \in J_i$. As u and v are positioned the same way with respect to the elements of B, we have $\bar{a} \in S$ iff $\bar{b} \in S$, and hence $\bar{a} \in P$ iff $\bar{b} \in P$. ∎

Lemma 4.3. *Any boundary point of P is an element of B or is adjacent to an element of B. In particular, ∂P is finite.*

Proof. If z is neither an element of B nor adjacent to an element of B, there are x, y such that $x < z < y$, and $[x, y] \cap B = \emptyset$. Then, by Lemma 4.2, there is no pair of P-separable elements in $[x, y]$. Hence $z \notin \partial P$. ∎

The following lemma is crucial in our consideration.

Lemma 4.4. *If $x < y$ and x, y are P-separable, $[x, y] \cap \partial P \neq \emptyset$.*

Proof. By Lemma 4.2, $[x, y] \cap B \neq \emptyset$. Let $[x, y] \cap B = \{b_1, \ldots, b_n\}$, $b_1 < \cdots < b_n$. Since x, y are P-separable, the pair x, b_1, or the pair b_n, y, or one of the pairs b_i, b_{i+1} is P-separable.

Indeed, there are $\bar{a}, \bar{d}$ in D and $i \in \{1, \ldots, s\}$ such that

$$(\bigwedge_{j \notin J_i} a_j = d_j \wedge \bigwedge_{j \in J_i} (a_j = x \wedge d_j = y)) \wedge (P(\bar{a}) \not\equiv P(\bar{d})).$$

Let b_0 be x, b_{n+1} be y, and for $k = 0, \ldots, n+1$, $\bar{a}(k)$ be such a sequence that

$$(\bigwedge_{j \notin J_i} a(k)_j = a_j \wedge \bigwedge_{j \in J_i} a(k)_j = b_k).$$

There is such a k that

$$(P(\bar{a}(k)) \not\equiv P(\bar{a}(k+1))).$$

It means that b_k, b_{k+1} is P-separable.

Towards a contradiction, suppose that neither x, y nor the b_i's are boundary points of P. Then b_i and b_{i+1} are P-inseparable, for all i. Indeed, there are $u > b_i$ and $v < b_{i+1}$ such that in $[b_i, u]$ and $[v, b_{i+1}]$ there are no P-separable pairs of elements. If $u \geqslant b_{i+1}$ or $v \leqslant b_i$, we are done. If $b_i < v \leqslant u < b_{i+1}$, the pairs b_i, v and v, b_{i+1} are P-inseparable, and hence the pair b_i, b_{i+1} is P-inseparable, too. In the case $b_i < u < v < b_{i+1}$ the pair u, v is P-inseparable, by Lemma 4.2; since the pairs b_i, u and v, b_{i+1} are P-inseparable, we have the P-inseparability of the pair b_i, b_{i+1}. Analogous arguments show the P-inseparability of each of the pairs x, b_1 and b_n, y. A contradiction. ∎

Denote by $\mathrm{Inv}(P)$ the least ∂P-invariant relation containing P. Clearly, $\mathrm{Inv}(P)$ consists of all k-tuples $\bar{b}$ for which there is $\bar{a} \in P$ such that a_i and b_i are positioned the same way with respect to ∂P, for all i. By Lemma 4.1, $\mathrm{Inv}(P)$ is a simple relation over ∂P.

Lemma 4.5. $P = \mathrm{Inv}(P) \cap D$.

Proof. We need to prove that there are no $\bar{a} \in P$ and $\bar{b} \in D \setminus P$ such that a_i and b_i are positioned the same way with respect to ∂P, for all i. Suppose there is a counterexample pair $\bar{a}, \bar{b}$. Let $i_1 \in J_1, \dots, i_s \in J_s$; we can assume that $x_{i_1} < \dots < x_{i_s}$ for $\bar{x} \in D$. As $\bar{a} \neq \bar{b}$, there is m such that $a_{i_m} \neq b_{i_m}$, but $a_{i_l} = b_{i_l}$ for $1 \leqslant l < m$. Choose a counterexample pair $\bar{a}, \bar{b}$ with the largest possible m. As a_{i_m} and b_{i_m} are positioned the same way with respect to ∂P, in the closed interval between them, there are no points in ∂P. So, by Lemma 4.4, a_{i_m} and b_{i_m} are P-inseparable.

Suppose $a_{i_m} < b_{i_m}$. Let $\bar{c} = (c_1, \dots, c_k)$ be the result of replacement in the tuple $\bar{b}$ the elements b_j with a_{i_m}, for all $j \in J_m$. Clearly, $\bar{c} \in D$. Due to the P-inseparability of a_{i_m} and b_{i_m}, we have $\bar{c} \notin P$. The elements a_i and c_i are positioned the same way with respect to ∂P, for all i; so $\bar{a}, \bar{c}$ is a counterexample pair. Since $a_{i_l} = c_{i_l}$ for $1 \leqslant l \leqslant m$, we have a contradiction with the maximality of m.

Now suppose $a_{i_m} > b_{i_m}$. Let $\bar{c} = (c_1, \dots, c_k)$ be the result of replacement in the tuple $\bar{a}$ the elements a_j with b_{i_m}, for all $j \in J_m$. Clearly, $\bar{c} \in D$. Due to the P-inseparability of a_{i_m} and b_{i_m}, we have $\bar{c} \in P$. The elements c_i and b_i are positioned the same way with respect to ∂P, for all i, so $\bar{c}, \bar{b}$ is a counterexample pair. Since $c_{i_l} = b_{i_l}$ for $1 \leqslant l \leqslant m$, we have a contradiction with the maximality of m. ∎

Any k-ary relation P is the disjoint union of all $P \cap D$, where D runs over the set of E_k-classes. Obviously, if P is finitely representable, every such $P \cap D$ is equal to $S \cap D$, for some simple relation S. So Lemma 4.5 implies

Corollary 4.6. *Any finitely represented k-ary relation P is the disjoint union of all* $\mathrm{Inv}(P \cap D) \cap D$*, where D runs over the set of equivalence classes of E_k.*

Let P be a finitely representable k-ary relation. Denote by ∂P the union of all $\partial(P \cap D)$. As $\mathrm{Inv}(P \cap D)$ is defined over $\partial(P \cap D)$, we have

Corollary 4.7. *Any finitely represented relation P is defined over ∂P.*

Thus, with every finitely represented relation P we have associated a certain canonical finite set of parameters ∂P, over which the relation is defined; the relation P is, in a sense, reduced to a finite family $\{\mathrm{Inv}(P \cap D)\}$ of simple relations over the set ∂P. Moreover, the set ∂P and the family $\{\mathrm{Inv}(P \cap D)\}$ can be found uniformly in P (by means of certain FO queries), and P can be uniformly recovered from the set ∂P and the family $\{\mathrm{Inv}(P \cap D)\}$ (by means of a certain FO query).

Now we are going to find a canonical representation for simple relations. Let S be a simple k-ary relation over a finite set B. For a non-empty B, we call a k-cell $I_1 \times \cdots \times I_k$ *B-minimal* if every I_j can be defined by a formula of one of the following forms:

$$x = b_i; \qquad x < b_1; \qquad b_n < x; \qquad b_i < x < b_{i+1},$$

where $B = \{b_1, \ldots, b_n\}$ and $b_1 < \cdots < b_n$. The only $\emptyset$-minimal k-sell is, by definition, the cell U^k. Obviously, the set of B-minimal k-cells is finite. Clearly, the B-minimal k-cells are pairwise disjoint. Moreover, if a k-cell C' is defined over B and a k-cell C is B-minimal then $C \subseteq C'$ provided $C \cap C' \neq \emptyset$. It follows that S can be decomposed into a disjoint union of B-minimal k-cells, namely, into the disjoint union of all B-minimal k-cells which are contained in S. Note that some of the B-minimal k-cells can be empty.

We show how to encode the simple relation S by a finite family of relations on the finite set B.

Every 1-cell over B is defined by a formula of one of the following forms:

$$(0)\ x = x; \qquad (1)\ x = b; \qquad (2)\ x < b; \qquad (3)\ b < x; \qquad (4)\ b < x < b',$$

where $b, b' \in B$. For $i < 5$, denote by n_i the number of constants from B in the formula (i); so $n_0 = 0$, $n_1 = n_2 = n_3 = 1$, and $n_4 = 2$.

For any $\tau = (\tau_1, \ldots, \tau_k) \in 5^k$, we are going to associate with S and B a relation S_τ on B of arity $n_\tau = n_{\tau_1} + \cdots + n_{\tau_k}$.

Let an n_τ-tuple of variables $\bar{y} = (y_1, \ldots, y_{n_\tau})$ is the concatenation of tuples $\bar{y}_1, \ldots, \bar{y}_k$, where the length of $\bar{y}_i$ is n_{τ_i}. For $i = 1, \ldots, k$, denote by $\phi_i(x_i, \bar{y}_i)$ the formula

- $x_i = x_i$ if $n_{\tau_i} = 0$,
- $x_i = v$ if $n_{\tau_i} = 1$ and $\bar{y}_i$ is v,
- $x_i < v$ if $n_{\tau_i} = 2$ and $\bar{y}_i$ is v,
- $x_i > v$ if $n_{\tau_i} = 3$ and $\bar{y}_i$ is v,
- $u < x_i < v$ if $n_{\tau_i} = 4$ and $\bar{y}_i$ is a pair (u, v).

This formula just says that x_i belongs to the 1-cell of type (i) defined by parameters $\bar{y}_i$. Denote by $\phi_\tau(\bar{x}, \bar{y})$ the conjunction of all the ϕ_i's; this formula says that $\bar{x}$ belongs to the k-cell $C_\tau(\bar{y}) = I_1 \times \cdots \times I_k$, where I_i is a k-cell of type τ_i defined by parameters $\bar{y}_i$.

For an n_τ-tuple $\bar{b}$ in U, we define $S_\tau(\bar{b})$ to be true if $\bar{b}$ is in B, and the k-cell $C_\tau(\bar{b})$ is B-minimal and is contained in S. Clearly, the B-minimality of the k-cell means exactly that, for $i \neq 1$, the interval $\phi_i(U, \bar{b}_i)$ has no common points with B.

It is easy to see that the relations S_τ can be uniformly obtained from S and B by means of certain FO queries σ_τ. As S is the union of all k-cells

$C_\tau(\bar{b})$ for which $S_\tau(\bar{b})$ holds ($\tau \in 5^k$, $\bar{b} \in B^{n_\tau}$), one can uniformly recover S from B and the family $\{S_\tau\}_{\tau \in 5^k}$ by means of a certain FO query π. Namely, π says that, for one of the τ's, there is an n_τ-tuple $\bar{b}$ in B such that both $S_\tau(\bar{b})$ and $\phi_\tau(\bar{x}, \bar{b})$ hold.

Later we will need the following observation concerning the definition of π.

Observation. *Suppose $A = \{a_1, \dots, a_m\} \subseteq U$ with $a_1 < \dots < a_m$, and R_τ are arbitrary finite relations of arity n_τ on A. Then it is easy to write down a quantifier-free formula $\rho(\bar{x}, z_1, \dots, z_m)$ in the pure order language depending only on the isomorphism type of the finite structure*

$$\mathfrak{A} = (A, R_\tau, <, a_1, \dots, a_m)_{\tau \in 5^k},$$

which says that, for one of the τ's, there is an n_τ-tuple $\bar{z}$ in the set

$$\{z_1, \dots, z_m\}$$

such that both $R_\tau(\bar{z})$ and $\phi_\tau(\bar{x}, \bar{z})$ hold. So we can assert not only that the relation $\pi(A, \{R_\tau\})$ is finitely representable over A but, moreover, that there is a certain 'standard' finite representation for it over A which depends only on the isomorphism type of $\mathfrak{A}$.

Based on the analysis above and taking into account Corollary 4.6, it is an easy exercise to prove

Lemma 4.8. *Consider the following two database schemes η and θ. The scheme η consists of one k-ary symbol P; the scheme θ consists of a unary symbol B and n_τ-ary symbols $S_{D\tau}$, for $\tau \in 5^k$ and E_k-classes D.*

(1) *There is a unary FO η-query δ which, for any finitely represented k-ary input P, yields a finite set ∂P as the answer*
(2) *For any $\tau \in 5^k$ and any E_k-class D, there is a n_τ-ary FO η-query $\sigma_{D\tau}$ which, for any finitely represented k-ary input P, yields the finite relation $\mathrm{Inv}(P \cap D)_\tau$ on ∂P as the answer*
(3) *There is a k-ary FO θ-query π which for any finite input*

$$B, \{S_{D\tau} : \tau \in 5^k, D \text{ is a } E_k\text{-class}\},$$

yields as an answer a relation finitely representable over B.
(4) *The family of queries δ, $\{\sigma_{D\tau}\}$ is an inverse for the query π in the following sense: for any finitely represented k-ary input P, we have $P = \pi(\delta(P), \{\sigma_{D\tau}(P)\})$*

We summarize the consideration above in the following main results.

Theorem 4.9. *For any finite database scheme $\eta = \{P_1, \dots, P_n\}$ there are:*

- *a database scheme $\theta = \{B, S_1, \dots, S_m\}$, where B is unary*
- *η-queries δ of arity 1 and σ_i of arity of S_i, for $1 \leqslant i \leqslant m$*
- *locally generic θ-queries π_j of arity of P_j, for $1 \leqslant j \leqslant n$*

such that

(a) *for any finitely representable η-input p, the family of η-queries $\bar{\sigma} = \{\delta, \sigma_i\}$ yields a finite θ-state as the answer*

(b) *for any finite θ-state s, the family of θ-queries $\bar{\pi} = \{\pi_i\}$ yields as the answer an η-state finitely representable over B*

(c) *for any η-query, which is locally generic over finitely representable states, the result of replacement of $P_1, \dots, P_n$ in it with $\pi_1, \dots, \pi_n$ is locally generic over finite θ-states*

(d) *$\bar{\pi}$ is an inverse of $\bar{\sigma}$ in the following sense: $\bar{\pi}(\bar{\sigma}(p)) = p$, for any finitely representable η-state p*

Note that (c) here immediately follows from the observation above concerning the definition of π.

Theorem 4.10. *For any expanded ordered universe U, if*

(1) *for any finite database scheme ρ, any locally generic over finite states extended ρ-query is equivalent over finite states over U to a restricted ρ-query*

then

(2) *for any finite database scheme η, any locally generic over finite representable states extended η-query is equivalent over finite representable states over U to a restricted η-query*

Proof. Suppose (1) is true. Fix a finite database scheme $\eta = \{P_1, \dots, P_n\}$. Let Φ be a locally generic over finitely representable states extended η-query.

We use Theorem 4.9. First we replace in Φ the relation names $P_1, \dots, P_n$ with the formulas $\pi_1, \dots \pi_n$; we obtain a locally generic over finite states extended θ-query Ψ. By (1), the query Ψ is equivalent over finite states to a restricted θ-query Ψ_1. Now we replace in Ψ_1 the relation names $B, S_1, \dots, S_m$ with the formulas $\delta, \sigma_1, \dots, \sigma_m$. Then we obtain a restricted η-query Φ_1 equivalent over finite representable states to Φ. So we have proven (2). ∎

5. Collapse of extended locally generic queries

In this section, we pursue collapse results over finite states. However, all the results can be transferred to finitely representable states by directly applying Theorem 4.10.

Our ultimate goal is to prove that, under certain conditions on the universe U, any locally generic extended query is equivalent over finite states over U to a restricted query. Hence, it suffices to prove such a result for Boolean queries (for database schemes with constant symbols).

For an arbitrary signature L, an L-theory is defined to be a set of first-order L-sentences (that is, formulas of signature L without free variables). For a class K of structures of an arbitrary signature L (in symbols, L-structures), the first-order L-theory of K (in symbols, $\mathrm{Th}(K)$) is defined to be the set of all first-order L-sentences which hold in every structure in K. Two L-structures M and N are called *elementarily equivalent* (in symbols, $M \equiv N$), if ϕ holds in M iff ϕ holds in N, for any L-sentence ϕ. An L-theory T is said to be complete if all its models are elementarily equivalent.

Let ρ be a database scheme $\{R_1, \ldots, R_n, c_1 \ldots, c_k\}$. We denote $L \cup \rho$ by $L(\rho)$. A ρ-state s over an L-structure W is said to be *pseudo-finite in* W if (W, s) is a model of the first-order $L(\rho)$-theory $F(W, \rho)$ of all (W, r), where r is a finite ρ-state over W.

For a first-order $L(\rho)$-sentence ψ and $m < \omega$, there is a first-order L-sentence ψ_m such that, for any L-structure V, the sentence ψ_m holds in V iff ψ holds for all ρ-states over V, whose active domain has cardinality at most m. Thus, $\psi \in F(W, \rho)$ iff $\{\psi_m : m < \omega\} \subseteq \mathrm{Th}(W)$. It follows that $W \equiv V$ implies $F(V, \rho) = F(W, \rho)$. For a complete L-theory T, the first-order $L(\rho)$-theory $F(T, \rho)$ is well-defined to be $F(W, \rho)$, where W is an arbitrary model of T.

Let ρ' be a disjoint copy $\{R'_1, \ldots, R'_n, c'_1 \ldots, c'_k\}$ of ρ. For an $L(\rho)$-sentence θ denote by $\theta(\rho')$ its $L(\rho')$-copy, that is, the result of replacement of every occurrence of R_i and c_j in θ with R'_i and c'_j, respectively. Let $\bar{\rho} = \rho \cup \rho'$.

As we will use the standard technique of so-called special models, we summarize its basic definitions and facts (see [CK90] for detail).

For a structure M of an arbitrary signature L and a subset A of M, denote by $L(A)$ the signature obtained by adjoining to L names for the elements of A. We do not normally distinguish between elements of A and their names.

We say that M is an *elementary substructure* of N (in symbols, $M \preceq N$ or $N \succeq M$), if M is a substructure of N, and ϕ holds in M iff ϕ holds in N, for any $L(M)$-sentence ϕ.

A set p of first-order $L(A)$-formulas with one free variable x is said to be a *type* over A in M if every finite subset $\{\phi_1(x), \ldots, \phi_k(x)\}$ of p is realized in M (that is, $(\exists x)(\phi_1(x) \& \ldots \& \phi_k(x))$ holds in M), and, for every $L(A)$-formula $\phi(x)$, either $\phi \in p$ or $\neg\phi \in p$. We say that a subset q of p *isolates* p if p is the only type over A in M containing q.

Let A be a subset of M. For any $N \succeq M$ and $a \in N$, the set of all $L(A)$-formulas $\phi(x)$ such that $\phi(a)$ holds in N forms a type over A in M; denote it by $\mathrm{tp}(a/A)$. For any type p over A in M, there are $N \succeq M$ and $a \in N$ such that $p = \mathrm{tp}(a/A)$. We denote $\mathrm{tp}(A)$ the set of all $L(A)$-sentences which hold in M.

For a cardinal λ, a structure M is said to be λ-*saturated* if any type p over any its subset A of power $< \lambda$ is realized in M; that is, $p = \mathrm{tp}(a/A)$, for some $a \in M$. For any infinite $\lambda \geqslant |L|$, every two elementarily equivalent λ-saturated structures of power λ are isomorphic. $|L|$ denotes the cardinality of L.

A structure M of power λ is called *special* if M is the union of a family $\{M_\mu : \mu \text{ is a cardinal } < \lambda\}$, where $M_\mu \preceq M_\nu \preceq M$ for $\mu < \nu < \lambda$, and each M_μ is μ^+-saturated. Here μ^+, as usual, denotes the least cardinal greater than μ. Every two elementarily equivalent special structures of the same power are isomorphic. For any infinite L-structure M and any cardinal $\lambda > |L|, |M|$ with $\lambda^* = \lambda$, there exists a special $N \succeq M$ of power λ. Here $\aleph^*_\alpha$ is defined to be $\sum_{\beta<\alpha} 2^{\aleph_\beta}$, and $|M|$ is the cardinality of M. It is easy to construct cardinals λ with $\lambda^* = \lambda$ of arbitrarily large cofinality.

Theorem 5.1. *For any universe U and any Boolean extended ρ-query ϕ the following conditions are equivalent:*

(1) *there is a restricted ρ-query ψ which is equivalent to ϕ over finite database states over U*
(2) *ϕ is generic for pseudo-finite states over V, for all $V \equiv U$*
(3) *for some uncountable power κ with $\kappa = \kappa^*$, the query ϕ is generic over pseudo-finite states over a special model $V \equiv U$ with $|V| = \kappa$*

The following proof works not only for $L_0 = \{<\}$ but for an arbitrary L_0. So the result can be considered as a finite model theory analog of Theorem 3.5.

Proof. (1)$\Rightarrow$(2). Suppose ϕ is equivalent to a restricted query ψ, for finite database states over U. Then $\phi \leftrightarrow \psi$ is in $F(U, \rho)$ and so in $F(V, \rho)$, for every $V \equiv U$. As ψ, being restricted, is generic even for all states over V, the genericity of ϕ for pseudo-finite states in V follows.

(2)$\Rightarrow$(3) is trivial.

(3)$\Rightarrow$(1). Let $T = Th(U)$. First we show that (3) implies the inconsistency of the following set of formulas

$$\Gamma = F(T, \bar{\rho}) \cup \{\theta(\rho) \leftrightarrow \theta(\rho') : \theta \text{ is an } L_0(\rho)\text{-sentence}\} \cup \{\phi(\rho), \neg\phi(\rho')\}.$$

Towards a contradiction, suppose (W, r, r') is a model of Γ. We can assume that it is a special model and $|W| = \kappa$. Then its reducts W, $(W|L_0, r)$ and $(W|L_0, r')$ are special, too. As W and V are elementarily equivalent special models of the same cardinality, W and V are isomorphic. Due to Γ, the models $(W|L_0, r)$ and $(W|L_0, r')$ are elementarily equivalent and hence isomorphic. In other words, there is an L_0-automorphism of W which transforms r to r'. As (r, r') is pseudo-finite in W, and $\phi(\rho)$ holds in (W, r) but $\phi(\rho')$ fails in (W, ρ'), the query ϕ is not generic for pseudo-finite states in W. Since V is isomorphic to W, we have a contradiction with (3).

By compactness, there is a finite inconsistent part of Γ. Then, for some $m < \omega$ and $L_0(\rho)$-sentences $\theta_0, \ldots, \theta_{m-1}$,

$$F(T, \bar{\rho}),\ \wedge_{i<m}(\theta_i(\rho) \leftrightarrow \theta_i(\rho')),\ \phi(\rho)\ \vdash\ \phi(\rho').$$

Denote $\theta^1 = \theta$, $\theta^0 = \neg\theta$. It easily follows that, for some $\tau_0, \ldots, \tau_{l-1} \in 2^m$, the extended query ϕ is equivalent, for finite states over U, to the restricted query $\psi = \vee_{i<l} \wedge_{j<m} \theta_i^{\tau_{ij}}$. ∎

Let I be a subset of a universe V. We say that a Boolean extended ρ-query ϕ is *locally generic over pseudo-finite states over I in V* if the following holds: if an $\bar{\rho}$-state (r, r') over I is pseudo-finite in V and r can be transformed into r' by a partial L_0-isomorphism in V then $\phi(\rho)$ holds in (V, r) iff $\phi(\rho')$ holds in (V, r').

A linearly ordered subset I of a structure M is said to be an *indiscernible sequence* in M if $\theta(\bar{a})$ holds in M iff $\theta(\bar{b})$ holds in M, for every first-order L-formula $\theta(x_1, \ldots, x_n)$ and any two n-tuples $\bar{a}$ and $\bar{b}$ in I with $a_1 < \cdots < a_n$ and $b_1 < \cdots < b_n$. A subset I of a structure M is said to be an *indiscernible* *set* in M if $\theta(\bar{a})$ holds in M iff $\theta(\bar{b})$ holds in M, for any first-order L-formula $\theta(x_1, \ldots, x_n)$ and any two n-tuples $\bar{a}$ and $\bar{b}$ of *distinct* elements in I.

We will need the following well-known fact (see [CK90]) which is an easy corollary of the classical Ramsey theorem.

Ramsey's Theorem. *Let M be an infinite L-structure, and $<$ a linear ordering of the universe of M. Then, for every L-formula $\theta(x_1, \ldots, x_n)$, there is an infinite subset I of M which is a θ-indiscernible sequence with respect to $<$.*

Theorem 5.2. *Let an extended Boolean ρ-query ϕ be locally generic for finite states over U. Suppose, for some uncountable κ with $\kappa = \kappa^*$, there is a special model $V \equiv U$ of power κ such that, for any infinite indiscernible sequence I in V, the query ϕ is locally generic over pseudo-finite states over I in V. Then ϕ is equivalent over finite states over U to a restricted ρ-query.*

Proof. Suppose the contrary. Let κ and V satisfy the conditions of the theorem. By Theorem 5.1, there exist a model $W \equiv U$ and a pseudo-finite $\bar{\rho}$-state (p, p') over W such that some $<$-automorphism h of W transforms p to p', but $(W, p) \models \phi(\rho)$ and $(W, p') \models \neg\phi(\rho')$. Clearly, we can assume that (W, p, p', h) is a special model of power κ; so we can assume $W = V$.

Let σ and σ' be copies of ρ such that $\rho, \rho', \sigma, \sigma'$ are pairwise disjoint. Let

$$\eta = \rho \cup \rho' \cup \sigma \cup \sigma' \cup \{F, F', I\},$$

where F, F' are new binary relation symbols and I is a new unary relation symbol.

Let $T = \mathrm{Th}(U)$. We will prove the consistency of the set of first-order $L(\eta)$-sentences Γ which say about an η-state (r, r', s, s', F, F', I) over a model of T the following:

(1) (r, r') satisfies $\mathrm{Th}(V, p, p')$
(2) (r, r', s, s', F, F', I) satisfies $F(T, \eta)$
(3) F and F' are partial $<$-isomorphisms transforming r to s and r' to s', respectively
(4) s and s' are states over I
(5) there are at least n distinct elements in I, for all $n < \omega$
(6) if $\bar{x}$ and $\bar{y}$ are k-tuples in I and $x_1 < \cdots < x_k$ and $y_1 < \cdots < y_k$, then $\theta(\bar{x}) \equiv \theta(\bar{y})$, for every $k < \omega$ and every k-ary L-formula θ
(7) s satisfies $\phi(\sigma)$, and s' satisfies $\neg\phi(\sigma')$

Let $(M, r, r', s, s', F, F', I)$ be a special model of power κ for Γ. Due to (1), we have (M, r, r') is isomorphic to (V, p, p'); so we can assume that $(M, r, r') = (V, p, p')$. Due to (2), the η-state (r, r', s, s', F, F', I) is pseudo-finite in V. Due to (3), the partial $<$-isomorphism $g = F' \circ h \circ F^{-1}$ transforms s to s', which, due to (4), are states over I. Due to (5–6), I is an infinite L-indiscernible set in V. Due to (7), s satisfies $\phi(\sigma)$, but s' does not satisfy $\phi(\sigma')$. It follows that ϕ is not locally generic over pseudo-finite states over I in V, a contradiction.

By compactness, it remains to prove the finite consistency of Γ. It suffices to show that, for every $\gamma \in \mathrm{Th}(V, p, p')$, every natural number $n < \omega$ and every L-formula $\theta(\bar{x})$, there is a finite η-state (r, r', s, s', F, F', I) over V, which satisfies γ, (3), (4), and (7), and such that I is a θ-indiscernible sequence of cardinality at least n.

By the corollary of Ramsey's Theorem mentioned above, we can find an infinite θ-indiscernible sequence J in V. As (p, p') is pseudo-finite, there is a finite $\bar{\rho}$-state (r, r') over V that satisfies $\gamma \wedge \phi(\rho) \wedge \neg\phi(\rho')$. Let I be an arbitrary finite subset of J of cardinality at least n. It is easy to find finite s, s', F, F' which satisfy (3) and (4). Since ϕ is locally generic, we have (7). The finite consistency of Γ is proved, and the proof is completed. ∎

As a side remark, note that this technique implies a result from [BDLW96] and [OV96] about the so-called *active semantics.* But the proof in [OV96] is shorter than our proof.

An FO formula is said to be *active* if it is obtained from some formula by relativization of every of its quantifiers with respect to the formula $AD(x)$ isolating the active domain.

Theorem 5.3. *For finite states, any locally generic active extended query is equivalent to a restricted query.*

Proof. By Theorem 5.2, it suffices to show that any active Boolean query is locally generic over *arbitrary* states over any indiscernible sequence in any V. We may assume that the signature L is relational. For a ρ-state r over V, denote by V_r the substructure of (V, r) with domain $AD(r)$. Clearly, an active ρ-query holds in V_r iff it holds in (V, r). Since, for any ρ-states r and r' over an indiscernible sequence I in V, any partial L_0-isomorphism transforming r into r' induces an $L(\rho)$-isomorphism between V_r and $V_{r'}$, the result follows. ∎

Now our aim is to give a general condition on a complete theory which ensures collapse of locally order-generic queries to pure order ones, over finite states over models of T.

Let M and N be L-structures, A be a subset of M, and B be a subset of N. A bijection $h : A \to B$ is said to be a *partial L-isomorphism* (an *elementary map*) from M to N if $\phi(\bar{a})$ holds in M iff $\phi(h(\bar{a}))$ holds in N, for any atomic L-formula (for any L-formula) $\phi(\bar{x})$ and any tuple $\bar{a}$ in A. A partial L-isomorphism (an elementary map) from M to M is called a partial L-isomorphism (an elementary map) in M. Clearly, every elementary map in M is a partial isomorphism in M; if M admits quantifier elimination, the converse is also true.

We recall the following well-known result (see [Sac72], [RW81]):

Fact. *Let T be a complete L-theory. Then the following are equivalent:*
(1) *T admits quantifier elimination;*

(2) *there is an infinite cardinal λ such that, for every finitely generated L-structure A, for every models M, N of T with $A \subseteq M, N$, for every $a \in N \backslash A$, if M is λ-saturated, the quantifier-free type of a over A is realized in M;*

(3) *there is an infinite cardinal λ such that every λ-saturated model M of T is finitely homogeneous in the following sense: for any partial isomorphism $h : A \to B$ in M with finite A and B, and any $a \in M$ there is $b \in M$ such that $h \cup \{(a, b)\}$ is a partial isomorphism in M.*

Let T be an L-theory, and T' be an L'-theory with $L \subseteq L'$ and $T \subseteq T'$. The theory T' is said to be a *definitional extension* of T if every L'-formula is equivalent in T' to an L-formula. Every L-theory T has a *standard* definitional extension admitting quantifier elimination: for every L-formula $\phi(\bar{x})$ with n free variables add a new n-ary relation symbol P_ϕ to L and a new axiom $\forall \bar{x}\, (P_\phi(\bar{x}) \leftrightarrow \phi(\bar{x}))$ to T.

We say that a complete L-theory T has the *Pseudo-finite Homogeneity Property* if there exist its definitional extension T' of a signature L' and an infinite cardinal λ such that, whenever A and B are pseudo-finite sets in a model M' of T', and $h : A \to B$ is a partial isomorphism in M' with λ-saturated (M', A, B, h), for any $a \in M'$ there is $b \in M'$ such that $h \cup \{(a, b)\}$ is a partial isomorphism in M'.

Note that, by the Fact above, the theory T' here automatically admits quantifier elimination, because any finite set in M' is pseudo-finite, and if M' is λ-saturated then (M', A, B, h) is λ-saturated, too, for finite A and B.

Therefore the definition of the Pseudo-finite Homogeneity Property admits the following equivalent formulation: a complete L-theory T has the Pseudo-finite Homogeneity Property iff there exists an infinite cardinal λ such that, whenever A and B are pseudo-finite sets in a model M of T, and $h : A \to B$ is an elementary map in M with λ-saturated (M, A, B, h), for any $a \in M$ there is $b \in M$ such that $h \cup \{(a, b)\}$ is an elementary map in M.

The Pseudo-finite Homogeneity Property makes sense not only for theories of ordered universes, and there are obvious examples of theories with this property. For example, in infinite structures of empty signature the pseudo-finite sets are exactly sets whose complements are infinite; hence the Pseudo-finite Homogeneity Property for the theory of these structures easily follows. It turned out that for ordered structures the property is especially interesting because it gives a sufficient condition for collapse results. Later we will give a series of examples of ordered universes with this property.

Theorem 5.4. *Suppose the first-order theory of a universe U has the Pseudo-finite Homogeneity Property. Let an extended query ϕ be locally generic over finite states over U. Then ϕ is equivalent over finite states over U to a restricted query.*

Proof. It suffices to prove that ϕ satisfies the condition of Theorem 5.2. Suppose λ witnesses that T has the Pseudo-finite Homogeneity Property. Let $\kappa = \kappa^* > |L| + \aleph_0$, and $\mathrm{cf}(\kappa) > \lambda$. Let $V \equiv U$ be a special model of power κ. Let I be an infinite L-indiscernible sequence in V. Suppose

$\bar{\rho}$-state (r, r') over I is pseudo-finite in V and r can be transformed to r' by a partial L_0-isomorphism g in V, whose domain is A, the active domain of r, and whose range is A', the active domain of r'. We need to show that $\phi(\rho)$ holds in (V, r) iff $\phi(\rho')$ holds in (V, r'). We may assume that (V, A, A', g) is λ-saturated. (Indeed, consider a special model $(V_0, r_0, r'_0, g_0, I_0)$ of power κ elementarily equivalent to (V, r, r', g, I). It suffices to prove the claim for $(V_0, r_0, r'_0, g_0, I_0)$; but it is λ-saturated as $\mathrm{cf}(\kappa) > \lambda$.) Using a Fraïssé-Ehrenfeucht game, we will show that g is an $L(\rho)$-elementary map from (V, r) to (V, r'). Due to the L-indiscernibility of I, the map g is an L-elementary map, and, in particular, a partial $L(\rho)$-isomorphism. Therefore, due to the Pseudo-finite Homogeneity Property, to complete the proof of the theorem, using Theorem 5.2, it suffices to prove the following lemmas:

Lemma 5.5. *The active domain of any pseudo-finite database state is a pseudo-finite set.*

Lemma 5.6. *Let A be a pseudo-finite set in V, and $a \in V$. Then $A \cup \{a\}$ is a pseudo-finite set.*

Lemma 5.7. *If $h : C \to D$ is a partial isomorphism in V, and $M = (V, C, D, h)$ is λ-saturated, then*

$$M' = (V, C \cup \{c\}, D \cup \{d\}, h \cup \{(c, d)\})$$

is λ-saturated, for any $c, d \in V$.

Proof of Lemma 5.5. Consider the database scheme $\tau = \{P\}$, where P is a unary relation name. For any $L(\tau)$-sentence γ, there is an $L(\rho)$-sentence γ^* such that $(V, s) \models \gamma^*$ iff $(V, AD(s)) \models \gamma$, for any ρ-state s. Suppose a state s is pseudo-finite in V, and $\gamma \in F(V, \tau)$. Since the active domain of any finite state is finite, we have $(V, r) \models \gamma^*$ for all finite ρ-states r. So $(V, s) \models \gamma^*$, and hence $(V, AD(s)) \models \gamma$. ∎

Proof of Lemma 5.6. Consider the database scheme $\tau = \{P\}$, where P is a unary relation name. Let $\theta \in F(V, \tau)$. Let $\theta^*(x)$ be the result of replacement of every occurrence of $P(y)$ in θ with $P(y) \vee y = x$, where x is a new variable. Then $\forall x \theta^*(x)$ belongs to $F(V, \tau)$ and so holds in (V, A). Hence θ holds in $(V, A \cup \{a\})$. Thus, $A \cup \{a\}$ is pseudo-finite. ∎

Proof of Lemma 5.7. M' is definable in the λ-saturated structure M with parameters c, d. ∎

Let A_0 be A, B_0 be B, and g_0 be g. It suffices to prove that Duplicator has a winning strategy in the game. The desired winning strategy is to ensure that, after each round i of the game, A_i and B_i are pseudo-finite sets in V, $g_i : A_i \to B_i$ is an elementary map in V with λ-saturated (V, A_i, B_i, g_i), $g \subseteq g_i$, and g_i is a partial $L(\rho)$-isomorphism from (V, r) to (V, r').

Suppose Spoiler starts a new round $i + 1$ and chooses an element $a_{i+1} \in V$ (or $b_{i+1} \in V$). By the Pseudo-finite Homogeneity Property, there is $b_{i+1} \in V$ (correspondingly, $a_{i+1} \in V$) such that $g_{i+1} = g_i \cup \{(a_{i+1}, b_{i+1})\}$ is an elementary map in V. Then Duplicator chooses b_{i+1} (correspondingly, a_{i+1}). Let $A_{i+1} = A_i \cup \{a_{i+1}\}$ and $B_{i+1} = B_i \cup \{b_{i+1}\}$. By the definition of g, g_{i+1} is a partial $L(\rho)$-isomorphism from (V, r) to (V, r'). By Lemma 5.6, A_{i+1} and B_{i+1} are pseudo-finite sets. By Lemma 5.7, $(V, A_{i+1}, B_{i+1}, g_{i+1})$ is λ-saturated.

The collapse result is proved. ∎

Now we introduce a certain property of complete theories which is strictly stronger than the Pseudo-finite Homogeneity Property, and so ensures the collapse result, too.

We say that a complete theory T has the *Isolation Property*, if there is a cardinal λ such that, for any pseudo-finite set A and any element a in a model of T, there is $A_0 \subseteq A$ with $|A_0| < \lambda$ such that $\mathrm{tp}(a/A_0)$ isolates $\mathrm{tp}(a/A)$.

Theorem 5.8. *The Isolation Property implies the Pseudo-finite Homogeneity Property.*

Proof. We show that λ witnessing that T has the Isolation Property witnesses that T has the Pseudo-finite Homogeneity Property. Let V be a λ-saturated model of T, and A, B pseudo-finite sets in V, and $a \in V$. Let $h : A \to B$ be an L-elementary map in V. We show that there is $b \in B$ such that $h \cup \{(a, b)\}$ is an L-elementary map in V. Choose $A_0 \subseteq A$ with $|A_0| < \lambda$ such that $p_0 = \mathrm{tp}(a/A_0)$ isolates $p = \mathrm{tp}(a/A)$. Since the map h is elementary, $h(p)$ is a type over B, and $h(p_0)$ isolates $h(p)$. (For a set $q(x)$ of formulas over A we denote by $h(q)$ the set $\{\theta(x, h(\bar{c})) : \theta(x, \bar{c}) \in q\}$.) As V is λ-saturated, there is $b \in V$ realizing $h(p_0)$ and hence $h(p)$. So $h \cup \{(a, b)\}$ is an L-elementary map. ∎

We will show that any theory with the Isolation Property is unstable. To prove the unstability of a theory, it suffices to show that there exists an infinite indiscernible sequence in one of its model which is not an indiscernible set (see [She90], Theorem II.2.13).

Theorem 5.9. *Any theory with the Isolation Property is unstable.*

Proof. First we show that, for any complete L-theory T with infinite models, there is an infinite indiscernible sequence in a model of T whose members form a pseudo-finite set. Consider the infinite set Γ of first-order sentences of the signature $L \cup \{<, P\}$ (where $<$ is a new binary relation symbol, and P is a new unary relation symbol) which says:

- $<$ is a linear order
- P is infinite
- $\theta(\bar{a}) \leftrightarrow \theta(\bar{b})$, for any L-formula $\theta(v_1, \dots, v_n)$ and $a_1, \dots, a_n, b_1, \dots, b_n$ in P with $a_1 < \dots < a_n$ and $b_1 < \dots < b_n$

For any finite $\Delta \subset \Gamma$, by standard using of the Ramsey Theorem, one can find a model of $T \cup \Delta$ in which P is finite. Therefore, by compactness,

$F(T, \{P\}) \cup \Gamma$ has a model $(M, <, I)$. Then I is an indiscernible sequence in M with respect to $<$, and the set I is pseudo-finite in M.

Now suppose T has the Isolation Property, and λ witnesses that. By compactness, we may assume that $|I| \geq \lambda$. Stretching I by compactness, we can find an elementary extension $(M', <')$ of $(M, <)$ and an indiscernible sequence I' in M' such that I is a proper subsequence of I'. We claim that I' is not an indiscernible set in M'. Indeed, let $a \in I' \setminus I$. By the choice of λ, there is $I_0 \subset I$ with $|I_0| < \lambda$ such that $\mathrm{tp}(a/I_0)$ isolates $\mathrm{tp}(a/I)$. Let $b \in I \setminus I_0$. As the types $\mathrm{tp}(a/I)$ and $\mathrm{tp}(b/I)$ are different—the formula $x = b$ discriminates them—the types $\mathrm{tp}(a/I_0)$ and $\mathrm{tp}(b/I_0)$ are different, too. So I' is not an indiscernible set. ∎

On the other hand, there are obvious examples of stable theories with Pseudo-finite Homogeneity Property (for example, the theory of infinite structures of empty signature). Later we will give examples of ordered universes which have the Pseudo-finite Homogeneity Property but do not have the Isolation Property (they are, of course, unstable). Theorem 5.12 below gives a broad class of theories with the Isolation Property and provides a lot of examples of collapse results.

A complete L-theory T is said to be *o-minimal* if in every its model any definable set is a finite union of singletons and open intervals. The following characterization of o-minimal theories is not hard to prove.

Theorem 5.10. *A complete L-theory T is o-minimal iff there exists T', a definitional expansion of T in a language L', such that any L'-formula $\theta(x, \bar{y})$ is T'-equivalent to a disjunction of formulas of the form $\psi(\bar{y}) \wedge \rho(x, \bar{y})$, where $\rho(x, \bar{y})$ has one of the following forms, for some L'-terms t and t' in the variables $\bar{y}$:*

$$x = x, \quad x = t, \quad x < t, \quad t < x, \quad t < x < t'.$$

Hint of the proof. The "if" part is trivial. To prove the "only if" part, note that, by compactness, for every L-formula $\theta(x, \bar{y})$, there is a uniform finite upper bound for the number of boundary points of the sets of the form $\theta(M, \bar{a})$, where M is a model of T and $\bar{a}$ is a tuple in M. Denote by $f_{i\theta}(\bar{a})$ the ith boundary point of $\theta(M, \bar{a})$. Expanding the models of T by the definable functions $f_{i\theta}$, we get the class of models whose theory T' is the desired extension of T by definitions. ∎

We call a complete L-theory T *quasi-o-minimal* iff there exists T', a definitional expansion of T in a language L', such that any L'-formula $\theta(x, \bar{y})$ is T'-equivalent to a disjunction of formulas of the form $\phi(x) \wedge \psi(\bar{y}) \wedge \rho(x, \bar{y})$, where $\rho(x, \bar{y})$ has one of the following forms, for some L'-terms t and t' in the variables $\bar{y}$:

$$x = x, \quad x = t, \quad x < t, \quad t < x, \quad t < x < t'.$$

By Theorem 5.10, every o-minimal theory is quasi-o-minimal. Clearly, in every model of a quasi-o-minimal theory any definable set is a finite union of singletons and sets of the form $I \cap D$, where I is an open interval and D is a set definable without parameters.

There exist quasi-o-minimal theories which are not o-minimal. The simplest example is the theory T of dense ordered sets without endpoints with a distinguished subset which is dense and whose complement is also dense in the universe. It can be easily shown that T is the theory of the structure $(\mathbb{R}, <, \mathbb{Q})$. The theory is not o-minimal because the distinguished subset is not a finite union of singletons and open intervals. Standard arguments show that T admits quantifier elimination; obviously, we can take T as T' from the definition of quasi-o-minimality.

Another example of a quasi-o-minimal theory is the theory T of $(\mathbb{Z}, <, +)$. Indeed, by Presburger's Theorem, the definitional expansion of the model by the constants 0, 1 and the unary predicates 'n divides x', for all positive integers n, admits quantifier elimination. For a positive integer n, define the function $f_n(x)$ by the condition $0 \leqslant x - nf_n(x) < n$. Since '$n$ divides x' iff $nf_n(x) = x$, and $nx = t$ iff $x = f_n(t)$, and $nx < t$ iff $x < f_n(t)$, and $nx > t$ iff $x > f_n(t)$, the theory T' of the definitional expansion of $(\mathbb{Z}, <, +)$ by the constants 0, 1 and all the functions $f_n(x)$ satisfies the condition of the definition of quasi-o-minimality. However, the theory of $(\mathbb{Z}, <, +)$ is not o-minimal because the definable subsets $n\mathbb{Z}$ are not finite unions of singletons and open intervals.

Similarly, the theory of $(\mathbb{N}, <, +)$ is quasi-o-minimal but not o-minimal.

New examples of quasi-o-minimal structures can be constructed using the *ordered union* operation. Let U_i be an L_i-structure, where L_i contains $<$, and U_i is linearly ordered by $<$, for $i = 1, 2$. We assume that the universes of U_1 and U_2 are disjoint. Let $L = L_1 \cup L_2 \cup \{P_1, P_2\}$, where P_1 and P_2 are new unary relation names. The ordered union of U_1 and U_2 is defined to be the L-structure U, whose universe is the union of the universes of U_1 and U_2, and

- P_i^U is defined to be the universe of U_i,
- $<^U$ is defined to be $<^{U_1} \cup <^{U_2} \cup (U_1 \times U_2)$,
- $Q^U = Q^{U_i}$ for any $Q \in L_i \setminus \{<\}$.

A routine induction on the complexity of a formula shows that any L-formula is equivalent in U to a positive Boolean combination of L-formulas in each of which all variables, free or bounded, are restricted to P_1 or are restricted to P_2. It follows, in particular, that the models of the theory of U are exactly the ordered unions of U_1' and U_2', where $U_1 \equiv U_1'$ and $U_2 \equiv U_2'$. So the ordered union of two complete theories of ordered structures can be well-defined. Another obvious consequence is

Theorem 5.11. *The ordered union of any two quasi-o-minimal theories is a quasi-o-minimal theory.*

Note that the ordered union of two o-minimal theories is an o-minimal theory iff any model of the first of them has the greatest element or any model of the second of them has the least element. So the construction gives a lot of quasi-o-minimal theories which are not o-minimal.

Theorem 5.12. *Any quasi-o-minimal theory has the Isolation Property.*

Proof. Let A be a pseudo-finite set in a model V of a quasi-o-minimal theory T, and $a \in V$. We will show that there is $A_0 \subseteq A$ with $|A_0| \leqslant |T|$ such that $\mathrm{tp}(a/A_0)$ isolates $\mathrm{tp}(a/A)$.

Let T' be the definitional expansion of T from the definition of quasi-o-minimality. Let $\eta(x,\bar{y})$ be an L-formula. It is T'-equivalent to a disjunction $\theta(x,\bar{y})$ of L'-formulas of the form $\phi(x) \wedge \psi(\bar{y}) \wedge \rho(x,\bar{y})$, where $\rho(x,\bar{y})$ has one of the following forms, for some L'-terms t and t' in the variables $\bar{y}$:

$$x = x, \quad x = t, \quad x < t, \quad t < x, \quad t < x < t'.$$

Denote by S_θ the finite set of all L'-terms $t,\ t'$ involved in ρ in some of the disjuncts of θ. Let F be a finite subset of V. Then, for every $d \in V$, in the finite subset of V

$$\{t(\bar{c}) : t \in S_\theta, \quad \bar{c} \in F, \quad t(\bar{c}) \leqslant d\}$$

there is a maximal element $m_\theta(F,d)$, provided the subset is not empty. Similarly, for every $d \in V$, in the finite subset of V

$$\{t(\bar{c}) : t \in S_\theta, \quad \bar{c} \in F, \quad t(\bar{c}) \geqslant d\}$$

there is a minimal element $m^\theta(F,d)$, provided the subset is not empty.

As the set A is pseudo-finite in V, the same holds for A instead of F. Denote $m_\theta = m_\theta(A,a)$, and $m^\theta = m^\theta(A,a)$.

Let m_θ be $t_\theta(\bar{c}_\theta)$, where $t_\theta \in S_\theta$, $\bar{c}_\theta \in A$, when m_θ is defined; otherwise we write $m_\theta = -\infty$. Let m^θ be $t^\theta(\bar{c}^\theta)$, where $t^\theta \in S_\theta$, $\bar{c}^\theta \in A$, when m^θ is defined; otherwise we write $m^\theta = \infty$. Clearly, $m_\theta = m^\theta$ iff $a = t(\bar{c})$ for some $t(\bar{y}) \in S_\theta$ and $\bar{c} \in A$. If $m_\theta \neq m^\theta$ then $m_\theta < a < m^\theta$.

Suppose $\eta(x,\bar{c})$ is in $\mathrm{tp}(a/A)$. Since $\eta(a,\bar{c})$ holds in V, one of the disjuncts $\phi(a) \wedge \psi(\bar{c}) \wedge \rho(a,\bar{c})$ of $\theta(a,\bar{c})$ holds; we have $\phi(x) \in \mathrm{tp}(a/\emptyset)$ and $\psi(\bar{y}) \in \mathrm{tp}(A)$.

If $m_\theta = m^\theta$ then $a = t(\bar{c})$, for some $t(\bar{y}) \in S_\theta$ and $\bar{c} \in A$. In this case $\mathrm{tp}(a/\bar{c})$ isolates $\mathrm{tp}(a/A)$, and we can take $\bar{c}$ as A_0.

Now suppose $m_\theta \neq m^\theta$. We will show that

$$\mathrm{tp}(A) \cup \mathrm{tp}(a/\emptyset) \vdash \ m_\theta < x < m^\theta \rightarrow \eta(x,\bar{c}).$$

It suffices to prove that

$$\mathrm{tp}(A) \vdash \ m_\theta < x < m^\theta \rightarrow \rho(x,\bar{c}).$$

The case when $\rho(x,\bar{y})$ is $x = x$ is trivial. The formula $\rho(x,\bar{y})$ cannot be of the form $x = t(\bar{y})$: otherwise $a = t(\bar{c})$, and we would have $m_\theta = m^\theta$.

Suppose $\rho(x,\bar{y})$ is of the form $t(\bar{y}) < x < t'(\bar{y})$; so we have $t(\bar{c}) < a < t'(\bar{c})$. Therefore the formulas $t(\bar{c}) \leqslant t_\theta(\bar{c}_\theta))$ and $t^\theta(\bar{c}^\theta) \leqslant t'(\bar{c})$ belong to $\mathrm{tp}(A)$. Then $\mathrm{tp}(A)$ together with $m_\theta < x < m^\theta$ implies $t(\bar{c}) < x < t'(\bar{c})$. The cases when $\rho(x,\bar{y})$ is of the form $x < t(\bar{y})$ or $t(\bar{y}) < x$ can be considered similarly.

So the set A_0 of all c_θ, c^θ satisfies the desired conditions. ∎

Now we give a series of examples of ordered structures which show that the Isolation Property is strictly weaker than quasi-o-minimality and strictly stronger than the Pseudo-finite Homogeneity Property.

Let $L = \{<, E\}$ and T_{dt} be the theory of all the structures of the form $(A, <, E)$, where $<$ is a dense ordering without endpoints, and E is an equivalence relation with two class both of which are dense in A. The structure $(\mathbb{R}, <, E)$, where E is the equivalence relation whose classes are $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$, is a model of T_{dt}; so T_{dt} is consistent. Standard back-and-forth arguments show that T_{dt} is countably categorical and hence complete. Also, standard arguments show that T_{dt} admits quantifier elimination.

Theorem 5.13. *T_{dt} is not quasi-o-minimal, but has the Isolation Property.*

Proof. Towards a contradiction, suppose T_{dt} is quasi-o-minimal. It easily follows from the quantifier elimination that every L-formula $\phi(x)$ is T_{dt}-equivalent to $x = x$ or $x \neq x$. Therefore every definable set in any model of T_{dt} is a disjoint union of finitely many open intervals and singletons. However, any E-class which is dense and whose complement is also dense, is not of this form, even though it is definable. A contradiction.

Now we show that T_{dt} has the Isolation Property. We prove that, for every pseudo-finite set A and any element a in a model of T_{dt}, there is $A_0 \subseteq A$ with $|A_0| \leqslant 2$ such that $\mathrm{tp}(a/A_0)$ isolates $\mathrm{tp}(a/A)$. If $a \in A$ we can take $\{a\}$ as A_0; so we assume $a \notin A$.

Since A is pseudo-finite, it has the least element b and the greatest element c, provided $A \neq \emptyset$. Suppose $a < b$. If $E(a, b)$ then, due to the quantifier elimination, the set of formulas

$$\{x < b,\ E(x, b)\}$$

isolates $\mathrm{tp}(a/A)$. Similarly, if $\neg E(a, b)$ then

$$\{x < b,\ \neg E(x, b)\}$$

isolates $\mathrm{tp}(a/A)$. So in both the cases we can take $\{b\}$ as A_0. Analogously, in the case $a > c$ we can take $\{c\}$ as A_0.

If neither $a < b$ nor $a > c$ then, due to the pseudo-finiteness of A, there are a_* and a^* in A such that $a_* < a < a^*$, and a^* is the successor of a_* in A. If $E(a, a_*)$ then, due to the quantifier elimination, the set of formulas

$$\{a_* < x < a^*,\ E(x, a_*)\}$$

isolates $\mathrm{tp}(a/A)$. Similarly, if $\neg E(a, a_*)$ then

$$\{a_* < x < a^*,\ \neg E(x, a_*)\}$$

isolates $\mathrm{tp}(a/A)$. So in both the cases we can take $\{a_*, a^*\}$ as A_0.

The Isolation Property for T_{dt} is proved. ∎

Consider the structures of the form $(A, <, E)$, where $(A, <)$ is a dense linearly ordered set without endpoints, and E is an equivalence relation on A with infinitely many classes all of which are dense. An example of such a structure is $(\mathbb{R}, <, E)$, where $E(x, y)$ means $x - y \in \mathbb{Q}$. Standard back-and-forth arguments show that the theory T_{de} of such structures is countably categorical and hence complete. Also, standard arguments show that T_{de} admits quantifier elimination.

Theorem 5.14. *T_{de} does not have the Isolation Property, but has the Pseudo-finite Homogeneity Property in the following strong sense: for every model M of T_{de}, whenever A and B are pseudo-finite sets in M, and $h : A \to B$ is a partial isomorphism in M, for any $a \in M$ there is $b \in M$ such that $h \cup \{(a, b)\}$ is a partial isomorphism in M.*

Proof. The following observation is crucial for the proof: for every pseudo-finite subset C of a model of T_{de}, every open interval in the model contains an element which is equivalent to no element of C. Clearly, it suffices to show that for an arbitrary finite C. The latter holds because there are infinitely many E-classes in V, and every E-class is dense in V.

To prove the first part of the theorem, we show that, for every infinite cardinal λ, there are a pseudo-finite set A and an element a in a model of T_{de} such that, for no $A_0 \subseteq A$ with $|A_0| < \lambda$, the type $\mathrm{tp}(a/A_0)$ isolates $\mathrm{tp}(a/A)$.

Since in every model of T_{de} there are finite sets of arbitrarily large size whose elements are pairwise non-equivalent, we can find, by compactness, a pseudo-finite set A in a model U of T_{de} with $|A| \geqslant \lambda$ whose elements are pairwise non-equivalent. Let c be the greatest element of A. Choose $a \in U$ such that $a > c$, and a is equivalent to no element of A. We claim that $\mathrm{tp}(a/A_0)$ does not isolate $\mathrm{tp}(a/A)$, for any proper subset A_0 of A. Indeed, due to the quantifier elimination, the set of formulas

$$q = \{d < x,\ \neg E(x, d) :\ d \in A_0\}$$

isolates $\mathrm{tp}(a/A_0)$. But q cannot isolate $\mathrm{tp}(a/A)$, because, for $e \in A \setminus A_0$, the set $q \cup \{E(x, e)\}$ is consistent (it is realized by any element which is E-equivalent to e and is greater than c), even though $\neg E(x, e)$ is in $\mathrm{tp}(a/A)$.

Now we prove the second part of the theorem.

In the case of empty A and B we can take as b an arbitrary element of V. So we assume that A is nonempty and $a \notin A$.

First suppose that $a > d$ for each $d \in A$. Since A is pseudo-finite, it has the greatest element. Denote it by c; then $h(c)$ is the greatest element of B. If a is E-equivalent to some element $a' \in A$, one can take as b an arbitrary element of V which is $> h(c)$ and E-equivalent to $h(a')$. If a is E-equivalent to no element in A, one can take as b an arbitrary element of V which is $> h(c)$ and E-equivalent to no element in B.

The case when $a < d$ for each $d \in A$ can be considered similarly.

Now suppose there are $a', a'' \in A$ such that $a' < a < a''$. Then, due to the pseudo-finiteness of A, there are a_* and a^* in A such that $a_* < a < a^*$, and a^* is the successor of a_* in A. If a is E-equivalent to some element $a' \in A$, one can take as b an arbitrary element of V which lies between $h(a_*)$ and $h(a^*)$, and is E-equivalent to $h(a')$. If a is E-equivalent to no element of A, one can take as b an arbitrary element of V which lies between $h(a_*)$ and $h(a^*)$, and E-equivalent to no element of B. ∎

Let F be an ordered division ring, T_0 be the first-order theory of ordered vector spaces over F with a distinguished subspace, and T_{ds} be the first-order theory of ordered nonzero vector spaces over F with a distinguished proper dense subspace. Here an ordered vector space over F is defined to be a vector space V over F whose additive group is linearly ordered so that αv is positive, for any positive $\alpha \in F$ and any positive $v \in V$. We consider T_0 and T_{ds} in the signature $\{+, <, f_\alpha, P\}_{\alpha \in F}$, where f_α is a name for the unary operation of multiplication by the scalar α, and P is a name for the distinguished subspace. The theory T_0 is obviously consistent. We will show the consistency of T_{ds} in the proof of Theorem 5.15 below.

A first-order theory T is said to be *model complete* iff for all models A and B of T, if $A \subset B$ then $A \preceq B$. Clearly, if a theory admits quantifier elimination, it is model complete. A theory T^* is said to be a *model completion* of it subtheory T if, first, any model of T can be embedded into a model of T^*, and, second, T^* is complete over any model of T, that is, for any model A of T and any models B, C of T^* with $A \subseteq B, C$, the structures $(B, a)_{a \in A}$ and $(C, a)_{a \in A}$ are elementarily equivalent.

Theorem 5.15. *T_{ds} admits quantifier elimination, is complete, and is a model completion of T_0.*

Proof. First we show that any model (U, U_0) of T_0 can be embedded into a model of T_{ds}. We may assume that $U \neq U_0$. It suffices to find an ordered vector space $V \supseteq U$ such that some direct complement U' of U in V is dense. Indeed, then $U_0 + U'$ is a proper dense subspace of V, and (U, U_0) is a submodel of $(V, U_0 + U')$. Note that, for any $u, v \in U$ with $u < v$, in the lexicographic product $U \times U$ we have $(u, 0) < (w, w) < (v, 0)$ where $w = \frac{u+v}{2}$. The vector space $U \times U$ is the direct sum of its subspace $U \times \{0\}$ and the diagonal subspace of $U \times U$; both of them are isomorphic to U as ordered spaces. Thus, if U_1 and U_2 are two copies of U, we can extend the orderings on U_1 and U_2 to an ordering on its direct sum $U_1 \oplus U_2$, which makes the direct sum an ordered vector space such that between any two distinct elements of U_1 there is an element from U_2. By iterating this construction, we can make the direct sum $U_1 \oplus U_2 \oplus \ldots$ of ω copies of U an ordered vector space, which extends each of U_i and in which $U_2 \oplus U_3 \oplus \ldots$ is a dense subspace.

Now we show that T_{ds} admits quantifier elimination, or, equivalently, is submodel complete [Sac72]. It suffices to prove that, for any model A of T_0 and models B and C of T_{ds} with $A \subseteq B$, C, if C is saturated over A then, for any $b \in B \setminus A$, the quantifier-free type of b over A is realized in C. Let $A = (U, U_0)$, $B = (V, V_0)$, and $C = (W, W_0)$. Denote by $q(x)$ the set of all formulas $\phi(x)$ of the forms $x < a$ and $a < x$ with $a \in U$ such that $\phi(b)$ holds in B.

First suppose $b \in V_0 + U$, and $b = d + a$, where $d \in V_0$, $a \in U$. It suffices to show that $q(x) \cup \{P(x-a)\}$ is realized in C. Due to the saturation of C over A, to show the latter, it suffices to prove that this set of formulas is finitely satisfiable in C. The latter means that, for any $a_1, a_2 \in A$ with $a_1 < b < a_2$, there is $c \in a + W_0$ such that $a_1 < c < a_2$. Since W_0 is dense in W, the result follows.

Now suppose $b \notin V_0 + U$. It suffices to show that $q(x) \cup \{\neg P(x-a) : a \in A\}$ is realized in C. Due to the saturation of C over A, to show the latter, it suffices to prove that this set of formulas is finitely satisfiable in C. The latter means that, for any $a_1, a_2 \in A$ with $a_1 < b < a_2$ and any $d_1, \ldots, d_n \in A$, there is $c \notin d_1 + W_0, \ldots, d_n + W_0$ such that $a_1 < c < a_2$. Since the index of W_0 in W is infinite, there is $d \in C$ such that $d + W_0$ meets no $d_i + W_0$. As W_0 is dense in W, there is $c \in d + W_0$ such that $a_1 < c < a_2$, and the result follows.

Thus, T_{ds} is a model completion of T_0. The theory T_0 has the Joint Embedding Property because any models (U, U_0) and (V, V_0) of T_0 are embedded into its model $(U \times V, U_0 \times V_0)$, where $\times$ denotes the lexicographic product. Therefore T_{ds} is complete. ∎

Note that for the structure $(\mathbb{R}, +, \times, <, \mathbb{Q})$ the collapse result fails.

First, the locally generic query "*the number of elements of P is even*" over $M = (\mathbb{R}, +, \times, <, \mathbb{Q})$ is expressible in the first-order extended language. For example, the cardinality of a subset P of $\mathbb{R}$ is even iff (M, P) satisfies the first-order sentence which says:

> *If $P \neq \emptyset$, there is a real number α such that the integral parts $[x]$ of elements $x \in \alpha P$ are pairwise distinct, and for some even positive integer n and some integer m, the remainders when m is divided by $n! + 1$, $2(n!) + 1$, $\ldots$, $n(n!) + 1$ are pairwise distinct and form the set $\{[x] : x \in \alpha P\}$.*

(The latter sentence is first-order indeed: as $\mathbb{Z}$ is first-order definable in the field of the rational numbers without parameters (see [Rob49]), it is definable in M, too.) If the sentence holds, the cardinality of P is obviously even. Suppose P is not empty, the cardinality of P is n, and n is even. Choose nonzero real α so that $|x - y| \geqslant 1$ for any different $x, y \in \alpha P$. Then $[x]$ are pairwise distinct, for $x \in \alpha P$. Let $\{[x] : x \in \alpha P\} = \{r_1, \ldots, r_n\}$. As $n! + 1$, $2(n!) + 1$, $\ldots$, $n(n!) + 1$ are pairwise coprime, by the Chinese remainder theorem, there is an integer m such that for $1 \leqslant i \leqslant n$, the remainder when m is divided by $i(n!) + 1$ is r_i, and we are done.

Secondly, on the other hand, the sentence cannot be expressed as a restricted query, otherwise, by compactness, we could construct two elementarily nonequivalent dense ordering without endpoints with distinguished infinite pseudo-finite subsets; this is impossible as shown in the proof of Theorem 3.4.

A modification of the argument above shows that the locally generic query "*the cardinality of P is even*" over the ordered field of rational numbers is expressible in the first-order extended language, even though we know that it cannot be expressed as a restricted query—the latter can be shown by the same arguments as above. The cardinality of a set of rational numbers P is even iff $(\mathbb{Q}, +, \times, <, P)$ satisfies the first-order sentence which says:

> *If $P \neq \emptyset$, there is a positive integer k such that $kP \subseteq \mathbb{Z}$ and, for some even positive integer n and some integer m, the remainders when m is divided by $n! + 1$, $2(n!) + 1$, $\dots$, $n(n!) + 1$ are pairwise distinct and form the set kP.*

Indeed, if the sentence holds, the cardinality of P is obviously even. Suppose P is not empty, the cardinality of P is n, and n is even. Let k be the product of the denominators of members of P; then $kP \subseteq \mathbb{Z}$. Let $kP = \{r_1, \dots, r_n\}$. There is an integer m such that for $1 \leqslant i \leqslant n$ the remainder when m is divided by $i(n!) + 1$ is r_i, and we are done.

Theorem 5.16. *The theory T_{ds} has the Pseudo-finite Homogeneity Property (with $\lambda = |F|^+$), but doesn't have the Isolation Property. For instance, the collapse result holds for any structure of the form $(\mathbb{R}, +, <, F, f_\alpha)_{\alpha \in F}$, where F is a subfield of $\mathbb{R}$.*

Proof. First we prove that T_{ds} has the Pseudo-finite Homogeneity Property with $\lambda = |F|^+$. Let $M = (W, W_0)$ be a model of T_{ds}, where W is an ordered vector space over F, and W_0 is its proper dense subspace. Let A, B be pseudo-finite subsets in M with λ-saturated (M, A, B). Let $h : A \to B$ be a partial L-isomorphism in M, and $a \in W$. We need to find $b \in W$ such that $h \cup \{(a, b)\}$ is a partial isomorphism in M.

If $a = t(\bar{a})$, where $t(\bar{x})$ is a term and $\bar{a}$ is a tuple in A, one can take $t(h(\bar{a}))$ as a b. So we will assume that $a \neq t(\bar{a})$, for every $t(\bar{x})$ and $\bar{a}$.

Fix a term $t(\bar{x})$. For every finite subset X of W, in (M, X) the following holds: for every $d \in W$ the set $\{t(\bar{c}) : \bar{c} \in X,\ t(\bar{c}) \leqslant d\}$ has a maximal element $m_t(d, X)$, and the set $\{t(\bar{c}) : \bar{c} \in X,\ d \leqslant t(\bar{c})\}$ has a minimal element $m^t(d, X)$, provided $X \neq \emptyset$. As A is pseudo-finite in M, the same holds for A instead of X. Let $m_t(a, A) = t(\bar{a}_t)$ and $m^t(a, A) = t(\bar{a}^t)$, where $\bar{a}_t$ and $\bar{a}^t$ are tuples in A. Denote $\bar{b}_t = h(\bar{a}_t)$ and $\bar{b}^t = h(\bar{a}^t)$.

Suppose $a - t_0(\bar{a}_0) \in W_0$, for some term $t_0(\bar{x})$ and tuple $\bar{a}_0$ in A. Denote $\bar{b}_0 = h(\bar{a}_0)$. It suffices to show that the set of formulas

$$\{P(x - t_0(\bar{b}_0)),\ t(\bar{b}_t) < x < t(\bar{b}^t) :\ t(\bar{x}) \text{ is a term}\}$$

is realized in M. Since M is λ-saturated, it suffices to verify that this set is finitely satisfiable in M. The latter holds as $t_0(\bar{b}_0) + W_0$ is dense in W.

Now suppose that $a - t(\bar{a}) \notin W_0$, for any term $t(\bar{x})$ and any tuple $\bar{a}$ in A. For a term $t(\bar{x})$ and a new unary relation name Q, denote by $\theta_t(x, Q)$ the formula

$$\forall \bar{x} (\textstyle\bigwedge_i Q(x_i) \to \neg P(x - t(\bar{x}))).$$

It suffices to show that the set of formulas

$$\{\theta_t(x, Q),\ t(\bar{b}_t) < x < t(\bar{b}^t) :\ t(\bar{x}) \text{ is a term}\}$$

is realized in (M, B). Since (M, B) is λ-saturated, it suffices to verify that this set is finitely satisfiable in M. To prove the latter, it suffices to show that, for any terms $t_1(\bar{x}), \ldots, t_n(\bar{x})$, the set S of solutions of the formula

$$\theta_{t_1}(x, Q) \wedge \cdots \wedge \theta_{t_n}(x, Q)$$

in (M, B) is dense in W. Since B is pseudo-finite, it suffices to prove that in the case of finite B. Put $C_i = \{t_i(\bar{b}) : \bar{b} \in B\}$ and $C = C_1 \cup \cdots \cup C_n$. Since B is finite, the set C is finite, too. Then there is $d \in W$ such that $d - c \notin W_0$ for all $c \in C$. We have $d + W_0 \subseteq S$ because if $\theta_{t_i}(d + w, B)$ fails for some $w \in W_0$ then $d - c \in W_0$ for some $c \in C_i$. Since $d + W_0$ is dense in W, the result follows.

Now we show that T_{ds} does not have the Isolation Property.

We will show that, for every cardinal λ, there are a pseudo-finite set A of power $\geqslant \lambda$ and an element a in a model of T_{ds} such that $\operatorname{tp}(a/A)$ is not isolated by $\operatorname{tp}(a/A_0)$, for any proper subset A_0 of A.

Let $p_n(x_0, \ldots, x_n)$ be a set of L-formulas which says that $x_0, \ldots, x_n$ are linearly independent modulo the subspace P. Let $q_n(x_0, \ldots, x_n)$ be a set of L-formulas which says that $x_0 > y$, for every linear combination y of $x_1, \ldots, x_n$. For a new unary relation name Q, let $F(T, Q)$ be the set of all $L(Q)$-sentences which hold in all (M, X), where X is a finite set in a model M of T_{ds}. Consider the signature L' which is obtained by adjoining to $L(Q)$ a set of new constant symbols $\{c, c_i : i < \lambda\}$. Consider the set Γ of L'-sentences, which is the union of $F(T, Q)$, $\{Q(c_i) : i < \lambda\}$, and all $p_n(c, c_{i_1}, \ldots, c_{i_n})$, $q_n(c, c_{i_1}, \ldots, c_{i_n})$, for $i_1 < \ldots i_n < \lambda$. Suppose $(M, A, a, a_i)_{i<\lambda}$ is a saturated over singletons model of Γ. Then

(1) A is a pseudo-finite set in M,

(2) $a_i \in A$ for all i,

(3) $\{a, a_i : i < \lambda\}$ is linearly independent modulo P,

(4) $a > c$, for any linear combination of the a_i's.

In particular, $|A| \geqslant \lambda$. We will show that $\operatorname{tp}(a/A)$ is not isolated by $\operatorname{tp}(a/A_0)$, for any proper subset A_0 of A.

Let $r(x)$ be the union of all $p_n(x, b_1, \ldots, b_n)$ and $q_n(x, b_1, \ldots, b_n)$, where $n < \omega$ and $b_1, \ldots, b_n$ are distinct elements of A_0. Clearly, $r(x) \subseteq t(a/A_0)$; moreover, $r(x)$ isolates $t(a/A_0)$ as T_{ds} admits quantifier elimination. We have $\neg P(x - a_0) \in \operatorname{tp}(a/A)$, for any $a_0 \in A \setminus A_0$. Therefore it suffices to prove that $r(x) \cup \{P(x - a_0)\}$ is consistent.

Since A is pseudo-finite, there $d \in M$ with $d \geqslant |a|$, for every $a \in A$. As M is saturated over d, there is $b \in M$ with $b > \alpha d$ for all $\alpha \in F$. Since $P + a_0$ is dense, there is $e \in P + a_0$ with $e > b$. Then e realizes $r(x)$. Indeed, let $b_1, \ldots, b_n \in A_0$. Then $p_n(e, b_1, \ldots, b_n)$ holds because $e \in P + a_0$ and $p_n(a_0, b_1, \ldots, b_n)$ holds. We have

$$\alpha_1 b_1 + \cdots + \alpha_n b_n \leqslant (|\alpha_1| + \cdots + |\alpha_n|)d < b < e,$$

for any $\alpha_i \in F$; therefore $q_n(e, a_1, \ldots, a_n)$ holds.

So, to complete the proof, it suffices to show that Γ is finitely satisfiable. To prove the latter, we show that, for every $n < \omega$, the set

$$\Gamma_n = F(T, Q) \cup \{Q(c_i) : \ i < n\} \cup p_n(c, c_0, \dots, c_{n-1}) \cup q_n(c, c_0, \dots, c_{n-1})$$

has a model. Consider an $\aleph_0$-saturated model (V, V_0) of T_{ds}. Due to the $\aleph_0$-saturation of (V, V_0), we can choose $c_0, \dots, c_{n-1}$ which are linearly independent over V_0.

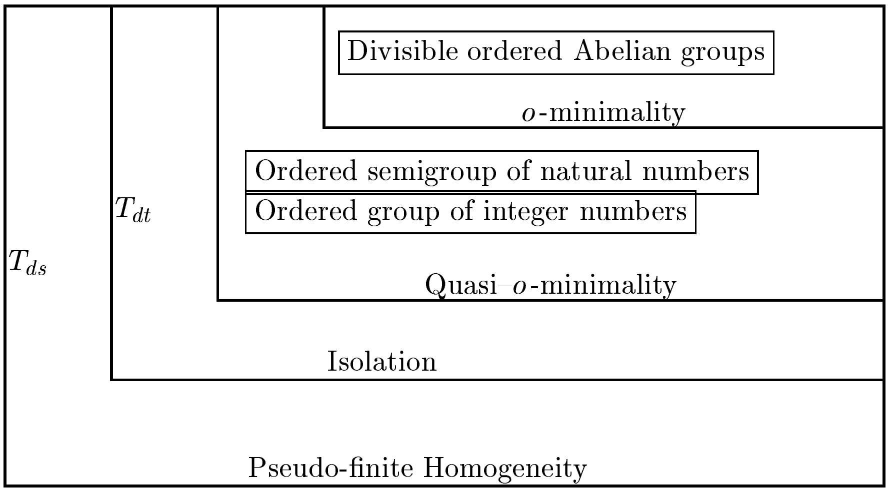
It suffices to show that there is c such that $p_n(c, c_0, \dots, c_{n-1})$ and $q_n(c, c_0, \dots, c_{n-1})$ hold because then we can take $\{c_0, \dots, c_{n-1}\}$ as Q.

Let d be the greatest element among $|c_0|, \dots, |c_{n-1}|$. Due to the saturation of (V, V_0) over d, there is $b \in V$ with $b > \alpha d$ for all $\alpha \in F$. Clearly, for every c, if $c > b$ then $q_n(x, c_0, \dots, c_{n-1})$ holds. For a subset S of F, consider the set of formulas

$$\{\neg P(x - (\sigma_1 c_1 + \dots + \sigma_n c_n)) : \sigma_1, \dots, \sigma_n \in S\} \cup \{x > b\}.$$

Since V_0 is of infinite index and dense in V, the set is realized in (V, V_0), for every finite S. Since (V, V_0) is saturated over b, there is $c \in V$ with $c > b$ such that $p_n(c, c_0, \dots, c_{n-1})$ holds, and we are done. ∎

The following picture presents our collapse results.



6. Open questions

6.1. **Problem 1.** Further work needs to be done for integer numbers. For instance, the authors are under impression that an effective translation algorithm for locally generic queries over $(\mathbb{Z}, <, +)$ can be extracted from a Fraïssé-Ehrenfeucht-style proof of decidability of Presburger Arithmetic.

6.2. **Problem 2.** In general, we **conjecture** that if, for an expansion of $(\mathbb{Z}, <)$ or another expanded ordered domain, the first-order theory of the expanded domain is decidable and locally generic extended queries do not express more than restricted ones, then there is an effective translation algorithm.

6.3. **Problem 3.** Then, how much higher than $+$ in $(\mathbb{Z}, <)$ can we go? Obviously, $+, \times$ make locally generic extended queries possible to express more than pure order ones because, for example, the query "the cardinality of P is even" is locally generic and expressible in the FO extended language but not in the restricted language. We **conjecture** that if, in an expansion of $(\mathbb{Z}, <)$, locally generic extended queries express more than restricted ones, the first-order theory of the expanded domain is undecidable.

6.4. **Problem 4.** Moreover, we **conjecture** that if, in an expansion of

$$(\mathbb{Z}, <),$$

locally generic extended queries express more than restricted ones, then there is a first-order formula $\Phi(x, y)$ in the language of the expansion such that for any finite subset A of integer numbers there is such a natural number b that

$$A = \{a \in \mathbb{Z} \mid \Phi(a, b)\}.$$

In particular, the Random Graph is interpretable in such an expansion.

6.5. **Problem 5.** On the side of finitization, a natural question is, can the results for finitely represented states be extended to more powerful constraint databases, say, to linear constraint databases? If yes, for which domains? A specific **open problem** is, can it be done for linear constraints with at least one constant over rational or real numbers.

6.6. **Comments.** Anuj Dawar has observed that if the Random Graph is interpretable in an expansion of $(\mathbb{Z}, <)$, then the first-order theory of this expansion is undecidable. Therefore, truth of our Conjecture 6.4 would imply truth of our Conjecture 6.3.

To support our Conjecture 6.4, we propose the following observation: for any expansion of the ordered set of integers by unary relations, the collapse result holds. In fact, the following more general result holds.

Let T_{un} be the first-order theory of a structure which is an expansion of a linearly ordered set by unary relations.

Theorem 6.1. *The theory T_{un} has the Isolation Property.*

Proof. Let M be a model of T_{un}, and A be a pseudo-finite set in it, $a \in M \setminus A$. We show that there is $A_0 \subseteq A$ with $|A_0| \leq 2$ such that $\text{tp}(a/A_0)$ isolates $\text{tp}(a/A)$. As A is pseudo-finite, one of the following holds:

- There are b and c in A such that there are no elements in A between them, and $b < a < c$
- A has the maximal element b, and $a > b$
- A has the minimal element c, and $a < c$

Then we can take $\{b, c\}$ as A_0 in the first case, $\{b\}$ in the second case, and $\{c\}$ in the third case. We will prove it only for the first case; in other two cases the arguments are similar. We need to show that, for any $a' \in M$, if $(M, a, b, c) \equiv (M, a', b, c)$ then $(M, a, d)_{d \in A} \equiv (M, a', d)_{d \in A}$. We may assume that M is ω-saturated. Let N be the substructure of M whose underlying set is the interval (b, c). Then (N, a) and (N, a') are elementarily equivalent ω-saturated structures. Therefore they are back-and-forth equivalent; in other words, in the game in N starting with the partial isomorphism $\{(a, a')\}$ the Duplicator has a winning strategy. Since M is an expansion of a linearly ordered set by unary relations, any map from M to M, which is the identity outside N and whose restriction on N is a partial isomorphism in N, is a partial isomorphism in M. It follows that $(M, a, d)_{d \in A}$ and $(M, a', d)_{d \in A}$ are back-and-forth equivalent, and so elementarily equivalent. ∎

Corollary 6.2. *The collapse result holds for T_{un}.*

Acknowledgments

We are grateful to the late Paris Kanellakis, and also to Michael Benedikt, Jan Van den Bussche, Yuri Gurevich, Leonid Libkin, Charles Steinhorn, and Jianwen Su for helpful comments and suggestions.

References

[BB97] J.T. Baldwin and M. Benedikt, *Stability theory, permutations of indiscernibles, and embedded finite models*, Preprint, 1997.

[BDLW96] M. Benedikt, G. Dong, L. Libkin, and L. Wong, *Relational expressive power of constraint query languages*, Proc. 15th ACM Symp. on Principles of Database Systems, 1996, pp. 5–16.

[BPW97] O. Belegradek, Y. Peterzil, and F. Wagner, *Quasi-o-minimal structures*, Preprint, 1997.

[BST96] O.V. Belegradek, A.P. Stolboushkin, and M.A. Taitslin, *On order-generic queries*, Technical report 96–01, DIMACS, 1996.

[BST97] O.V. Belegradek, A.P. Stolboushkin, and M.A. Taitslin, *Generic queries over quasi-o-minimal domains*, Logical Foundation of Computer Science, 4th International Symposium, LFCS'97, Yaroslavl, Russia, July 1997, Proceedings, LNCS, vol. 1234, Springer-Verlag, 1997, pp. 21–32.

[CH80] A. Chandra and D. Harel, *Computable queries for relational databases*, Journal of Computer and System Sciences **21** (1980), 156–178.

[CK90] C.C. Chang and H.J. Keisler, *Model theory*, 3rd ed., North Holland, 1990.

[FZ97] J. Flum and M. Ziegler, *Pseudo-finite homogeneity and saturation*, Preprint, 1997.

[GS94] S. Grumbach and J. Su, *Finitely representable databases*, Proc. 13th ACM Symp. on Principles of Database Systems, 1994.

[GS95] S. Grumbach and J. Su, *Dense-order constraint databases*, Proc. 14th ACM Symp. on Principles of Database Systems, 1995, pp. 66–77.

[GST95] S. Grumbach, J. Su, and C. Tollu, *Linear constraint databases*, Proc. Logic and Computational Complexity (LCC'94), LNCS, Springer-Verlag, 1995.

[Gur90] Yu. Gurevich, Private communication, 1990.

[Hod93] W. Hodges, *Model theory*, Cambridge University Press, 1993.

[KG94] P.C. Kanellakis and D.Q. Goldin, *Constraint programming and database query languages*, Proc. International Symposium on Theoretical Aspects of Computer Software (TACS'94), 1994, pp. 96–120.

[KKR90] P.C. Kanellakis, G.M. Kuper, and P.Z. Revesz, *Constraint query languages*, Proc. 9th ACM Symp. on Principles of Database Systems, 1990, pp. 299–313.

[KKR95] P.C. Kanellakis, G.M. Kuper, and P.Z. Revesz, *Constraint query languages*, Journal of Computer and System Sciences **51** (1995), no. 1, 26–52.

[KPS86] J.F. Knight, A. Pillay, and C. Steinhorn, *Definable sets in ordered structures, II*, Transactions of American Mathematical Society **295** (1986), no. 2, 593–605.

[OV96] M. Otto and J. Van den Bussche, *First-order queries on databases embedded in an infinite structure*, Information Processing Letters **60** (1996), 37–41.

[PS86] A. Pillay and C. Steinhorn, *Definable sets in ordered structures, I*, Transactions of American Mathematical Society **295** (1986), no. 2, 565–592.

[PS88] A. Pillay and C. Steinhorn, *Definable sets in ordered structures, III*, Transactions of American Mathematical Society **309** (1988), no. 2, 469–476.

[PVV94] J. Paradaens, J. Van den Bussche, and D. Van Gucht, *Toward a theory of spatial database queries*, Proc. 13th ACM Symp. on Principles of Database Systems, ACM Press, 1994, pp. 279–288.

[PVV95] J. Paradaens, J. Van den Bussche, and D. Van Gucht, *First-order queries on finite structures over reals*, Proc. 10th IEEE Symp. on Logic in Computer Science, IEEE Computer Society Press, 1995, pp. 79–87.

[Rob49] Julia Robinson, *Definability and decision problems in arithmetic*, Journal of Symbolic Logic **14** (1949), 98–114.

[RW81] B.I. Rose and R.E. Woodrow, *Ultrahomogeneous structures*, Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **27** (1981), no. 1, 23–30.

[Sac72] G.E. Sacks, *Saturated model theory*, W.A. Benjammin, Inc., Reading, Massachusetts, 1972.

[She90] S. Shelah, *Classification theory and the number of non-isomorphic models*, 2rd ed., North Holland, Amsterdam, 1990.

[ST96] A.P. Stolboushkin and M.A. Taitslin, *Linear vs. order constraint queries over rational databases*, Proc. 15th ACM Symp. on Principles of Database Systems, 1996, pp. 17–27.

O.V.Belegradek, Department of Mathematics, Kemerovo State University, Kemerovo, Russia 650043

E-mail address: beleg@kaskad.uucp.stanet.ru

Alexei P. Stolboushkin, Fourth Dimension Software, 555 Twin Dolphin Dr., Suite 500, Redwood City, CA 94065, and UCLA Department of Mathematics, Los Angeles, CA 90095-1555

E-mail address: aps@4ds.com

M.A.Taitslin, Department of Computer Science, Tver State University, Tver, Russia 170000

E-mail address: Michael.Taitslin@tversu.ru