# A Collapse Result for Constraint Queries over Structures of Small Degree

Leonid Libkin*

**Abstract**

Collapse results, which are central for understanding constraint database queries, show that in terms of the expressive power, a large class of queries collapses to a much smaller one, typically involving only a restricted form of quantification. Most collapse results have been proved over constraints involving a linear order, and proofs are typically rather nontrivial. In this short note we give an easy proof of a powerful form of collapse for a large class of constraints without a linear order, namely those in which all basic relations are of small degree.

## 1 Introduction

A typical setting for database constraint queries is as follows [5]. We have an infinite first-order structure $\mathcal{M} = \langle U, \Omega \rangle$ with $U$ being the universe, and $\Omega$ a set of constants, predicates and functions on $U$. Furthermore, we have a relational vocabulary $\sigma$ whose symbols are interpreted as finite relations; the intended interpretation of those is finite database relations whose elements come from $U$. The standard query language in this setting is $\mathrm{FO}(\mathcal{M}, \sigma)$ – first-order logic in the language $\Omega \cup \sigma$.

Consider, for example, $\mathcal{M} = \langle \mathbb{R}, +, \cdot, 0, 1, < \rangle$, $\sigma = \{S\}$, where $S$ is binary, and a Boolean query (sentence)

$$\Phi \quad = \quad \exists a \exists b \; \forall x \forall y \; (S(x, y) \rightarrow a \cdot x + b = y),$$

saying that $S$ lies on a line. It is not immediately clear how to evaluate such a query due to the presence of quantifiers of the form $\exists a$, $\exists b$, meaning: $\exists a \in \mathbb{R}$, $\exists b \in \mathbb{R}$.

However, often one can restrict quantification to a finite set. By the *active domain* of a finite $\sigma$-structure $D$ we mean the (finite) set $adom(D)$ of all elements of $U$ that occur in $D$. When we write $\exists x \in adom$ or $\forall y \in adom$, we mean that quantification is restricted to $adom(D)$. For example, the above query can be written in such a form: note that $S$ lies on a line iff every triple of points in $S$ is collinear. It is easy to write a formula $\mathrm{collinear}(x_1, y_1, x_2, y_2, x_3, y_3)$ over $\langle \mathbb{R}, +, \cdot, 0, 1, < \rangle$ which holds iff $(x_1, y_1), (x_2, y_2)$ and $(x_3, y_3)$ are collinear. Hence, $\Phi$ is equivalent to

$$\Phi' \;=\; \forall x_1, y_1, x_2, y_2, x_3, y_3 \in adom \; (S(x_1, y_1) \wedge S(x_2, y_2) \wedge S(x_3, y_3)) \;\rightarrow\; \mathrm{collinear}(x_1, y_1, x_2, y_2, x_3, y_3).$$

Notice that $\sigma$-relations occur only in the scope of restricted quantifiers $\forall x \in adom$ and $\exists x \in adom$. Hence, if the theory of the underlying structure $\mathcal{M}$ is decidable, such queries can be evaluated in the

*Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, Ontario M5S 3H5, Canada. E-mail: libkin@cs.toronto.edu.

usual bottom-up fashion, as normal relational calculus database queries (at least if all the elements in the database are definable by formulae over $\mathcal{M}$). Since $\langle \mathbb{R}, +, \cdot, 0, 1, < \rangle$ has quantifier-elimination, in our example unrestricted quantifiers ranging over $\mathbb{R}$ can be eliminated altogether. We shall consider this type of collapse results in this note.

## 2 Definitions and Main Theorem

We shall refer to quantification over $U$ as *unrestricted* quantification, and to quantification over $adom(\cdot)$ as *restricted* quantification.

We say that $\mathcal{M}$ admits *restricted quantifier collapse (RQC)* if for any $\sigma$, every FO($\mathcal{M}, \sigma$) formula is equivalent to a formula in which no $\sigma$-symbol appears in the scope of an unrestricted quantifier.

If $\mathcal{M}$ has quantifier-elimination, RQC implies that all unrestricted quantifiers can be eliminated – this is a very strong condition known as the natural-active collapse. It was shown in [6] for the additive group of the reals; since then, many extensions and other examples appeared in the literature (see [5] for a survey).

Assume without loss of generality that only relation symbols appear in $\Omega$ (we can replace functions by their graphs, and constants by unary relations having one element). Given $\mathcal{M} = \langle U, \Omega \rangle$, its *Gaifman graph* $\mathcal{G}(\mathcal{M})$ is an undirected graph $(U, E)$, such that $(a, b) \in U$ iff there is an $\Omega$-relation $R$ such that both $a$ and $b$ occur in the same tuple of $R$ (for example, if $\mathcal{M}$ itself is a graph, then $\mathcal{G}(\mathcal{M})$ is its reflexive-symmetric closure). We say that $\mathcal{M}$ is of small degree if there is a constant $d \in \mathbb{N}$ such that every $a \in U$ has degree at most $d$ in $\mathcal{G}(\mathcal{M})$.

From results in [1], one can conclude the collapse over such structures only for *generic* queries: those that commute with any permutation on $U$ (for example, $\Phi$ shown above is *not* generic). In general, checking if a query is generic is undecidable. Here we prove a general result on the expressiveness of FO($\mathcal{M}, \sigma$) for structures of small degree.

**Theorem 1** *Every structure of small degree admits the restricted quantifier collapse.* □

Any structure that has a linear order, or makes it possible to define one, is not of small degree. However, there are important examples of structures of small degrees, most often involving strings. One such example is the set $\Sigma^*$ of all finite strings over a finite alphabet $\Sigma$ considered as term algebras: that is, there is a constant $\epsilon$ for the empty string, and functions $x \mapsto x \cdot a, a \in \Sigma$ that add alphabet symbols at the end of a string. Since we deal with relational signatures, a proper representation of this structure is $\langle \Sigma^*, U_\epsilon, (L_a)_{a \in \Sigma} \rangle$, where $U_\epsilon$ is interpreted as the singleton-set $\{\epsilon\}$, and $L_a = \{(a \cdot x, x) \mid x \in \Sigma^*\}$. This structure is of degree $|\Sigma| + 1$. Another example is a model-theoretic representation of queues used in [7]: a structure $\langle \Sigma^*, U_\epsilon, (L_a, R_a)_{a \in \Sigma} \rangle$, where $R_a = \{(x \cdot a, x) \mid x \in \Sigma^*\}$ (that is, symbols could be added both on the left and on the right).

In general, a first-order theory $T$ has the RQC if every model of the theory has it.

**Corollary 1** *Let $T$ be a complete theory that has a model of small degree. Then $T$ has the RQC.*

As another corollary to the proof, we shall see the following. We say that a function $P : 2^U \to 2^U$ is $\mathcal{M}$-definable if there is a first-order formula $\alpha_P(x, y)$ such that $P(X) = \{b \mid \mathcal{M} \models \alpha_P(a, b), a \in X\}$.

**Corollary 2** *If $\mathcal{M}$ is of small degree, then every* $\mathrm{FO}(\mathcal{M},\sigma)$ *sentence* $\Phi$ *is equivalent to a sentence* $\Phi'$ *in which all quantifiers range over* $P_\Phi(adom(D))$*, where* $P_\Phi$ *is* $\mathcal{M}$*-definable. Furthermore, there exist a constant* $c$ *that depends on* $\Phi$ *only such that* $|P_\Phi(X)| \leq c \cdot |X|$ *for any finite* $X \subset U$. □

That is, it suffices to restrict quantification to a finite set whose size is linear in $|adom(D)|$.

# 3 Proof of the theorem

Fix $\mathcal{M}$ of degree at most $d$. Given $\vec{a}$ and $b$ over $U$, by $\delta(\vec{a}, b)$ we mean the shortest distance, in $\mathcal{G}(\mathcal{M})$, between $b$ and an element of $\vec{a}$. Let $B_r(\vec{a})$ be the ball of radius $r$ around $\vec{a}$, that is, $\{b \mid \delta(\vec{a}, b) \leq r\}$. The $r$-neighborhood of $\vec{a}$ is the substructure of $\mathcal{M}$ whose universe is $B_r(\vec{a})$; it is denoted by $N_r(\vec{a})$. For $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$, we write $N_r(\vec{a}) \cong N_r(\vec{b})$ if there exists an isomorphism $h : N_r(\vec{a}) \to N_r(\vec{b})$ such that $h(a_i) = b_i, 1 \leq i \leq n$.

Note that there is an upper bound $M(d, r)$ on the size of $B_r(a)$ that is determined by $r$ and $d$ only. Consequently, there is an upper bound $I(d, r, m)$ on the number of isomorphism types of $N_r(\vec{a})$, $\vec{a}$ of length $m$, that depends on $d, r$ and $m$ only.

We will use the following result.

**Fact 1 (Gaifman [3])** *For every formula* $\varphi(\vec{x})$ *over* $\mathcal{M}$ *there exists a number* $r \geq 0$ *such that* $N_r(\vec{a}) \cong N_r(\vec{b})$ *implies* $\mathcal{M} \models \varphi(\vec{a}) \leftrightarrow \varphi(\vec{b})$.

Next, consider any infinite set $X \subseteq U$, and any formula $\varphi(x, y)$, and let $r$ be given by Gaifman's theorem. Since $d$ is fixed, $X$ is infinite, and the number of isomorphism types of $r$-neighborhoods is bounded by $I(d, r, 1)$, there are $a, a' \in X$ such that $N_r(a) \cong N_r(a')$ and $\delta(a, a') > 2r + 1$. Hence, $N_r(a, a') \cong N_r(a', a)$ and $\varphi(a, a') \leftrightarrow \varphi(a', a)$, which shows that no formula can define a linear ordering on an infinite set (that is, the theory of $\mathcal{M}$ is stable [4]).

We now use a sufficient condition for RQC, given in [2]. We say that $\mathcal{M}$ has the *finite cover property* if there is a formula $\varphi(x, \vec{y})$ such that for every $n$, one can find $n$ tuples $\vec{a}_1, \ldots, \vec{a}_n$ such that $\exists x \bigwedge_{i \neq j} \varphi(x, \vec{a}_i)$ holds for every $j \leq n$, but $\exists x \bigwedge_i \varphi(x, \vec{a}_i)$ does not hold.

We shall use the following result.

**Fact 2 (Flum-Ziegler [2])** *If* $\mathcal{M}$ *does not have the finite cover property, then it has the RCQ.*

To show that $\mathcal{M}$ does not have the finite cover property, we need the following, due to [8]. Given a formula $\alpha(\vec{x}, \vec{y})$ and a tuple of constants $\vec{b}$, by $\alpha(\cdot, \vec{b})$ we mean $\{\vec{a} \mid \alpha(\vec{a}, \vec{b}) \text{ holds}\}$. A formula $\varphi(x, y, \vec{z})$ is an *equivalence* formula if for every $\vec{b}$, $\varphi(\cdot, \cdot, \vec{b})$ is an equivalence relation. The result that we need, applied to our setting (that is, in the case of stable theories), says the following.

**Fact 3 (Shelah [8])** *Assume that* $\mathcal{M}$ *has the following property: for every equivalence formula* $\varphi(x, y, \vec{z})$*, there is a number* $l$ *that depends on* $\varphi$*, such that if* $\varphi(\cdot, \cdot, \vec{b})$ *is of finite index for every* $\vec{b}$*, then, for every* $\vec{b}$*,* $\varphi(\cdot, \cdot, \vec{b})$ *has at most* $l$ *equivalence classes. Then* $\mathcal{M}$ *does not have the finite cover property.*

Thus, it remains to prove the uniform bounds on the number of equivalence classes of definable equivalence relations of finite index. Consider an equivalence formula $\varphi(x, y, \vec{z})$ with $\vec{z}$ of length $m$. Let $r$ be given by Fact 1. Fix $\vec{c}$, and let $\approx_{\vec{c}}$ be the equivalence relation given by $\varphi(\cdot, \cdot, \vec{c})$. Assume that it is of finite index. We calculate an upper bound on the number of equivalence classes of $\approx_{\vec{c}}$ that are disjoint from $B_{2r+1}(\vec{c})$. For those equivalence classes, whether $a \approx_{\vec{c}} b$, is determined by the isomorphism type of the neighborhood $N_r(a, b)$.

We call an equivalence class sparse if there are two elements, $a$ and $b$ in it, such that $\delta(a, b) > 2r + 1$. Let $K_1, \ldots, K_s$ be all the distinct sparse equivalence classes of $\approx_{\vec{c}}$, and let $(a_i, b_i)$ be a pair of elements from $K_i$ with $\delta(a_i, b_i) > 2r + 1$. Let $l$ be $M(d, 2r + 1)$, the upper bound on the size of $B_{2r+1}(e)$, for a single element $e$. Assume that $s > I(d, r, 2) \cdot (l + 1)$. Then for $l + 1$ classes (without loss of generality, $K_1, \ldots, K_{l+1}$), $(a_i, b_i)$ realize the same isomorphism type of an $r$-neighborhood, for $i \leq l + 1$. In particular, for some $i_0 \leq l + 1$, $\delta(a_1, b_{i_0}) > 2r + 1$, by definition of $l$. Since $\delta(a_1, b_1) > 2r + 1$, $\delta(a_{i_0}, b_{i_0}) > 2r + 1$, and $N_r(a_1, b_1) \cong N_r(a_{i_0}, b_{i_0})$, we have $N_r(a_1, b_1) \cong N_r(a_1, b_{i_0})$. Hence $a_1 \approx_{\vec{c}} b_{i_0}$ and thus $K_1 = K_{i_0}$, which is impossible. This contradiction shows that $s \leq I(d, r, 2) \cdot (M(d, 2r+1) + 1)$.

Now consider nonsparse equivalence classes. Since $\approx_{\vec{c}}$ is of finite index, there are finitely many of them, say $C_1, \ldots, C_t$. Let $a_1, \ldots, a_t$ be representatives of those classes. We know that $C_i \subseteq B_{2r+1}(a_i)$, $i = 1, \ldots, t$.

Let $\tau$ be the isomorphism type of one of $N_{2r+1}(a_i)$s (without loss of generality, take $N_{2r+1}(a_1)$). Assume that there are infinitely many points that realize $\tau$ (that is, whose $(2r + 1)$-neighborhoods are isomorphic to $N_{2r+1}(a_1)$). Since $\mathcal{M}$ is infinite and of small degree, we can find an infinite sequence of elements $e_1, e_2, \ldots$ such that:

1. $B_{2r+1}(e_i)$s are pairwise disjoint, and all of them are disjoint from $B_{2r+1}(a_1)$; and
2. the equivalence class of $e_i$ is not contained in $B_{2r+1}(e_i)$.

In particular, for each $e_i$ we can find $p_i \notin B_{2r+1}(e_i)$ such that $e_i \approx_{\vec{c}} p_i$. If $p_i \notin B_{2r+1}(a_1)$, then $N_r(e_i, p_i) \cong N_r(a_1, p_i)$, thus showing $p_i \approx_{\vec{c}} a_1$, which is impossible since the $\approx_{\vec{c}}$-equivalence class of $a_1$ is contained in $B_{2r+1}(a_1)$. Thus, we can assume that $p_i \in B_{2r+1}(a_1)$ for all $i$. Then there exist two indexes $i, j$ such that $p_i = p_j$, and consequently $N_r(e_i, p_i) \cong N_r(e_j, p_j)$. This shows $e_i \approx_{\vec{c}} p_i \approx_{\vec{c}} p_j \approx_{\vec{c}} e_j$. Since $N_r(e_i, a_1) \cong N_r(e_i, e_j)$ we conclude $e_i \approx_{\vec{c}} a_1$, which again contradicts the assumption that the $\approx_{\vec{c}}$-equivalence class of $a_1$ is contained in $B_{2r+1}(a_1)$.

This contradiction shows that the number of points that realize $\tau$ is finite. Let $\tau_1, \ldots, \tau_{I(d,2r+1,1)}$ enumerate all the isomorphism types of $2r + 1$-neighborhoods of a single point, with $\tau_1, \ldots, \tau_q$ being those with finitely many elements realizing them. Suppose the number of elements realizing $\tau_i$, $i \leq q$, is $m_i$; then $m_1 + \ldots + m_q$ is the total number of points realizing $2r + 1$-neighborhood with finitely many realizers. Since we chose representatives of classes $C_1, \ldots, C_t$ arbitrarily, this gives us a $(m_1 + \ldots + m_q) \cdot M(d, 2r + 1)$ upper bound for $t$.

Summing up, the number of equivalence classes of $\approx_{\vec{c}}$ that do not intersect $B_{2r+1}(\vec{c})$ is at most

$$I(d, r, 2) \cdot M(d, 2r + 1) \ + \ (m_1 + \ldots + m_q) \cdot M(d, 2r + 1),$$

and the number of classes that intersect $B_{2r+1}(\vec{c})$ is at most $M(d, 2r+1) \cdot m$. Hence, we have an upper bound on the number of equivalence classes of $\approx_{\vec{c}}$ that depends on $\varphi$ and $\mathcal{M}$ only, but not on $\vec{c}$. This concludes the proof.

It remains to prove Corollaries 1 and 2. For Corollary 1, note that for each $d$, there is a first-order sentence saying that $\mathcal{M}$ is of degree at most $d$; hence if $T$ is complete and has a model of degree $d$, all its models are such. This shows that the bounds $M(d,r)$ and $I(d,r,m)$ are parameters of the theory, not a particular model. Furthermore, since $d$ is fixed, for each $r, k \geq 0$ and each isomorphism type $\tau$ of an $r$-neighborhood, there are first-order sentences stating that are are fewer than $k$ (exactly $k$, more than $k$) elements realizing $\tau$. Hence, if such a number is finite, it is the same in every model of $T$, and likewise if there are infinitely many realizers, this is too true in every model of $T$. This shows that the uniform bound on the number of equivalence classes of $\varphi(\cdot,\cdot,\vec{z})$ will be the same in every model of $T$ (that is, it depends only on $\varphi$), thus showing that every model of $T$ has the RQC.

For Corollary 2, take a sentence $\Phi$ and apply the RQC. Let $\Phi'$ be the resulting sentence. We look at the subformulae $\alpha(\vec{x})$ not involving restricted quantification. Note that all the variables $\vec{x}$ are bound by restricted quantifiers (hence, they range over the active domain). By [3] we can assume that quantification in $\alpha$ is restricted to the $r_\alpha$-neighborhood of its free variables, for some fixed $r_\alpha$. Choose the maximum $r$ among $r_\alpha$'s; clearly it is determined by $\Phi$. Then quantification in $\Phi'$ can be assumed to be over $B_r(adom(D))$, which is of the size at most $I(d,r,1) \cdot |adom(D)|$. This completes the proof. $\square$

We conclude by offering a couple of remarks on the effectiveness of the collapse result. In general, the mere existence of an equivalent query with restricted quantification does not imply that such a query can be found effectively, and some of the collapse results (e.g., over the real field) were first proved non-constructively, and later constructive proofs were found, see [5]. The restriction to structures of small degree does not change this general situation. Our proof relies heavily on a theorem by Flum and Ziegler [2], which was proved non-constructively. In some case, the RQC cannot be made effective even with small degrees. To see this, let $\pi : \mathbb{N} \to \mathbb{N}$ be a nonrecursive bijection; we then define a relation $E = \{(2i, 2\pi(i)+1) \mid i > 0\}$ of degree 1. Even though $\langle \mathbb{N}, E\rangle$ admits the RQC, it cannot be made effective. As another example, the collapse over $\langle \Sigma^*, U_\epsilon, (R_a)_{a\in\Sigma}\rangle$ is effective, which can be shown by a tedious inductive proof that calculates the size of the neighborhood to which quantification is restricted. In general, whether the RQC is effective, depends on every particular structure, and the general proof presented here is unlikely to help answer this question.

**Acknowledgment** Thanks to Michael Benedikt and Markus Junker for discussions on the finite cover property, unary term algebras, and collapse, that gave rise to this note.

# References

[1] J. Baldwin, M. Benedikt. Embedded finite models, stability theory and the impact of order. *LICS 1998*, pages 490–500.

[2] J. Flum and M. Ziegler. Pseudo-finite homogeneity and saturation. *J. Symbolic Logic* 64 (1999), 1689–1699.

[3] H. Gaifman. On local and non-local properties. In *Proceedings of the Herbrand Symposium, Logic Colloquium '81*, North-Holland, 1982.

[4] W. Hodges. *Model Theory.* Cambridge, 1993.

[5] G. Kuper, L. Libkin, and J. Paredaens (editors). *Constraint Databases.* Springer-Verlag, 2000.

[6] J. Paredaens, J. Van den Bussche, and D. Van Gucht. First-order queries on finite structures over the reals. *SIAM J. Comput.* 27 (1998), 1747–1763.

[7] T. Rybina, A. Voronkov. A decision procedure for term algebras with queues. *ACM TOCL* 2 (2001), 155–181.

[8] S. Shelah. *Classification Theory.* North-Holland, 1990.