Shrinking Games and Local Formulas

H. Jerome Keisler

Dept. of Mathematics, University of Wisconsin, Madison WI 53706

Wafik Boulos Lotfallah

Dept. of Engineering. Math. and Physics, Cairo University, Cairo, Egypt 11451

Abstract

Gaifman's normal form theorem showed that every first order sentence of quantifier rank n is equivalent to a Boolean combination of "scattered local sentences", where the local neighborhoods have radius at most 7^{n-1}. This bound was improved by Lifsches and Shelah to $3 \cdot 4^{n-1}$. We use Ehrenfeucht-Fraïssé type games with a "shrinking horizon" to get a spectrum of normal form theorems of the Gaifman type, depending on the rate of shrinking. This spectrum includes the result of Lifsches and Shelah, with a more easily understood proof and with the bound on the radius improved to 4^{n-1}. We also obtain bounds for a normal form theorem of Schwentick and Barthelmann.

Key words: Ehrenfeucht-Fraïssé games, finite model theory, local formulas, quantifier rank

1 Introduction

Gaifman [Gai82] proved a normal form theorem for first order sentences using formulas which are local with respect to the path length connecting elements by atomic formulas. This theorem states that every first order sentence of quantifier rank n is equivalent to a finite Boolean combinations of "scattered local sentences" saying that there exist disjoint local neighborhoods of some first order type, with radius at most 7^{n-1}. This bound on the radius was improved by Lifsches and Shelah [LS96] to $3 \cdot 4^{n-1}$.

Email addresses: keisler@math.wisc.edu (H. Jerome Keisler), lotfalla@uwalumni.com (Wafik Boulos Lotfallah).

[1] Supported in part by the Vilas Trust Fund

Since the local neighborhoods are disjoint in Gaifman's theorem, the centers of the neighborhoods are separated by at least twice the radius. It is natural to ask whether there are other normal form theorems with different relationships between the radii of the neighborhoods and the distances separating their centers.

In this paper we use Ehrenfeucht-Fraïssé type games with a shrinking horizon between two relational structures to obtain a spectrum of normal form theorems of the Gaifman type, which roughly correspond to the rate of shrinking. Taking one particular level of this spectrum, we get a more easily understood proof of the result of Lifsches and Shelah, while slightly improving the bound to 4^{n-1}. We also apply the shrinking games to get bounds for another normal form theorem which was proved by Schwentick and Barthelmann [SB99] as a consequence of Gaifman's theorem.

Shrinking Ehrenfeucht-Fraïssé type games were also applied in another direction by Schwentick [Sch96].

2 Basic Definitions

We fix a finite relational vocabulary ν. $\mathcal{A}, \mathcal{B}$ will always stand for ν-structures, $\mathbf{a}$, $\mathbf{b}$ will always stand for finite sequences in $\mathcal{A}$, $\mathcal{B}$, respectively, and $(\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b})$ will stand for the structures with distinguished elements. $\mathcal{A} \cong \mathcal{B}$ means that the structures $\mathcal{A}$ and $\mathcal{B}$ are isomorphic. Abusing notation we let $\mathcal{A}$ denote either the structure or its universe.

The shrinking game to be introduced here shares the following features with the basic Ehrenfeucht-Fraïssé game in [Ehr61].

The game is played on two structures $\mathcal{A}$ and $\mathcal{B}$ by two players, Spoiler and Duplicator. Roughly speaking, Spoiler tries to prove that the two structures look different, while Duplicator tries to prove that they look alike. By a **position** *we will mean a triple $((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), n)$ where $|\mathbf{a}| = |\mathbf{b}| < \omega$, and n is a natural number which represents the number of rounds yet to be played.*

When $n = 0$ the game ends, and $((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), 0)$ is a **winning position** *for Duplicator if and only if $(\mathcal{A}, \mathbf{a})$ and $(\mathcal{B}, \mathbf{b})$ satisfy the same atomic formulas.*

When $n > 0$, the ordinary **Ehrenfeucht-Fraïssé game** proceeds from the position $((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), n)$ according to the following rules:

- Spoiler chooses an element c in one structure (say $\mathcal{A}$).
- Duplicator chooses an element d in the other structure ($\mathcal{B}$).

- The game continues from the new position $((\mathcal{A}, \mathbf{a}, c), (\mathcal{B}, \mathbf{b}, d), n-1)$.

We write $(\mathcal{A}, \mathbf{a}) \equiv_n (\mathcal{B}, \mathbf{b})$ if Duplicator has a winning strategy in the Ehrenfeucht-Fraïssé game starting from the position $((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), n)$. We say that $(\mathcal{A}, \mathbf{a})$ and $(\mathcal{B}, \mathbf{b})$ **agree** on a set of (first order) formulas $\mathcal{F}$, in symbols $(\mathcal{A}, \mathbf{a}) \equiv_{\mathcal{F}} (\mathcal{B}, \mathbf{b})$, if for every formula $\psi(\mathbf{x}) \in \mathcal{F}$, $\mathcal{A} \models \psi(\mathbf{a})$ if and only if $\mathcal{B} \models \psi(\mathbf{b})$. The importance of the Ehrenfeucht-Fraïssé game stems from the basic result that $(\mathcal{A}, \mathbf{a}) \equiv_n (\mathcal{B}, \mathbf{b})$ if and only if $(\mathcal{A}, \mathbf{a})$ and $(\mathcal{B}, \mathbf{b})$ agree on all formulas of quantifier rank at most n.

We say that a formula φ is **Boolean over** a set of formulas $\mathcal{F}$ if φ is logically equivalent to a finite Boolean combination of formulas from $\mathcal{F}$ which have the same free variables as φ. Some of the results in this paper will show that one equivalence relation implies another. In view of the following simple lemma, results of this kind often lead to normal form theorems which say that every formula in one set is Boolean over another set.

Lemma 2.1 *Let $\mathcal{F}$ and $\mathcal{G}$ be two sets of formulas with the same free variables. Suppose that whenever $(\mathcal{A}, \mathbf{a}) \equiv_{\mathcal{F}} (\mathcal{B}, \mathbf{b})$ we have $(\mathcal{A}, \mathbf{a}) \equiv_{\mathcal{G}} (\mathcal{B}, \mathbf{b})$. Then every formula in $\mathcal{G}$ is Boolean over $\mathcal{F}$.* ∎

3 The Shrinking Game

We first need some notation concerning neighborhoods and distances in relational structures.

The **Gaifman graph** over a structure $\mathcal{A}$ is the graph over the universe of $\mathcal{A}$ whose edges are the pairs (a, a') of elements of $\mathcal{A}$ such that both a and a' occur in some atomic formula which holds in $\mathcal{A}$. The Gaifman graph over $\mathcal{A}$ is undirected and contains all pairs (a, a).

If $a, c \in \mathcal{A}$ we let $\delta(a, c)$ be the natural distance between a and c in the Gaifman graph over $\mathcal{A}$, i.e. the length of the shortest path connecting a and c. Thus $\delta(a, a) = 0$, and $\delta(a, c) = 1$ if and only if $a \neq c$ and there is an edge connecting a and c. Clearly δ is a metric (possibly taking the value ∞) on $\mathcal{A}$. For sequences $\mathbf{a}, \mathbf{c}$ in $\mathcal{A}$ we also define $\delta(\mathbf{a}, \mathbf{c})$ to be the minimum distance between elements of $\mathbf{a}$ and elements of $\mathbf{c}$. The **degree** of an element $a \in \mathcal{A}$ is the degree of a in the Gaifman graph, that is, the number of elements at distance 1 from a.

If each predicate symbol of ν is at most binary, then $\delta(x, y) = 1$ can be defined by a quantifier-free formula. This leads to a simple and natural relationship between quantifier rank and the Gaifman graph. For example, it implies that for any $s \leq 2^n$, the inequality $\delta(x, y) \leq s$ can be expressed by a formula of quantifier rank at most n.

Things are not as nice if the largest arity of a relation symbol in ν is $d > 2$. In that case, $d-2$ quantifiers are needed to express $\delta(x, y) = 1$, and an extra $d-2$ quantifiers are also needed to express larger distances.

If $\mathbf{a} = \langle a_1, \ldots, a_k \rangle$ are k elements in $\mathcal{A}$ and $k > 0$, $r \geq 0$, the r-**neighborhood** $\mathcal{N}_r^{\mathcal{A}}(\mathbf{a})$ **around a** is the substructure of $(\mathcal{A}, \mathbf{a})$ whose universe is the set of elements at distance $\leq r$ from one of $a_1, \ldots, a_k$. In the case $k = 1$, i.e. $\mathbf{a} = \langle a \rangle$, the r-neighborhood $\mathcal{N}_r^{\mathcal{A}}(a)$ is called **simple**, and the element a is called the **center** of the neighborhood. When $k = 0$, i.e. $\mathbf{a}$ is the empty sequence, we define $\mathcal{N}_r^{\mathcal{A}}(\mathbf{a})$ to be the whole structure $\mathcal{A}$.

A set $C \subseteq \mathcal{A}$ is called s-**scattered** if $\delta(c, d) > s$ for any pair of distinct elements $c, d \in C$. The cardinality $|C|$ will be called the **width** of the s-scattered set C.

Definition 3.1 *For the remainder of this paper, we fix a sequence* $\langle s_n : n \geq 0 \rangle$ *of natural numbers called the* **scattering parameters***, and we define another sequence* $\langle r_n : n \geq 0 \rangle$*, called the* **local radii***, as follows:*

$$r_0 = 1, \qquad\qquad r_{n+1} = 2r_n + s_n.$$

Roughly speaking, we will decompose a structure $\mathcal{A}$ into simple neighborhoods of radius r_n whose centers are s_n-scattered. A straightforward induction shows that

$$r_{n+1} = 2^{n+1} + \sum_{i=0}^{n} 2^{n-i} s_i.$$

A simple example of a scattering parameter is **base** t **exponential growth**, where t is a fixed natural number, and for all $m < n$,

$$r_m = t^m \text{ and } s_m = (t-2) r_m.$$

We now define the shrinking game, which depends on the underlying sequence s_n of scattering parameters. The notation $(\mathcal{A}, \mathbf{a}) \approx_n (\mathcal{B}, \mathbf{b})$ will mean that Duplicator has a winning strategy in the shrinking game at the position $((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), n)$. We define the possible moves and the relation $\approx_n$ by a simultaneous induction on n.

The **shrinking game:** The winning positions for Duplicator in the shrinking game are the same as in the Ehrenfeucht-Fraïssé game. The rules proceeding from the position $((\mathcal{A},\mathbf{a}),(\mathcal{B},\mathbf{b}),n)$, $n>0$ are as follows. Spoiler chooses one structure (say $\mathcal{A}$) and an integer $m<n$, and makes either a **local move** or a **scattered move**.

Local move:

- Spoiler chooses an element $c\in\mathcal{N}^{\mathcal{A}}_{r_m+s_m}(\mathbf{a})$.
- Duplicator chooses an element $d\in\mathcal{N}^{\mathcal{B}}_{r_m+s_m}(\mathbf{b})$.
- The new position is $((\mathcal{A},\mathbf{a},c),(\mathcal{B},\mathbf{b},d),m)$.

Scattered move:

- Spoiler chooses a nonempty finite s_m-scattered set $C\subseteq\mathcal{N}^{\mathcal{A}}_{r_m}(\mathbf{a})$ such that $(\mathcal{A},c)\approx_m(\mathcal{A},e)$ for all $c,e\in C$, and if $|\mathbf{a}|=0$ then $|C|\leq n-m$.
- Duplicator chooses an s_m-scattered set $D\subseteq\mathcal{N}^{\mathcal{B}}_{r_m}(\mathbf{b})$ such that $|D|=|C|$.
- Spoiler chooses an element $d\in D$.
- Duplicator chooses an element $c\in C$.
- The new position is $((\mathcal{A},c),(\mathcal{B},d),m)$.

Note that Spoiler can shorten the shrinking game by choosing $m<n-1$. The reason for giving Spoiler this freedom is to insure that the set of scattered moves available to Spoiler increases as n increases, as stated in the following easy lemma.

Lemma 3.2 *(i) $\approx_n$ is an equivalence relation for each n.*

(ii) If $(\mathcal{A},\mathbf{a})\approx_n(\mathcal{B},\mathbf{b})$, then $(\mathcal{A},\mathbf{a})\approx_m(\mathcal{B},\mathbf{b})$ for all $m\leq n$. ■

In the Ehrenfeucht-Fraïssé game, n is always the number of moves which remain to be played, while in the shrinking game there will be n or fewer moves to be played, depending on the choices of Spoiler.

The following lemma explains the role of the local radii r_n.

Lemma 3.3 *Let $\mathbf{a},\mathbf{c}\in\mathcal{A}$ and $\mathbf{b},\mathbf{d}\in\mathcal{B}$ with $|\mathbf{a}|=|\mathbf{b}|>0$ and $|\mathbf{c}|=|\mathbf{d}|>0$. Suppose $(\mathcal{A},\mathbf{a})\approx_n(\mathcal{B},\mathbf{b})$, $(\mathcal{A},\mathbf{c})\approx_n(\mathcal{B},\mathbf{d})$, $\delta(\mathbf{a},\mathbf{c})>r_n$, and $\delta(\mathbf{b},\mathbf{d})>r_n$. Then $(\mathcal{A},\mathbf{a},\mathbf{c})\approx_n(\mathcal{B},\mathbf{b},\mathbf{d})$.*

Proof: The proof is by induction on n. The case $n=0$ follows easily since $r_0=1$. For the induction step, Duplicator has winning strategies at the positions $((\mathcal{A},\mathbf{a}),(\mathcal{B},\mathbf{b}),n)$ and $((\mathcal{A},\mathbf{c}),(\mathcal{B},\mathbf{d}),n)$. By definition one has $r_n\geq 2r_m+s_m$ for each $m<n$, and this is just what is needed to construct a winning strategy for Duplicator at the combined position $((\mathcal{A},\mathbf{a},\mathbf{c}),(\mathcal{B},\mathbf{b},\mathbf{d}),n)$. The details are omitted. ■

4 n-Quantifier Equivalence

Definition 4.1 *Fix a finite sequence* $\mathbf{s} = (s_0, \ldots, s_{n-1})$ *of scattering parameters. We say that* **s shrinks rapidly** *if* $2r_m \leq s_m$ *for all* $m < n$.

For example, $\mathbf{s}$ shrinks rapidly when $r_m = t^m$ and $s_m = (t-2)r_m$ for some fixed $t \geq 4$. Note that if $\mathbf{s}$ shrinks rapidly, then for each $m < n$, each s_m-scattered set contains at most one point in each simple r_m-neighborhood.

We assume throughout this section that $\mathbf{s}$ *shrinks rapidly.*

Theorem 4.2 *For all structures* $\mathcal{A}$ *and* $\mathcal{B}$, $\mathcal{A} \approx_n \mathcal{B}$ *implies* $\mathcal{A} \equiv_n \mathcal{B}$.

Proof: We prove by induction that Duplicator can play the Ehrenfeucht-Fraïssé game starting from the position $(\mathcal{A}, \mathcal{B}, n)$ so that for each $i = 1 \ldots n$, she maintains the property

$$(\mathcal{A}, \mathbf{a}) \approx_{n-i} (\mathcal{B}, \mathbf{b}) \tag{1}$$

after i rounds, where $\mathbf{a}, \mathbf{b}$ are the elements chosen in the Ehrenfeucht-Fraïssé game.

Induction Base ($i = 1$): Without loss of generality let $a_1 \in \mathcal{A}$ be chosen by Spoiler in the Ehrenfeucht-Fraïssé game. Since every local neighborhood of the empty sequence is the whole structure, this is just a local move for Spoiler in the shrinking game at the initial position $(\mathcal{A}, \mathcal{B}, n)$. The winning strategy of Duplicator in the shrinking game gives a response $b_1 \in \mathcal{B}$ such that (1) holds for $i = 1$.

Induction Step: Suppose that $1 \leq i < n$ and (1) is true for i.

We can assume without loss of generality that Spoiler chooses $a_{i+1} \in \mathcal{A}$ in the Ehrenfeucht-Fraïssé game. Let $p = n - i$. Thus (1) says that Duplicator has a winning strategy in the shrinking game at the position

$$((\mathcal{A}, \mathbf{a}), (\mathcal{B}, \mathbf{b}), p). \tag{2}$$

We distinguish between two cases.

Case a: $a_{i+1} \in \mathcal{N}^{\mathcal{A}}_{r_{p-1}+s_{p-1}}(\mathbf{a})$.

Here Duplicator pretends that a_{i+1} is a local move by Spoiler in the shrinking game at position (2). Duplicator uses her winning strategy to respond with an element $b_{i+1} \in \mathcal{N}^{\mathcal{B}}_{r_{p-1}+s_{p-1}}(\mathbf{b})$. This insures that (1) holds for $i+1$.

Case b: $a_{i+1} \notin \mathcal{N}^{\mathcal{A}}_{r_{p-1}+s_{p-1}}(\mathbf{a})$.

Here Duplicator hopes to find in the structure $\mathcal{B}$ a duplicate $d \notin \mathcal{N}^{\mathcal{B}}_{r_{p-1}}(\mathbf{b})$ such that

$$(\mathcal{A}, a_{i+1}) \approx_{p-1} (\mathcal{B}, d). \tag{3}$$

Then choosing $b_{i+1} = d$ will guarantee (1) for $i+1$ by Lemmas 3.2 and 3.3.

For the sake of contradiction we assume that this is not possible. So any d satisfying (3) will be inside $\mathcal{N}^{\mathcal{B}}_{r_{p-1}}(\mathbf{b})$. Let D be an s_{p-1}-scattered set in $\mathcal{N}^{\mathcal{B}}_{r_{p-1}}(\mathbf{b})$ of maximal width w consisting of elements $d \in \mathcal{B}$ satisfying (3). The above argument for Case a shows that D is nonempty, i.e. $w \geq 1$. Since $2r_{p-1} \leq s_{p-1}$, the set D contains at most one element in each neighborhood $\mathcal{N}^{\mathcal{B}}_{r_{p-1}}(b_j)$, so $w = |D| \leq |\mathbf{b}| = i$.

We now form a set C in $\mathcal{A}$ as follows. Let Spoiler choose $m = p-1$ and the set D as a scattered move in the shrinking game at the position (2). Take $C \subseteq \mathcal{N}^{\mathcal{A}}_{r_{p-1}}(a_1, \ldots, a_i)$ to be the response of Duplicator in her winning strategy. Then $|C| = w$, and $(\mathcal{A}, c) \approx_{p-1} (\mathcal{B}, d) \approx_{p-1} (\mathcal{A}, a_{i+1})$ for all $c \in C, d \in D$.

We will now consider another play of the shrinking game starting from the original position $(\mathcal{A}, \mathcal{B}, n)$. By the hypothesis $\mathcal{A} \approx_n \mathcal{B}$, Duplicator has a winning strategy for this game. Noting that $C \cup \{a_{i+1}\}$ is s_{p-1}-scattered, we let Spoiler choose $m = p-1$ and the set $C \cup \{a_{i+1}\}$ of width $w+1$ as a scattered move in $\mathcal{A}$. This is a legal move since $w+1 \leq i+1 = n-m$. Duplicator must then respond with an s_{p-1}-scattered set consisting of $w+1$ elements of $\mathcal{B}$ which satisfy (3). By assumption, all these elements must be inside $\mathcal{N}^{\mathcal{B}}_{r_{p-1}}(\mathbf{b})$, contradicting the maximality of w. ∎

Here is a generalization of Theorem 4.2 to structures with distinguished elements.

Corollary 4.3 *If* $(\mathcal{A}, \mathbf{a}) \approx_n (\mathcal{B}, \mathbf{b})$, *and* $\mathcal{A} \approx_n \mathcal{B}$, *then* $(\mathcal{A}, \mathbf{a}) \equiv_n (\mathcal{B}, \mathbf{b})$.

Proof: The argument is exactly as in the proof of Theorem 4.2 but starting after $|\mathbf{a}|$ rounds. ∎

5 Shrinking Formulas

In this section we continue to assume that $\mathbf{s}$ *shrinks rapidly.*

We will apply the shrinking games to obtain normal form theorems of the type introduced in Gaifman [Gai82] and improved by Lifsches and Shelah [LS96].

We will introduce a hierarchy of first order formulas which corresponds in a natural way to the shrinking games. This hierarchy depends on a given sequence of scattering parameters $\mathbf{s}$.

Definition 5.1 *The set* $\mathcal{SH}_n(\mathbf{x})$ *of* **shrinking formulas** *in* $\mathbf{x}$ *of rank at most* n *is defined inductively as follows.*

$\mathcal{SH}_0(\mathbf{x})$ *is the set of all quantifier free formulas in* $\mathbf{x}$*.*

For each $m < n$, $\mathcal{SH}_{m+1}(\mathbf{x})$ *is the set of all finite Boolean combinations of formulas in* $\mathcal{SH}_m(\mathbf{x})$ *and formulas of the forms*

$$(\exists y \in \mathcal{N}_{s_m+r_m}(\mathbf{x}))\, \psi(\mathbf{x}, y), \tag{4}$$

$$(\exists y_1, \ldots, y_l \in \mathcal{N}_{r_m}(\mathbf{x})) \left(\bigwedge_{i \le l} \theta(y_i) \wedge \bigwedge_{i<j\le l} \delta(y_i, y_j) > s_m \right), \tag{5}$$

where $\psi(\mathbf{x}, y) \in \mathcal{SH}_m(\mathbf{x}, y)$, $\theta(y) \in \mathcal{SH}_m(y)$, *and* $l \le n - m$ *when* $|\mathbf{x}| = 0$.

In the case that $\mathbf{x}$ is empty, shrinking formulas in $\mathbf{x}$ are called **shrinking sentences**, formula (5) simplifies to

$$(\exists y_1, \ldots, y_l) \left(\bigwedge_{i \le l} \theta(y_i) \wedge \bigwedge_{i<j\le l} \delta(y_i, y_j) > s_m \right),$$

and formula (4) is not needed because it is the special case of formula (5) where $l = 1$.

Lemma 5.2 *There are only finitely many shrinking formulas in* $\mathbf{x}$ *of rank at most* n*, up to logical equivalence.*

Proof: This follows by an easy induction on n, since the language has a finite vocabulary. ∎

We now establish the connection between shrinking formulas and the shrinking game.

Lemma 5.3 *If* $(\mathcal{A}, \mathbf{a})$ *and* $(\mathcal{B}, \mathbf{b})$ *agree on all shrinking formulas of rank at most* n*, then* $(\mathcal{A}, \mathbf{a}) \approx_n (\mathcal{B}, \mathbf{b})$*.*

Proof: This is proved by induction on n. Assume the result for all $m < n$, and suppose Spoiler makes a scattered move, choosing $m < n$ and an s_m-scattered set $C \subseteq \mathcal{N}^{\mathcal{A}}_{r_m}(\mathbf{a})$, with $|C| \leq n - m$ if $|\mathbf{a}| = 0$. Then all the $c \in C$ belong to the same $\approx_m$-equivalence class. By inductive hypothesis, there is a formula $\theta(y) \in \mathcal{SH}_m(y)$ which defines this equivalence class. Then $(\mathcal{A}, \mathbf{a})$ satisfies the formula (5) in Definition 5.1 with $l = |C|$. This formula belongs to $\mathcal{SH}_n(\mathbf{x})$, and thus is also satisfied by $(\mathcal{B}, \mathbf{b})$. Therefore Duplicator has a winning response.

The argument is similar when Spoiler makes a local move, so Duplicator has a winning strategy for the shrinking game. ∎

We now put our results together to get a normal form theorem.

Theorem 5.4 *If* $\mathcal{A}, \mathcal{B}$ *agree on all shrinking sentences of rank at most* n*, then* $\mathcal{A} \equiv_n \mathcal{B}$*.*

Proof: By Theorem 4.2 and Lemma 5.3. ∎

To get normal form theorems of the Gaifman type, we examine the locality properties of shrinking formulas.

Definition 5.5 *By an* s**-scattered** r**-local sentence of width** l *we mean a sentence of the form*

$$\exists y_1 \cdots \exists y_l \left(\bigwedge_{i \leq l} \theta(y_i) \wedge \bigwedge_{i < j \leq l} \delta(y_i, y_j) > s \right)$$

where $\theta(y)$ *is* r*-local.*

Proposition 5.6 *(i) Every shrinking formula of rank at most* n *is* $(r_n - 1)$*-local.*

(ii) Every shrinking sentence of rank at most n *is a finite Boolean combination sentences each of which is* s_m*-scattered and* $(r_m - 1)$*-local for some* $m < n$*.*

Proof: By induction on n, using the fact that for each $m < n$, $(s_m + r_m) + (r_m - 1) \leq r_n - 1$. ∎

We can now state a normal form theorem of the Gaifman type.

Theorem 5.7 *Fix a scattering sequence* $\mathbf{s}$ *which shrinks rapidly. Then each first order sentence of quantifier rank at most* n *is logically equivalent to a finite Boolean combination of sentences each of which is* s_m*-scattered and* $(r_m - 1)$*-local of width at most* $n - m$*, for some* $m < n$*.*

In the case that $r_m = 4^m$, we get the result of Lifsches and Shelah [LS96] with an improved bound on the radius.

Corollary 5.8 *Every first order sentence of quantifier rank at most* n *is logically equivalent to a finite Boolean combination of sentences each of which is* $2 \cdot 4^m$*-scattered and* $(4^m - 1)$*-local of width at most* $n - m$*, for some* $m < n$*.*

Proof: Take $r_m = 4^m$ and $s_m = 2r_m$ in Theorem 5.7. ∎

By examining the proofs, one can readily extend the normal form results in this section from sentences to formulas. We will leave these extensions to the reader.

Finally, we give an upper bound for the quantifier rank of the scattered local sentences in Theorem 5.7.

Theorem 5.9 *Let* d *be the least upper bound of* 2 *and the maximum number of arguments of the relation symbols of the vocabulary* ν*. In Theorem 5.7, the local formulas inside the* s_m*-scattered* $(r_m - 1)$*-local sentences can be taken to have quantifier rank at most* $\log_2(r_m) + d - 1$*.*

In the case $r_m = 4^m$, the local formulas inside can be taken to have quantifier rank at most $2m + d - 1$.

Proof: One proves by induction on n that each shrinking formula in $\mathbf{x}$ of rank n has quantifier rank at most $|\mathbf{x}| + \log_2(r_n) + d - 2$. To do this, a bound is needed on the quantifier rank of the distance inequality $\delta(x, y) \leq r$. In the case $d = 2$, this inequality is expressible by a first order formula of quantifier rank $\log_2(r)$. In the case $d > 2$, it is expressible by a first order formula of quantifier rank $\log_2(r) + d - 2$. The remaining details are left to the reader. ∎

6 The Schwentick-Barthelmann Normal Form

In this section we continue to assume that $\mathbf{s}$ *shrinks rapidly.*

In [SB99] Schwentick and Barthelmann modified Gaifman's normal form by proving that every first order sentence is logically equivalent to a sentence of the form $\exists x_1 \ldots \exists x_l \forall y \varphi(\mathbf{x}, y)$ where φ is a first order local formula around y. We will use the shrinking game to give another proof of this fact which makes it easier to keep track of width and locality bounds in the normal form.

Theorem 6.1 *Let $n > 0, r = n \cdot 4^n, q = log(r) + d - 1$. Suppose that every sentence of the form*

$$\exists x_1 \ldots \exists x_l \forall y \varphi(\mathbf{x}, y) \tag{6}$$

which holds in $\mathcal{A}$ holds in $\mathcal{B}$, where $l \leq n$ and φ is a first order local formula around y of radius at most r and quantifier rank at most q. Then $\mathcal{A} \approx_n \mathcal{B}$.

Proof: The hypotheses say that Duplicator has a winning strategy in the following **one-sided game**, where Spoiler must start in $\mathcal{A}$:

- Spoiler chooses a tuple $\mathbf{a}$ in $\mathcal{A}$ of width $\leq n$.
- Duplicator chooses $\mathbf{b}$ in $\mathcal{B}$ with $|\mathbf{b}| = |\mathbf{a}|$.
- Spoiler chooses $d \in \mathcal{B}$.
- Duplicator chooses $c \in \mathcal{A}$.
- The game proceeds with an ordinary Ehrenfeucht-Fraïssé game from

  $$((\mathcal{N}_r^{\mathcal{A}}(c), \mathbf{a}, c), (\mathcal{N}_r^{\mathcal{B}}(d), \mathbf{b}, d), q).$$

Using arguments like those in the preceding sections of this paper, one can now show that Duplicator has a winning strategy in the shrinking game. ∎

Corollary 6.2 *Every first order sentence of quantifier rank n is logically equivalent to a finite conjunction of sentences of the form (6) in Theorem 6.1.*

7 Narrow Sentences

For an arbitrary scattering sequence $\mathbf{s}$, we obtain a normal form theorem for sentences in which there is a uniform bound on the size of an s_m-scattered set in each r_m neighborhood.

Definition 7.1 *A first order sentence φ is* **s-narrow** *if there exists a finite bound k such that for each $m < n$, φ logically implies that for all x, every s_m-scattered set in $\mathcal{N}^{r_m}(x)$ has width at most k.*

Note that if $\mathbf{s}$ shrinks rapidly, then every sentence is $\mathbf{s}$-narrow with bound $k = 1$. In the next theorem, the interesting case is where $\mathbf{s}$ does not shrink rapidly.

Theorem 7.2 *Fix a scattering sequence $\mathbf{s}$. Let φ be an $\mathbf{s}$-narrow sentence with bound k. Then each first order sentence of quantifier rank at most n is φ-*

equivalent to a finite Boolean combination of first order s_m-scattered $(r_m - 1)$-local sentences of width at most $k(n-m)$, for $m < n$.

We omit the proof, which is a direct generalization of the proof of Theorem 5.7.

8 Conclusion

We introduced a shrinking Ehrenfeucht-Fraïssé game, in which the players move in neighborhoods whose radii shrink at a rate which depends on a sequence $\mathbf{s}$ of scattering parameters. The main result shows that if $\mathbf{s}$ shrinks rapidly and Duplicator has a winning strategy for the n-round shrinking game on two structures, then Duplicator has a winning strategy for the n-round Ehrenfeucht-Fraïssé game.

This leads, as a special case, to a more easily understood proof of a Gaifman type normal form theorem of Lifsches and Shelah, with a slightly improved bound on the local radius. The shrinking game is also used to obtain bounds for a normal form theorem of Schwentick and Barthelmann showing that each first order sentence is equivalent to a sentence of the form $\exists x_1 \cdots \exists x_l \forall y \varphi(\mathbf{x}, y)$ where $\varphi(\mathbf{x}, y)$ is local around y.

For an arbitrary scattering sequence $\mathbf{s}$, a spectrum of normal form theorems is obtained for $\mathbf{s}$-narrow sentences. The method gives bounds on the radius of the local neighborhoods, the number of local neighborhoods and distance between them, and on the quantifier rank of the scattered local sentences.

References:

[Ehr61] A. Ehrenfeucht, An application of games to the completeness problem for formalized theories, **Fund. Math.**, vol. 49 (1961), pp. 129-141.

[Gai82] H. Gaifman, On local and nonlocal properties, J. Stern, editor, **Logic Colloquium '81**, North Holland, Amsterdam (1982), pp. 105-135.

[Lib97] L. Libkin, On the forms of locality over finite models, In **Proc. 12th IEEE Symp. on Logic in Computer Science** (1997).

[Lib00] L. Libkin, On counting logics and local properties, In **ACM TOCL**, vol. 1 (1) (2000).

[LS96] S. Lifsches and S. Shelah, Distorted sums of models, Publication #603, Shelah's Archive.

[Sch96] T. Schwentick, On winning Ehrenfeucht games and monadic NP, **Annals of Pure and Applied Logic**, vol. 79 (1996), pp. 61-92.

[SB99] T. Schwentick and K. Barthelmann, Local normal forms for first-order logic with applications to games and automata, **Discrete Math. Theor. Comput. Sci.**, vol. 3 (1999), no. 3, pp. 109-124 (electronic).