On a syntactic approximation to logics that capture complexity classes

Argimiro Arratia*
Dpto. de Matemática Aplicada
Facultad de Ciencias
Universidad de Valladolid
Valladolid 47005, Spain
arratia@mac.uva.es

Carlos E. Ortiz†
Department of Mathematics
and Computer Science
Arcadia University
450 S. Easton Road,
Glenside, PA 19038-3295, U.S.A.
ortiz@arcadia.edu

Abstract

We formulate a formal syntax of approximate formulas for the logic with counting quantifiers, $\mathcal{SOLP}$, studied by us in [1], where we showed the following facts: (i) In the presence of a built–in (linear) order, $\mathcal{SOLP}$ can describe **NP**–complete problems and fragments of it capture classes like **P** and **NL**; (ii) weakening the ordering relation to an almost order (in the sense of [7]) we can separate meaningful fragments, using a combinatorial tool suited for these languages.

The purpose of the approximate formulas is to provide a syntactic approximation to logics contained in $\mathcal{SOLP}$ with built-in order, that should be complementary of the semantic approximation based on almost orders, by producing approximating logics where problems are described within a small counting error. We introduce a concept of *strong expressibility* based on approximate formulas, and show that for many fragments of $\mathcal{SOLP}$ with built-in order, including ones that capture **P** and **NL**, expressibility and strong expressibility are equivalent. We state and prove a Bridge Theorem that links expressibility in fragments of $\mathcal{SOLP}$ over almost-ordered structures to strong expressibility with respect to approximate formulas for the corresponding fragments over ordered structures. A consequence of these results is that proving inexpressibility results over fragments of $\mathcal{SOLP}$ with built-in order can be done by proving inexpressibility over the corresponding fragments with built-in almost order, where separation proofs are easier.

Subject Classification: Logic in computer science; Descriptive Complexity.
Keywords: Proportional quantifiers, approximate formulas, almost order, expressiveness, computational complexity, **P**, **NL**.

1 Introduction

In Descriptive Complexity the difficulty of a problem is measured in terms of the syntactic resources, such as number of quantifiers, number of variables and other symbols,

*Supported by Spain's MEC, projects MOISES (TIC2002-04019-C03-03) and SINGACOM (MTM-00958)

†Supported by a Faculty Award Grant from the Christian R. & Mary F. Lindback Foundation, and a Visiting Research Fellowship from University of Valladolid, Spain

needed to describe the problem in some logical formalism. These syntactic measures are intimately related to the complexity of resources needed to solve a problem by various models of computation. Problems are considered as sets of finite structures, and of the few techniques from classical model theory for showing inexpressibility in logics that survive the passing from the infinite to the finite structures is the method of Ehrenfeucht-Fraïssé games (cf. [2]).

However this combinatorial technique of games, has been unsuccessful in showing meaningful separation, the most sought after being among logics capturing the classes **P** and **NP**. One of the reason for this failure is that, as of today, all known logics that capture **P** need a relation of linear order built–in into the semantics, and in the presence of a built–in linear order, an Ehrenfeucht-Fraïssé game has little power for telling structures apart (e.g. see [5, § 6.6]). To overcome this difficulty, a natural idea is to study *approximations* to logics with built–in order, where techniques like Ehrenfeucht-Fraïssé games become effective in showing separability results, and hopefully these separations in the approximate setting will give a clue on how to go about separating the associated logics with order.

There are two main approaches to define approximate logics in model theory. One is to play with the semantics, where constructs as built–in order is weakened to an "*almost–order*", and, frequently, some counting operator is added to compensate for the loss of expressive power (e.g. [3], [7] among others). This approach has some limitations. To mention one, the paper by Libkin and Wong [7] shows that a very powerful extension of first order logic (FO) with additional counting quantifiers, denoted $\mathcal{L}^*_{\infty\omega}(C)$, which subsumes all known "pure" counting extensions of FO (meaning that fixpoint operators are not considered), in the presence of almost–orders, has the *bounded number of degrees property* (or BNDP) and thus cannot express the transitive closure of a binary relation.

The other approach is syntactic and is found in classical model theory as in, for example, Keisler's logic of probability quantifiers (see [6]), who conceived it as a logic appropriate for his investigations on *probability hyperfinite spaces*, or infinite structures suitable for approximating large finite phenomena of applied mathematics. Under this approach, for each formula φ of a logic and every rational ϵ one construct an approximate formula φ_ϵ with the property that in every model $\mathcal{A}$, if $\epsilon_1 < 0 < \epsilon_2$ then $\varphi_{\epsilon_1} \to \varphi \to \varphi_{\epsilon_2}$, and as ϵ tends to 0, the interpretation of φ_ϵ should be closer to φ. This approach has been developed with success for the theory of classical metric spaces but not, to our knowledge, in Computational Complexity theory.

In this paper we develop a syntactic approach to the task of approximating logics with built–in order based on a notion of approximate formulas and show how it relates to the semantic approach based on almost orders. This approach is potentially relevant to the problem of separating logics with built-in order, since we obtain a bridge result that implies that separation of logics with built-in almost-order can be translated into separation of corresponding logics with built-in order.

The framework for our results is the second order logic of proportionality quantifiers, $\mathcal{SOLP}$, defined in [1]. The quantifiers for this logic are counting quantifiers acting upon second order terms. When restricted to built-in almost orders, this logic avoids the BNDP, has non trivial expressive power, and general separations results of combinatorial nature can be obtained. More specifically, $\mathcal{SOLP}$ consists of quantifiers of the form $(P(X) \geq r)$ and $(P(X) \leq r)$ for rational $r \in (0, 1)$, and whose meaning is that the cardinality of the set X, say of arity $k > 0$, is greater than or equal to (or less than or equal to) r times the cardinality of the set of k–tuples in the model. We review the definition of $\mathcal{SOLP}$ and summarise facts found in [1] about its expressive power in the presence of almost orders in section 2.

The proportional quantifiers $(P(X) \geq r)$ and $(P(X) \leq r)$ are suitable for allowing approximations, which in the case of monadic second order variables, are defined in the following way: For a formula $\psi \in \mathcal{SOLP}$ and every $\epsilon > 0$, the approximate formula ψ_ϵ is obtained by replacing every quantifier $(P(X) \geq r)$ by $(P(X) \geq r - \epsilon)$, and every quantifier $(P(X) \leq r)$ by $(P(X) \leq r + \epsilon)$ (X is of arity 1). Our definition for any arity of X is more elaborate, but it is the right one for establishing a correspondence between satisfaction of formulas in $\mathcal{SOLP}$ in almost ordered structures and satisfaction of the corresponding approximate formulas in ordered structures. This result we call Bridge Theorem and it is shown in section 3.

In section 4 we introduce the notion of the ϵ–approximate logic $\mathcal{L}_\epsilon$, for every fragment $\mathcal{L}$ of $\mathcal{SOLP}$; a logic that should have an expressive power "almost" similar to the expressive power of $\mathcal{L}$. This notion in turn generates the notions of strong expressibility and ϵ–relaxed fragments. An ϵ–relaxed fragment is one for which $\mathcal{L}_\delta = \mathcal{L}$ (in terms of expressive power) for every $\delta \in (-\epsilon, \epsilon)$. Surprisingly, fragments of $\mathcal{SOLP}$ with built–in order that capture **P** and **NL** are ϵ–relaxed. A nice property of ϵ–relaxed logics is that for them strong expressibility and expressibility are "almost" equivalent (an idea that we will formalise). A consequence of this is Theorem 4.9 that shows that to prove inexpressibility of problems in ϵ–relaxed logics *with built–in order* it is enough to prove inexpressibility of the same problem in the δ-approximate logics ($\delta \in (-\epsilon, \epsilon)$) with respect to almost ordered structures. Since proving inexpressibility for logics over almost orders is, in practice, easier than the usual checking of satisfaction in ordered structures, this last result has potential applicability for studying separation of well known logics with built-in order, such as the ones that capture **NL** and **P**.

We end the paper arguing why strong inexpressibility should imply inexpressibility in ϵ–relaxed logics, and in the presence of order. For if it is not the case then the behaviour of the approximating formulas is very strange: their complexity (based on number of variables and arity of second order variables) tend to infinity as their ϵ–error approaches 0, that is, as the approximate formulas tend to the exact formula.

2 The second order logic of proportional quantifiers

Definition 2.1 *The Second Order Logic of Proportional quantifiers, denoted $\mathcal{SOLP}$, is the set of formulas of the form*

$$Q_1 \cdots Q_u \theta(x_1, \ldots, x_s, X_1, \ldots, X_r) \tag{1}$$

where $\theta(x_1, \ldots, x_s, X_1, \ldots, X_r)$ is a first order formula over some vocabulary τ with first order variables $x_1, \ldots, x_k$ and second order variables, X_1, $\ldots$, X_r; each Q_j ($j \leq u$) is either $(P(X_i) \geq t_i)$ or $(P(X_i) \leq t_i)$, where t_i is a rational in $(0, 1)$, for some $i \leq r$. Whenever we want to make the underlying vocabulary τ explicit we will write $\mathcal{SOLP}(\tau)$.

We also define $\mathcal{SOLP}(\tau)[r_1, \ldots, r_k]$, for a given vocabulary τ and sequence r_1, r_2, $\ldots$, r_k of distinct natural numbers, as the sublogic of $\mathcal{SOLP}(\tau)$ where the proportional *quantifiers can only be of the form* $(P(X) \leq q/r_i)$ *or* $(P(X) \geq q/r_i)$*, for* $i = 1, \ldots, k$ *and* q *a natural number such that* $0 \leq q < r_i$*.*

Another fragment of $\mathcal{SOLP}$ *which will be of interest for us is the* Second Order Monadic Logic of Proportional quantifiers*, denoted* $\mathcal{SOMLP}$*, which is* $\mathcal{SOLP}$ *with the arity of the second order variables in* (1) *being all equal to 1.*

The interpretation for the proportional quantifiers is very natural: Let X be a second order variable of arity k, $\overline{Y}$ a vector of second order variables, $\overline{x} = x_1, \ldots, x_m$ first order variables and $\phi(\overline{x}, \overline{Y}, X)$ a formula in $\mathcal{SOLP}(\tau)$ over some (finite) vocabulary τ (which does not contains X or any of the variables in $\overline{Y}$ as a relation symbol). Let r be a rational in $(0, 1)$. Then

$$(P(X) \geq r)\phi(\overline{x}, \overline{Y}, X) \;\text{ and }\; (P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$$

have the following semantics. For an appropriate finite τ–structure $\mathcal{A}$, elements $\overline{a} = (a_1, \ldots, a_m)$ in A and an appropriate vector of relations $\overline{B}$ over A, we have

$\mathcal{A} \models (P(X) \geq r)\phi(\overline{a}, \overline{B}, X) \iff$ there exists $S \subseteq A^k$ such that $\mathcal{A} \models \phi(\overline{a}, \overline{B}, S)$ and $|S| \geq r \cdot |A|^k$

Similarly for $(P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$, substituting in the above definition $\geq$ for $\leq$.

Example 2.2 *Let* $\tau = \{R, s, t\}$ *where* R *is a ternary relation symbol, and* s *and* t *are constant symbols. Let* r *be a rational with* $0 < r < 1$*. We define*

NOT-IN-CLOS$_{\leq r}$ **:=** $\{\mathcal{A} = \langle A, R, s, t\rangle :$ *A has a set containing s but not t, closed under R, and of size at most a fraction r of $|A|$* $\}$.

Let $\beta_{nclos}(X)$ *be the following formula*

$$\begin{aligned} \beta_{nclos}(X) \;\; := \;\; & \forall x \forall u \forall v\, [X(s) \wedge \neg X(t) \\ & \wedge\, (X(u) \wedge X(v) \wedge R(u, v, x) \rightarrow X(x))] \end{aligned}$$

Then

$$\mathcal{A} \in \text{NOT-IN-CLOS}_{\leq r} \iff \mathcal{A} \models (P(X) \leq r)\beta_{nclos}(X)$$

In [1] it is shown that, for $r = 1/n$*, this problem is complete for* **P** *under first order reductions.* □

Example 2.3 *Let* $\tau = \{E, s\}$ *where* E *is a binary relation symbol and* s *is a constant symbol. We think of* τ*-structures as graphs with a specify vertex* s *(the source). Let* r *be a rational with* $0 < r < 1$*. We define*

NCON$_{\geq r}$ **:=** $\{\mathcal{A} = \langle A, E, s\rangle : \langle A, E\rangle$ *is a digraph and at least a fraction r of the vertices are* **not** *connected to* $s\}$

Let $\alpha_{ncon}(Y)$ *be the following formula*

$$\alpha_{ncon}(Y) \;\; := \;\; \neg Y(s) \wedge \forall x \forall y (E(x, y) \wedge Y(x) \rightarrow Y(y))$$

Then $\mathcal{A} \in \text{NCON}_{\geq r} \iff \mathcal{A} \models (P(Y) \geq r)\alpha_{ncon}(Y)$*.*

The problem NCON$_{\geq 1/2}$ *is complete for* **NL** *under first order reductions (see [1]).* □

2.1 Summary of facts about a semantic approximation to $\mathcal{SOLP}$

In [1] we study the expressive power of $\mathcal{SOLP}$ in the presence of built–in order and when this external predicate is weakened to an almost order (see [5] for the notion and use of built–in numerical predicates in Descriptive Complexity). We summarise below the facts from [1] that we need about what we view as "semantic approximations" to definability in $\mathcal{SOLP}$ and some of its fragments. Besides those fragments mentioned in Definition 2.1 we are interested in the logics $\mathcal{SOLP}Horn$ and $\mathcal{SOLP}Krom$, which were defined in [1] after Grädel's definitions of the Horn and Krom subsets of Second Order logic in [4].

A first order formula α over a vocabulary τ plus second order variables $X_1, \ldots, X_r$ of arities $k_1, \ldots, k_r$, respectively, plus possibly a binary relation symbol $=$ (equality) and the constant $\perp$ (standing for false), is a *universal Horn formula*, if α is a universally quantified conjunction of formulas over $\tau \cup \{X_1, \ldots, X_r\}$ of the form $(\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_s) \rightarrow \varphi$, where φ is either $X_i(\overline{u}_i)$ (where $\overline{u}_i$ denotes a k_i-tuple of first order terms, $i = 1, \ldots, r$) or $\perp$, and $\psi_1, \ldots, \psi_s$ are atomic or negation of atomic $(\tau \cup \{X_1, \ldots, X_r\})$-formulas except that any occurrence of the variables X_i must be positive (there are no restrictions on the predicates in τ or $=$). The logic $\mathcal{SOLP}Horn$ is the set of formulas of the form

$$(P(X_1) \leq t_1) \cdots (P(X_r) \leq t_r)\alpha$$

where each t_i is a rational in $(0, 1)$, and α is a universal Horn formula over some vocabulary τ and second order variables $X_1, \ldots, X_r$. For example, the problem NOT-IN-CLOS$_{\leq r}$ presented in Example 2.2 is definable in $\mathcal{SOLP}Horn[r]$.

A first order formula α over $\tau \cup \{X_1, \ldots, X_r\} \cup \{=, \perp\}$ is a *universal Krom formula*, if α is a universally quantified conjunction of clauses, where each clause is a disjunction of literals with at most two occurrences (positive or not) of the predicates $X_1, \ldots, X_r$, i.e. α is a 2-CNF formula with respect to the variables $X_1, \ldots, X_r$. The logic $\mathcal{SOLP}Krom$ is the set of formulas of the form

$$(P(X_1) \geq t_1) \cdots (P(X_r) \geq t_r)\alpha$$

where each t_i is a rational in $(0, 1)$, and α is a universal Krom formula over some vocabulary τ and second order variables $X_1, \ldots, X_r$. For example, the problem NCON$_{\geq r}$ presented in Example 2.3 is definable in $\mathcal{SOLP}Krom[r]$.

We have shown in [1] that:

(1) In the presence of order (at least a built–in successor), $\mathbf{P} \subseteq \mathcal{SOLP}[2]$ (in the sense that any class of structures decidable in $\mathbf{P}$ is definable by a sentence of $\mathcal{SOLP}[2]$) and, furthermore, it is captured by the fragment $\mathcal{SOLP}Horn[2]$, consisting of formulas of the form $(P(X_1) \leq 1/2) \cdots (P(X_r) \leq 1/2)\alpha$, where α is a universal Horn formula.

(2) In the presence of order, $\mathbf{NL}$ is captured by $\mathcal{SOLP}Krom[2]$, a fragment consisting of formulas of the form $(P(X_1) \geq 1/2) \cdots (P(X_r) \geq 1/2)\alpha$, where α is a universal Krom formula. (This and the previous capturing of $\mathbf{P}$ by fragments of $\mathcal{SOLP}$ are inspired on Grädel's [4], but taking into account the limitations in the cardinalities of second order variables imposed by our counting quantifiers.)

(3) With respect to almost ordered structures we have an infinite hierarchy within the monadic fragment $\mathcal{SOMLP}$, namely,

$$\mathcal{SOMLP}[2] \subsetneq \mathcal{SOMLP}[2,3] \subsetneq \mathcal{SOMLP}[2,3,5] \subsetneq \ldots$$

(4) With respect to almost ordered structures and unbounded arity we have that

$$\mathcal{SOLP}Horn[2] \subsetneq \mathcal{SOLP}[2,3].$$

The separation results listed in (3) and (4) were obtained with appropriate Ehrenfeucht–Fraïssé type of games.

The concept of almost order (taken from [7]) constitute the core of our "semantic approximations", around which we work our syntactic approximations, and thus we pause to review this concept and further constructions from [1].

Definition 2.4 *A function $g : \mathbb{N} \to \mathbb{N}$ is sublinear if, for all $n \in \mathbb{N}$, $g(n) < n$. For a fixed positive integer k, a k-preorder over a set A is a binary, reflexive and transitive relation P in which every induced equivalence class of $P \cap P^{-1}$ has size at most k. An almost linear order over A, determined by a sublinear function $g : \mathbb{N} \to \mathbb{N}$, is a binary relation $\leq_g$ over A with a partition of the universe A into two sets B, C, such that B has cardinality $n - g(n)$ and $\leq_g$ restricted to B is a linear order, $\leq_g$ restricted to C is a 2-preorder, and for every $x \in C$ and every $y \in B$, $x \leq_g y$.*

Note that for any function $g : \mathbb{N} \to \mathbb{N}$, the almost linear order $\leq_g$ over a set A induces an equivalence relation $\sim_g$ in A defined by $a \sim_g b$ iff $a \leq_g b$ and $b \leq_g a$. For $a \in A$, let $[a]_g$ denote its $\sim_g$–equivalence class, and $[A]_g := \{[a]_g : a \in A\}$.

Definition 2.5 *Fix a sublinear $g : \mathbb{N} \to \mathbb{N}$ and let R be an n-ary relation on a set A. Let $\leq_g$ be an almost order determined by g in A. We say that R is consistent with $\leq_g$ if for every pair of vectors $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ of elements in A with $a_i \sim_g b_i$ for every $i \leq n$, we have that*

$$R(a_1, \dots, a_n) \text{ holds if and only if } R(b_1, \dots, b_n) \text{ holds.}$$

Let $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \dots, R_k^{\mathcal{A}}, C_1^{\mathcal{A}}, \dots, C_s^{\mathcal{A}} \rangle$ be a τ-structure. We say that $\mathcal{A}$ is consistent with $\leq_g$ if and only if for every $i \leq k$, $R_i^{\mathcal{A}}$ is consistent with $\leq_g$.

For a τ-structure $\mathcal{A}$, consistent with $\leq_g$, it makes sense to define the *quotient structure* $\mathcal{A}/_{\sim_g}$, as a τ-structure consisting of $[A]_g$ as its universe, and for a k-ary relation $R \in \tau$,

$$R^{\mathcal{A}/\sim_g} := \{([a_1]_g, \dots, [a_k]_g) : (a_1, \dots, a_k) \in R^{\mathcal{A}}\}$$

Furthermore, for a subset $B \subseteq A$ we define its $\leq_g$-*contraction* as $[B]_g := \{[b]_g : b \in B\}$. All these terms will play their role in a theorem below that bridges from satisfaction in almost ordered structures to satisfaction in quotient structures, where the order turns linear.

By $\mathcal{SOLP} + \leq_g$, for an almost order $\leq_g$, we understand the logic $\mathcal{SOLP}$ with the almost order $\leq_g$ as additional built-in relation, and where we only consider models $\mathcal{A}$ that are consistent with $\leq_g$. Furthermore, for the formulas of the form $(P(X) \geq r)\phi(\overline{x}, \overline{Y}, X)$ and $(P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$, we require the following modification of the semantics: For an appropriate finite model $\mathcal{A}$ consistent with $\leq_g$, for elements $\overline{a} = (a_1, \dots, a_m)$ in A and an appropriate vector of relations $\overline{B}$, consistent with $\leq_g$, we should have

$$\mathcal{A} \models (P(X) \geq r)\phi(\overline{a}, \overline{B}, X) \iff \begin{array}{l}\text{there exists } S \subseteq A^k \text{, consistent with } \leq_g,\\ \text{such that } \mathcal{A} \models \phi(\overline{a}, \overline{B}, S) \text{ and } |S| \geq r \cdot |A|^k\end{array}$$

Similarly for $(P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$, substituting in the above condition $\geq$ for $\leq$.

Remark 2.6 *In general, given a logic $\mathcal{L} \subseteq \mathcal{SOLP}$, we use $\mathcal{L}+ \leq_g$ to indicate that all possible (finite) models of $\mathcal{L}$ have an almost order $\leq_g$, determined by a sublinear function g. Also $\mathcal{L}+ \leq$ indicates that the models have an additional linear order.*

The property of being consistent for $\leq_g$ holds for all the formulas in $\mathcal{SOLP}(\tau)_{\leq_g}$.

Lemma 2.7 *Let $\mathcal{A}$ be a structure which is consistent with $\leq_g$. Then, for every formula $\psi(\overline{x})$ in $\mathcal{SOLP} + \leq_g$, the set $\psi^{\mathcal{A}} := \{\overline{a} \in \mathcal{A} : \mathcal{A} \models \psi(\overline{a})\}$ is consistent with $\leq_g$.*

Proof: The proof is an easy induction in formulas. $\square$

3 A syntax of approximate formulas

We now introduce the notion of approximate formulas for $\mathcal{SOLP}$. The purpose of these formulas is to provide a link between satisfaction in almost ordered structures and satisfaction in their corresponding quotient structures. This we will make precise in the Bridge Theorem (Theorem 3.5 below). The general conclusion will be that whatever we can say about a class of almost ordered structures we can "approximately" say about a class of their quotient structures (which are fully linearly ordered structures), and vice versa.

Definition 3.1 (Approximate Formulas) *For every rational $\epsilon \in [0, 1)$ and for every formula $\theta(\overline{x}, \overline{X}) \in \mathcal{SOLP}(\tau)$, we define the ϵ-approximation of $\theta(\overline{x}, \overline{X})$, denoted $\theta(\overline{x}, \overline{X})_\epsilon$, as follows:*

Atomic formulas *If $\theta(\overline{x}) := R(\overline{x})$ then $\theta(\overline{x})_\epsilon := R(\overline{x})$, for R relation symbol in τ. If $\theta(\overline{x}, X) := X(\overline{x})$, with X a second order variable of arity $k \geq 1$, then $\theta(\overline{x}, X)_\epsilon := X(\overline{x})$.*

Negation of atomic formulas *If $\theta(\overline{x}) := \neg R(\overline{x})$ then $\theta(\overline{x})_\epsilon := \neg R(\overline{x})$, for $R \in \tau$. Likewise, $(\neg X(\overline{x}))_\epsilon := \neg X(\overline{x})$.*

Conjunction, Disjunction *If $\theta(\overline{x}, X) := \phi(\overline{x}, X) \star \psi(\overline{x}, X)$ then $\theta(\overline{x}, X)_\epsilon := \phi_\epsilon(\overline{x}, X) \star \psi_\epsilon(\overline{x}, X)$, with $\star \in \{\wedge, \vee\}$.*

First order quantifiers *If $\theta(\overline{x}, \overline{X}) := Qz\varphi(\overline{x}, z, \overline{X})$ then $\theta(\overline{x}, \overline{X})_\epsilon := Qz(\varphi(\overline{x}, z, \overline{X})_\epsilon)$, with $Q \in \{\exists, \forall\}$.*

Proportional quantifiers *If $\theta(\overline{x}, \overline{X}) := (P(Y) \geq r)\varphi(\overline{x}, \overline{X}, Y)$, where Y is of arity $k \geq 1$, then $\theta(\overline{x}, \overline{X})_\epsilon$ is*

$$\begin{cases} (P(Y) \geq (1-\epsilon)^{k-1}[r - k\epsilon])(\varphi(\overline{x}, \overline{X}, Y)_\epsilon) & \textit{if } r - k\epsilon > 0 \\ (P(Y) \geq 0)(\varphi(\overline{x}, \overline{X}, Y)_\epsilon) & \textit{otherwise} \end{cases}$$

If $\theta(\overline{x}, \overline{X}) := (P(Y) \leq r)\varphi(\overline{x}, \overline{X}, Y)$ then $\theta(\overline{x}, \overline{X})_\epsilon$ is

$$\begin{cases} (P(Y) \leq (1+\epsilon)^{k-1}[r + k\epsilon])(\varphi(\overline{x}, \overline{X}, Y)_\epsilon) & \textit{if } (1+\epsilon)^{k-1}(r + k\epsilon) < 1 \\ (P(Y) \leq 1)(\varphi(\overline{x}, \overline{X}, Y)_\epsilon) & \textit{otherwise} \end{cases}$$

Remark 3.2 *We can (and will) always assume that ϵ is small enough so that the ϵ–approximation for formulas with proportional quantifiers is the first option in their definition, e.g., $(P(Y) \leq (1+\epsilon)^{k-1}[r + k\epsilon])\varphi(\overline{x}, \overline{X}, Y)_\epsilon$.*

The previous definition describes syntactic approximations "from the right" or "positive". We can also have approximations from the left or negative (our intuition of right or left approximation will be formalised by lemma 3.4 below). What we want for $\phi_{-\epsilon}$ to have is the property that $(\phi_{-\epsilon})_\epsilon := \phi$. With this in mind we have the following definition.

Definition 3.3 (Approximate Formulas for Negative Values) *For every rational $\epsilon \in [0, 1)$ and for every formula $\theta(\overline{x}, \overline{X}) \in \mathcal{SOLP}(\tau)$, we define the $-\epsilon$-approximation of $\theta(\overline{x}, \overline{X})$ by induction in the complexity of the formulas as follows:*

First order formulas *If $\theta(\overline{x}, \overline{X})$ is a first order formula with free second order variables among the $\overline{X}$ and free first order variables among the $\overline{x}$, then $\theta(\overline{x}, \overline{X})_{-\epsilon} := \theta(\overline{x}, \overline{X})$.*

Proportional quantifiers *If $\theta(\overline{x}, \overline{X}) := (P(Y) \geq r)\varphi(\overline{x}, \overline{X}, Y)$, where Y is of arity $k \geq 1$, then $\theta(\overline{x}, \overline{X})_{-\epsilon}$ is*

$$\begin{cases} (P(Y) \geq \frac{1}{(1-\epsilon)^{k-1}}[r + k\epsilon(1-\epsilon)^{k-1}])(\varphi(\overline{x}, \overline{X}, Y)_{-\epsilon}) & \textit{if } \frac{r}{(1-\epsilon)^{k-1}} + k\epsilon < 1 \\ (P(Y) \geq 1)(\varphi(\overline{x}, \overline{X}, Y)_{-\epsilon}) & \textit{otherwise} \end{cases}$$

If $\theta(\overline{x}, \overline{X}) := (P(Y) \leq r)\varphi(\overline{x}, \overline{X}, Y)$ then $\theta(\overline{x}, \overline{X})_{-\epsilon}$ is

$$\begin{cases} (P(Y) \leq \frac{1}{(1+\epsilon)^{k-1}}[r - k\epsilon(1-\epsilon)^{k-1}])(\varphi(\overline{x}, \overline{X}, Y)_{\epsilon}) & \textit{if } \frac{r}{(1+\epsilon)^{k-1}} - k\epsilon > 0 \\ (P(Y) \leq 0)(\varphi(\overline{x}, \overline{X}, Y)_{-\epsilon}) & \textit{otherwise} \end{cases}$$

The basic link between positive and negative approximate formulas, and the formula that they approximate is given by the following lemma.

Lemma 3.4 *For every formula $\theta(\overline{x}, \overline{X}) \in \mathcal{SOLP}(\tau)$, for every finite τ–structure $\mathcal{A}$, for every collection of sets $\overline{A}$ in $\mathcal{A}$, for every tuple of elements $\overline{a}$ in $\mathcal{A}$ and for ϵ and δ such that $0 < \delta < \epsilon < 1$, we have that:*

$$\mathcal{A} \models \theta(\overline{a}, \overline{A})_{-\epsilon} \to \theta(\overline{a}, \overline{A})_{-\delta} \to \theta(\overline{a}, \overline{A}) \to \theta(\overline{a}, \overline{A})_{\delta} \to \theta(\overline{a}, \overline{A})_{\epsilon}.$$

Furthermore, for every formula $\theta(\overline{x}, \overline{X}) \in \mathcal{SOLP}(\tau)$, for every ϵ with $0 < \epsilon < 1$

$$(\theta(\overline{x}, \overline{X})_{-\epsilon})_\epsilon = \theta(\overline{x}, \overline{X}) = (\theta(\overline{x}, \overline{X})_\epsilon)_{-\epsilon}.$$

Proof: If $\theta := (P(X) \geq r)\psi(X)$, with X of arity $k \geq 1$, then the chain of implications hold because, for $0 < \delta < \epsilon < 1$,

$$P(X) \geq \frac{r}{(1-\epsilon)^{k-1}} + \epsilon k > \frac{r}{(1-\delta)^{k-1}} + \delta k > r > (1-\delta)^{k-1}(r - \delta k) > (1-\epsilon)^{k-1}(r - \epsilon k)$$

and, if $\theta := (P(X) \leq r)\psi(X)$,

$$P(X) \leq \frac{r}{(1+\epsilon)^{k-1}} - \epsilon k < \frac{r}{(1+\delta)^{k-1}} - \delta k < r < (1+\delta)^{k-1}(r+\delta k) < (1+\epsilon)^{k-1}(r+\epsilon k)$$

The second part follows by easy substitution. $\square$

We will now show that it is possible to jump from satisfaction in almost order (respectively, linearly ordered) structures to satisfaction of approximate formulas in linearly ordered (respectively, almost ordered) structures.

Theorem 3.5 (Bridge Theorem) *Fix a sublinear function g and an almost order $\leq_g$. For every formula $\theta(x_1, \ldots, x_k, X) \in \mathcal{SOLP}(\tau)$, for every τ-structure $\mathcal{A}$ of size m and consistent with $\leq_g$, for every $\overline{a} = (a_1, \ldots, a_k) \in A^k$, for every predicate S of arity $t \geq 1$, the following holds:*

(i) $\mathcal{A} \models \theta(\overline{a}, S)$ *implies* $\mathcal{A}/_{\sim_g} \models \theta([\overline{a}]_g, [S]_g)_{\gamma(m)}$, *where* $\gamma(m) = \dfrac{g(m)}{2m - g(m)}$

(ii) $\mathcal{A}/_{\sim_g} \models \theta([\overline{a}]_g, [S]_g)$ *implies* $\mathcal{A} \models \theta(\overline{a}, S)_{\beta(m)}$, *where* $\beta(m) = \dfrac{g(m)}{2m}$

(iii) $\mathcal{A} \models \theta(\overline{a}, S)_{-\gamma(m)}$ *implies* $\mathcal{A}/_{\sim_g} \models \theta([\overline{a}]_g, [S]_g)$

(iv) $\mathcal{A}/_{\sim_g} \models \theta([\overline{a}]_g, [S]_g)_{-\beta(m)}$ *implies* $\mathcal{A} \models \theta(\overline{a}, S)$

Proof: By induction in the syntactic complexity of the formula.

Atomic formulas and negation of atomic formulas The result clearly follows because for atomic formulas and their negation θ_ϵ $(-1 < \epsilon < 1)$ is the same as θ.

Conjunction, disjunction Direct.

First order quantification The key tool is Lemma 2.7 which guarantees that it is indistinct which representative of a $\sim_g$-class we take as witnesses for the existentially or universally quantified variables, together with the fact that, for any ϵ, the ϵ–approximation coincides with the original formula.

Proportional quantifiers (i): Suppose that $\mathcal{A}$ satisfies the formula $(P(Y) \geq r)\theta(\overline{a}, \overline{S}, Y)$ for $0 < r < 1$ and Y of arity $k \geq 1$. Then, for some $B \subseteq A^k$, $|B| \geq rm^k$ and $\mathcal{A} \models \theta(\overline{a}, \overline{S}, B)$. By inductive hypothesis $\mathcal{A}/_{\sim_g} \models \theta([\overline{a}]_g, [\overline{S}]_g, [B]_g)_{\gamma(m)}$, where $\gamma(m) = g(m)/(2m - g(m))$. In the worst case, B contains elements from every $\leq_g$–2-preorder, and when passing to its $\leq_g$-contraction, all possible equivalent k–tuples determined by elements in the same class are removed. There are at most $k(g(m)/2)m^{k-1}$ of these and thus, $|[B]_g| \geq rm^k - k\dfrac{g(m)}{2}m^{k-1}$ provided $rm > k(g(m)/2)$; otherwise, we can only say $|[B]_g| \geq 0$. The proportion of this set of $\sim_g$–classes with respect to the totality of k–tuples in $[A]_g$ is

$$\begin{aligned}
P([B]_g) &\geq \left(\frac{2m}{2m - g(m)}\right)^{k-1} \left[r\left(\frac{2m}{2m - g(m)}\right) - k\frac{g(m)}{2m - g(m)}\right] \\
&= (1+\gamma(m))^{k-1}[r(1+\gamma(m)) - k\gamma(m)] \\
&= (1+\gamma(m))^{k-1}[r - (k-r)\gamma(m)] \\
&\geq (1-\gamma(m))^{k-1}[r - k\gamma(m)]
\end{aligned}$$

Thus,

$$\mathcal{A}/\!\sim_g \models (P(Y) \geq (1-\gamma(m))^{k-1}[r - k\gamma(m)])\theta([\overline{a}]_g, [\overline{S}]_g, Y)_{\gamma(m)}$$

which is the desired result.

Now, suppose that $\mathcal{A}$ satisfies the formula $(P(Y) \leq r)\theta(\overline{a}, \overline{S}, Y)$, with r and Y as above. We argue similarly as before, but now the witness set B is such that, in the worst case,

$$|[B]_g| \leq rm^k$$

which is the following proportion of $|[A]_g|^k = (m - g(m)/2)^k$:

$$\begin{aligned} P([B]_g) &\leq \left(\frac{2m}{2m-g(m)}\right)^{k-1}\left[r\left(\frac{2m}{2m-g(m)}\right)\right] \\ &= (1+\gamma(m))^{k-1} r(1+\gamma(m)) \\ &\leq (1+\gamma(m))^{k-1}[r + k\gamma(m)] \end{aligned}$$

Thus,

$$\mathcal{A}/\!\sim_g \models \Big(P(Y) \leq (1+\gamma(m))^{k-1}[r + k\gamma(m)]\Big)\, \theta([\overline{a}]_g, [\overline{S}]_g, Y)_{\gamma(m)}$$

(ii): Suppose that $\mathcal{A}/\!\sim_g$ satisfies the formula $(P(Y) \geq r)\theta([\overline{a}]_g, [\overline{S}]_g, Y)$ for $0 < r < 1$ and Y of arity $k \geq 1$. Then, for some set C of k–tuples of $[A]_g$, $|C| \geq r(m - g(m)/2)^k$ and $\mathcal{A}/\!\sim_g \models \theta([\overline{a}]_g, [\overline{S}]_g, C)$. By inductive hypothesis $\mathcal{A} \models \theta(\overline{a}, \overline{S}, (C)^g)_{\beta(m)}$, where $\beta(m) = g(m)/2m$, and in the worst case we add nothing new to the expansion of C, that is, $|(C)^g| = |C|$. The proportion of this set with respect to the set of k–tuples over $\mathcal{A}$ is

$$\begin{aligned} P((C)^g) &\geq \left(\frac{2m-g(m)}{2m}\right)^{k-1} r\left(\frac{2m-g(m)}{2m}\right) \\ &= (1-\beta(m))^{k-1} r(1-\beta(m)) \\ &\geq (1-\beta(m))^{k-1}[r - k\beta(m)] \end{aligned}$$

Thus,

$$\mathcal{A} \models \Big(P(Y) \geq (1-\beta(m))^{k-1}[r - k\beta(m)]\Big)\, \theta(\overline{a}, \overline{S}, Y)_{\beta(m)}$$

which is the desired result.

Now suppose that $\mathcal{A}/\!\sim_g$ satisfies the formula $(P(Y) \leq r)\theta([\overline{a}]_g, [\overline{S}]_g, Y)$. By inductive hypothesis $\mathcal{A} \models \theta(\overline{a}, \overline{S}, (C)^g)_{\beta(m)}$, where $(C)^g$ is the expansion of $C \subseteq [A]_g^k$, with $|C| \leq r(m - g(m)/2)^k$, and in the worst case $|(C)^g| \leq r(m-g(m)/2)^k + k(g(m)/2)m^{k-1}$. The proportion of this set with respect to m^k is

$$\begin{aligned} P((C)^g) &\leq \left(\frac{2m-g(m)}{2m}\right)^{k-1}\left[r\left(\frac{2m-g(m)}{2m}\right) + k\frac{g(m)}{2m}\right] \\ &= (1-\beta(m))^{k-1}[r(1-\beta(m)) + k\beta(m)] \\ &\leq (1+\beta(m))^{k-1}[r + k\beta(m)] \end{aligned}$$

Thus,

$$\mathcal{A} \models \Big(P(Y) \leq (1+\beta(m))^{k-1}[r + k\beta(m)] \Big)\, \theta(\overline{a}, \overline{S}, Y)_{\beta(m)}$$

which is the desired result.

(iii) and (iv): Follow from parts (i) and (ii) and that $(\theta_{-\epsilon})_\epsilon = \theta$. For example, if $\mathcal{A}/_{\sim_g} \not\models \theta$ then $\mathcal{A}/_{\sim_g} \not\models (\theta_{-\gamma(m)})_{\gamma(m)}$, and by part (i) we get $\mathcal{A} \not\models \theta_{-\gamma(m)}$. This shows (iii). $\square$

The picture that we have relating satisfaction in the almost ordered world with satisfaction in the ordered world is the following (the horizontal arrows are given by Lemma 3.4 and the diagonal arrows by the Bridge Theorem):

$$\begin{tikzcd}
\mathcal{A} \models \theta_{-\gamma} \arrow[r] \arrow[dr] & \theta \arrow[r] \arrow[dr] & \theta_\beta & \text{(almost order)} \\
\mathcal{A}/_{\sim_g} \models \theta_{-\beta} \arrow[r] \arrow[ur] & \theta \arrow[r] \arrow[ur] & \theta_\gamma & \text{(order)}
\end{tikzcd}$$

We will be needing the following additional property on approximate formulas.

Lemma 3.6 *For every formula $\theta(\overline{x}, \overline{X}) \in \mathcal{SOLP}(\tau)$, for every rationals γ and λ, with $-1 \leq \gamma < \lambda \leq 1$, for every $\delta \in (\gamma, \lambda) \cap \mathbb{Q}$, there exists a rational $\mu > 0$ such that:*

- $(\delta - \mu, \delta + \mu) \subseteq (\gamma, \lambda)$*, and*
- *for every τ–structure $\mathcal{A}$ and for every collection of sets $\overline{A}$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:*

$$\mathcal{A} \models \theta(\overline{a}, \overline{A})_\gamma \rightarrow (\ \theta(\overline{a}, \overline{A})_\delta\ )_{-\mu} \rightarrow \theta(\overline{a}, \overline{A})_\delta \rightarrow (\ \theta(\overline{a}, \overline{A}\ )_\delta)_\mu \rightarrow \theta(\overline{a}, \overline{A})_\lambda$$

Proof: The proof is by induction in formulas. The first order case is direct. We shall then analyse formulas with proportional quantifiers.

Assume that the desired property holds for $\theta(\overline{x}, \overline{X}, Y)$.

Case 1: Consider the formula $\Psi(\overline{x}, \overline{X}) := (P(Y) \geq r)\theta(\overline{x}, \overline{X}, Y)$. Let

$$f(r,\omega) := \begin{cases} 1 & \text{if } \frac{r}{(1+\omega)^{k-1}} - k\omega \geq 1 \text{ and } \omega < 0 \\ \frac{r}{(1+\omega)^{k-1}} - k\omega & \text{if } \frac{r}{(1+\omega)^{k-1}} - k\omega \leq 1 \text{ and } \omega < 0 \\ (1-\omega)^{k-1}[r - k\omega] & \text{if } 0 \leq (1-\omega)^{k-1}[r-k\omega] \text{ and } \omega \geq 0 \\ & \\ 0 & \text{if } (1-\omega)^{k-1}[r-k\omega] < 0 \text{ and } \omega \geq 0 \end{cases}$$

be a function from $[0,1] \times (-1,1)$ onto $[0,1]$. Note that this function is continuous and for every $r \in [0,1]$ and $\epsilon \in (-1,1)$,

$$(\ (P(Y) \geq r)\theta(\overline{x}, \overline{X}, Y)\ )_\epsilon := (P(Y) \geq f(r,\epsilon)\ )(\theta(\overline{x}, \overline{X}, Y))_\epsilon.$$

Furthermore, for every $r \in [0,1]$, $f(r,\ )$ is a decreasing function with the property that $f(r,0) = r$. Fix then a nonempty interval $(\gamma, \lambda) \subseteq (-1,1)$ and a $\delta \in (\gamma, \lambda)$. By induction hypothesis there exists a μ_1 with $(\delta - \mu_1, \delta + \mu_1) \subseteq (\gamma, \lambda)$ and such that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}, B$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\mathcal{A} \models \theta(\overline{a}, \overline{A}, B)_\gamma \rightarrow (\ \theta(\overline{a}, \overline{A}, B)_\delta\ )_{-\mu_1} \rightarrow \theta(\overline{a}, \overline{A}, B)_\delta \rightarrow (\ \theta(\overline{a}, \overline{A}, B)_\delta\ )_{\mu_1} \rightarrow \theta(\overline{a}, \overline{A}, B)_\lambda.$$

Note that $f(f(r,\delta),0) = f(r,\delta)$. Note also that $f(r,\lambda) \leq f(r,\gamma)$. Then, since f is continuous, there exists a μ_2 such that, for all $\epsilon \in (\delta - \mu_2, \delta + \mu_2)$,

$$f(\ f(r,\delta), \epsilon) \in [f(r,\lambda), f(r,\gamma)]$$

Let $\mu = \min\{\mu_1, \mu_2\}$. From the previous remarks we know that $(\delta - \mu, \delta + \mu) \subseteq (\gamma, \lambda)$ and that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\begin{aligned} \mathcal{A} \models (P(Y) \geq f(r,\gamma))[\theta(\overline{a}, \overline{A}, Y)]_\gamma &\rightarrow (P(Y) \geq f(f(r,\delta), -\mu))[\theta(\overline{a}, \overline{A}, Y)_\delta\ ]_{-\mu} \\ \rightarrow (P(Y) \geq f(f(r,\delta),0))[\theta(\overline{a}, \overline{A}, Y)]_\delta &\rightarrow\ (P(Y) \geq f(f(r,\delta),\mu))[\theta(\overline{a}, \overline{A}, Y)_\delta\ ]_\mu \\ &\rightarrow (P(Y) \geq f(r,\lambda))[\theta(\overline{a}, \overline{A}, Y)]_\lambda. \end{aligned}$$

but this is exactly the desired result that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\mathcal{A} \models \Psi(\overline{a}, \overline{A})_\gamma \rightarrow (\ \Psi(\overline{a}, \overline{A})_\delta\ )_{-\mu} \rightarrow \Psi(\overline{a}, \overline{A})_\delta \rightarrow (\ \Psi(\overline{a}, \overline{A}\ )_\delta)_\mu \rightarrow \Psi(\overline{a}, \overline{A})_\lambda$$

Case 2: Consider now the formula $\Psi(\overline{x}, \overline{X}) := (P(Y) \leq r)\theta(\overline{x}, \overline{X}, Y)$. Let

$$h(r,\omega) := \begin{cases} 0 & \text{if } \frac{r}{(1-\omega)^{k-1}} + k\omega \leq 0 \text{ and } \omega < 0 \\ \frac{r}{(1-\omega)^{k-1}} + k\omega & \text{if } \frac{r}{(1-\omega)^{k-1}} + k\omega > 0 \text{ and } \omega < 0 \\ (1+\omega)^{k-1}[r + k\omega] & \text{if } (1+\omega)^{k-1}[r+k\omega] \leq 1 \text{ and } \omega \geq 0 \\ 1 & \text{if } (1+\omega)^{k-1}[r+k\omega] > 1 \text{ and } \omega \geq 0 \end{cases}$$

be a function from $[0,1] \times (-1,1)$ onto $[0,1]$. Note that this function is continuous and for every $r \in [0,1]$, $\epsilon \in (-1,1)$,

$$(\ (P(Y) \leq r)\theta(\overline{x}, \overline{X}, Y)\ )_\epsilon := (P(Y) \leq h(r,\epsilon)\ )(\theta(\overline{x}, \overline{X}, Y))_\epsilon.$$

Furthermore, for every $r \in [0,1]$, $h(r,\ )$ is an increasing function with the property that $h(r,0) = r$.

Fix then a nonempty interval $(\gamma, \lambda) \subseteq (-1,1)$ and a $\delta \in (\gamma, \lambda)$. By induction hypothesis there exists a μ_1 with $(\delta - \mu_1, \delta + \mu_1) \subseteq (\gamma, \lambda)$ and such that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}, B$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\mathcal{A} \models \theta(\overline{a}, \overline{A}, B)_\gamma \rightarrow (\ \theta(\overline{a}, \overline{A}, B)_\delta\ )_{-\mu_1} \rightarrow \theta(\overline{a}, \overline{A}, B)_\delta \rightarrow (\ \theta(\overline{a}, \overline{A}, B)_\delta\ )_{\mu_1} \rightarrow \theta(\overline{a}, \overline{A}, B)_\lambda.$$

Note that $h(\ h(r,\delta)\ ,0) = h(r,\delta)$. Note also that $h(r,\gamma) \leq h(r,\lambda)$. Then, since h is continuous, there exists a μ_2 such that, for all $\epsilon \in (\delta - \mu_2, \delta + \mu_2)$,

$$h(\ h(r,\delta)\ ,\epsilon) \in [h(r,\gamma), h(r,\lambda)]$$

Let $\mu = \min\{\mu_1, \mu_2\}$. From the previous remarks we know that $(\delta - \mu, \delta + \mu) \subseteq (\gamma, \lambda)$ and that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\begin{aligned}\mathcal{A} \models (P(Y) \leq h(r,\gamma))[\theta(\overline{a},\overline{A},Y)]_\gamma \rightarrow\ & (P(Y) \leq h(h(r,\delta),-\mu))[\theta(\overline{a},\overline{A},Y)_\delta\ ]_{-\mu} \\ \rightarrow (P(Y) \leq h(h(r,\delta),0))[\theta(\overline{a},\overline{A},Y)]_\delta \rightarrow\ & (P(Y) \leq h(h(r,\delta),\mu))[\theta(\overline{a},\overline{A},Y)_\delta\ ]_\mu \\ & \rightarrow (P(Y) \leq h(r,\lambda))[\theta(\overline{a},\overline{A},Y)]_\lambda.\end{aligned}$$

but this is exactly the desired result that for every model $\mathcal{A}$ and for every collection of sets $\overline{A}$ in $\mathcal{A}$ and elements $\overline{a}$ in $\mathcal{A}$, we have that:

$$\mathcal{A} \models \Psi(\overline{a},\overline{A})_\gamma \rightarrow (\ \Psi(\overline{a},\overline{A})_\delta\ )_{-\mu} \rightarrow \Psi(\overline{a},\overline{A})_\delta \rightarrow (\ \Psi(\overline{a},\overline{A}\ )_\delta)_\mu \rightarrow \Psi(\overline{a},\overline{A})_\lambda$$

This completes the proof of the theorem. $\square$

4 Strong expressibility

We define the idea of strong equivalence for two formulas as follows.

Definition 4.1 *Fix two sentences $\phi, \psi \in \mathcal{SOLP}$. We say that ϕ* **is strongly equivalent** *to ψ (in symbols $\phi \Leftrightarrow_S \psi$) iff there exists $\epsilon \in (0,1)$ such that in every model $\mathcal{A}$:*

$$\mathcal{A} \models \phi_\epsilon \rightarrow \psi_{-\epsilon} \quad \textit{and} \quad \mathcal{A} \models \psi_\epsilon \rightarrow \phi_{-\epsilon}.$$

The intuition is that two sentences that are strongly equivalent can be syntactically approximate as much as we like. Formally what this means is that, if $\phi \Leftrightarrow_S \psi$ then there exists an $\epsilon > 0$ such that for every $\beta, \gamma \in (-\epsilon, \epsilon)$, $\models \phi \leftrightarrow \phi_\beta \leftrightarrow \psi_\gamma \leftrightarrow \psi$. This follows from $\phi \Leftrightarrow_S \psi$ because, for every model $\mathcal{A}$:

$$\mathcal{A} \models \phi_\beta \rightarrow \phi_\epsilon \rightarrow \psi_{-\epsilon} \rightarrow \psi_\gamma \rightarrow \psi_\epsilon \rightarrow \phi_{-\epsilon} \rightarrow \phi_\beta.$$

Note that if ϕ is strongly equivalent to ψ then for every model $\mathcal{A}$, $\mathcal{A} \models \phi \leftrightarrow \psi$ (i.e. ϕ and ψ are equivalent). This holds because if $\phi \Leftrightarrow_S \psi$, then in every model $\mathcal{A}$, using Lemma 3.4 we get

$$\mathcal{A} \models \phi \rightarrow \phi_\epsilon \rightarrow \psi_{-\epsilon} \rightarrow \psi \rightarrow \psi_\epsilon \rightarrow \phi_{-\epsilon} \rightarrow \phi.$$

Conversely, if ϕ is not equivalent to ψ then ϕ is not strongly equivalent to ψ.

Note also that it is not clear at all that $\phi \Leftrightarrow_S \phi$. The next example proves that this happens sometimes.

Example 4.2 *The property of being 2-colorable can be expressed in $\mathcal{SOMLP}[2]$ as follows. Let X and Y be two unary second order variables, and let $\theta(X,Y)$ be a formula that says that*

$$\begin{array}{rcl} (X \textit{ and } Y \textit{ are disjoint }) & \wedge & \\ \forall x \forall y(((X(x) \vee Y(x)) \wedge E(x,y)) & \rightarrow & \neg(X(y) \vee Y(y))) \ \wedge \\ ((\neg(X(x) \vee Y(x)) \wedge E(x,y)) & \rightarrow & (X(y) \vee Y(y))) \end{array}$$

Then the $\mathcal{SOMLCP}[2](\{E\})$ sentence

$$\phi := (P(X) \leq (1/2))(P(Y) \leq (1/2))\theta(X, Y)$$

holds in a graph, if and only if the graph is 2-colorable. (The idea is that X and Y constitute a partition of one of the possible two colors; in fact the color applied to the fewest number of vertices.)

Now observe that for every $\epsilon << 1/2$, if $\mathcal{A} \models \phi_\epsilon$ then

$$\mathcal{A} \models (P(X) \leq (1/2 + \epsilon))(P(Y) \leq (1/2 + \epsilon))\theta(X, Y)$$

It follows that $\mathcal{A}$ is 2-colorable, and in consequence

$$\mathcal{A} \models \phi_{-\epsilon} := (P(X) \leq (1/2 - \epsilon))(P(Y) \leq (1/2 - \epsilon))\theta(X, Y)$$

We proceed to define the approximate logics.

Definition 4.3 *Fix a logic $\mathcal{L} \subseteq \mathcal{SOLP}$ and an $\epsilon \in (-1, 1) \cap \mathbb{Q}$. The ϵ-approximation of $\mathcal{L}$, denoted $\mathcal{L}_\epsilon$, is the following fragment of $\mathcal{SOLP}$:*

$$\{\phi_\epsilon : \phi \in \mathcal{L}\}$$

By convention we define $\mathcal{L}_0 = \mathcal{L}$. The approximation of $\mathcal{L}$ (or the approximate logic corresponding to $\mathcal{L}$) is the set of formulas $\mathcal{L}_A := \bigcup_{\epsilon \in (-1,1) \cap \mathbb{Q}} \mathcal{L}_\epsilon$

We are interested in fragments of $\mathcal{SOLP}$ that behave "decently" for the notion of strong equivalence, i.e. where at least we can ask that for every formula ϕ in the fragment $\phi \Leftrightarrow_S \phi$.

Definition 4.4 *We say that a fragment $\mathcal{L}$ of $\mathcal{SOLP}$ is* ϵ**-relaxed** *if:*

- *For every $\delta \in (-\epsilon, \epsilon) \cap \mathbb{Q}$, $\mathcal{L}_\delta = \mathcal{L}$ (i.e., their expressive power is the same).*

Two important examples of ϵ–relax logics are the languages $\mathcal{SOLP}Horn[2]+\leq$ and $\mathcal{SOLP}Krom[2]+\leq$, which were defined and studied in [1] (see also section 2.1 above), and which capture **P** and **NL**, respectively. A summary of the reasons why these languages are ϵ–relax is as follows: For any ϵ, the problem NOT-IN-CLOS$_{\leq 1/2+\epsilon}$ (example 2.2) is expressible in $(\mathcal{SOLP}Horn[2]+\leq)_\epsilon$, and it is complete for **P** via quantifier free first order reductions (same proof as in [1]). Therefore, any problem in **P** has a definition in $(\mathcal{SOLP}Horn[2]+\leq)_\epsilon$. Conversely, the satisfaction of sentences in $(\mathcal{SOLP}Horn[2]+\leq)_\epsilon$ can be decided in **P** by the algorithm described in [1] for $(\mathcal{SOLP}Horn[2]+\leq)$. Thus, $(\mathcal{SOLP}Horn[2]+\leq)_\epsilon = \mathbf{P} = (\mathcal{SOLP}Horn[2]+\leq)$. The argument for $(\mathcal{SOLP}Krom[2]+\leq)_\epsilon = \mathbf{NL} = (\mathcal{SOLP}Krom[2]+\leq)$ is similar.

The main property of relaxed fragments is the following:

Lemma 4.5 *Let $\mathcal{L}$ be a ϵ-relaxed fragment of $\mathcal{SOLP}$. Then for every sentence $\phi \in \mathcal{L}$, there exists a $\lambda \in (-\epsilon, \epsilon) \cap \mathbb{Q}$ and sentence $\theta \in \mathcal{L}_\lambda$ such that $\phi \leftrightarrow \theta$ and $\theta \Leftrightarrow_S \theta$.*

Proof: Fix a sentence $\phi \in \mathcal{L}$. For every $\delta \in (-\epsilon, \epsilon)$ there exists then a sentence $\phi[\delta] \in \mathcal{L}$ such that $(\phi[\delta])_\delta \leftrightarrow \phi$. The cardinality of all the sentences in $\mathcal{L}$ is countable. Hence by the pigeonhole principle there exists a sentence $\theta \in \mathcal{L}$ and two rational numbers $\gamma < \delta \in (-\epsilon, \epsilon)$ such that $\theta_\gamma \leftrightarrow \phi \leftrightarrow \theta_\delta$. By the properties of approximate formulas (Lemma 3.6) we know that there exists rationals λ, μ such that $\lambda \in (\gamma, \delta) \subseteq (-\epsilon, \epsilon)$, $\mu > 0$, and

$$\phi \rightarrow \theta_\gamma \rightarrow (\theta_\lambda)_{-\mu} \rightarrow \theta_\lambda \rightarrow (\theta_\lambda)_\mu \rightarrow \theta_\delta \rightarrow \phi.$$

hence $\theta_\lambda \Leftrightarrow_S \theta_\lambda \leftrightarrow \phi$. $\square$

The previous lemma motivates our notion of strong expressibility.

Definition 4.6 *Let $\mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{SOLP}$ and fix $\phi \in \mathcal{SOLP}$ a sentence. We say that the fragment $\mathcal{L}$* **strongly expresses** *a sentence ϕ with respect to $\mathcal{L}'$ iff there exists a formula $\psi \in \mathcal{L}$ and a formula $\theta \in \mathcal{L}'$ such that $\theta \Leftrightarrow_S \psi$ and $\theta \leftrightarrow \phi$.*

Clearly, if a fragment $\mathcal{L}$ strongly expresses a sentence ϕ (with respect to any extension), then $\mathcal{L}$ expresses the sentence ϕ (because $\theta \Leftrightarrow_S \psi$ implies $\theta \leftrightarrow \psi$). Conversely, if a fragment $\mathcal{L}$ does not expresses ϕ then the fragment $\mathcal{L}$ does not strongly expresses ϕ.

When we are working with relaxed fragments, we get the following strengthening of the above observations.

Theorem 4.7 *Let $\mathcal{L}, \mathcal{L}'$ be ϵ-relaxed fragments of $\mathcal{SOLP}$ such that $\mathcal{L} \subseteq \mathcal{L}'$ and let $\phi \in \mathcal{SOLP}$ a sentence. Then the following statements are equivalent:*

- *ϕ is expressible in $\mathcal{L}$;*
- *There exists a $\mu \in (-\epsilon, \epsilon) \cap \mathbb{Q}$ such that ϕ is strongly expressible in $\mathcal{L}_\mu$ with respect to $\mathcal{L}'_\mu$, i.e. there exists sentences $\rho \in \mathcal{L}', \theta \in \mathcal{L}_\mu$ such that $\phi \leftrightarrow \rho_\mu$ and $\rho_\mu \Leftrightarrow_S \theta$.*

Proof: Suppose first that there exists a $\mu \in (-\epsilon, \epsilon)$ and sentences $\rho \in \mathcal{L}', \theta \in \mathcal{L}_\mu$ such that $\phi \leftrightarrow \rho_\mu$ and $\rho_\mu \Leftrightarrow_S \theta$. We can conclude then that $\phi \leftrightarrow \theta$. Since the expressive power of $\mathcal{L}_\mu$ is the same as the expressive power of $\mathcal{L}$ we can conclude that ϕ is expressible in $\mathcal{L}$.

For the other direction, assume that ϕ is expressible in $\mathcal{L}$. Note first that from lemma 4.5, since $\mathcal{L}'$ is an ϵ-relaxed fragment, we know that there exists a sentence $\rho \in \mathcal{L}'_\lambda$ for some $\lambda \in (-\epsilon, \epsilon)$ such that $\phi \leftrightarrow \rho$ and $\rho \Leftrightarrow_S \rho$. More specifically there exists γ such that $(\lambda - \gamma, \lambda + \gamma) \subseteq (-\epsilon, \epsilon)$ and $\rho_\gamma \leftrightarrow \rho_{-\gamma}$.

From hypothesis we know that there exists $\theta[\lambda] \in \mathcal{L}_\lambda$ such that $\theta[\lambda] \leftrightarrow \phi \leftrightarrow \rho$. Applying again lemma 4.5 to $\theta[\lambda]$ and using the fact that $\mathcal{L}_\lambda$ is a γ-relaxed fragment, we know that there exists a sentence $\theta \in \mathcal{L}_\mu$ for some μ in $(\lambda - \gamma, \lambda + \gamma) \cap \mathbb{Q}$ such that $\theta[\lambda] \leftrightarrow \theta$ and $\theta \Leftrightarrow_S \theta$. More specifically there exists ω such that $(\mu - \omega, \mu + \omega) \subseteq (\lambda - \gamma, \lambda + \gamma) \subseteq (-\epsilon, \epsilon)$ and $\theta_\omega \leftrightarrow \theta_{-\omega}$.

We have then the following sequences of implications:

$$\rho_\gamma \rightarrow \rho_{-\gamma} \rightarrow \rho \rightarrow \theta[\lambda] \rightarrow \theta \rightarrow \theta_\omega \rightarrow \theta_{-\omega}$$

and symmetrically,

$$\theta_\omega \rightarrow \theta_{-\omega} \rightarrow \theta \rightarrow \theta[\lambda] \rightarrow \rho \rightarrow \rho_\gamma \rightarrow \rho_{-\gamma}$$

These two sequences of implications imply that $\rho_\mu \Leftrightarrow_S \theta$, with $\rho_\mu \in \mathcal{L}'_\mu$, $\theta \in \mathcal{L}_\mu$ and $\phi \leftrightarrow \rho_\mu$. $\Box$

The importance of this lemma is that it shows the equivalence of the notion of expressibility and strong expressibility in the context of ϵ-relaxed fragments. This suggest that any tool that helps us prove strong inexpressibility may be transformed into a tool that proves inexpressibility. The rest of the paper is devoted to the exploration of this idea.

We begin by obtaining a tool, based in almost orders and ϵ–relaxed fragments, that proves strong inexpressibility, over ordered structures.

Theorem 4.8 *Fix fragments $(\mathcal{L}+\leq) \subseteq (\mathcal{L}'+\leq)$ of $\mathcal{SOLP}+\leq$ and a sentence $\phi \in (\mathcal{L}'+\leq)$. Suppose that $(\mathcal{L}+\leq)$ and $(\mathcal{L}'+\leq)$ are ϵ-relaxed. Assume finally that for every formula $\theta \in (\mathcal{L}+\leq)$ and every $0 < \omega < \epsilon$ there exists a sublinear function g and two models $\mathcal{A}, \mathcal{B}$ in $(\mathcal{L}+\leq_g)$ with the following properties:*

- *If $\mathcal{A} \models \theta$ then $\mathcal{B} \models \theta$;*
- *$\mathcal{A}/_{\sim_g} \models \phi$ and $\mathcal{B}/_{\sim_g} \not\models \phi$;*
- *if $|\mathcal{A}| = m_1$ and $|\mathcal{B}| = m_2$ then $g(m_i)/(2m_i - g(m_i)) < \omega$, for $i = 1, 2$.*

Then ϕ is not strongly expressible by $\mathcal{L}+\leq$.

Proof: Assume in order to get a contradiction that ϕ is strongly expressed in $(\mathcal{L}+\leq)$. Then there exists sentences $\rho \in (\mathcal{L}'+\leq)$, $\theta \in (\mathcal{L}+\leq)$ with $\rho \leftrightarrow \phi$ and $\theta \Leftrightarrow_S \rho$.

We know then that there exists a rational $\omega \in (0, \epsilon)$ such that for every model $\mathcal{C}$ in $\mathcal{SOLP}+\leq$ the following property $(*)$ holds:

- $\mathcal{C} \models \theta_\omega \to \rho_{-\omega}$,
- $\mathcal{C} \models \rho_\omega \to \theta_{-\omega}$.

Consider then the two models $\mathcal{A}, \mathcal{B}$ and the sublinear function g associated with $\phi, \epsilon, \theta, \omega$ by the hypothesis of the theorem. We consider two cases.

- If $\mathcal{A} \models \theta$ then by hypothesis we have that $\mathcal{B} \models \theta$. Applying now the Bridge Theorem we get that $(\mathcal{B}/_{\sim_g}) \models \theta_\omega$. However, since $(\mathcal{B}/_{\sim_g}) \not\models \phi$ and $\phi \leftrightarrow \rho$, we get that $(\mathcal{B}/_{\sim_g}) \not\models \rho_{-\omega}$, but this contradicts property $(*)$.
- If $\mathcal{A} \not\models \theta$ then by the Bridge Theorem we have that $(\mathcal{A}/_{\sim_g}) \not\models \theta_{-\omega}$. But by hypothesis $(\mathcal{A}/_{\sim_g}) \models \phi$ and $\phi \leftrightarrow \rho$, hence we get that $(\mathcal{A}/_{\sim_g}) \models \rho_\omega$, which is a contradiction with property $(*)$. $\Box$

The conditions in the preceding theorem are analogous to those in the tool we developed in [1] to separate logics with almost orders. This means we now have a tool that, working with the almost orders, helps us prove that statements are not strongly expressible in fragments of $\mathcal{SOLP}+\leq$. Actually, our previous work with almost orders gives us a notion of games that implies the hypothesis of the previous theorem. We have then a very general game theoretic tool that can prove strong inexpressibility in formulas with built-in order based on games in formulas with almost order.

The last question, naturally, is to see when strong inexpressibility is the same as inexpressibility. What we will do now is to use the previous theorem and Theorem 4.7 to obtain a tool, based on separation of almost orders that will prove inexpressibility in formulas with built-in order.

Theorem 4.9 *Fix fragments* $(\mathcal{L}+\leq) \subseteq (\mathcal{L}'+\leq)$ *of* $(\mathcal{SOLP}+\leq)$ *and a sentence* $\phi \in (\mathcal{L}'+\leq)$. *Suppose that* $(\mathcal{L}+\leq)$ *and* $(\mathcal{L}'+\leq)$ *are* ϵ*-relaxed. Assume finally that for every* $\mu \in (-\epsilon,\epsilon)\cap\mathbb{Q}$*, for every* $\omega > 0$ *such that* $(\mu-\omega,\mu+\omega) \subseteq (-\epsilon,\epsilon)$*, for every formula* $\theta \in (\mathcal{L}_\mu+\leq)$ *there exists a sublinear function* g *and two models* $\mathcal{A},\mathcal{B}$ *in* $(\mathcal{L}+\leq_g)$ *with the following properties:*

- *If* $\mathcal{A}\models\theta$ *then* $\mathcal{B}\models\theta$*;*
- $\mathcal{A}/_{\sim_g}\models\phi$ *and* $\mathcal{B}/_{\sim_g}\not\models\phi$*;*
- *if* $|\mathcal{A}| = m_1$ *and* $|\mathcal{B}| = m_2$ *then* $g(m_i)/(2m_i - g(m_i)) < \omega$*, for* $i = 1,2$*.*

then ϕ *is not expressed by* $(\mathcal{L}+\leq)$*.*

Proof: Assume in order to get a contradiction that ϕ is expressible in $(\mathcal{L}+\leq)$. Since $(\mathcal{L}+\leq) \subseteq (\mathcal{L}'+\leq)$ and $(\mathcal{L}+\leq)$ and $(\mathcal{L}'+\leq)$ are ϵ-relaxed we can invoke Theorem 4.7 to obtain that there exists a $\mu\in(-\epsilon,\epsilon)\cap\mathbb{Q}$ such that ϕ is strongly expressible in $(\mathcal{L}_\mu+\leq)$ with respect to $(\mathcal{L}'_\mu+\leq)$, i.e. there exists sentences $\rho\in(\mathcal{L}'+\leq),\theta\in(\mathcal{L}_\mu+\leq)$ such that $\phi\leftrightarrow\rho_\mu$ and $\rho_\mu\Leftrightarrow_S\theta$.

We know then that there exists a $0<\omega<1$ such that for every model $\mathcal{C}$ of $(\mathcal{SOLP}+\leq)$:

- $\mathcal{C}\models\theta_\omega\to(\rho_\mu)_{-\omega}$,
- $\mathcal{C}\models(\rho_\mu)_\omega\to\theta_{-\omega}$.

Note that we can select ω such that $(\mu-\omega,\mu+\omega)\subseteq(-\epsilon,\epsilon)$.

Consider then the two models $\mathcal{A},\mathcal{B}$ and the sublinear function g associated to ϕ,ϵ,μ,θ by the hypothesis of the theorem. We consider two cases.

- If $\mathcal{A}\models\theta$ then by hypothesis we have that $\mathcal{B}\models\theta$. Applying now the bridge theorem we get that $(\mathcal{B}/_{\sim_g})\models\theta_\omega$. However, since $(\mathcal{B}/_{\sim_g})\not\models\phi$ and $\phi\leftrightarrow\rho_\mu$, we get that $(\mathcal{B}/_{\sim_g})\not\models\rho_\mu$, but this contradicts the hypothesis that $\rho_\mu\Leftrightarrow_S\theta$.
- If $\mathcal{A}\not\models\theta$ then by the Bridge Theorem we have that $(\mathcal{A}/_{\sim_g})\not\models\theta_{-\omega}$. But by hypothesis $(\mathcal{A}/_{\sim_g})\models\phi$ and $\phi\leftrightarrow\rho_\mu$, hence we get that $(\mathcal{A}/_{\sim_g})\models(\rho_\mu)_\omega$, which is a contradiction with the hypothesis that $\psi\Leftrightarrow_S\theta$. $\square$

Note that the tool just developed works by checking properties of models in $\mathcal{SOLP}$ $+\leq_g$, i.e. models with almost orders, where separation proofs are easier. Note also that our prime examples of relaxed fragments where we could applied this tool are the $\mathcal{SOLP}Krom[2]+\leq$ and $\mathcal{SOLP}Horn[2]+\leq$ that correspond to the classes **NL** and **P**. Hence we have a tool, based on almost orders and approximate formulas that tells us that, in order to separate $\mathcal{SOLP}Krom[2]+\leq$ from $\mathcal{SOLP}Horn[2]+\leq$ with built-in order (which is hard), we only need to separate related logics in the context of almost orders. As we have seen from our previous work in almost orders we already have nice tools that do that (although in some limited context).

5 Further remarks on the complexity of expressibiliy

In this section we consider the situation where we are able to strongly separate $\mathcal{L}$ from $\mathcal{L}'$ by a sentence ϕ but ϕ is still expressible in $\mathcal{L}$. How is the behavior of the approximations of ϕ?

What we will present is a condition that says, basically, that if a sentence is not strongly expressible in a fragment but is expressible, is because something very ugly occurs. In the rest of the section we are going to formalize this idea.

Definition 5.1 *A sentence in $\mathcal{SOLP}$ is equivalent to one of the form*

$$\phi := Q_1X_1Q_2X_2\dots Q_rX_rA_1x_1A_2x_2\dots A_fx_f\bigvee_{i=1}^{q}\bigwedge_{j=1}^{t_i}\theta_{i,j}(\overline{X},\overline{x})$$

where the Q_sX_s are proportionality quantifiers over the second order variable X_s and the A_sx_s are either $\exists x_s$ or $\forall x_s$ with x_s a first order variable, and $\theta_{i,j}(\overline{X},\overline{x})$ is an atomic formula or negation of atomic formula with its first order free variable being members of $\overline{x}=(x_1,\dots,x_f)$, and its second order free variable (if any) being member of $\overline{X}=(X_1,\dots,X_r)$. Let m_ϕ be the maximum arity of the second order variables $X_1,\dots,X_r$. Then, the **complexity of the sentence** *ϕ is defined as the sum $r+f+m_\phi$.*

Let us return again to the scenario where we consider two ϵ-relaxed fragments $\mathcal{L}\subseteq\mathcal{L}'$ and a sentence $\phi\in\mathcal{L}'$ that is expressible in $\mathcal{L}$. Let ψ the sentence in $\mathcal{L}$ equivalent to ϕ. We want to see which condition will ensure that ϕ is strongly expressible in $\mathcal{L}$ with respect to $\mathcal{L}'$, i.e. there exists $\theta\in\mathcal{L}',\rho\in\mathcal{L}$ such that for every model $\mathcal{A}$, $\mathcal{A}\models\phi\leftrightarrow\theta$ and $\mathcal{A}\models\theta\Leftrightarrow_S\rho$. We know that for every $\delta\in(-\epsilon,\epsilon)$ there exists sentences $\theta[\delta]\in\mathcal{L},\rho[\delta]\in\mathcal{L}'$ such that $\theta[\delta]_\delta\leftrightarrow\phi\leftrightarrow\psi\leftrightarrow\rho[\delta]_\delta$. Suppose that we select the $\theta[\delta],\rho[\delta]$ to have minimal complexity, and furthermore, suppose that we have the following property (**):

> There exists a natural number M such that:
>
> - $\forall\delta$, with $0<\delta<\epsilon$, there exists α,β with $0<\alpha<\delta$ and $-\delta<-\beta<0$ such that the complexity of the sentences $\theta[\alpha]$ and $\theta[-\beta]$ is bounded by M.
> - $\forall\delta$, with $0<\delta<\epsilon$, there exists α',β' with $0<\alpha'<\delta$ and $-\delta<-\beta'<0$ such that the complexity of the sentences $\rho[\alpha']$ and $\rho[-\beta']$ is bounded by M.

Then, the pigeonhole principle implies that there exists sentences $\theta\in\mathcal{L}'$, $\rho\in\mathcal{L}$ such that:

- for every δ and ϵ, $0<\delta<\epsilon$, there exists α,β with $0<\alpha<\delta$ and $-\delta<-\beta<0$ with $\theta=\theta[-\beta]$ and $\theta=\theta[\alpha]$.
- for every δ and ϵ, $0<\delta<\epsilon$, there exists α',β' with $0<\alpha'<\delta$ and $-\delta<-\beta'<0$ with $\rho=\rho[-\beta']$ and $\rho=\rho[\alpha']$.

It follows then that there exists $\alpha_1, \beta_1 < \epsilon$ such that in every model $\mathcal{A}$:

$$\mathcal{A} \models \theta_{\alpha_1} \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow \rho_{-\beta_1}$$

Similarly we get that there exists $\alpha_2, \beta_2 < \epsilon$ such that in every model $\mathcal{A}$:

$$\mathcal{A} \models \rho_{\alpha_2} \leftrightarrow \psi \leftrightarrow \phi \leftrightarrow \theta_{-\beta_2}$$

Let $\delta = min(\alpha_1, \alpha_2, \beta_1, \beta_2)$. We have then that in every model $\mathcal{A}$:

$$\mathcal{A} \models \rho_\delta \rightarrow \rho_{\alpha_2} \rightarrow \theta_{-\beta_2} \rightarrow \theta_{-\delta},$$

and similarly we have that in every model $\mathcal{A}$:

$$\mathcal{A} \models \theta_\delta \rightarrow \theta_{\alpha_1} \rightarrow \rho_{-\beta_1} \rightarrow \rho_{-\delta},$$

The two statements above imply that $\phi \leftrightarrow \theta$ and $\psi \leftrightarrow \rho$ and $\rho \Leftrightarrow_S \theta$. In other words, we have the following lemma.

Lemma 5.2 *Consider two ϵ-relaxed fragments $\mathcal{L} \subseteq \mathcal{L}'$ and a sentence $\phi \in \mathcal{L}'$ that is expressible in $\mathcal{L}$. Let ψ be the sentence in $\mathcal{L}$ equivalent to ϕ. We know that for every $\delta \in (-\epsilon, \epsilon)$ there exists sentences $\theta[\delta] \in \mathcal{L}, \rho[\delta] \in \mathcal{L}'$ such that $\theta[\delta]_\delta \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow \rho[\delta]_\delta$.*

Suppose that there exists a natural number M such that:

- *$\forall \delta\, (0 < \delta < \epsilon)$ there exists α, β with $0 < \alpha < \delta$ and $-\delta < -\beta < 0$ such that the minimal complexity of the sentences $\theta[\alpha]$ and $\theta[-\beta]$ is bounded by M.*
- *$\forall \delta\, (0 < \delta < \epsilon)$ there exists α', β' with $0 < \alpha' < \delta$ and $-\delta < -\beta' < 0$ such that the minimal complexity of the sentences $\rho[\alpha']$ and $\rho[-\beta']$ is bounded by M.*

Then ϕ is strongly expressible in $\mathcal{L}$ with respect to $\mathcal{L}'$. $\Box$

The counterpositive of the above lemma is actually the result we are interested in.

Corollary 5.3 *Consider two ϵ-relaxed fragments $\mathcal{L} \subseteq \mathcal{L}'$ and a sentence $\phi \in \mathcal{L}'$ that is expressible in $\mathcal{L}$. Let ψ be the sentence in $\mathcal{L}$ equivalent to ϕ. We know that for every $\delta \in (-\epsilon, \epsilon)$ there exists sentences $\theta[\delta] \in \mathcal{L}, \rho[\delta] \in \mathcal{L}'$ such that $\theta[\delta]_\delta \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow \rho[\delta]_\delta$. Suppose that ϕ is not strongly expressible in $\mathcal{L}$ with respect to $\mathcal{L}'$. Then,*

- *there exists $\delta \in (0, \epsilon)$ such that for all $\alpha \in (0, \delta)$ the minimal complexity of the formulas $\theta[\alpha]$ is bigger than M; or*
- *there exists $\delta \in (0, \epsilon)$ such that for all $\alpha \in (-\delta, 0)$ the minimal complexity of the formulas $\theta[-\alpha]$ is bigger than M; or*
- *there exists $\delta \in (0, \epsilon)$ such that for all $\alpha' \in (0, \delta)$ the minimal complexity of the formulas $\rho[\alpha']$ is bigger than M; or*
- *there exists $\delta \in (0, \epsilon)$ such that for all $\alpha' \in (-\delta, 0)$ the minimal complexity of the formulas $\rho[-\alpha']$ is bigger than M.*

Here is a direct consequence of the above corollary. We know that $(\mathcal{SOLP}Krom[2]+\leq) \subseteq (\mathcal{SOLP}Horn[2]+\leq)$ are ϵ-relaxed fragments of $\mathcal{SOLP}$ that capture **NL** and **P** respectively. Suppose that you can prove that a problem Q in **P** is not strongly expressible in $(\mathcal{SOLP}Krom[2]+\leq)$ with respect to $(\mathcal{SOLP}Horn[2]+\leq)$ by using any of the tools at our disposal. Then if still Q was expressible in $(\mathcal{SOLP}Krom[2]+\leq)$ the previous corollary implies that there exists a $\delta \in (-\epsilon,\epsilon) - \{0\}$ such that the minimal complexity of the sentences $\theta_\omega \in (\mathcal{SOLP}Krom[2]+\leq)_\omega$ (for $\omega \in (0,\delta)$ or $(\delta,0)$) that capture Q tends to infinity, or there exists a $\delta \in (-\epsilon,\epsilon) - 0$ such that the minimal complexity of the sentences $\rho_\omega \in (\mathcal{SOLP}Horn[2]+\leq)_\omega$ (for $\omega \in (0,\delta)$ or $(\delta,0)$) that capture Q tends to infinity. That is a very strange phenomena!

References

[1] Arratia, A., and Ortiz, C. (2005) Counting proportions of sets: expressive power with almost order. (*To appear in: Latin American Theoretical INformatics, LATIN'06, Lecture Notes in Comp. Sci., Springer, 2006.*)

[2] Ebbinghaus, H.D., and Flum, J., *Finite Model Theory* (Springer-Verlag, 1995).

[3] Etessami, K., and Immerman, N. (1995), Reachability and the power of local ordering, *Theoretical Comp. Sci.* **148**, 2, 261–279.

[4] Grädel, E. (1992), Capturing complexity classes by fragments of second order logic, *Theoretical Comp. Sci.* **101**, 35–57.

[5] Immerman, N., *Descriptive Complexity* (Springer, 1998).

[6] H.J. Keisler, Hyperfinite model theory. In: R.C. Gandy and J.M. E. Hyland (eds) Logic Colloquium 76 (North-Holland,1977).

[7] Libkin, L., and Wong, L., (2002) Lower bounds for invariant queries in logics with counting. *Theoretical Comp. Sci.* **288**, 153-180.