Proceedings ISSAC 97

Uniform Quantifier Elimination and Constraint Query Processing

Saugata Basu*

Mathematical Sciences Research Institute, Berkeley, CA 94720, U.S.A

saugata@msri.org
http://www.msri.org/members/bio/saugata

Abstract

In this paper we introduce a variant of the quantifier elimination problem for the first order theory of real closed fields. Instead of considering a single quantified formula, we consider a uniform sequence of such formulas and eliminate quantifiers to obtain another uniform sequence. Our immediate motivation comes from a problem in the theory of constraint databases with real polynomial constraints. Using the uniform quantifier elimination algorithm, we give an algorithm for converting a query with natural domain semantics to an equivalent one with active domain semantics. A non-constructive version of this result was proved in [2]. Very recently, a constructive proof was also given independently in [5].

In the last part of the paper, using some elementary tools from first order logic and results described above, we prove that parity is not expressible in the constraint query language over the reals. This was proved by Benedikt et al [2]. However unlike ours, their proof uses difficult techniques from non-standard analysis and model theory of ordered structures.

1 Introduction

In this paper we consider the following problem which we call *uniform quantifier elimination* over real closed fields. Our primary motivation for looking at this problem comes from the theory of constraint databases. However, we believe that the problem is interesting in its own right.

We first recall the problem of eliminating quantifiers from first order formulas in the theory of real closed fields which has received a lot of attention in recent years [21, 12, 20, 6]. The following complexity result which is the best known for this problem appears in [6].

Let $\mathcal{P} = \{P_1, \ldots, P_s\}$ be a family s polynomials in $k + l$ variables, $X_1, \ldots, X_k, Y_1, \ldots, Y_l$, whose degrees are bounded by d.

*Supported by a postdoctoral fellowship from the Mathematical Sciences Research Institute, Berkeley. Research at MSRI is supported in part by NSF grant DMS-9022140.

Also let

$$\phi = (Q_\omega X^{[\omega]}) \ldots (Q_1 X^{[1]}) F(P_1, \ldots, P_s),$$

be a first-order formula, where, $Q_i \in \{\forall, \exists\}$, $Q_i \neq Q_{i+1}$, $Y = (Y_1, \ldots, Y_l)$ is a block of l free variables, $X^{[i]}$ is a block of k_i variables, $\sum_{1 \leq i \leq \omega} k_i = k$, and $F(P_1, \ldots, P_s)$ is a quantifier-free Boolean formula, with atomic predicates of the form,

$$P_i(Y, X^{[\omega]}, \ldots, X^{[1]})\ \sigma_j\ 0, 1 \leq i \leq s,$$

where $\sigma_j \in \{>, <, =\}$.

The quantifier elimination problem is to construct a quantifier free Boolean formula, $\psi(Y)$, such that $y \in R^l$ satisfies the given formula $\phi(Y)$, if and only if $\psi(y)$ is true.

The algorithm presented in [6] has a complexity of $s^{(l+1)\Pi(k_i+1)} d^{l \Pi O(k_i)}$. and produces a quantifier free formula in the following form:

$$\psi(Y) = \vee_{i=1}^{I} \wedge_{j=1}^{J_i} (P_{ij}(Y)\ \sigma_{ij}\ 0),$$

where $P_{ij}(Y)$ are polynomials in the variables Y, $\sigma_{ij} \in \{>, <, =\}$

$$I \leq s^{(l+1)\Pi_i (k_i+1)} d^{l \Pi_i O(k_i)},$$
$$J \leq s^{\Pi_i (k_i+1)} d^{\Pi_i O(k_i)},$$

and the degrees of the polynomials $P_{ij}(y)$ are bounded by $d^{\Pi_i O(k_i)}$.

The problem that we consider in this paper is a variant of the problem described above. We call a sequence, $\{\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) \mid n > 0\}$ of first-order formulas ϕ_n in the language of ordered fields, to be a *uniform sequence* if each ϕ_n has the form,

$$\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) =$$

$$Q^1_{1 \leq k_1 \leq n} \ldots Q^\omega_{1 \leq k_\omega \leq n} \phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}),$$

where $Q^i \in \{\bigvee, \bigwedge\}, 1 \leq i \leq \omega$ and ϕ is some fixed $(l+\omega)$-ary quantifier-free first-order formula.

Thus for every n, ϕ_n is a first order formula with $l + n$ free variables. We will refer to the variables $T_1, \ldots, T_l$ as *parameters*.

Given a uniform sequence of formulas,

$$\Phi = \{\phi_n \mid n > 0\},$$

where

$$\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) =$$
$$Q^1_{1 \le k_1 \le n} \cdots Q^\omega_{1 \le k_\omega \le n} \phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}),$$

we define the *size* of Φ henceforth denoted $||\Phi|| = |\phi|$ where $|\phi|$ is the length of the formula ϕ.

Before proceeding further we give an example.

Example:

Consider the uniform sequence of formulas,

$$\phi_n(T_1, Y_1, \ldots, Y_n) = \bigwedge_{1 \le k_1 \le n} (Y_{k_1} - T_1 = 0), \quad n > 0.$$

Now consider the sequence of quantified formulas, $(\exists T_1)\phi_n(T_1, Y_1, \ldots, Y_n)$. From the fact that there exists quantifier elimination for the first order theory of real closed fields it is clear that for every $n > 0$, there exists a quantifier free formula equivalent to $(\exists T_1)\phi_n(T_1, Y_1, \ldots, Y_n)$. However, it is not clear whether there exists a uniform sequence of quantifier-free formulas, $\{\Psi_n(Y_1, \ldots, Y_n) \mid n > 0\}$ such that for every $n > 0$ $\Psi_n(Y_1, \ldots, Y_n) \Leftrightarrow (\exists T_1)\phi_n(T_1, Y_1, \ldots, Y_n)$. In this example, it is easily seen that letting

$$\Psi_n = \bigwedge_{1 \le k_1 \le n} \bigwedge_{1 \le k_2 \le n} (Y_{k_1} - Y_{k_2} = 0),$$

we get a uniform sequence of quantifier-free formulas satisfying,

$$\Psi_n(Y_1, \ldots, Y_n) \Leftrightarrow (\exists T_1)\phi_n(T_1, Y_1, \ldots, Y_n)$$

for every $n > 0$.

In this paper we address the question whether in general it is possible to eliminate quantifiers from a uniform sequence of formulas and obtain another uniform sequence of quantifier free formulas. We give an algorithm for solving this problem. We call this procedure *uniform quantifier elimination*.

We now describe the application of *uniform quantifier elimination* to the theory of *constraint databases*.

In recent years, constraint databases, first introduced by Kanellakis, Kuper and Revesz [18] has attracted a lot of attention. Unlike traditional databases which are finite collections of data items, constraint databases permit infinite collections of items to be stored in the database. They are a powerful generalization of Codd's relational model [22]. different types of constraints have been considered by various authors [18, 17, 14] including dense linear order inequalities, real polynomial inequalities etc.

Even though constraint databases permit infinite number of elements in the database, for the purposes of this paper it is enough to consider our database to be a *finite* set S of real numbers.

We first introduce some notation. Let $L = \langle 0, 1, +, \times, \le \rangle$ be the language of ordered fields. We extend L to $L' = \langle 0, 1, +, \times, \le, R \rangle$ by including a unary relation symbol R where $R(x) \Leftrightarrow x \in S$. A *query* is a first order L'-sentence. For instance the query,

$$\mathcal{Q} = (\forall T_1)(\forall T_2)((T_1 \ne T_2) \wedge R(T_1) \wedge R(T_2)) \Rightarrow (T_1 - T_2)^2 \ge 4,$$

expresses the fact that any two distinct members of S differ by at least 2.

Thus, $\mathcal{Q}$ will be true when $S = \{1, 5, 11\}$ but false for $S = \{1, 2\}$.

Let $\langle \Re, S \rangle$ denote the L'-structure consisting of the reals (with the usual identification of $+$ with addition, $\times$ with multiplication and so on) and a subset $S \subset \Re$ representing the unary relation R. We call two queries $\mathcal{Q}_1, \mathcal{Q}_2$ to be *equivalent* iff for every finite subset S of real numbers, $\langle \Re, S \rangle \models \mathcal{Q}_1 \Leftrightarrow \langle \Re, S \rangle \models \mathcal{Q}_2$.

Note that the quantified variables in the query vary over all the reals. This interpretation of a query is also called *natural domain semantics.* Unless otherwise stated all queries in this paper will be interpreted in this manner. Another, interpretation called the *active domain semantics*, interpret the quantified variables to vary only over the set S. Clearly, a query with active domain semantics is equivalent to another one with natural domain semantics. However, the converse is not quite obvious. Given a query with natural domain semantics, such as in the example above, does there exist a query with active domain semantics that is equivalent to it ? Actually, in the above example $\mathcal{Q}$ with active domain semantics is equivalent to itself with natural domain semantics. However, this is not always the case. Consider the query,

$$\mathcal{Q} = (\exists T_1) \neg R(T_1).$$

Clearly, the natural and active domain semantics give different interpretation to this query. The query is always true with natural domain semantics but clearly always false with active domain semantics.

We apply our algorithmic result on uniform quantifier elimination to give an algorithm that given a query with natural domain semantics, outputs another query equivalent to it with active domain semantics. The advantage of having a query with active domain semantics over a finite database is clear. In order to evaluate such a query it suffices to go over the elements of the finite database as all the quantified variables vary only over the database. Thus, we have a uniform algorithm to evaluate such a query independent of the set S. This will allow, for example, compile time query optimization that are possible in classical database setting [4].

Another important problem that has been well-studied in the literature is determining the expressive power of constraint queries. This problem has spurred a lot of research leading to several interesting inexpressibility results for constraint queries over various structures [15, 14, 2]. The *parity* of the cardinality of a set and the *connectivity* of a finite graph are well-known examples of queries that are not definable over finite models in a first order language with equality. Their undefinability has been shown using different proof techniques such as locality [11], 0/1 laws [9, 10], Ehrenfeucht-Fraisse games [8]. These undefinability results also hold in the presence of an order relation [13].

Grumbach and Su [14] proved that parity is not definable over constraint databases with linear constraints. They also conjectured that parity is not definable for constraint databases over structures with decidable theories. One interesting case of this conjecture is that parity is not definable over the reals with queries allowed to have polynomial equalities and inequalities.

Let us pose this question more concretely.

Question: Does there exist a query (that is a first order L'-sentence) σ such that for every finite set $S \subset \Re$ with $|S|$ even, $\langle \Re, S \rangle \models \sigma$ and for every finite set $S \subset \Re$ with $|S|$ odd, $\langle \Re, S \rangle \models \neg\sigma$?

This was answered in the negative by Benedikt et al.

[2] who used difficult techniques from non-standard analysis and model theory of ordered structures.

Using the fact that any query with natural domain semantics is equivalent to one with active domain semantics we give a simpler and elementary proof for the conjecture over the reals.

2 Prior Work

Grumbach and Su [14] first conjectured that queries such as transitive closure, connectivity test, and parity are not definable over structures with decidable theories. They proved that parity is definable over the natural numbers or the rationals, with addition and multiplication. However, the theories for these structures are undecidable. On the other hand they show that these queries are not expressible over natural numbers with only addition (also called Presburger arithmetic). This proof involves Ehrenfeucht-Fraisse games.

One of the most interesting decidable theories is that of the class of real closed fields. Examples of real closed fields are the familiar real numbers, but also real algebraic numbers, the field of Puiseux series with real coefficients and so on. The problem of designing efficient algorithms for solving the decision problem for the theory of reals and performing quantifier elimination, has attracted a lot of attention (see [20, 12, 6]). Constraint databases over the reals with queries allowed to have polynomial inequalities can be of use in various areas such as database systems for storing geographical information, where the manipulation of geometrical and topological information is important.

Benedikt et al [2] proved the Grumbach-Su conjecture for constraint queries over the reals. However, their proof is rather complex and involve Ramsey theoretic techniques and also results from non-standard analysis and the theory of o-minimal structures.

The equivalence of natural and active domain semantics over real closed fields was proved in [2], [4]. However, their proofs are essentially non-constructive and one does not obtain an effective algorithm for converting a query with the natural domain semantics into an equivalent one with active domain semantics. Very recently, in [5] an algorithm was given for this problem, but the paper contains no complexity analysis.

The main contributions of this paper are as follows. We introduce the problem of *uniform quantifier elimination* over the reals and give an algorithm for performing uniform quantifier elimination. As a result we obtain an algorithm for converting a constraint query with natural domain semantics to an equivalent one with active domain semantics. Using this we give a simple, elementary proof of the Grumbach-Su conjecture for the reals.

In the concluding section, we identify a certain finiteness condition, which we call the *locality property* such that any structure whose theory admits quantifier elimination and also satisfies this condition will also satisfy the Grumbach-Su conjecture. We show that several other theories which admit quantifier elimination also satisfies this condition and indicate a general method to prove the conjecture for these theories too.

The rest of the paper is organized as follows. In section 3 we identify a key property of real closed fields that we call the *locality property* that is important in our proofs.

In section 4 we state and prove our result on uniform quantifier elimination. In section 5 we show how to use the previous result to convert a query with natural domain semantics to an equivalent one with active domain semantics.

In section 6 we give our proof of the parity conjecture over the reals and finally in section 7 we pose some open questions.

3 The Locality Property

In this section we identify one important property satisfied by real closed fields that will be used later on.

Let $\mathcal{M}$ be any structure (for example, $\mathcal{M}$ can be any real closed field).

We say that $\mathcal{M}$ has the locality property if the following holds.

Let $\phi_i(Y,T), 1 \leq i \leq n$ be any collection of first-order formulas with parameters $Y = (Y_1, \ldots, Y_l)$ such that $max_i|\phi_i| = L$, where $|\phi|$ denotes the length of ϕ. For any values of the parameters, $y = (y_1, \ldots, y_l)$, let $S_i(y) = \{t \in \mathcal{M} \mid \phi_i(y,t)\}$, and let $W(y) = \bigcap_i S_i(y)$.

Then, there exists a number M which depends only on L, and for each i a finite set of size M of definable points, parametrized by Y, $\{t_{i,1}(Y), \ldots, t_{i,M}(Y)\} \subset S_i(Y)$, such that $W(y) = \emptyset$ iff $t_{i,j}(y) \notin W(Y)$ for all $1 \leq i \leq n$, and $1 \leq j \leq M$.

3.1 Locality Property for Real Closed Fields

We now explain why the locality property is satisfied by real closed fields. We remark that this property has been studied in detail (also w.r.t. complexity) under the name of "Skolem theory" in [23, 24].

The basic reason is that a definable subset of the reals consists of a union of finite number of points and intervals. Now, if a finite collection of such sets have a point in common, then there must be a point in common which is an "extreme" point of one of the intervals. More precisely, let $\mathcal{M}$ be a real closed field. Then the sets $S_i(y)$ are each a finite union of open segments and points. The number of connected components of $S_i(y)$ is bounded by some function of L. Now define the extreme points of $S_i(y)$ as follows:

Let, $S_i(y) = I_1(y) \cup I_2(y) \cup \ldots I_{M'}(y)$, where each $I_j(y)$ is a connected component of $S_i(y)$ and $I_j(y) < I_{j+1}(y)$ in the usual ordering. For each interval $I_j(y)$ define its extreme points as follows:

- If $I_j(y)$ is a point then it is an extreme point.
- If $I_j(y) = [a_i(y), b_i(y)]$ then $a_i(y)$ and $b_i(y)$ are its extreme points.
- If $I_j(y) = (a_i(y), b_i(y))$ then $a_i(y) + \epsilon$ and $b_i(y) - \epsilon$ are its extreme points for a positive infinitesimal ϵ.
- Similarly, if $I_j(y) = (a_i(y), b_i(y)]$ then $a_i(y) + \epsilon$ and $b_i(y)$ are its extreme points for a positive infinitesimal ϵ and so on.

By known bounds (see [1]) on the complexity of semi-algebraic sets, and by quantifier elimination results for the theory of the reals [21, 6], the number of extreme points for every $S_i(y)$ is bounded by some function of L. This also shows that these extreme points are definable. To see this observe that we can define the first extreme point of the set $S_i(y)$, the second extreme point and so, till the M^{th} extreme point where M is the bound on the number of extreme points.

The use of the infinitesimal ϵ does not create any problem. We can remove its occurence using techniques similar to those used in [6]. See also [23, 24].

Moreover, it is easy to check that if the intersection of the $S_i(y)$ is non-empty then it must contain one of the extreme points for some $S_i(y)$.

4 Our Results

We now prove the main theorem which states that we can eliminate quantifiers from a uniform sequence of formulas (with parameters) and still have a uniform sequence. The proof is algorithmic and we can bound the complexity of the algorithm in terms of the size of the input sequence.

Theorem 1 *Let, $\Phi = \{\phi_n(Y_1, \ldots, Y_n, T_1, \ldots, T_l) \mid n > 0\}$ be a uniform sequence of formulas with parameters $T_1, \ldots, T_l$, where*

$$\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) =$$
$$Q^1_{1 \le k_1 \le n} \cdots Q^\omega_{1 \le k_\omega \le n} \phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}),$$

and $Q^i \in \{\bigvee, \bigwedge\}$. Let the number of different $(l+\omega)$-variate polynomials appearing in ϕ be s and let their degrees be bounded by d.

Then, there exists an algorithm that outputs a quantifier-free first order formula,

$$\psi(T_2, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_{\omega'}}),$$

along with $Q^i \in \{\bigvee, \bigwedge\}, 1 \le i \le \omega'$, such that for every $n > 0$,

$$\psi_n(T_2, \ldots, T_l, Y_1, \ldots, Y_n) =$$
$$Q^1_{1 \le k_1 \le n} \cdots Q^{\omega'}_{1 \le k_{\omega'} \le n} \psi(T_2, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_{\omega'}}),$$
$$\equiv (\exists T_1)\phi_n(Y_1, \ldots, Y_n, T_1, \ldots, T_l).$$

The complexity of the algorithm is bounded by $s^2 d^{O(l+\omega)}$, and the size of the formula ψ is bounded by $s^2 d^{O(l+\omega)}|\phi|$.

Proof: Let the $(l+\omega)$-variate polynomials appearing in the formula ϕ be $\{P_1, \ldots, P_s\}$.

We treat these polynomials as polynomials in T_1 and consider the other variables as parameters.

For some fixed values of the parameters, say $(t_2, \ldots, t_l, y_1, \ldots, y_n)$, let $S_n(t_2, \ldots, t_l, y_1, \ldots, y_n) = \{t_1 \mid \phi_n(t_1, \ldots, t_l, y_1, \ldots, y_n)\}$.

We call a sequence of lists of parametrized points $W_n(Z, Y_1, \ldots, Y_n)$ to be a uniform sequence if,

$$W_n(Z, Y) = \bigcup_{1 \le i_i, i_2, \ldots, i_k \le n} \{t_j(Z, Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}) \mid 1 \le j \le M\},$$

for some constants k and M.

We compute a uniform sequence of witness points (using the locality property),

$$W_n(T_2, \ldots, T_l, Y_1, \ldots, Y_n) =$$
$$\bigcup_{1 \le i_i, i_2, \ldots, i_\omega \le n} \{t_j(T_2, \ldots, T_l, Y_{i_1}, Y_{i_2}, \ldots, Y_{i_\omega}) \mid 1 \le j \le M\},$$

such that for every fixed value of the parameters, say $(t_2, \ldots, t_l, y_1, \ldots, y_n)$, the set $S_n(t_2, \ldots, t_l, y_1, \ldots, y_n)$ is non-empty iff it contains a point in $W_n(t_2, \ldots, t_l, y_1, \ldots, y_n)$.

Note, that M is a constant depending only on the length of ϕ independent of n.

Let, $\mu_j(T_2, \ldots, T_l, Y_{i_1}, Y_{i_2}, \ldots, Y_{i_\omega}, Y_{k_1}, Y_{k_2}, \ldots, Y_{k_\omega})$ denote the quantifier-free formula expressing the fact that $T_1 = t_j(T_2, \ldots, T_l, Y_{i_1}, Y_{i_2}, \ldots, Y_{i_\omega})$ satisfies $\phi(T_1, \ldots, T_l, Y_{k_1}, Y_{k_2}, \ldots, Y_{k_\omega})$. Since the points t_j are definable μ_j exists and we will show later how to construct μ_j.

We define,

$$\psi_n(T_2, \ldots, T_l, Y_1, \ldots, Y_n) =$$
$$\bigvee_{1 \le i_1, \ldots, i_\omega \le n} \bigvee_{1 \le j \le M} Q^1_{1 \le k_1 \le n} \cdots Q^\omega_{1 \le k_\omega \le n} \mu_j.$$

We note that one can always push the "small" disjunctions in the formula above inside to obtain a uniform formula.

In order to prove this, notice for example that,

$$\bigvee_{1 \le i \le n} \psi^1_n(Y_i) \wedge \bigwedge_{1 \le i \le n} \psi^2_n(Y_i) =$$
$$\bigvee_{1 \le i \le n} \bigwedge_{1 \le j \le n} (\psi^1_n(Y_i) \wedge \psi^2_n(Y_j)).$$

Since M is a constant, $\Psi = \{\psi_n \mid n > 0\}$ is a uniform sequence of formulas.

Also it is clear that,

$$(\exists T_1)\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) \equiv \psi_n(T_2, \ldots, T_l, Y_1, \ldots, Y_n)$$

for every $n > 0$.

It remains to show how we define the uniform sequence of witness points W_n. We rely heavily on techniques from [6]. We assume that the reader is familiar with the algorithm for performing ordinary quantifier elimination over the reals as presented in [6] (section 6.2, page 1042), as well as the notions of Thom encodings of real roots of polynomials and the use of infinitesimals [6] (section 2, page 1009).

Consider the formula, $\phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$. Let the set of polynomials appearing in $\phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$ be $\{P_1, \ldots, P_s\}$.

Let $\mathcal{P} = \cup_{1 \le i \le s}\{P_i, P_i + \epsilon, P_i - \epsilon\}$, where ϵ is a positive infinitesimal.

Now, for every polynomial $Q_i \in \mathcal{P}$, using the Block Elimination Subroutine [6] (section 5.2, page 1037) we compute a family of polynomials $\mathcal{Q}_i \subset D[T_2, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}][T_1]$ such that over every cell of this family the polynomial Q_i has the same number of real roots.

The degrees of the polynomials in $\mathcal{Q}_i$ is bounded by $d^{O(1)}$.

Compute a point in every cell of the partition of $R^{l-1+\omega}$ induced by $\mathcal{Q}_i$ and compute the list of Thom encodings of the real roots of Q_i over each such point. The complexity of doing this is $d^{O(l+\omega)}$.

Now using the Inverse Sign Determination Subroutine [6] (section 6.1, page 1041) compute for every list of Thom encodings (say L) obtained above a formula, $F_L(T_2, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$ such that for all $(t_2, \ldots, t_l, y_{k_1}, \ldots, y_{k_\omega})$ satisfying F_L the list of Thom encodings of the roots of $P_i(T_1, t_2, \ldots, t_l, y_{k_1}, \ldots, y_{k_\omega})$ is L.

Finally, we can construct the predicates $s_{ij}(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}), 1 \leq i \leq 3s, 1 \leq j \leq d$, such that $s_{ij}(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$ is true iff T_1 is the j^{th} largest root of $Q_i(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$ considered as a polynomial in T_1.

We compute the $s_{ij}(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})$ from the F_L's in the obvious way.

Each s_{ij} will be of the form, $s_{ij} = \wedge_L(F_L \wedge \sigma_L)$, where σ_L is the Thom encoding of the jth root encoded by the list L. Note that F_L is of size $d^{O(l+\omega)}$ and σ is of the form, $(Q_i = 0) \wedge Q_i'\{>,=,<\}0, \ldots, Q_i^{(deg(Q_i))}\{>,=,<\}0$ where all the derivatives are with respect to T_1.

The μ_j's are now computed by separately eliminating T_1 from each quantified formula of the form,

$$(\exists T_1) P_i(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega})\{>,=,<\}0 \wedge$$

$$s_{jk}(T_1, \ldots, T_2, Y_{i_1}, \ldots, Y_{i_\omega})$$

and substituting the quantifier-free formula so obtained in place of the original atoms $P_i\{>,=,<\}0$.

The total complexity of this last step is $s^2 d^{O(l+\omega)}$.

We can now bound the complexity of the algorithm described above. From the algorithm it is clear that $M = O(sd)$. Thus the size of the output formula ψ is bounded by $(s^2 d^{O(l+\omega)})|\phi|$. Also the complexity of the algorithm is $s^2 d^{O(l+\omega)}$.

$\diamond$

We can generalize the above result for elimination of blocks of quantifiers rather than a single one. We have the following result which we state here without proof.

Theorem 2 *Let,* $\Phi = \{\phi_n(Y_1, \ldots, Y_n, T_1, \ldots, T_l) \mid n > 0\}$ *be a uniform sequence of formulas with parameters* $T_1, \ldots, T_l$, *where*

$$\phi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n) =$$

$$Q^1_{1 \leq k_1 \leq n} \ldots Q^\omega_{1 \leq k_\omega \leq n} \phi(T_1, \ldots, T_l, Y_{k_1}, \ldots, Y_{k_\omega}).$$

Let the number of different $(l + \omega)$*-variate polynomials appearing in* ϕ *be* s *and let their degrees be bounded by* d.

Let $R_1, \ldots, R_m \in \{\exists, \forall\}$, $R_i \neq R_{i+1}$, *and let* $T^{[1]}, \ldots, T^{[m]}$ *be a partition of the variables,* $T_1, \ldots, T_l$ *into* m *blocks of size* $l1, \ldots, l_m$, *wehere* $\sum_{1 \leq i \leq m} l_i = l$.

Then, there exists an algorithm that outputs a quantifier-free first order formula,

$$\psi(Y_{k_1}, \ldots, Y_{k_{\omega'}}),$$

along with $Q^i \in \{\bigvee, \bigwedge\}, 1 \leq i \leq \omega'$, *such that for every* $n > 0$,

$$\psi_n(Y_1, \ldots, Y_n) =$$

$$Q^1_{1 \leq k_1 \leq n} \ldots Q^{\omega'}_{1 \leq k_{\omega'} \leq n} \psi(Y_{k_1}, \ldots, Y_{k_{\omega'}}),$$

$$\equiv (R_1 T^{[1]}) \ldots (R_m T^{[m]})(\exists T_1)\phi_n(Y_1, \ldots, Y_n, T_1, \ldots, T_l).$$

The complexity of the algorithm is bounded by $s^{\Pi_i(l_i+1)} d^{\omega \Pi_i O(l_i^2)}$, *and the size of the formula* ψ *is bounded by* $s^{\Pi_i(l_i+1)} d^{\omega \Pi_i O(l_i^2)}|\phi|$.

5 Natural to Active Domain Semantics

Our algorithm for uniform quantifier elimination can be used to give an algorithm for converting a query with natural domain semantics to an equivalent one with active domain semantics.

Consider a query σ with natural domain semantics. Let

$$\sigma = (Q_1 T_1)(Q_2 T_2) \ldots (Q_l T_l)\psi(T_1, \ldots, T_l),$$

where each $Q_i \in \{\exists, \forall\}$, where ψ is a L'-formula.

For every positive integer n we obtain a L-formula $\sigma_n(Y_1, \ldots, Y_n)$ with n free variables $Y_1, \ldots, Y_n$ from σ by replacing every occurence of $R(T_j)$ in ψ by the disjunct,

$$R_n(T_j, Y) = \bigvee_{1 \leq i \leq n} (T_j = Y_i),$$

for $1 \leq j \leq l$.

Let,

$$\sigma_n(Y_1, \ldots, Y_n) =$$

$$(Q_1 T_1)(Q_2 T_2) \ldots (Q_l T_l)\psi_n(T_1, \ldots, T_l, Y_1, \ldots, Y_n).$$

It is clear that $\{\psi_n \mid n > 0\}$ is a uniform sequence of formulas.

Using theorem 2 we can eliminate the quantifiers from σ_n, to obtain formulas, ψ_n.

Now, let

$$\psi_n(Y_1, \ldots, Y_n) = Q^1_{1 \leq i_1 \leq n} \cdots Q^k_{1 \leq i_k \leq n} \psi(Y_{i_1}, \ldots, Y_{i_k}),$$

$Q^j \in \{\bigvee, \bigwedge\}$.

We syntactically replace every occurence of $\bigvee_{1 \leq i_j \leq n}$ by $(\exists T_j)$ and replace every occurence of $\bigwedge_{1 \leq i_j \leq n}$ by $(\forall T_j)$, and replace Y_{i_j} by the variable T_j.

Let the L'-sentence obtained in this manner from the sequence of formulas ψ_n^0 be denoted by Ψ. It is clear that Ψ with active domain semantics is equivalent to ϕ with natural domain semantics.

Let the total number of polynomials appearing in σ be s and their degrees be bounded by d. Moreover, let the number of blocks of quantifiers be m with the i^{th} block of length l_i, with $\sum_i l_i = l$. Then from the complexity of the above algorithm is $s^{\Pi_i(l_i+1)} d^{\omega \Pi_i O(l_i^2)}$. This follows from the complexity bound in theorem 2.

6 Parity

We next use theorem 1 to prove that parity is not expressible by a first order sentence over the reals.

First we choose a real closed field $\mathcal{R}$, with elements $\Omega_1, \Omega_2, \ldots, \Omega_n$, where Ω_1 is larger than all the integers, Ω_2 is larger than all powers of Ω_1 and so on. We can prove that such models exist by the compactness theorem. However, to be more specific we can consider the field, $R\langle\Omega_1\rangle\langle\Omega_2\rangle \cdots \langle\Omega_n\rangle$, where R is any real closed field (say the reals) and for every i, $R\langle\Omega_1\rangle\langle\Omega_2\rangle \ldots \langle\Omega_i\rangle$ is the field of Puiseux series in $\frac{1}{\Omega_i}$ with coeffecients in $R\langle\Omega_1\rangle\langle\Omega_2\rangle \ldots \langle\Omega_{i-1}\rangle$. It is well known that the resulting field is real closed [1].

Then, for any l-ary L-formula, ψ, $\psi(Y_{i_1}, \ldots, Y_{i_l})$ is equivalent to a finite Boolean formula whose atoms are of the form, $Y_j \leq Y_k, j,k \in \{i_1,\ldots,i_l\}$, for the Y_j's restricted to the set $\{\Omega_1,\ldots,\Omega_n\}$. This is clear, because the sign of any polynomial $P(\Omega_{i_1},\ldots,\Omega_{i_l})$ with integral coefficients is determined only by the ordering of the i_j's.

We now prove that the parity is not expressible over real closed fields. Note that it suffices to prove this for any one real closed field, as we can then deduce this for all other real closed fields by the transfer principle.

Theorem 3 *Parity is not expressible over real closed fields.*

Proof: Assume that there exists a first order sentence σ expressing parity.

Let

$$\sigma = (Q_1 T_1)(Q_2 T_2)\ldots(Q_l T_l)\psi(T_1,\ldots,T_l),$$

where each $Q_i \in \{\exists,\forall\}$.

For every positive integer n we obtain a L-formula $\sigma_n(Y_1,\ldots,Y_n)$ with n free variables $Y_1,\ldots,Y_n$ from σ by replacing every occurence of $R(T_j)$ in ψ by the disjunct,

$$R_n(T_j,Y) = \bigvee_{1\leq i\leq n}(T_j = Y_i),$$

for $1 \leq j \leq l$.

Let,

$$\sigma_n(Y_1,\ldots,Y_n) =$$
$$(Q_1T_1)(Q_2T_2)\ldots(Q_lT_l)\psi_n(T_1,\ldots,T_l,Y_1,\ldots,Y_n).$$

It is clear that $\{\psi_n \mid n > 0\}$ is a uniform sequence of formulas.

Using theorem 2 we can eliminate the quantifiers from σ_n to obtain a uniform sequence $\phi_n(Y_1,\ldots,Y_n)$.

Now $(y_1,\ldots,y_n)$ satisfy ϕ_n iff $|\bigcup_{1\leq i\leq n}\{y_i\}|$ is even.

Let,

$$\phi_n(T_1,\ldots,T_l,Y_1,\ldots,Y_n) =$$
$$Q^1_{1\leq k_1\leq n}Q^2_{1\leq k_2\leq n}\ldots Q^{\omega}_{1\leq k_\omega\leq n}\phi(T_1,\ldots,T_l,Y_{k_1},\ldots,Y_{k_\omega}).$$

Now consider the set, $S = \{\Omega_1,\ldots,\Omega_n\} \subset \mathcal{R}$. If we restrict our attention to S^n, then in ϕ_n we can replace the predicate $\phi(Y_{i_1},\ldots,Y_{i_k})$ by the another k-ary predicate with atoms of the form $Y_j \leq Y_k, j,k \in \{i_1,\ldots,i_k\}$ as in the preceding paragraph. Call this formula, $\phi'_n(Y_1,\ldots,Y_n)$. Note that, for $(y_1,\ldots,y_n) \in S^n$, $\phi'_n(y_1,\ldots,y_n)$ is satisfied iff $|\bigcup_{1\leq i\leq n}\{y_i\}|$ is even. However, ϕ'_n contains no additions or multiplications. This would imply that parity is expressible in $\langle N,\leq\rangle$ (i.e in a discrete order without end-points) which is not true (see [14]).

$\diamond$

7 Concluding Remarks

In this paper we have introduced a new variant of the quantifier elimination problem which we call *uniform quantifier elimination*. We give an algorithm for performing uniform quantifier elimination and analyse its complexity. We next give applications of the uniform quantifier elimination algorithm in constraint database theory.

Finally, we note that our uniform quantifier elimination algorithm was based only on the locality property of real closed fields. Two key properties being used for proving the locality property for the reals are the fact that definable subsets only have a finite number of connected components, and that if there is a common point for a collection of such subsets there is a common point to be found at an extremity.

There are analogous properties for certain other theories admitting quantifier elimination. We discuss two of them.

The definable subsets of an algebraically closed field of char 0 are either finite sets of points or complements of finite sets of points. These are also called the Zariski constructible sets. Moreover, if it is a finite set of points the number of points is again uniformly bounded. If a definable subset is a finite set of points we let these points be the witness points. Given a collection of definable subsets, if all of them are infinite then they always have a common point. If, however, some of them are finite then if there is at all a common point it must be one of these points.

The theory of p-adic fields is another interesting theory admitting quantifier elimination. The definable subsets (see [7] for terminology and details) are unions of a finite collection of finite cells and possibly an infinite cell. The number of finite cells is again uniformly bounded by the parameters of the defining formula. Moreover, if two finite cells intersect, one contains the other. Again, by choosing witness points in the finite cells we can ensure the locality property.

Thus the same technique can be used to prove the conjecture for a wider class of theories which admit quantifier elimination.

8 Acknowledgement

The author would like to thank an anonymous referee for comments that helped to improve the presentation of the paper considerably.

References

[1] J. Bochnak, M. Coste, M.-F. Roy Géométrie algébrique réelle. Springer-Verlag (1987).

[2] M. Benedikt, G. Dong, L. Libkin, L. Wong Relational Expressive Power of Constraint Query Languages *Proc. of 15th ACM Symposium on Principles of Database Systems, 1996.*

[3] M. Ben-Or, D. Kozen , J. Reif The complexity of elementary algebra and geometry. *J. of Computer and Systems Sciences*, 32:251– 264, (1986).

[4] M. Benedikt, L. Libkin On the Structure of Queries in Constraint Query Languages *Proc. of 11th Annual IEEE Symposium on Logic in Computer Science, 1996.*

[5] M. Benedikt and L. Libkin Relational Languages over Interpreted Structure *Proceedings of PODS, 1997* (to appear).

[6] S. Basu, R. Pollack, M.-F. Roy On the combinatorial and algebraic complexity of Quantifier Elimination, Journal of the ACM, Nov 1996, Vol 43, Number 6, 1002-1046. (extended abstract appeared in *Proc. 35th IEEE Symp. on Foundations of Computer Science*, 632-642, (1994).)

[7] P. J. Cohen *Decision Procedures for Real and p-Adic Fields,* Comm. of Pure and Applied Mathematics, Vol. XXII, 131-151 (1969).

[8] A. EHRENFEUCHT *An application of games to the completeness problem for formalized theories,* Fund. Math, 49, 1961.

[9] R. FAGIN *Probabilities on finite models,* Journal of Symbolic Logic, 41(1):50-58, 1976.

[10] R. FAGIN *Finite model theory - a personal perspective,* Theoretical Computer Science, 116(1)3-31, 1993.

[11] H. GAIFMAN *On local and non-local properties,* Proc. Herbrand Symposium Logic Colloquium, 105-35, North Holland, 1981.

[12] D. GRIGOR'EV *The Complexity of deciding Tarski algebra* Journal of Symbolic Computation 5 (1988), 65-108.

[13] Y. GUREVICH *Logic and the challenge of computer science,* Current Trends in Theoretical Computer Science, 1-57, Computer Science Press, 1988.

[14] S. GRUMBACH, J. SU Queries with Arithmetical Constraints, *Proc. of Int. Conf. on Principles and Practice of Constraint Programming, 1995.*

[15] S. GRUMBACH, J. SU Finitely Representable databases, *Proc. of 13th ACM Symposium on Principles of Database Systems, 1994.*

[16] S. GRUMBACH, J. SU Dense Order Constraint Databses, *Proc. of 14th ACM Symposium on Principles of Database Systems, 1995.*

[17] P. KANELLAKIS, D. Q. GOLDIN Constraint Programming and Database Query Languages, *Proc. 2nd Conf on Theoretical Aspects of Computer Software (TACS),* LNCS Vol 789, Springer-Verlag, 1994.

[18] P. KANELLAKIS, G. KUPER, P. REVESZ Constraint Query Languages, *Proc. of 9th ACM Symposium on Principles of Database Systems, 1990.*

[19] J. PAREDAENS, J. VAN DEN BUSSCHE, D. VAN GUCHT First Order Queries over finite structures over the reals, *Proc. IEEE Symposium on Logic In Computer Science, 1996.*

[20] J. RENEGAR *On the computational complexity and geometry of the first order theory of the reals*, Journal of Symbolic Computation, (1992).

[21] A. TARSKI *A Decision method for elementary algebra and geometry*, University of California Press (1951).

[22] J.D. ULLMAN *Principles of Database Systems*, Computer Science Press, 1983.

[23] V. WEISPFENNING *The complexity of linear problems in fields,* Journal of Symbolic Computation 5(1988).

[24] V. WEISPFENNING Quantifier elimination for real algebra - the quadratic case and beyond, AAECC 8(1997).