
Approximate formulae for a logic that capture classes of computational complexity

ARGIMIRO ARRATIA*, *Dpto. de Matemática Aplicada Facultad de Ciencias Universidad de Valladolid Valladolid 47005, Spain*
E-mail: 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.*
E-mail: ortiz@arcadia.edu

Abstract

This paper presents a syntax of approximate formulae suited for the logic with counting quantifiers $SO\mathcal{L}\mathcal{P}$. This logic was formalised by us in [1] where, among other properties, we showed the following facts: (i) In the presence of a built-in (linear) order, $SO\mathcal{L}\mathcal{P}$ can describe \mathbf{NP} -complete problems and some of its fragments capture the classes \mathbf{P} and \mathbf{NL} ; (ii) weakening the ordering relation to an almost order we can separate meaningful fragments, using a combinatorial tool adapted to these languages.

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

Subject Classification: Logic in computer science; Descriptive Complexity.

Keywords: Proportional quantifiers, approximate formulae, almost order, expressiveness, computational complexity, \mathbf{P} , \mathbf{NL} .

1 Introduction

Descriptive Complexity deals mainly with producing logics that define all problems of particular computational complexity, and adapting the classical tools for showing inexpressibility of queries in logics to the context of finite models, in the hope to obtain worthy lower bounds for computational classes such as \mathbf{P} or \mathbf{NP} . The limitations of this logical approach

*Supported by grants Ramón y Cajal (MEC+FEDER-FSE); MOISES (TIN2005-08832-C03-02) and SINGACOM (MTM2004-00958), MEC-Spain

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

to showing computational complexity bounds for classes like say, \mathbf{P} , \mathbf{NL} (*nondeterministic logspace*), and others within \mathbf{NP} , boils down to the fact that, as of today, all known logics that define problems in these classes need a relation of linear order built into their semantics; and in the presence of a built-in linear order it has been shown that logical inexpressibility tools such as Ehrenfeucht-Fraïssé games have little power for telling structures apart (e.g. see [6, § 6.6]). (The reader should be aware that this need of a built-in linear order and the issues that its presence or absence produces, does not concern the class \mathbf{NP} or the computational complexity classes above \mathbf{NP} , since by the well known result of Fagin \mathbf{NP} can be described by sentences of Existential Second Order logic, and in this logic one can express the existence of an arbitrary linear order.) On the other hand, in the absence of a built-in linear order, logics lose significantly expressive power: for example, first order logic (FO) extended with a least fixed point operator (LFP(FO)) with order captures all of \mathbf{P} (in the sense that it is capable of defining all polynomial time computable properties), but without order can not express the parity of the size of a set. 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 such as built-in orders are weakened to *almost-orders*, and, frequently, some counting operator is added to compensate for the loss of expressive power. This has been the typical approach within the Descriptive Complexity community (e.g. [3], [8] among others), and it has some severe limitations: for example, the paper by Libkin and Wong [8] shows that a very powerful extension of first order logic with additional counting quantifiers, known as $\mathcal{L}_{\infty\omega}^*(C)$, which subsumes various counting extensions of FO, 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.

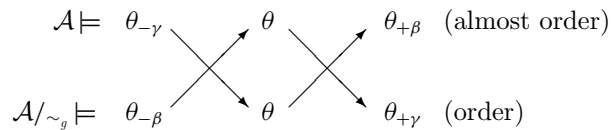
The other approach is *syntactic* and is found in classical model theory as in, for example, Keisler's logic of probability quantifiers (see [7]), who conceived it as a logic appropriate for his investigations on *probability hyperfinite spaces*, or infinite structures suitable for approximating large finite phenomena of applied mathematics. Under this approach, for each formula φ of a logic and every real number ϵ one constructs an approximate formula φ_ϵ with the property that in every model \mathcal{A} , if $\epsilon_1 < 0 < \epsilon_2$ then $\varphi_{\epsilon_1} \rightarrow \varphi \rightarrow \varphi_{\epsilon_2}$, and as ϵ tends to 0, the interpretation of φ_ϵ should be closer to φ . This approach has been developed with success in 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 the notion of approximate formulae *à la* Keisler, 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 result that implies that separation of logics with built-in almost-order can be translated into separation of corresponding approximate logics with built-in order.

The framework for our results is the second order logic of proportionality quantifiers, *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 bounded number of degrees property, 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 $0 < r < 1$, 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. Of particular interest will be the *Horn* and *Krom* fragments of \mathcal{SOLP} , which are defined after Gradel's Horn and Krom fragments of Second Order logic [4], and consisting of formulae formed with a block of our proportional quantifiers applied to formulae of type Horn with respect to the the second order variables (the Horn fragment) or to a Krom (or 2-CNF) type formula (for the Krom fragment). We review the definition of \mathcal{SOLP} , and its Horn and Krom fragments, and summarise facts found in [1] about their 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 $\psi_{+\epsilon}$ 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 formulae in \mathcal{SOLP} in almost ordered structures and satisfaction of the corresponding approximate formulae in ordered structures. This result we call Bridge Theorem (see section 3), and its contents is illustrated by the following picture:



What this says, for example, is that satisfaction in almost ordered structure \mathcal{A} of $\theta_{-\gamma}$, a (negative) approximation of formula θ , implies satisfaction of θ in corresponding ordered structure \mathcal{A}/\sim_g ; and the latter implies satisfaction back in \mathcal{A} of (positive) approximation $\theta_{+\beta}$. Similar path of satisfaction goes through beginning in ordered structure \mathcal{A}/\sim_g , as shown in the picture above.

In section 4 we introduce the notion of the ϵ -approximate logic \mathcal{L}_ϵ , for every fragment \mathcal{L} of \mathcal{SOLP} , which consists of all approximate formulae of formulae in \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 δ such that $-\epsilon < \delta < \epsilon$. Surprisingly, the Horn and Krom fragments of \mathcal{SOLP} with built-in order, which were shown in [1] to capture **P** and **NL** respectively, 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 4.13 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 ($-\epsilon < \delta < \epsilon$) 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 mentioned 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 formulae 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 formulae tend to the exact formula.

2 The second order logic of proportional quantifiers

Throughout this paper we use standard notation and concepts of Finite Model Theory as presented in the books by Ebbinghaus and Flum [2] and by Immerman [6]. Our vocabularies are finite and consist of relation symbols and constant symbols. Our structures are all finite, and if \mathcal{A} is a structure over vocabulary τ , or τ -structure, and A is its universe, we either use $|\mathcal{A}|$ or $|A|$ to denote its size, that is, the number of elements in A .

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

$$Q_1 \cdots Q_u \theta(x_1, \dots, x_s, X_1, \dots, X_r) \quad (1)$$

where $\theta(x_1, \dots, x_s, X_1, \dots, X_r)$ is a first order formula over some vocabulary τ with (free) first order variables x_1, \dots, x_s and second order variables, X_1, \dots, 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 such that $0 < t_i < 1$, for some $i \leq r$. Whenever we want to make the underlying vocabulary τ explicit we will write $\mathcal{SO}\mathcal{LP}(\tau)$.

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

Another fragment of $\mathcal{SO}\mathcal{LP}$ which will be of interest for us is the Monadic Second Order Logic of Proportional quantifiers, denoted $\mathcal{MSO}\mathcal{LP}$, which is $\mathcal{SO}\mathcal{LP}$ 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 , \bar{Y} a vector of second order variables, $\bar{x} = x_1, \dots, x_m$ first order variables and $\phi(\bar{x}, \bar{Y}, X)$ a formula in $\mathcal{SO}\mathcal{LP}(\tau)$ over some (finite) vocabulary τ (which does not contain X or any of the variables in \bar{Y} as a relation symbol). Let r be a rational such that $0 < r < 1$. Then the formula

$$(P(X) \geq r)\phi(\bar{x}, \bar{Y}, X)$$

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

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

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

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 a built-in order and when this external predicate is weakened to an almost order (see [6] 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 $\mathcal{SOLPHorn}$ and $\mathcal{SOLPKrom}$, which were defined in [1] after Grädel’s definitions of the Horn and Krom fragments of Second Order logic in [4].

A first order formula α over a vocabulary τ plus second order variables X_1, \dots, X_r of arities k_1, \dots, 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 formulae over $\tau \cup \{X_1, \dots, X_r\}$ of the form $(\psi_1 \wedge \psi_2 \wedge \dots \wedge \psi_s) \rightarrow \varphi$, where φ is either $X_i(\bar{u}_i)$ (where \bar{u}_i denotes a k_i -tuple of first order terms, $i = 1, \dots, r$) or \perp , and ψ_1, \dots, ψ_s are atomic or negation of atomic ($\tau \cup \{X_1, \dots, X_r\}$)-formulae except that any occurrence of the variables X_i must be positive (there are no restrictions on the predicates in τ or $=$). The logic $\mathcal{SOLPHorn}$ is the set of formulae 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, \dots, X_r .

Example 2.2 *We present a problem definable in $\mathcal{SOLPHorn}$. 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*

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

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

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

Then

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

In [1] it is shown that, for $r = 1/n$, this problem is complete for \mathbf{P} under first order reductions with built-in successor. ■

Remark 2.3 *The problem NOT-IN-CLOS is related to the complement of the Path System Accessibility problem in the sense that one is reducible to the other via first order definable reductions. We shall make this fact precise in the proof of Theorem 4.6, where we shall be needing it. An instance of the Path System Accessibility problem (abbreviated in the literature as PS, e.g. [9]) is a finite structure $\mathcal{A} = \langle A, R, s, t \rangle$ or a path system, where the universe A consists of, say, n vertices, a relation $R \subseteq A \times A \times A$ (the rules of the system), a source $s \in A$, and a target $t \in A$ such that $s \neq t$. A positive instance of PS is a path system \mathcal{A} where the target is accessible from the source, where a vertex v is accessible if it is the source s or if*

$R(x, y, v)$ holds for some accessible vertices x and y , possibly equal. In [9] Stewart shows that PS is complete for \mathbf{P} via quantifier free first order reductions that include built-in successor relation (see [9] for details). Since \mathbf{P} is closed under complement and also closed under the aforementioned first order reductions, it follows that the complement of PS is also complete for \mathbf{P} (and by the opening comments in this remark this also holds for NOT-IN-CLOS).

A first order formula α over $\tau \cup \{X_1, \dots, X_r\} \cup \{=, \perp\}$ 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, \dots, X_r , i.e. α is a 2-CNF formula with respect to the variables X_1, \dots, X_r . The logic *SOLPKrom* is the set of formulae 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, \dots, X_r .

Example 2.4 We now present a problem definable in *SOLPKrom*. 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 specified vertex s (the source). Let r be a rational with $0 < r < 1$. We define

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

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

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

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

The problem $\text{NCON}_{\geq 1/2}$ has been shown to be complete for \mathbf{NL} under first order reductions with built-in successor (see [1]). ■

Remark 2.5 The problem NCON is related to the complement of the Transitive Closure (or TC) problem (see [6]) in the sense that one is reducible to the other via first order definable reductions. We shall use this fact in the proof of Theorem 4.7. The problem TC is known to be complete for \mathbf{NL} via first order reductions [6]. On the other hand, \mathbf{NL} is closed under complement by a remarkable result of Immerman [5], and independently by Szelepcsényi [10], and it is also closed under first order reductions. Then it follows that the complement of TC and the problem NCON are complete for \mathbf{NL} .

We have shown in [1] that:

- (1) In the presence of order (at least a built-in successor), $\mathbf{P} \subseteq \text{SOLP}[2]$ (in the sense that any class of structures decidable in \mathbf{P} is definable by a sentence of *SOLP*[2]) and, furthermore, it is captured by the fragment *SOLPHorn*[2], consisting of formulae 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 *SOLPKrom*[2], a fragment consisting of formulae 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 *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 a strict hierarchy within the monadic fragment \mathcal{MSOLP} , namely,

$$\mathcal{MSOLP}[2] \not\subseteq \mathcal{MSOLP}[2,3] \not\subseteq \mathcal{MSOLP}[2,3,5] \not\subseteq \dots$$

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

$$\mathcal{SOLPHorn}[2] \not\subseteq \mathcal{SOLP}[2,3].$$

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

The concept of almost order (inspired from [8]) constitute the core of our “semantic approximations”, around which we work our syntactic approximations.

Definition 2.6 (Almost order). *Let $g: \mathbb{N} \rightarrow \mathbb{R}$ be a sublinear and non-decreasing function (that is, for all n in \mathbb{N} , $g(n) < n$ and for all n, m in \mathbb{N} , if $n \leq m$ then $g(n) \leq g(m)$).*

An almost order over a structure \mathcal{A} induced by g is a binary relation \leq_g over A such that there is a partition of A into two sets B and C satisfying:

- *The cardinality of B is at least $n - g(n)$, where n is the size of A ;*
- *The restriction of \leq_g to B is a linear order;*
- *The restriction of \leq_g to C is reflexive and transitive where every equivalence class of \sim_g has size at most 2 (we write $x \sim_g y$ iff both $x \leq_g y$ and $y \leq_g x$); and,*
- *for any b in B and any c in C , $b \leq_g c$ holds and $c \leq_g b$ does not.*

Note that for any function $g: \mathbb{N} \rightarrow \mathbb{R}$, 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 $[A]_g := \{[a]_g : a \in A\}$. Observe that if $|A| = n$ then $|[A]_g| = n - g(n)/2$.

Definition 2.7 *Fix a sublinear $g: \mathbb{N} \rightarrow \mathbb{R}$ and let R be a k -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_k) and (b_1, \dots, b_k) of elements in A with $a_i \sim_g b_i$ for every $i \leq k$, we have that*

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

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

Let $\text{cons-ao}(\mathcal{A}, g)$ denotes the set of almost orders over \mathcal{A} induced by the function g that are consistent with \mathcal{A} .

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 a function g , we understand the logic where we consider models \mathcal{A} together with a built-in *arbitrary* almost order \leq_g in $\text{cons-ao}(\mathcal{A}, g)$. Furthermore, for the formulae of the form $(P(X) \geq r)\phi(\bar{x}, \bar{Y}, X)$ and $(P(X) \leq r)\phi(\bar{x}, \bar{Y}, X)$, we require the following modification of the semantics: For an appropriate finite model \mathcal{A} , for a sublinear function g and an almost order \leq_g in $\text{cons-ao}(\mathcal{A}, g)$, for elements $\bar{a} = (a_1, \dots, a_m)$ in A and an appropriate vector of relations \bar{B} , consistent with \leq_g , we should have

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

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

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.

Remark 2.8 *Our use of \leq_g as a built-in construct may seem to differ from common knowledge and usage of built-in relations (as numeric relations whose value only depends on the size of the structure), for it seems tied up to the particular characteristics of the working structures as it requires that every relation be consistent with \leq_g (Definition 2.7). We argue here that is not the case, and make some provisos that will clarify this matter. Indeed, any sentence ϕ defining some relation is satisfied by an almost ordered structure (a structure with the additional built-in \leq_g), provided the truth of ϕ is conditioned to that part of the structure consistent with \leq_g . However, we can explicitly free our logic from this apparent dependency, by noting that our logic (and all fragments we consider) is strong enough to express that the almost order is consistent with respect to the input structure (this can be done within first order logic); thus we can assume (and ask the reader to assume) that every sentence like ϕ comes joint with a guard that says “the almost order is consistent”. In this way, our built-in \leq_g has the numeric interpretation of any other built-in, depending only on the size of the input structure.*

Finally, for two logics \mathcal{L} and \mathcal{L}' , whenever we write the inclusion $\mathcal{L} \subseteq \mathcal{L}'$ this is meant in terms of expressive power.

3 A syntax of approximate formulae

We now introduce the notion of approximate formulae for \mathcal{SOLP} . The purpose of these formulae 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 3.6 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 Formulae). *For every ϵ such that $0 \leq \epsilon < 1$, and for every formula $\theta(\bar{x}, \bar{X}) \in \mathcal{SOLP}(\tau)$, we define the (positive) ϵ -approximation of $\theta(\bar{x}, \bar{X})$, denoted $\theta(\bar{x}, \bar{X})_{+\epsilon}$, as follows:*

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

Proportional quantifiers If $\theta(\bar{x}, \bar{X}) := (Q_1 \dots Q_u)\varphi(\bar{x}, \bar{X})$, where $\varphi(\bar{x}, \bar{X})$ is a first-order formula and Q_1, \dots, Q_u are proportional quantifiers, its ϵ -approximation is the \mathcal{SOLP} -formula $(\theta(\bar{x}, \bar{X}))_{+\epsilon} := (Q'_1 \dots Q'_u)\varphi(\bar{x}, \bar{X})$, where, for each j , the proportional quantifier Q'_j is chosen as follows:

(a) If Q_j is of the form $(P(Y) \geq r)$, where Y is of arity $k \geq 1$, then Q'_j is of the form

$$\begin{cases} (P(Y) \geq (1-\epsilon)^{k-1}[r - k\epsilon]) & \text{if } r - k\epsilon > 0 \\ (P(Y) \geq 0) & \text{otherwise} \end{cases}$$

(b) If Q_j is of the form $(P(Y) \leq r)$, then Q'_j is of the form

$$\begin{cases} (P(Y) \leq (1+\epsilon)^{k-1}[r + k\epsilon]) & \text{if } (1+\epsilon)^{k-1}(r + k\epsilon) < 1 \\ (P(Y) \leq 1) & \text{otherwise} \end{cases}$$

(The 0-approximation of $\theta(\bar{x}, \bar{X})$ is clearly itself. In this case we will always drop the 0 in $\theta(\bar{x}, \bar{X})_0$.)

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

The previous definition describes syntactic approximations “from the right” or “positive”. We can also have approximations from the left or negative (our intended meaning for right or left approximations 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 propose the following definition.

Definition 3.3 (Approximate Formulae for Negative Values). For every ϵ such that $0 \leq \epsilon < 1$, and for every formula $\theta(\bar{x}, \bar{X}) \in \mathcal{SOLP}(\tau)$, we define the $-\epsilon$ -approximation of $\theta(\bar{x}, \bar{X})$ by induction in the complexity of the formulae as follows:

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

Proportional quantifiers If $\theta(\bar{x}, \bar{X}) := (Q_1 \dots Q_u)\varphi(\bar{x}, \bar{X})$, where $\varphi(\bar{x}, \bar{X})$ is a first-order formula and Q_1, \dots, Q_u are proportional quantifiers, then $\theta(\bar{x}, \bar{X})_{-\epsilon} := (Q'_1 \dots Q'_u)\varphi(\bar{x}, \bar{X})$, where, for each j , the proportional quantifier Q'_j is chosen as follows:

(a) If Q_j is of the form $(P(Y) \geq r)$, where Y is of arity $k \geq 1$, then Q'_j is of the form

$$\begin{cases} (P(Y) \geq \frac{1}{(1-\epsilon)^{k-1}}[r + k\epsilon(1-\epsilon)^{k-1}]) & \text{if } \frac{r}{(1-\epsilon)^{k-1}} + k\epsilon < 1 \\ (P(Y) \geq 1) & \text{otherwise} \end{cases}$$

(b) If Q_j is of the form $(P(Y) \leq r)$, then Q'_j is of the form

$$\begin{cases} (P(Y) \leq \frac{1}{(1+\epsilon)^{k-1}}[r - k\epsilon(1+\epsilon)^{k-1}]) & \text{if } \frac{r}{(1+\epsilon)^{k-1}} - k\epsilon > 0 \\ (P(Y) \leq 0) & \text{otherwise} \end{cases}$$

Observe that when our proportional quantifiers are of monadic type, that is, they act upon second order variables of arity 1 like, say, $(P(Y) \geq r)$, with arity of Y equal to 1, then its ϵ -approximation, according to our definition, is what one would naturally expect to be, namely, $(P(Y) \geq r - \epsilon)$ (just set $k=1$ in the corresponding definition). Thus, our definition just generalises this natural notion of approximation for monadic predicates to the general case of quantifiers of any arity k . Furthermore, these definitions for syntactic approximations are adequate for establishing a continuous process for syntactically approaching a formula (a fact that we will formally state in the next lemma), and, as we shall see in Theorem 3.6, constitute the right syntactic associate for the semantic notion of satisfaction over almost ordered structures in \mathcal{SOLP} .

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

$$\mathcal{A} \models \theta(\bar{a}, \bar{A})_{-\epsilon} \rightarrow \theta(\bar{a}, \bar{A})_{-\delta} \rightarrow \theta(\bar{a}, \bar{A}) \rightarrow \theta(\bar{a}, \bar{A})_{+\delta} \rightarrow \theta(\bar{a}, \bar{A})_{+\epsilon}.$$

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

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

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. ■

We now want to show that it is possible to jump from satisfaction in almost order (respectively, linearly ordered) structures to satisfaction of approximate formulae in linearly ordered (respectively, almost ordered) structures. For that we need as a preliminary step to show that, for a sublinear function g and a structure \mathcal{A} , the property of being consistent for an almost order \leq_g holds for all the formulae in $(\mathcal{SOLP} + \leq_g)$.

Lemma 3.5 *Let g be a sublinear function. Let \mathcal{A} be a structure together with a built-in almost order \leq_g in $\text{cons-ao}(\mathcal{A}, g)$. Then, for every formula $\psi(\bar{x})$ in $(\mathcal{SOLP} + \leq_g)$, the set $\psi^{\mathcal{A}} := \{\bar{a} \in \mathcal{A} : \mathcal{A} \models \psi(\bar{a})\}$ is consistent with \leq_g .*

Proof: The proof is an easy induction in formulae. ■

Theorem 3.6 (Bridge Theorem). *Fix a sublinear function g . For every formula $\theta(x_1, \dots, x_k, X) \in \mathcal{SOLP}(\tau)$, for every τ -structure \mathcal{A} of size m and for all almost order \leq_g in $\text{cons-ao}(\mathcal{A}, g)$, for every $\bar{a} = (a_1, \dots, a_k) \in A^k$, for every predicate S of arity $t \geq 1$, the following holds:*

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

- (ii) $\mathcal{A}/\sim_g \models \theta([\bar{a}]_g, [S]_g)$ implies $\mathcal{A} \models \theta(\bar{a}, S)_{+\beta(m)}$, where $\beta(m) = \frac{g(m)}{2m}$
 (iii) $\mathcal{A} \models \theta(\bar{a}, S)_{-\gamma(m)}$ implies $\mathcal{A}/\sim_g \models \theta([\bar{a}]_g, [S]_g)$
 (iv) $\mathcal{A}/\sim_g \models \theta([\bar{a}]_g, [S]_g)_{-\beta(m)}$ implies $\mathcal{A} \models \theta(\bar{a}, S)$

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

First order formulae The key tool is Lemma 3.5 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 $(P(Y) \geq r)\theta(\bar{a}, \bar{S}, Y)$ for $0 < r < 1$ and Y of arity $k \geq 1$. Then, for some $B \subseteq A^k$, $|B| \geq rm^k$ and $\mathcal{A} \models \theta(\bar{a}, \bar{S}, B)$. By inductive hypothesis $\mathcal{A}/\sim_g \models \theta([\bar{a}]_g, [\bar{S}]_g, [B]_g)_{+\gamma(m)}$, where $\gamma(m) = g(m)/(2m - g(m))$. Recall that $|[A]_g| = m - g(m)/2$, where $m = |A|$. Thus, we aim to prove that

$$|[B]_g| \geq \begin{cases} (1 - \gamma(m))^{k-1} (r - k\gamma(m)) \left(m - \frac{g(m)}{2}\right)^k, & \text{when } r > k\gamma(m) \\ 0 & \text{, otherwise.} \end{cases} \quad (2)$$

Note that with the suitable choice of $\gamma(m) = \frac{g(m)}{2m - g(m)}$, the non trivial case of equation (2) can be simplified as follows.

$$\begin{aligned} & (1 - \gamma(m))^{k-1} (r - k\gamma(m)) \left(m - \frac{g(m)}{2}\right)^k \\ &= \left(1 - \frac{g(m)}{2m - g(m)}\right)^{k-1} \left(r - k \frac{g(m)}{2m - g(m)}\right) \left(m - \frac{g(m)}{2}\right)^k \\ &= \frac{2^{k-1} (m - g(m))^{k-1}}{(2m - g(m))^{k-1}} \left(r - k \frac{g(m)}{2m - g(m)}\right) \frac{(2m - g(m))^k}{2^k} \\ &= (m - g(m))^{k-1} \left(r - k \frac{g(m)}{2m - g(m)}\right) \frac{(2m - g(m))}{2} \\ &= (m - g(m))^{k-1} \left(r \frac{2m - g(m)}{2} - k \frac{g(m)}{2}\right) \end{aligned}$$

Thus, we need to prove that

$$|[B]_g| \geq \begin{cases} (m - g(m))^{k-1} \left(r \frac{2m - g(m)}{2} - k \frac{g(m)}{2}\right) & \text{, when } r > k \frac{g(m)}{2m - g(m)} \\ 0 & \text{, otherwise.} \end{cases} \quad (3)$$

In the worst case, B contains every two elements from every \sim_g -class, 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 tuples, and therefore we have that

$$|[B]_g| \geq \begin{cases} m^{k-1}(rm - k\frac{g(m)}{2}) & , \text{ when } r > k\frac{g(m)}{2m} \\ 0 & , \text{ otherwise.} \end{cases}$$

This clearly implies (3). Thus,

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

which is the desired result.

Now, suppose that \mathcal{A} satisfies the formula $(P(Y) \leq r)\theta(\bar{a}, \bar{S}, Y)$, with r and Y as above. We argue inductively, as in the preceding case, but this time observe that the witness set B is such that, in the worst case, $|[B]_g| \leq rm^k$. Thus, for the non trivial case, the proportion that this set represents, with respect to $|[A]_g|^k = (m - g(m)/2)^k$ is:

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

Thus,

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

(ii): Suppose that \mathcal{A}/\sim_g satisfies the formula $(P(Y) \leq r)\theta([\bar{a}]_g, [\bar{S}]_g, Y)$. By inductive hypothesis $\mathcal{A} \models \theta(\bar{a}, \bar{S}, (C)^g)_{+\beta(m)}$, where $(C)^g$ is the expansion of $C \subseteq [A]_g^k$. Our aim now is to show that (in the non trivial case)

$$|(C)^g| \leq (1 + \beta(m))^{k-1} [r + k\beta(m)]m^k \quad (4)$$

for $\beta(m) = g(m)/2m$. For this choice of β , equation (4) is equivalent to

$$|(C)^g| \leq \left(\frac{2m + g(m)}{2m}\right)^{k-1} \left[r + k\frac{g(m)}{2m}\right]m^k \quad (5)$$

We note that $|C| \leq r(m - g(m)/2)^k$, and therefore when we expand to $(C)^g$, in the worst case, we throw in all possible k -tuples determined by elements in the same class, and hence

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

which obviously imply (5) since $(2m - g(m))/2m < 1 < (2m + g(m))/2m$. Thus,

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

which is the desired result.

Now suppose that \mathcal{A}/\sim_g satisfies the formula $(P(Y) \geq r) \theta([\bar{a}]_g, [\bar{S}]_g, Y)$ 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 $\mathcal{A}/\sim_g \models \theta([\bar{a}]_g, [\bar{S}]_g, C)$. By inductive hypothesis $\mathcal{A} \models \theta(\bar{a}, \bar{S}, (C)^g)_{+\beta(m)}$, 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

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

Thus,

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

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). ■

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

$$\begin{array}{ccccc} \mathcal{A} \models & \theta_{-\gamma} & \longrightarrow & \theta & \longrightarrow & \theta_{+\beta} & \text{(almost order)} \\ & \searrow & & \nearrow & & \searrow & \\ \mathcal{A}/\sim_g \models & \theta_{-\beta} & \longrightarrow & \theta & \longrightarrow & \theta_{+\gamma} & \text{(order)} \end{array}$$

Now the ground is set. From previous experiences with weak forms of order (e.g. [1], [3], [8], and many others) we learnt that inexpressibility results are easy to accomplish in the presence of almost order, but to transfer these separations to the truly (linearly) ordered world is hard. Our picture shows that, in fact, the passing from the almost ordered world to a corresponding ordered world (or vice versa) changes the syntactic description of some problem for an approximate description. Is this the best we can get? To put it another way, is an approximate description as good as an exact description for determining inexpressibility of a class of ordered structures? We feel that the answer to this last question is “yes in almost all cases”, and in the remainder of this paper we give formal support to this intuition.

4 Strong expressibility

Our idea of sentences that are strongly equivalent is that their respective approximation for some small error should be equivalent. The consequence is that within an interval of radius the given error, all approximations are equivalent, and so are the sentences. This will then lead us to a stronger concept of expressibility, which we obviously call strong expressibility.

Definition 4.1 *Let ϕ and ψ be two sentences of $\mathcal{SO}\mathcal{LP}$. We say that ϕ and ψ are strongly equivalent (and we write $\phi \Leftrightarrow_S \psi$) if, and only if, there exists $0 \leq \epsilon \leq 1$ such that for every $0 < \eta \leq \epsilon$ and for every finite structure \mathcal{A} :*

$$\mathcal{A} \models \phi_{+\eta} \rightarrow \psi_{-\eta} \quad \text{and} \quad \mathcal{A} \models \psi_{+\eta} \rightarrow \phi_{-\eta}.$$

We define similarly the strong equivalence between two approximate sentences in $\mathcal{SO}\mathcal{LP}$.

Observe that two sentences that are strongly equivalent can be syntactically approximate among themselves 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 β and γ , with $-\epsilon < \beta < \epsilon$ and $-\epsilon < \gamma < \epsilon$, and for any finite structure \mathcal{A} , $\mathcal{A} \models \phi_{+\beta} \leftrightarrow \psi_{+\gamma}$. This follows from $\phi \Leftrightarrow_S \psi$ and Lemma 3.4 because, for every model \mathcal{A} :

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

In particular (taking $\beta = \gamma = 0$), 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. Put the other way around, 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 $\phi \Leftrightarrow_S \phi$. As a matter of fact, we next show that our notion of strong equivalence can behave very badly.

Proposition 4.2 *There is a sentence that is not strongly-equivalent to itself.*

Proof: We prove that the sentence $\Theta := (P(X) \geq 1/2)(P(Y) \geq 1/2)\varphi(X, Y)$, where

$$\varphi(X, Y) := \forall x(X(x) \vee Y(x)) \wedge \forall y(X(y) \rightarrow \neg Y(y))$$

is not strongly equivalent to itself. Observe that, for any finite structure \mathcal{A} ,

$$\mathcal{A} \models \Theta \iff |\mathcal{A}| \text{ is even}$$

Now, note that since $\Theta_{+\epsilon}$ always follows from $\Theta_{-\epsilon}$ (by Lemma 3.4) we must prove that for every $0 < \epsilon < 1$ there is a structure \mathcal{A} such that $\mathcal{A} \not\models \Theta_{+\epsilon} \implies \Theta_{-\epsilon}$.

For $\epsilon \geq 1/2$, this is clearly the case since $\Theta_{+\epsilon}$ is true on all finite structures but $\Theta_{-\epsilon}$ is false on all structures.

For $\epsilon < 1/2$, we choose some even sized structure for \mathcal{A} . We have $\mathcal{A} \models \Theta_{+\epsilon}$ but $\mathcal{A} \not\models \Theta_{-\epsilon}$ as $\Theta_{-\epsilon}$ is clearly false on all structures. \blacksquare

On the contrary, we can give an example of a formula which is strongly equivalent to itself.

Example 4.3 *Consider the following sentence of $\mathcal{MSO}\mathcal{LP}$:*

$$\Theta := (P(X) \geq 1/3)\xi(X, x, y)$$

where

$$\xi(X, x, y) = \forall x \forall y (\neg E(x, y) \vee X(x) \vee X(y)) \wedge (\neg E(x, y) \vee \neg X(x) \vee \neg X(y)).$$

This sentence of Monadic $SO\mathcal{L}\mathcal{P}$ captures 2-colourability in a graph.

Now observe that if we choose $\epsilon < 1/3$, then for all $\eta \leq \epsilon$, if $\mathcal{A} \models \Theta_{+\eta}$ then \mathcal{A} is a 2-colourable graph with a colour X of size $|X| \geq 1/3 + \eta$. Then, certainly,

$$\mathcal{A} \models \Theta_{-\eta} := (P(X) \geq (1/3 - \eta)) \xi(X, x, y)$$

Thus $\models \Theta_{+\eta} \rightarrow \Theta_{-\eta}$.

In view of the preceding example and proposition, we want to identify those fragments of $SO\mathcal{L}\mathcal{P}$ that behave “decently” for the notion of strong equivalence, i.e. where at least we can ask that every formula ϕ in the logic is strongly equivalent to itself. This motivates our definitions below of *approximate logic* and ϵ -relaxed fragments.

Definition 4.4 Fix a logic $\mathcal{L} \subseteq SO\mathcal{L}\mathcal{P}$ and an $-1 < \epsilon < 1$. The ϵ -approximation of \mathcal{L} , denoted \mathcal{L}_ϵ , is the following fragment of $SO\mathcal{L}\mathcal{P}$:

$$\{\phi_{+\epsilon}, \phi_{-\epsilon} : \phi \in \mathcal{L}\}$$

By convention we define $\mathcal{L}_0 = \mathcal{L}$. Also we will distinguish the positive fragment, $\mathcal{L}_{+\epsilon} := \{\phi_{+\epsilon} : \phi \in \mathcal{L}\}$, from the negative fragment, $\mathcal{L}_{-\epsilon} := \{\phi_{-\epsilon} : \phi \in \mathcal{L}\}$. The approximation of \mathcal{L} (or the approximate logic corresponding to \mathcal{L}) is the set of formulae $\mathcal{L}_A := \bigcup_{-1 < \epsilon < 1} \mathcal{L}_\epsilon$

Definition 4.5 We say that a fragment \mathcal{L} of $SO\mathcal{L}\mathcal{P}$ is ϵ -relaxed if, for every δ such that $-\epsilon < \delta < \epsilon$, $\mathcal{L}_\delta = \mathcal{L}$ (i.e., their expressive power is the same).

Two important examples of ϵ -relaxed logics are the languages $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)$ and $(SO\mathcal{L}\mathcal{P}Krom[2] + \leq)$, which were defined and studied in [1] (see also section 2.1 above), and which capture **P** and **NL**, respectively. In the following lines we give a summary of the reasons why these languages are ϵ -relaxed, and afterwards we give more technical details. The reader who feels that he does not need more arguments to believe these facts, may skip to Lemma 4.9.

For many $\epsilon < 1/2$, the problem $NOT\text{-}IN\text{-}CLOS_{\leq 1/2 + \epsilon}$ (Example 2.2) is expressible in $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)_{+\epsilon}$, and it is complete for **P** via quantifier free first order reductions (analogous proof as in [1]). Therefore, any problem in **P** has a definition in $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)_{+\epsilon}$. Conversely, the satisfaction of sentences in $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)_{+\epsilon}$ can be decided in **P** by the algorithm described in [1] for $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)$. Thus, $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)_{+\epsilon} = \mathbf{P} = (SO\mathcal{L}\mathcal{P}Horn[2] + \leq)$. The arguments for $(SO\mathcal{L}\mathcal{P}Krom[2] + \leq)_{+\epsilon} = \mathbf{NL} = (SO\mathcal{L}\mathcal{P}Krom[2] + \leq)$ are similar.

Now follows some details of the preceding sketch.

Theorem 4.6 For all $k \geq 4$, $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)$ is $1/k$ -relaxed.

Proof:

- For all $\epsilon < 1/2$, $NOT\text{-}IN\text{-}CLOS_{\leq 1/2 + \epsilon}$ is expressible in $(SO\mathcal{L}\mathcal{P}Horn[2] + \leq)_{+\epsilon}$ by the sentence $(P(X) \leq 1/2 + \epsilon) \beta_{nclos}(X)$, where $\beta_{nclos}(X)$ is the formula in Example 2.2.

- For all $k \geq 4$, if $\epsilon = 1/k$, then $\text{NOT-IN-CLOS}_{\leq 1/2+\epsilon}$ is complete for \mathbf{P} via quantifier free first order reductions (with successor).

We will define a reduction from the complement of the problem PS or *Path System Accessibility* (see Remark 2.3) to $\text{NOT-IN-CLOS}_{\leq 1/2+1/k}$, using quantifier free first order formulae. Recall that the signature for the problem $\text{NOT-IN-CLOS}_{\leq 1/2+\epsilon}$ is exactly that of PS, namely, $\tau = \{R, s, t\}$, consisting of a relation symbol of arity 3, and two constant symbols, s (the source) and t (the sink). Using numeric constant symbols $\mathbf{0}$ and \mathbf{max} , we can define the numbers $\mathbf{0}, \mathbf{1}, \dots, \mathbf{k-1}$. Given a τ -structure \mathcal{A} , we define a τ -structure \mathcal{A}' as follows: The universe of \mathcal{A}' consist of $2k$ disjoint copies of A ,

$$\begin{aligned} \mathcal{A}' := & A \times \{\mathbf{0}\} \times \{\mathbf{0}\} \cup \dots \cup A \times \{\mathbf{0}\} \times \{\mathbf{k-1}\} \\ & \cup A \times \{\mathbf{max}\} \times \{\mathbf{0}\} \cup \dots \cup A \times \{\mathbf{max}\} \times \{\mathbf{k-1}\} \end{aligned}$$

The constants are interpreted as $s^{\mathcal{A}'} = (\mathbf{0}, \mathbf{0}, \mathbf{0})$ and $t^{\mathcal{A}'} = (\mathbf{max}, \mathbf{max}, \mathbf{k-1})$. The relation $R^{\mathcal{A}'}$ consists of:

- a replica of $R^{\mathcal{A}}$ in $A \times \{\mathbf{0}\} \times \{\mathbf{0}\}$, taking as source $(\mathbf{0}, \mathbf{0}, \mathbf{0})$ and as sink $(\mathbf{max}, \mathbf{0}, \mathbf{0})$;
- a replica of $R^{\mathcal{A}}$ in $A \times \{\mathbf{max}\} \times \{\mathbf{0}\}$, taking as source $(\mathbf{0}, \mathbf{max}, \mathbf{0})$ and as sink $(\mathbf{max}, \mathbf{max}, \mathbf{0})$;
- the rule $((\mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{max}, \mathbf{0}))$;
- for each $\mathbf{i} \in \{\mathbf{0}, \mathbf{1}, \dots, \mathbf{k-1}\}$ and each $a \in A$,
the rule $((\mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}), (a, \mathbf{0}, \mathbf{i}))$;
- for each $\mathbf{i} \in \{\mathbf{0}, \mathbf{1}, \dots, \mathbf{k-1}\}$ and each $a \in A$,
the rule $((\mathbf{max}, \mathbf{max}, \mathbf{0}), (\mathbf{max}, \mathbf{max}, \mathbf{0}), (a, \mathbf{max}, \mathbf{i}))$.

Given a τ -structure \mathcal{A} and its associated structure \mathcal{A}' obtained by the above reduction, we have:

$$\mathcal{A} \notin \text{PS} \iff \mathcal{A}' \in \text{NOT-IN-CLOS}_{\leq 1/2+1/k}$$

Assume $\mathcal{A} \notin \text{PS}$. Then $(\mathbf{max}, \mathbf{0}, \mathbf{0})$ (and $(\mathbf{max}, \mathbf{max}, \mathbf{0})$) is not accessible from $(\mathbf{0}, \mathbf{0}, \mathbf{0})$, and hence $t^{\mathcal{A}'} = (\mathbf{max}, \mathbf{max}, \mathbf{k-1})$ is not accessible from $s^{\mathcal{A}'}$, and the subset of the $R^{\mathcal{A}'}$ -closure that contains $s^{\mathcal{A}'}$ has size $< 2|A| = \frac{2}{2k}|A'| < \frac{2+k}{2k}|A'|$. Hence, $\mathcal{A}' \in \text{NOT-IN-CLOS}_{\leq 1/2+1/k}$. Conversely, if $t^{\mathcal{A}'}$ is accessible from $s^{\mathcal{A}'}$, then $(\mathbf{max}, \mathbf{0}, \mathbf{0})$ and $(\mathbf{max}, \mathbf{max}, \mathbf{0})$ are accessible from $s^{\mathcal{A}'} = (\mathbf{0}, \mathbf{0}, \mathbf{0})$, and from these two vertices all other vertex (including $t^{\mathcal{A}'}$) is accessible; so the $R^{\mathcal{A}'}$ -closure that contains $s^{\mathcal{A}'}$ has size $> (2k-2)|A| = \frac{k-1}{k}|A'| \geq \frac{2+k}{2k}|A'|$, when $k \geq 4$. Hence, $\mathcal{A}' \notin \text{NOT-IN-CLOS}_{\leq 1/2+1/k}$.

- $(\text{SOLPHorn}[2]+\leq)$ is closed under *quantifier free projections* [6, § 11.2], because it captures \mathbf{P} (as shown in [1]), and this class is closed under such reductions.
- For any $k > 1$, the satisfaction of sentences in $(\text{SOLPHorn}[2]+\leq)_{1/k}$ can be decided in \mathbf{P} by the algorithm described in [1] for deciding $(\text{SOLPHorn}[2]+\leq)$. This is so, because the aforementioned algorithm works for all SOLPHorn formulae with proportional quantifiers of the form $(P(X) \leq t)$, with t any rational such that $0 < t < 1$; hence, in particular for $t = 1/2 + 1/k$.

All four points together give that, for $k \geq 4$,

$$(\text{SOLPHorn}[2]+\leq)_{1/k} = \mathbf{P} = (\text{SOLPHorn}[2]+\leq) \quad \blacksquare$$

Theorem 4.7 *For all $k \geq 4$, $(\text{SOLPKrom}[2]+\leq)$ is $1/k$ -relaxed.*

Proof: This time we work with the problem $\text{NCON}_{\geq r}$ of Example 2.4, whose underlying signature is $\tau = \{E, s\}$.

- For all $\epsilon < 1/2$, $\text{NCON}_{\geq 1/2-\epsilon}$ is expressible in $(\text{SOLPKrom}[2]+\leq)_{+\epsilon}$ by the sentence $(P(Y) \geq 1/2-\epsilon)\alpha_{\text{ncon}}(Y)$, where $\alpha_{\text{ncon}}(Y)$ is the formula in Example 2.4.
- For all $k \geq 4$, if $\epsilon = 1/k$, then $\text{NCON}_{\geq 1/2-\epsilon}$ is complete for \mathbf{NL} via quantifier free first order reductions (with successor).

We define a reduction from the complement of the Transitive Closure (TC) problem to $\text{NCON}_{\geq 1/2-1/k}$, using quantifier free first order formulae. The construction is very similar to the one above for $\text{NOT-IN-CLOS}_{\leq 1/2+\epsilon}$, but instead of joining vertices by the 3-ary relation R , we join them by the binary edge relation E .

Given a τ -structure \mathcal{A} and its associated structure \mathcal{A}' obtained by the reduction, we have:

$$\mathcal{A} \notin \text{TC} \iff \mathcal{A}' \in \text{NCON}_{\geq 1/2-1/k}$$

- The remainder of the proof follows similarly as the proof of Theorem 4.6 ■

We show below a useful property of approximate formulae: Given one, we can find another that refines the approximation. This will allow us to define “good neighbourhoods” of approximations around formulae, where we can equate equivalence of these formulae with equivalence of the approximations within these neighbourhoods (this is the content of Lemma 4.9).

Lemma 4.8 *For every formula $\theta(\bar{x}, \bar{X}) \in \text{SOLP}(\tau)$, for every γ and λ , with $-1 \leq \gamma < \lambda \leq 1$, for every δ verifying $\gamma < \delta < \lambda$, there exists a $\mu > 0$ such that:*

- $\gamma < \delta - \mu < \delta + \mu < \lambda$, and
- for every τ -structure \mathcal{A} and for every interpretation \bar{A} of relation symbols \bar{X} in \mathcal{A} , and elements \bar{a} in \mathcal{A} , we have that:

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

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

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

Case 1: Consider the formula $\Psi(\bar{x}, \bar{X}) := (P(Y) \geq r)\theta(\bar{x}, \bar{X}, Y)$. 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 $\epsilon \in (-1, 1)$,

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

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

$$\begin{aligned} \mathcal{A} \models \theta(\bar{a}, \bar{A}, B)_{+\gamma} &\rightarrow (\theta(\bar{a}, \bar{A}, B)_{+\delta})_{-\mu_1} \rightarrow \theta(\bar{a}, \bar{A}, B)_{+\delta} \\ &\rightarrow (\theta(\bar{a}, \bar{A}, B)_{+\delta})_{\mu_1} \rightarrow \theta(\bar{a}, \bar{A}, B)_{+\lambda}. \end{aligned}$$

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 ϵ , $-\mu_2 \leq \epsilon \leq \mu_2$,

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

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 interpretations \bar{A} in \mathcal{A} and elements \bar{a} in \mathcal{A} , we have that:

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

but this is exactly the desired result that for every model \mathcal{A} and for every interpretation \bar{A} of relation symbols in \mathcal{A} and elements \bar{a} in \mathcal{A} , we have that:

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

Case 2: Consider now the formula $\Psi(\bar{x}, \bar{X}) := (P(Y) \leq r)\theta(\bar{x}, \bar{X}, Y)$. 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 $r \in [0, 1]$, $\epsilon \in (-1, 1)$,

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

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

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

$$\begin{aligned} \mathcal{A} \models \theta(\bar{a}, \bar{A}, B)_{+\gamma} &\rightarrow (\theta(\bar{a}, \bar{A}, B)_{+\delta})_{-\mu_1} \rightarrow \theta(\bar{a}, \bar{A}, B)_{+\delta} \\ &\rightarrow (\theta(\bar{a}, \bar{A}, B)_{+\delta})_{\mu_1} \rightarrow \theta(\bar{a}, \bar{A}, B)_{+\lambda}. \end{aligned}$$

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 ϵ , $-\mu_2 \leq \epsilon \leq \mu_2$,

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

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 \bar{A} in \mathcal{A} and elements \bar{a} in \mathcal{A} , we have that:

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

but this is exactly the desired result that for every model \mathcal{A} and for every interpretation \bar{A} of relation symbols in \mathcal{A} and elements \bar{a} in \mathcal{A} , we have that:

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

This completes the proof of the lemma. ■

The main property of relaxed fragments is the following:

Lemma 4.9 *Let \mathcal{L} be a ϵ -relax fragment of $SO\mathcal{L}\mathcal{P}$. Then for every sentence $\phi \in \mathcal{L}$, there exists a $-\epsilon < \delta < \epsilon$ and sentence $\theta \in \mathcal{L}_\delta$ such that $\phi \leftrightarrow \theta$ and $\theta \leftrightarrow_S \theta$.*

Proof: Fix a sentence $\phi \in \mathcal{L}$. For every $-\epsilon < \lambda < \epsilon$ there exists a sentence $\tilde{\phi} \in \mathcal{L}$ such that $(\tilde{\phi})_\lambda \leftrightarrow \phi$. 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 real numbers γ and λ , with $-\epsilon < \gamma < \lambda < \epsilon$ such that $\theta_\gamma \leftrightarrow \phi \leftrightarrow \theta_\lambda$. By the properties of approximate formulae (Lemma 4.8) we know that there exists δ and μ such that $-\epsilon < \gamma < \delta < \lambda < \epsilon$ and $\mu > 0$, and

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

hence $\theta_\delta \leftrightarrow_S \theta_\delta \leftrightarrow \phi$. ■

The previous lemma motivates our notion of strong expressibility.

Definition 4.10 *Let $\mathcal{L} \subseteq \mathcal{L}' \subseteq SO\mathcal{L}\mathcal{P}$ and fix $\phi \in SO\mathcal{L}\mathcal{P}$ 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 express ϕ then the fragment \mathcal{L} does not strongly express ϕ .

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

Theorem 4.11 *Let $\mathcal{L}, \mathcal{L}'$ be ϵ -relaxed fragments of $SO\mathcal{L}\mathcal{P}$ such that $\mathcal{L} \subseteq \mathcal{L}'$ and let ϕ be an $SO\mathcal{L}\mathcal{P}$ -sentence. Then the following statements are equivalent:*

- ϕ is expressible in \mathcal{L} ;
- There exists a μ , with $\epsilon < \mu < \epsilon$, such that ϕ is strongly expressible in \mathcal{L}_μ with respect to \mathcal{L}'_μ , 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 $\epsilon < \mu < \epsilon$ 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}_μ 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 4.9, since \mathcal{L}' is an ϵ -relaxed fragment, we know that there exists a sentence $\rho \in \mathcal{L}'_\lambda$ for some $-\epsilon < \lambda < \epsilon$ such that $\phi \leftrightarrow \rho$ and $\rho \leftrightarrow_S \rho$. More specifically there exists γ such that $-\epsilon < \lambda - \gamma < \lambda + \gamma < \epsilon$ and $\rho_\gamma \leftrightarrow \rho_{-\gamma}$.

From hypothesis we know that there exists $\tilde{\theta} \in \mathcal{L}_\lambda$ such that $\tilde{\theta} \leftrightarrow \phi \leftrightarrow \rho$. Applying again Lemma 4.9 to $\tilde{\theta}$ and using the fact that \mathcal{L}_λ is a γ -relaxed fragment, we know that there exists a sentence $\theta \in \mathcal{L}_\mu$, for some $\lambda - \gamma < \mu < \lambda + \gamma$, such that $\tilde{\theta} \leftrightarrow \theta$ and $\theta \leftrightarrow_S \theta$. More specifically there exists ω such that $\lambda - \gamma < \mu - \omega < \mu + \omega < \lambda + \gamma$ and $\theta_\omega \leftrightarrow \theta_{-\omega}$.

We have then the following sequences of implications:

$$\rho_\gamma \rightarrow \rho_{-\gamma} \rightarrow \rho \rightarrow \tilde{\theta} \rightarrow \theta \rightarrow \theta_\omega \rightarrow \theta_{-\omega}$$

and symmetrically,

$$\theta_\omega \rightarrow \theta_{-\omega} \rightarrow \theta \rightarrow \tilde{\theta} \rightarrow \rho \rightarrow \rho_\gamma \rightarrow \rho_{-\gamma}$$

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 $\phi \leftrightarrow \rho_\mu$. ■

The importance of this theorem 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 help 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.

As a first approach to our challenging goal, we present a theorem that proves strong inexpressibility, over ordered structures, albeit under somewhat strong hypothesis.

Theorem 4.12 *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, and moreover that for every formula $\theta \in (\mathcal{L} + \leq)$ and every ω , $0 < \omega < \epsilon$, there exists a sublinear function g and two models \mathcal{A}, \mathcal{B} in $(\mathcal{L} + \leq_g)$ (i.e. almost ordered models), 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 in $(\mathcal{L} + \leq)$ with respect to $(\mathcal{L}' + \leq)$.

Proof: In order to get a contradiction assume that ϕ is strongly expressible 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 an $0 < \omega < \epsilon$ such that for every model \mathcal{C} in $(\mathcal{SOLP} + \leq)$ the following property (*) holds:

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

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 (*). \blacksquare

Observe that the previous theorem gives strong inexpressibility over ordered structures. 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 4.11 to produce a result that shows inexpressibility in formulae with built-in order, based on separation of almost orders.

Theorem 4.13 *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, and further that for every $-\epsilon < \mu < \epsilon$, for every $\omega > 0$ such that $-\epsilon < \mu - \omega < \mu + \omega < \epsilon$, 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 expressible in $(\mathcal{L}+\leq)$.

Proof: Assume 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 4.11 to obtain that there exists a μ , with $-\epsilon < \mu < \epsilon$, 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 $(\mathcal{SOLP}+\leq)$:

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

Note that we can select $\omega > 0$ such that $-\epsilon < \mu - \omega < \mu + \omega < \epsilon$.

Consider then the two models \mathcal{A}, \mathcal{B} and the sublinear function g associated to $\phi, \epsilon, \mu, \theta$ 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 $\psi \leftrightarrow_S \theta$. \blacksquare

Note that the result just developed works by checking properties of models in $(\mathcal{SOLP}+\leq_g)$, i.e. models with almost orders, where separation proofs have been shown in practice to be easier. Note also that our prime examples of relaxed fragments where we could applied this are the $(\mathcal{SOLPKrom}[2]+\leq)$ and $(\mathcal{SOLPHorn}[2]+\leq)$ that correspond to the classes **NL** and **P**. Hence we have a result, based on almost orders and approximate formulae that tells us that, in order to separate $(\mathcal{SOLPKrom}[2]+\leq)$ from $(\mathcal{SOLPHorn}[2]+\leq)$ 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 expressibility

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 formalise this idea.

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

$$\phi := Q_1 X_1 Q_2 X_2 \dots Q_r X_r A_1 x_1 A_2 x_2 \dots A_f x_f \bigvee_{i=1}^q \bigwedge_{j=1}^{t_i} \theta_{i,j}(\overline{X}, \overline{x})$$

where the $Q_s X_s$ are proportionality quantifiers over the second order variable X_s and the $A_s x_s$ 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 be 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 δ , such that $-\epsilon < \delta < \epsilon$ there exists sentences $\theta^{(\delta)} \in \mathcal{L}, \rho^{(\delta)} \in \mathcal{L}'$ such that

$$\models (\theta^{(\delta)})_\delta \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow (\rho^{(\delta)})_\delta. \quad (6)$$

Suppose that we select the sentences $\theta^{(\delta)}$ and $\rho^{(\delta)}$ to have minimal complexity among all the sentences in \mathcal{L} satisfying (6), 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 pigeon hole principle implies that there exists sentences $\Theta \in \mathcal{L}', \Lambda \in \mathcal{L}$ such that:

- for every δ and ϵ , $0 < \delta < \epsilon$, there exists α, β with $0 < \alpha < \delta$ and $-\delta < -\beta < 0$ with $\models \Theta \leftrightarrow \Theta^{(-\beta)}$ and $\models \Theta \leftrightarrow \Theta^{(\alpha)}$.
- for every δ and ϵ , $0 < \delta < \epsilon$, there exists α', β' with $0 < \alpha' < \delta$ and $-\delta < -\beta' < 0$ with $\models \Lambda \leftrightarrow \Lambda^{(-\beta')}$ and $\models \Lambda \leftrightarrow \Lambda^{(\alpha')}$.

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

$$\mathcal{A} \models \Theta_{\alpha_1} \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow \Lambda_{-\beta_1}$$

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

$$\mathcal{A} \models \Lambda_{\alpha_2} \leftrightarrow \psi \leftrightarrow \phi \leftrightarrow \Theta_{-\beta_2}$$

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

$$\mathcal{A} \models \Lambda_{\delta} \rightarrow \Lambda_{\alpha_2} \rightarrow \Theta_{-\beta_2} \rightarrow \Theta_{-\delta},$$

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

$$\mathcal{A} \models \Theta_{\delta} \rightarrow \Theta_{\alpha_1} \rightarrow \Lambda_{-\beta_1} \rightarrow \Lambda_{-\delta},$$

The two statements above imply that $\models \phi \leftrightarrow \Theta$ and $\models \psi \leftrightarrow \Lambda$ and $\Lambda \leftrightarrow_S \Theta$. 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 δ , such that $-\epsilon < \delta < \epsilon$, there exists sentences $\theta^{(\delta)} \in \mathcal{L}$ and $\rho^{(\delta)} \in \mathcal{L}'$ such that $\models (\theta^{(\delta)})_{\delta} \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow (\rho^{(\delta)})_{\delta}$. Suppose, additionally, 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}' . ■

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 δ , such that $-\epsilon < \delta < \epsilon$, there exists sentences $\theta^{(\delta)} \in \mathcal{L}$ and $\rho^{(\delta)} \in \mathcal{L}'$ such that $\models (\theta^{(\delta)})_{\delta} \leftrightarrow \phi \leftrightarrow \psi \leftrightarrow (\rho^{(\delta)})_{\delta}$. Suppose that ϕ is not strongly expressible in \mathcal{L} with respect to \mathcal{L}' . Then, for every natural number M ,*

- there exists δ , $0 < \delta < \epsilon$, such that for all α , $0 < \alpha < \delta$, the minimal complexity of the formula $\theta^{(\alpha)}$ is bigger than M ; or
- there exists δ , $0 < \delta < \epsilon$, such that for all α , $-\delta < \alpha < 0$, the minimal complexity of the formula $\theta^{(-\alpha)}$ is bigger than M ; or
- there exists δ , $0 < \delta < \epsilon$, such that for all α' , $0 < \alpha' < \delta$, the minimal complexity of the formula $\rho^{(\alpha')}$ is bigger than M ; or
- there exists δ , $0 < \delta < \epsilon$, such that for all α' , $-\delta < \alpha' < 0$, the minimal complexity of the formula $\rho^{(-\alpha')}$ is bigger than M . ■

Here is a direct consequence of the above corollary. We know that $(SOLPKrom[2]+ \leq) \subseteq (SOLPHorn[2]+ \leq)$ are ϵ -relaxed fragments of $SOLP$ that capture **NL** and **P** respectively. Suppose that you can prove that a problem Q in **P** is not strongly expressible

in $(\text{SOLPKrom}[2]+\leq)$ with respect to $(\text{SOLPHorn}[2]+\leq)$ by using any of the tools at our disposal. Then if still Q was expressible in $(\text{SOLPKrom}[2]+\leq)$ the previous corollary implies that there exists a $\delta \neq 0$, with $-\epsilon < \delta < \epsilon$, such that the minimal complexity of the sentences $\theta_\omega \in (\text{SOLPKrom}[2]+\leq)_\omega$ that capture Q tends to infinity, or there exists a $\delta \neq 0$, with $-\epsilon < \delta < \epsilon$, such that the minimal complexity of the sentences $\rho_\omega \in (\text{SOLPHorn}[2]+\leq)_\omega$ that capture Q tends to infinity (for either $0 < \omega < \delta$ or $\delta < \omega < 0$). This is indeed a very strange phenomena, which lead us to conjecture that expressibility implies strong expressibility, in the context of ϵ -relaxed logics.

Acknowledgement

The authors are grateful to the anonymous referee for his/her thoroughly revision of the original version of our paper, helping us improve it in form and contents.

References

- [1] Arratia, A., and Ortiz, C., Expressive power and complexity of a logic with quantifiers that count proportions of sets, *J. of Logic and Computation*, **16** (6): 817–840, 2006.
- [2] Ebbinghaus, H.D., and Flum, J., *Finite Model Theory* (Springer-Verlag, 1995).
- [3] Etessami, K., and Immerman, N., Reachability and the power of local ordering, *Theoretical Comp. Sci.* **148**, 2, 261–279, 1995.
- [4] Grädel, E., Capturing complexity classes by fragments of second order logic, *Theoretical Comp. Sci.* **101**, 35–57, 1992.
- [5] Immerman, N., Nondeterministic space is closed under complement. *SIAM Journal on Computing* **17**, 935–938, 1988.
- [6] Immerman, N., *Descriptive Complexity* (Springer, 1998).
- [7] Keisler, H.J., Hyperfinite model theory. In: R.C. Gandy and J.M. E. Hyland (eds) *Logic Colloquium 76* (North-Holland, 1977).
- [8] Libkin, L., and Wong, L., Lower bounds for invariant queries in logics with counting. *Theoretical Comp. Sci.* **288**, 153–180, 2002.
- [9] Stewart, I., Logical description of monotone NP problems, *J. Logic and Computation.* **4** (4), 337–357, 1994.
- [10] Szelepcsényi, R., The method of forced enumeration for nondeterministic automata. *Acta Informatica* **26**, 279–284, 1988.

Received 21 May 2007