

# UPCommons

## Portal del coneixement obert de la UPC

<http://upcommons.upc.edu/e-prints>

---

Aquesta és una còpia de la versió *author's final draft* d'un article publicat a la revista *Lecture notes in computer science*.

La publicació final està disponible a Springer a través de <http://dx.doi.org/10.1007/b95852>

This is a copy of the author 's final draft version of an article published in the journal *Lecture notes in computer science*.

The final publication is available at Springer via <http://dx.doi.org/10.1007/b95852>

---

### Article publicat / Published article:

Arratia, A., Ortiz, C.E. (2004) Approximating the expressive power of logics in finite models. "Lecture notes in computer science". Vol. 2976. p.540-556. Doi: 10.1007/b95852

# Approximating the Expressive Power of Logics in Finite Models

Argimiro Arratia<sup>1\*\*</sup> and Carlos E. Ortiz<sup>2</sup>

<sup>1</sup> Departamento de Matemáticas, Universidad Simón Bolívar, Venezuela, and  
Depto. de Matemáticas Aplicadas y Computación, Universidad de Valladolid, España  
`arratia@mac.cie.uva.es`,

<sup>2</sup> Department of Mathematics and Computer Science, Arcadia University, 450 S.  
Easton Road, Glenside, PA 19038-3295, U.S.A. `ortiz@arcadia.edu`

**Abstract.** We present a *probability logic* (essentially a first order language extended with quantifiers that count the fraction of elements in a model that satisfy a first order formula) which, on the one hand, captures uniform circuit classes such as  $AC^0$  and  $TC^0$  over *arithmetic models*, namely, finite structures with linear order and arithmetic relations, and, on the other hand, their semantics, with respect to our arithmetic models, can be closely approximated by giving interpretations of their formulas on finite structures where all relations (including the order) are restricted to be “modular” (i.e. to act subject to an integer modulo). In order to give a precise measure of the proximity between satisfaction of a formula in an arithmetic model and satisfaction of the same formula in the “approximate” model, we define the *approximate formulas* and work on a notion of approximate truth. We also indicate how to enhance the expressive power of our probability logic in order to capture polynomial time decidable queries,

There are various motivations for this work. As of today, there is not known logical description of any computational complexity class below **NP** which does not requires a built-in linear order. Also, it is widely recognized that many model theoretic techniques for showing definability in logics on finite structures become almost useless when order is present. Hence, if we want to obtain significant lower bound results in computational complexity via the logical description we ought to find ways of by-passing the ordering restriction. With this work we take steps towards understanding how well can we approximate, without a true order, the expressive power of logics that capture complexity classes on ordered structures.

## 1 Introduction

The logical description of many computational complexity classes is based on the fact that the possible domains of interpretations must be at least partially ordered. This is certainly the case for logics meant for describing complexity classes

---

<sup>\*\*</sup> Supported by grant *Ramón y Cajal 2003* from Ministerio Ciencia y Tecnología, España

below **NP**, for it is still unknown whether such classes can be described without any order, and it is further believed that is not the case (further comments in [5] and see also [4]). However, a negative aspect of describing low complexity classes by logics with built-in order is that model theoretic techniques for showing inexpressibility, such as Ehrenfeucht–Fraïssé games and its variations, becomes almost useless; thus, in turn, hopeless for leading into significant complexity lower bounds. (For an illustration of how difficult is to play Ehrenfeucht–Fraïssé games on ordered structures see Section 6.6 of [5].)

This dichotomy with the order had led researchers into exploring ways of keeping some order in the models for various forms of extensions of first order logic, and yet obtain some significant lower bound results (for example, see [3] and [7]). The results presented in this paper are inscribed in that line of research. We introduce a probability logic  $\mathcal{LP}$ , which is, essentially, first order logic extended with quantifiers that count the fraction of elements in a model that satisfy a first order formula. Our definition of the logic  $\mathcal{LP}$  is inspired on the probability logic of Keisler (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. In order to suit our need of this logic for describing computability problems, we restrict our use of relation symbols to a finite set and mainly of the arithmetic type: addition, multiplication and order. With this ability to approximately count and in the presence of built-in order, addition, and multiplication, fragments of this  $\mathcal{LP}$  logic are capable of fully describing circuit classes such as  $AC^0$  and  $TC^0$ , since they coincide with known logics that capture these computational complexity classes, for example, first order logic extended with threshold quantifiers. Following our programme of studying possible ways of reducing the scope of the order and other arithmetic relations within our models, we group in the same set of witnesses of a formula all those elements that are congruent modulo the value of a sublinear function  $F$ , and define the concept of an  $F$ -modular approximation of a finite structure  $\mathcal{A}$ . The  $F$ -modular approximation of  $\mathcal{A}$  thus obtained do not have the order built-in but just approximations of it, and subject to these interpretations we do get separation results among fragments of the corresponding logic  $\mathcal{LP}_F$ , for a particular family of (sublinear) functions  $F$ .

Having satisfied our goal of obtaining inexpressibility results within our probability logic under a weaker interpretation of the atomic symbols, we wonder how to translate that result to an inexpressibility of the same query (or similar query) in the logic with the unrestricted interpretation of symbols (e.g. full linear order). As a partial answer to this question we introduce the notion of approximate formulas and through them we establish a bridge between satisfaction in the structures with natural interpretations of the symbols and their corresponding  $F$ -modular approximation. In the last section of the paper we show how to extend this probability logic and approximations to capture **P**.

## 2 Logic of probability quantifiers

We work with finite vocabularies and finite models. A vocabulary or signature  $\tau$  is a set of relation symbols and constant symbols. The models for  $\tau$  will be denoted by  $\mathcal{A}_m, \mathcal{B}_n, \mathcal{C}_k$ , etc. where the subscripts refer to the cardinality of the model. A logic over the vocabulary  $\tau$  will be denoted  $\mathcal{L}(\tau)$ . In particular  $\text{FO}(\tau)$  is the set of first order formulas over  $\tau$  (or  $\tau$ -formulas). The logic we are mainly concerned in this paper is the logic of probability quantifiers which we define below. Given a natural number  $m$  and a set  $C \subseteq \{0, \dots, m-1\}$  we can define the natural probability  $\mu_m(C)$  as just the cardinality of  $C$  divided by  $m$ . Likewise, for  $s > 0$ , we can define, for every set  $C \subseteq \{0, \dots, m-1\}^s$  the natural probability  $\mu_m^s(C)$  as the cardinality of  $C$  divided by  $m^s$ .

**Definition 1.** *For a vocabulary  $\tau$ , we define the logic of probability quantifiers (or probability logic) over  $\tau$ , as the set of formulas  $\mathcal{LP}(\tau)$  formed as follows:*

**Atomic formulas** *Formulas of the form  $R(\bar{x}, \bar{c})$ , where  $R$  is a relation symbol in  $\tau$ ,  $\bar{x}$  is a vector of variables,  $\bar{c}$  is a vector of constants from  $\tau$ , are in  $\mathcal{LP}(\tau)$ .*

**Conjunction** *If  $\phi_1(\bar{x}), \phi_2(\bar{x}) \in \mathcal{LP}(\tau)$  then  $\phi_1(\bar{x}) \wedge \phi_2(\bar{x})$  is in  $\mathcal{LP}(\tau)$ .*

**Negation** *If  $\phi(\bar{x}) \in \mathcal{LP}(\tau)$  then  $\neg\phi(\bar{x}) \in \mathcal{LP}(\tau)$ .*

**Existential quantification** *If  $\phi(\bar{x}, z) \in \mathcal{LP}(\tau)$  and  $z$  is a variable not appearing in  $\bar{x}$ , then  $\exists z\phi(\bar{x}, z) \in \mathcal{LP}(\tau)$ .*

**Probability quantification** *Fix a rational number  $r$ ,  $0 \leq r < 1$ . If  $\phi(\bar{x}, z) \in \mathcal{LP}(\tau)$  and  $z$  is a variable not appearing in  $\bar{x}$ , then*

$$(P(z) > r)\phi(\bar{x}, z) \text{ and } (P(z) \geq r)\phi(\bar{x}, z) \text{ are in } \mathcal{LP}(\tau).$$

We define the following abbreviations:  $(P(z) < r)\phi(\bar{x}, z)$  stands for  $\neg(P(z) \geq r)\phi(\bar{x}, z)$ , and  $(P(z) \leq r)\phi(\bar{x}, z)$  stands for  $\neg(P(z) > r)\phi(\bar{x}, z)$ . Likewise  $\forall z\phi(\bar{x}, z)$  stands for  $\neg\exists z\neg\phi(\bar{x}, z)$  and  $\phi \vee \psi$  stands for  $\neg(\neg\phi \wedge \neg\psi)$ .

We define the interpretation of the formulas in  $\mathcal{LP}(\tau)$  in a finite structure  $\mathcal{B}_m$  ( $m \in \mathbb{N}$ ) by induction in formulas, with the usual interpretations for conjunction, negation and existential quantifier. The interpretation for a formula  $(P(z) > r)\phi(\bar{x}, z)$  in  $\mathcal{B}_m$  is as follows:

$$\mathcal{B}_m \models (P(z) > r)\phi(\bar{a}, z) \text{ iff } \mu_m(\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}) > r$$

Likewise, the interpretation of the formula  $(P(z) \geq r)\phi(\bar{x}, z)$  is as follows:

$$\mathcal{B}_m \models (P(z) \geq r)\phi(\bar{a}, z) \text{ iff } \mu_m(\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}) \geq r$$

Observe that under this interpretation,  $\neg(P(z) \geq r)\phi(\bar{x}, z)$  is equivalent to  $(P(z) > 1 - r)\neg\phi(\bar{x}, z)$ , and  $\neg(P(z) > r)\phi(\bar{x}, z)$  is equivalent to  $(P(z) \geq 1 - r)\neg\phi(\bar{x}, z)$ .

By  $\mathcal{LP}$  we denote the union of all probability logics  $\mathcal{LP}(\tau)$  taken over all finite vocabularies. We shall also deal with the following fragments of  $\mathcal{LP}$ :

**Definition 2.** Let  $\tau$  be a finite vocabulary. Let  $r_1, r_2, \dots, r_k$  be distinct natural numbers. By  $\mathcal{LP}(\tau)[r_1, r_2, \dots, r_k]$  we understand the smallest subset of  $\mathcal{LP}(\tau)$  containing the atomic formulas and closed under conjunction, negation, existential quantification and the probability quantifiers  $P(z) > q_{ij}/r_i$ ,  $P(z) \geq q_{ij}/r_i$  where  $i \leq k$  and  $q_{ij}$  are natural numbers such that  $0 \leq q_{ij} < r_i$ .

We had in mind using this type of logic for describing computational properties and for that matter we restrict semantics to finite models and also the kind of relation symbols for building our formulas. In general we restrict our symbols to be numerical (in a sense as explained in [5]), and in particular we fix throughout this paper the vocabulary  $\Gamma = \{\oplus, \otimes, \triangleleft, 0, 1\}$ , where  $\oplus, \otimes$  are ternary relation symbols and  $\triangleleft$  is a binary relation symbol and 0 and 1 are constant symbols. Furthermore, we fix a generic vocabulary  $\Gamma^+$  that contains  $\Gamma$  and a set  $\{R_s\}_{s=1}^k$  of other numerical relation symbols and a set  $\{c_w\}_{w=1}^u$  of other constant symbols. We define the **arithmetic** structures over  $\Gamma^+$  as the finite structures  $\mathcal{A}_m$  of the form:  $\mathcal{A}_m = \langle \{0, 1, \dots, m-1\}, \oplus, \otimes, \triangleleft, \{R_s\}_{s=1}^k, \{c_w\}_{w=1}^u, 0, 1 \rangle$ , where the relation symbols  $\oplus, \otimes, \triangleleft$  are interpreted as the usual addition, multiplication and order in the set  $\{0, 1, \dots, m-1\}$ .

We will refer to the probability logic restricted to finite structures that are arithmetic as  $\mathcal{LP}_A$ . The following examples show that the logic  $\mathcal{LP}_A$  contains fragments that are relevant to Descriptive Complexity Theory.

*Example 1.* Let  $FO(\Gamma)$  be the first order logic over  $\Gamma$  and consider the interpretation of the symbols in  $\Gamma$  as natural addition, multiplication and linear order. It is shown in [1] (see also [5]) that this logic captures the complexity class DLOGTIME-uniform  $AC^0$ , where  $AC^0$  is defined as the class of problems accepted by polynomial size, constant depth circuits with unbounded fan-in. This logic corresponds to the smallest subset of  $\mathcal{LP}_A$  that contains the atomic  $\Gamma$ -formulas and is closed under  $\wedge, \neg$  and  $\exists z$ .

*Example 2.* Let  $FO(\Gamma) + M$  be the first order logic over  $\Gamma$  with the interpretations of the symbols in  $\Gamma$  fixed as in the previous example, extended with the majority quantifier  $M$  which is defined as follows: If  $\phi(\bar{x}, z)$  is a formula with one free variable  $z$ , then  $(Mz)\phi(\bar{a}, z)$  is a well defined sentence, which is true if and only if  $\phi(\bar{a}, z)$  is true for more than half of the possible values for  $z$ . It is shown in [1] (see also [5]) that this logic captures the complexity class DLOGTIME-uniform  $TC^0$ , where  $TC^0$  is the class of problems accepted by circuits of polynomial size, constant depth and unbounded fan-in threshold gates (gates which counts its Boolean inputs of value 1 and compares the total with some prefixed number to determine its output). Note that this logic is the fragment of  $\mathcal{LP}_A$  that contains the atomic  $\Gamma$ -formulas and is closed under  $\wedge, \neg, \exists z$  and the quantifier  $P(z) > \frac{1}{2}$ ; that is  $\mathcal{LP}(\Gamma)[2]$ .

Our purpose is to approximate the expressive power of arithmetic relations occurring naturally in finite model theory by arithmetic relations that are “weaker” yet perform better under definability tools such as Ehrenfeucht–Fraïssé

games. Our choice of candidates for relations to approximate the natural arithmetic relations are those that are “modular” in a number theoretic sense.

By  $a \equiv_q b$  we mean that the number  $a$  is congruent to the number  $b$  modulo  $q$ . Furthermore, given  $\bar{a} = (a_1, \dots, a_k)$  and  $\bar{b} = (b_1, \dots, b_k)$  two vectors of natural numbers of equal length, we write  $\bar{a} \equiv_q \bar{b}$  as an abbreviation of  $a_1 \equiv_q b_1, a_2 \equiv_q b_2, \dots, a_k \equiv_q b_k$ . Also, whenever we write  $\bar{a} < m$  for some number  $m$ , we mean that  $a_1 < m, \dots, a_k < m$ .

We understand that a function  $F : \mathbb{N} \rightarrow \mathbb{N}$  is **sublinear**, if for every natural  $m > 0$ ,  $0 < F(m) \leq m$

**Definition 3.** Fix a sublinear function  $F$ , a formula  $\theta(\bar{x}) \in \mathcal{LP}(\Gamma^+)$  and a  $\Gamma^+$ -model  $\mathcal{B}_m$ . The formula  $\theta(\bar{x})$  is  **$F$ -modular** in  $\mathcal{B}_m$  iff the following condition holds:

- For every  $\bar{a}, \bar{b} < m$ , if  $\bar{a} \equiv_{F(m)} \bar{b}$  then  $(\mathcal{B}_m \models \theta(\bar{a}) \text{ iff } \mathcal{B}_m \models \theta(\bar{b}))$ .

We will say that a collection of formulas  $\{\theta_i(x)\}_{i=1}^r \subseteq \mathcal{LP}(\Gamma^+)$  is  **$F$ -modular** in  $\mathcal{B}_m$  iff every formula  $\theta_i$  is  $F$ -modular in  $\mathcal{B}_m$ .

The next lemma states that modularity is preserved by the logical operations and quantification of  $\mathcal{LP}(\Gamma^+)$ . The proof is an easy induction on formulas.

**Lemma 1.** If the collection of atomic  $\Gamma^+$ -formulas is  $F$ -modular for a structure  $\mathcal{B}_m$  then every formula in  $\mathcal{LP}(\Gamma^+)$  is  $F$ -modular for  $\mathcal{B}_m$ .  $\square$

The direct consequence of the above lemma is that the  $F$ -modularity of the formulas in  $\mathcal{LP}(\Gamma^+)$  in a model  $\mathcal{B}_m$  depends only on the modularity of the interpretation of the relation symbols in  $\mathcal{B}_m$ . Because of this fact, every model where all the interpretations of the relation symbols are  $F$ -modular will be called an  **$F$ -modular structure**.

*Remark 1.* For every natural numbers  $e$  and  $f > 0$  we understand  $[e]_f$  to be the remainder of dividing  $e$  by  $f$ . For any vector of natural numbers  $\bar{a} = (a_1, a_2, \dots, a_d)$ , we understand by  $[\bar{a}]_f$  the vector  $([a_1]_f, [a_2]_f, \dots, [a_d]_f)$ .

**Definition 4.** Fix a sublinear function  $F$  and an arithmetic structure  $\mathcal{A}_m$ . The  **$F$ -modular approximation** of  $\mathcal{A}_m$  is a structure

$$\mathcal{A}_m^F = \langle \{0, 1, \dots, m-1\}, \oplus, \otimes, \triangleleft, \{R_s\}_{s=1}^k, \{c_w\}_{w=1}^u, 0, 1 \rangle$$

such that for every  $a, b, c, a_1, \dots, a_r < m$ ,

- $\mathcal{A}_m^F \models \oplus(a, b, c)$  iff  $\mathcal{A}_m \models \oplus([a]_{F(m)}, [b]_{F(m)}, [c]_{F(m)})$ .
- $\mathcal{A}_m^F \models \otimes(a, b, c)$  iff  $\mathcal{A}_m \models \otimes([a]_{F(m)}, [b]_{F(m)}, [c]_{F(m)})$ .
- $\mathcal{A}_m^F \models \triangleleft(a, b)$  iff  $\mathcal{A}_m \models \triangleleft([a]_{F(m)}, [b]_{F(m)})$ .
- $\mathcal{A}_m^F \models R_s(a_1, \dots, a_r)$  iff  $\mathcal{A}_m \models R_s([a_1]_{F(m)}, \dots, [a_r]_{F(m)})$ .

It is easy to see that for every arithmetic structure  $\mathcal{A}_m$ , the structure  $\mathcal{A}_m^F$  is  $F$ -modular. We also remark that for every  $s$ , for every relation symbol  $R_s$ , the set  $\{(a_1, \dots, a_r) < m : \mathcal{A}_m^F \models R_s(a_1, \dots, a_r)\}$  and the set  $\{(a_1, \dots, a_r) < m : \mathcal{A}_m \models R_s(a_1, \dots, a_r)\}$  coincide in the set  $\{(a_1, \dots, a_r) : a_1, \dots, a_r < F(m)\}$ . These two remarks justify the name of  $F$ -modular approximation of  $\mathcal{A}_m$ .

### 3 Modular logics

Here is an example of a class of sublinear functions with some nice properties. These functions will play an important role in the rest of this paper.

*Example 3.* Fix a natural  $n > 0$ . For every natural number  $m$ , let  $t, r$  be the unique natural numbers such that  $m = tn + r$  and  $0 \leq r < n$ . Define the function  $g_n : \mathbb{N} \mapsto \mathbb{N}$  by  $g_n(m) = \begin{cases} tn & \text{if } m \geq n \\ 1 & \text{otherwise} \end{cases}$ .

For every  $n$ ,  $g_n$  is sublinear. Furthermore,  $\lim_{m \rightarrow \infty} \frac{g_n(m)}{m} = 1$ . Also, for every  $n$ ,  $g_n$  is first order definable in the following sense: there exists a formula  $\theta_n(x) \in FO$  with built-in order, addition and multiplication such that for every arithmetic model  $\mathcal{A}_m$ , for any  $a < m$ ,  $\mathcal{A}_m \models \theta_n(a)$  iff  $a+1 = g_n(m)$ . Here is why: Note first that in every  $\mathcal{A}_m$  it is possible to capture the property that  $x$  is the maximal element with a formula  $Max(x) \in FO(\{\oplus, \otimes, \triangleleft, 0, 1\})$  that says that  $\neg \exists z \oplus(x, 1, z)$ . Likewise, we can say that “the size of the model =  $tn+r$ ”, with  $r < n < (\text{size of the model})$ , by a formula  $DIVSIZE(t, n, r) \in FO(\{\oplus, \otimes, \triangleleft, 0, 1\})$  that says that there exists  $x$  such that  $Max(x)$  and

$$\begin{aligned} 0 < r < n \text{ and } x = tn + (r - 1) \text{ or} \\ r = 0 \text{ and } x = (t - 1)n + (n - 1). \end{aligned}$$

It follows then that the statement  $g_n(\text{size of model}) = h + 1$ , for  $n < m$ , is definable in the models  $\mathcal{A}_m$  by a formula in  $FO(\Gamma)$  that says that:

$$\begin{aligned} G(h, n) := \exists t, r, z (DIVSIZE(t, n, r) \wedge \\ [(\oplus(h, 1, z) \wedge \neg \oplus(0, 0, r) \wedge \otimes(t, n, z)) \vee (\oplus(0, 0, r) \wedge Max(h))]) \end{aligned}$$

For the case when  $n = m$ , we know that  $g_n(m) = m$  in which case we can define  $h$  as  $Max(h)$ . Finally, if  $m < n$  we know that  $g_n(m) = 1$  and we can define  $h = 0$ .

Recall that we refer to the probability logic restricted to finite structures that are arithmetic as  $\mathcal{LP}_A$ . The related logic restricted to modular approximations of arithmetic structures is formalise below.

**Definition 5.** We denote by  $\mathcal{LP}_F$  the probability logic restricted to structures that are  $F$ -modular approximations of arithmetic structures, for  $F$  a sublinear function. Likewise, by  $FO_F$  we understand the smallest fragment of  $\mathcal{LP}_F$  that contains the atomic formulas and is closed under  $\exists, \neg$  and  $\wedge$ . Similarly, we define  $\mathcal{LP}_F[r_1, \dots, r_k]$  as the smallest fragment of  $\mathcal{LP}_F$  that is closed under  $\exists, \neg, \wedge$  and  $(P(z) \geq q_{ij}/r_i)$  and  $(P(z) > q_{ij}/r_i)$  for  $i \leq k$  and natural numbers  $0 \leq q_{ij} < r_i$ . In particular, we define the **modular** probability logic

$$\mathcal{LP}_{MOD} = \bigcup_{n \in \mathbb{N}} \mathcal{LP}_{g_n}.$$

Likewise, we define

$$FO_{MOD} = \bigcup_{n \in \mathbb{N}} FO_{g_n} \quad \text{and} \quad \mathcal{LP}_{MOD}[r_1, \dots, r_k] = \bigcup_{n \in \mathbb{N}} \mathcal{LP}_{g_n}[r_1, \dots, r_k].$$

Note that the logics  $FO_{MOD}$ ,  $\mathcal{LP}_{MOD}[r_1, \dots, r_k]$ ,  $\mathcal{LP}_{MOD}$  do not have built-in order nor built-in addition nor built-in multiplication. Instead, for each  $n$ ,  $FO_{g_n}$ ,  $\mathcal{LP}_{g_n}[r_1, \dots, r_k]$ ,  $\mathcal{LP}_{g_n}$  have built-in  $g_n$ -modular approximations of the order, addition and multiplication.

We now show that the expressive power of  $\mathcal{LP}_{MOD}$  (respectively  $\mathcal{LP}_{MOD}[r_1, \dots, r_k]$ ,  $FO_{MOD}$ ) is contained in the expressive power of  $\mathcal{LP}_A$  (respectively  $\mathcal{LP}_A[r_1, \dots, r_k]$ ,  $FO$ ). Before proceeding, however, we need to clarify the meaning of a boolean query in the context of modular logics.

**Definition 6.** Fix a vocabulary  $\Gamma^+ = \Gamma \cup \{R_s\}_{s=1}^k \cup \{c_w\}_{w=1}^u$ . A **boolean query** for the modular logic  $\mathcal{LP}_{MOD}(\Gamma^+)$  is a map  $I : \{\mathcal{A}_m^{g_n} : m, n \in \mathbb{N}\} \rightarrow \{0, 1\}$ , with the additional property that for every  $1 < n_1 < n_2$ , for every  $m > n_2$ ,  $I(\mathcal{A}_m^{g_{n_1}}) = I(\mathcal{A}_m^{g_{n_2}})$ . We say that a boolean query is expressible in  $\mathcal{LP}_{MOD}(\Gamma^+)$  (respectively  $FO_{MOD}(\Gamma^+)$ ) iff there exists a sentence  $\theta \in \mathcal{LP}(\Gamma^+)$  (respectively  $FO$ ) such that for  $n \in \mathbb{N}$ , for every arithmetic structure  $\mathcal{A}_m$  with  $m > n$ ,  $I(\mathcal{A}_m^{g_n}) = 1$  iff  $\mathcal{A}_m^{g_n} \models \theta$ .

The idea behind the above definition of a boolean query for  $\mathcal{LP}_{MOD}$  is to capture the notion that a query does not depend on the built-in order or arithmetic predicates, instead it depends on notions that remain constant for all the approximations  $\mathcal{A}_m^{g_n}$ . For the rest of this section we fix again a vocabulary of the form  $\Gamma^+ = \Gamma \cup \{R_s\}_{s=1}^k \cup \{c_w\}_{w=1}^u$ , where  $R_s$  and  $c_w$  are numeric relations and constants.

**Lemma 2.** There exists formulas  $ADD(x_1, x_2, x_3, y)$ ,  $PRODUCT(x_1, x_2, x_3, y)$ ,  $ORDER(x_1, x_2, y)$  and for every  $s$ , formulas  $PRED_s(\bar{x}, y)$  in  $FO(\Gamma^+)$ , such that for natural  $n$ , for every arithmetic structure  $\mathcal{A}_m$  with  $m > n$ ,

- for every  $a, b, c < m$ ,  $\mathcal{A}_m^{g_n} \models \oplus(a, b, c)$  iff  $\mathcal{A}_m \models ADD(a, b, c, n)$ .
- For every  $a, b, c < m$ ,  $\mathcal{A}_m^{g_n} \models \otimes(a, b, c)$  iff  $\mathcal{A}_m \models PRODUCT(a, b, c, n)$ .
- For every  $a, b, c < m$ ,  $\mathcal{A}_m^{g_n} \models \triangleleft(a, b)$  iff  $\mathcal{A}_m \models ORDER(a, b, n)$ .
- For every index  $s$  and every  $\bar{a} < m$ ,  $\mathcal{A}_m^{g_n} \models R_s(\bar{a})$  iff  $\mathcal{A}_m \models PRED_s(\bar{a}, n)$ .

□

The previous lemma allow us to translate modular interpretations to natural interpretations.

**Corollary 1.** Let  $B$  be a boolean query expressible in  $\mathcal{LP}_{MOD}$ . Then this query is also expressible in  $\mathcal{LP}_A$ . Likewise, any boolean query expressible in  $FO_{MOD}$  (respectively  $\mathcal{LP}_{MOD}[r_1, \dots, r_k]$ ) is also expressible in  $FO_A$  (respectively  $\mathcal{LP}_A[r_1, \dots, r_k]$ ). □



The logic  $\mathcal{LP}_{MOD}$  is capable of expressing queries as the evenness of the cardinality of a set, as we show in the next example.

*Example 4.* We claim that there exists a sentence  $\theta_2$  in  $\mathcal{LP}(\{\oplus, \otimes, \triangleleft, 0, 1\})$  such that for all  $n$ , for every arithmetic structure  $\mathcal{A}_m$ , with  $m > n$ ,

$$\mathcal{A}_m^{g_n} \models \theta_2 \text{ iff } m \text{ is even}$$

To prove this, note first that for every naturals  $m > n > 1$  and every  $c$  such that  $g_n(m) > c > m - g_n(m)$ ,

$$\{y < m : \mathcal{A}_m^{g_n} \models c \triangleleft y \vee \oplus(0, y, c)\} = \{y < m : c \leq y \leq g_n(m) - 1\}$$

and this implies that for every  $c$  such that  $g_n(m) > c > m - g_n(m)$ ,

$$\mu_m(\{y < m : \mathcal{A}_m^{g_n} \models c \triangleleft y \vee \oplus(0, y, c)\}) = \frac{g_n(m) - c}{m}. \quad (1)$$

Fix now a natural  $n$ . Then there exists a natural  $k$  such that for every  $m > k$ ,  $g_n(m) > (3/4)m$  (since  $\lim_{m \rightarrow \infty} \frac{g_n(m)}{m} = 1$ ). Let  $m > k$  and consider the formula

$$\theta_2 := \exists x[(P(y) \geq 1/2)(x \triangleleft y \vee \oplus(0, x, y)) \wedge (P(y) \leq 1/2)(x \triangleleft y \vee \oplus(0, x, y))]$$

We claim that for  $m > n$ ,  $\mathcal{A}_m^{g_n} \models \theta_2$  iff  $m$  is even. One direction goes as follows: If  $m = 2s$  and  $g_n(m) > (3/4)m$  then  $m - g_n(m) < \frac{1}{2}s$ . Taking  $c$  as  $s - m + g_n(m)$ , we have by equation (1) that

$$\mu_m(\{y < m : \mathcal{A}_m^{g_n} \models (c \triangleleft y \vee \oplus(0, y, c))\}) = \frac{m - s}{m} = \frac{1}{2}$$

For the other direction, suppose that there exists a  $d < m$  such that

$$\mu_m(\{y < m : \mathcal{A}_m^{g_n} \models d \triangleleft y \vee \oplus(0, y, d)\}) = \frac{1}{2}$$

From the fact that  $\mathcal{A}_m^{g_n}$  is  $g_n$ -modular we obtain that there exists an  $a < g_n(m)$  such that

$$\mu_m(\{y < m : \mathcal{A}_m^{g_n} \models a \triangleleft y \vee \oplus(0, y, a)\}) = \frac{1}{2}$$

which implies that

$$\begin{aligned} & \mu_m(\{y < m : \mathcal{A}_m^{g_n} \models \neg(a \triangleleft y) \wedge \neg(\oplus(0, y, a))\}) \\ &= \mu_m(\{y < m : \mathcal{A}_m^{g_n} \models y \triangleleft a\}) = \frac{1}{2} \end{aligned} \quad (2)$$

Note now that  $a$  cannot be  $\leq m - g_n(m)$  because, if this was the case then from  $g_n$ -modularity we have that

$$\begin{aligned} \mu_m(\{y < m : \mathcal{A}_m^{g_n} \models y \triangleleft a\}) &\leq \mu_m(\{y < m : \mathcal{A}_m^{g_n} \models y \triangleleft m - g_n(m)\}) \\ &\leq \frac{2(m - g_n(m))}{m} = 2(1 - \frac{g_n(m)}{m}) < 2(1 - \frac{3}{4}) = \frac{1}{2} \end{aligned}$$

since for sufficiently large  $m$ ,  $\frac{g_n(m)}{m} > \frac{3}{4}$ , but this contradicts (2).

Thus  $m - g_n(m) < a < g_n(m)$ . We can apply now equation (1) to obtain that

$$\frac{1}{2} = \mu_m(\{y < m : \mathcal{A}_m^{g_n} \models a \triangleleft y \vee y = a\}) = \frac{g_n(m) - a}{m}$$

Hence  $\frac{1}{2} = (g_n(m) - a)/m$ , that is,  $g_n(m) - a = m/2$ , so  $m$  must be even.

In a similar way, one can prove that for every natural  $d > 2$ , there exists a formula  $\theta_d$  in  $FO + \{P(z) \geq 1/d, P(z) > (d-1)/d\}(\{\oplus, \otimes, \triangleleft, 0, 1\})$  such that for every natural  $n$ , for every arithmetic structure  $\mathcal{A}_m$  with  $m > n$ ,  $\mathcal{A}_m^{g_n} \models \theta_d$  iff  $m$  is a multiple of  $d$ .

A consequence of the above example is that the boolean query “the size of the model is divisible by  $d$ ”, for  $d > 1$ , is expressible in  $(FO + \{P(z) \geq 1/d, P(z) > (d-1)/d\})_{MOD}$ .

## 4 Separation results for modular logics

In this section we prove separation results between fragments of  $\mathcal{LP}_{MOD}$  defined in Definition 2. Since a formula such as  $\neg((P(z) > \epsilon)\varphi)$  is equivalent to  $(P(z) \geq 1 - \epsilon)\neg\varphi$  (and  $\neg((P(z) \geq \epsilon)\varphi)$  is equivalent to  $(P(z) > 1 - \epsilon)\neg\varphi$ ) we can push all negation symbols inside and together with all well known ways of manipulating quantifiers in a formula, we get the following prenex normal form for formulas in  $\mathcal{LP}$ .

**Theorem 1.** *For every formula  $\phi(\bar{x}) \in \mathcal{LP}(\Gamma^+)[r_1, r_2, \dots, r_k]$  there exists a quantifier free formula  $\theta(y_1, \dots, y_w, \bar{x}) \in \mathcal{LP}(\Gamma)[r_1, r_2, \dots, r_k]$  such that for every structure  $\mathcal{B}_m$ , for every vector of naturals  $\bar{a} < m$ ,*

$$\mathcal{B}_m \models \phi(\bar{a}) \leftrightarrow Q_1 y_1 Q_2 y_2 \dots Q_w y_w \theta(\bar{y}, \bar{a}),$$

where each quantifier  $Q_s$  is either  $\exists$  or  $\forall$  or  $(P(z) > q_{ij}/r_i)$  or  $(P(z) \geq q_{ij}/r_i)$ , for some  $i \in \{1, \dots, k\}$  and some  $0 \leq q_{ij} < r_i$ .  $\square$

We proceed now to define the notion of an  $F$ -chain of models and the stronger notion of a chain.

**Definition 7.** *Fix a sublinear function  $F$ . An  $F$ -chain of models  $\mathbf{C}$  is a collection of finite structures for  $\Gamma^+ = \Gamma \cup \{R_s\}_{s=1}^n \cup \{c_r\}_{r=1}^t$  with the following property:*

- *For every relation symbol  $R(\bar{x})$  of  $\Gamma^+$ , for every two models  $\mathcal{B}_m, \mathcal{B}_n$  in  $\mathbf{C}$  with  $m \leq n$  and  $F(m) = F(n)$ , and for every  $\bar{a} < F(m)$ ,  $\mathcal{B}_m \models R(\bar{a})$  iff  $\mathcal{B}_n \models R(\bar{a})$ .*

*A chain of models  $\mathbf{C}$  is a collection of finite structures for  $\Gamma^+$  with the following property:*

- For every relation symbol  $R(\bar{x})$  of  $\Gamma^+$ , for every two models  $\mathcal{B}_m, \mathcal{B}_n$  in  $\mathbf{C}$  with  $m \leq n$  and for every  $\bar{a} < m$ ,  $\mathcal{B}_m \models R(\bar{a})$  iff  $\mathcal{B}_n \models R(\bar{a})$ .

In other words, chains are collections of models with inter-compatibility for its predicates.

*Remark 2.* If  $\mathbf{C}$  is a chain of arithmetic models then, for every sublinear function  $F$ ,  $\mathbf{C}^F = \{\mathcal{A}_m^F : \mathcal{A}_m \in \mathbf{C}\}$  is an  $F$ -chain.

*Example 5.* Let  $\{\mathcal{A}_m\}_{m=1}^\infty$  be the collection of arithmetic models for  $\Gamma = \{\oplus, \otimes, \triangleleft, 0, 1\}$ . It is easy to check that this collection is a chain.

We are ready to obtain separation results for the expressive power of the different modular logics. Our main tool is the following lemma which establish conditions for elementary equivalence. It states that for every sentence  $\phi$  in  $\mathcal{LP}(\Gamma^+)$ , models that are in the same chain and have almost the same size can not distinguish  $\phi$ .

**Lemma 3.** *Let  $F$  be a sublinear function and  $\mathbf{C}$  an  $F$ -chain of models. Let  $r_1, r_2, \dots, r_k$  be distinct non zero natural numbers. Let  $\phi(x_1, \dots, x_s)$  be any formula in  $\mathcal{LP}(\Gamma^+)[r_1, r_2, \dots, r_k]$ . Then one of the following two possibilities hold:*

1. *For every two  $F$ -modular models  $\mathcal{B}_m$  and  $\mathcal{B}_{m+1}$  in  $\mathbf{C}$  such that  $m+1 > r_i$  and  $m \equiv_{r_i} -1$ , for every  $i \leq k$  and  $F(m) = F(m+1)$ , we have that, for every  $a_1, \dots, a_s < m$ ,  $\mathcal{B}_m \models \phi(a_1, \dots, a_s)$  implies  $\mathcal{B}_{m+1} \models \phi(a_1, \dots, a_s)$ , or*
2. *For every two  $F$ -modular models  $\mathcal{B}_m$  and  $\mathcal{B}_{m+1}$  in  $\mathbf{C}$  such that  $m+1 > r_i$  and  $m \equiv_{r_i} -1$ , for every  $i \leq k$  and  $F(m) = F(m+1)$ , we have that, for every  $a_1, \dots, a_s < m$ ,  $\mathcal{B}_{m+1} \models \phi(a_1, \dots, a_s)$  implies  $\mathcal{B}_m \models \phi(a_1, \dots, a_s)$ .*

*Proof.* We proceed by induction on the quantifier rank of  $\phi$ .

**Quantifier Free Formulas:** By definition of  $F$ -chain, if  $\phi(x_1, \dots, x_s)$  is quantifier free and  $a_1, \dots, a_s < F(m)$ , we have that

$$\mathcal{B}_m \models \phi(a_1, \dots, a_s) \text{ if and only if } \mathcal{B}_{m+1} \models \phi(a_1, \dots, a_s),$$

We prove that this equivalence holds for  $a_1, \dots, a_s < m$ . For each coordinate  $a_i$  such that  $F(m) \leq a_i < m$ , pick  $b_i < F(m)$  such that  $b_i \equiv_{F(m)} a_i$ , and otherwise take  $b_i = a_i$ . Since  $\mathcal{B}_m$  and  $\mathcal{B}_{m+1}$  are  $F$ -modular,  $\mathcal{B}_k \models \phi(a_1, \dots, a_s) \iff \mathcal{B}_k \models \phi(b_1, \dots, b_s)$  for  $k = m, m+1$ . From this it follows the desired equivalence for  $a_1, \dots, a_s < m$  and  $F(m) = F(m+1)$ .

**Existentially or Universally Quantified Formulas:** These two cases are not difficult to prove and we omit the proofs for lack of space. (Hint: the direction from  $\mathcal{B}_{m+1}$  to  $\mathcal{B}_m$  use that  $\mathcal{B}_{m+1}$  is  $F$ -modular.)

**Probability Quantifiers:** We assume that case 1. holds, that is, for  $F, m, r_1, \dots, r_k$  as in the hypothesis and for every  $a_1, \dots, a_s < m$  and every  $b < m$ :

$$\mathcal{B}_m \models \phi(a_1, \dots, a_s, b) \text{ implies } \mathcal{B}_{m+1} \models \phi(a_1, \dots, a_s, b).$$

We have two cases to consider under these hypothesis.

We consider first the formula,  $(P(z) \geq q_{ij}/r_i)\phi(\bar{a}, z)$ . Fix an arbitrary  $m$  satisfying that  $F(m) = F(m+1)$ ,  $m+1 > r_i$  and  $m \equiv_{r_i} -1$  for every  $i \leq k$ , fix  $a_1, \dots, a_s < m$ . Let  $t$  be a natural number such that  $m = tr_i + r_i - 1$ . Now, if  $\mathcal{B}_m \models (P(z) \geq q_{ij}/r_i)\phi(\bar{a}, z)$  and since  $\gcd(r_i, m) = 1$ , then

$$|\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}| > \frac{q_{ij}m}{r_i} = \frac{q_{ij}(tr_i + r_i - 1)}{r_i} = q_{ij}(t+1) - \frac{q_{ij}}{r_i}$$

and since  $q_{ij} < r_i$ , we obtain that  $|\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}| \geq q_{ij}(t+1)$ . By induction hypothesis we get that

$$|\{z < m+1 : \mathcal{B}_{m+1} \models \phi(\bar{a}, z)\}| \geq q_{ij}(t+1) = \frac{q_{ij}}{r_i}(t+1)(r_i) = \frac{q_{ij}}{r_i}(m+1),$$

which implies that  $\mu(\{z < m+1 : \mathcal{B}_{m+1} \models \phi(\bar{a}, z)\}) \geq q_{ij}/r_i$ , that is  $\mathcal{B}_{m+1} \models (P(z) \geq q_{ij}/r_i)\phi(\bar{a}, z)$ , which is the desired result.

Next we consider the formula  $(P(z) > q_{ij}/r_i)\phi(\bar{a}, z)$  and we shall prove that case 2. holds for this formula. Fix an arbitrary  $m$  satisfying that  $F(m) = F(m+1)$ ,  $m+1 > r_i$  and  $m \equiv_{r_i} -1$  for every  $i \leq k$ , fix  $a_1, \dots, a_s < m$ . Let  $t$  be a natural number such that  $m = tr_i + r_i - 1$ . If  $\mathcal{B}_m \models (P(z) > q_{ij}/r_i)\phi(\bar{a}, z)$  and since  $\gcd(r_i, m) = 1$ , then

$$|\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}| < \frac{q_{ij}m}{r_i} = \frac{q_{ij}(tr_i + r_i - 1)}{r_i} = q_{ij}(t+1) - \frac{q_{ij}}{r_i}$$

and since  $q_{ij} < r_i$ , we obtain that

$$|\{z < m : \mathcal{B}_m \models \phi(\bar{a}, z)\}| \leq q_{ij}(t+1).$$

By induction hypothesis we get that

$$|\{z < m+1 : \mathcal{B}_{m+1} \models \phi(\bar{a}, z)\}| \leq q_{ij}(t+1) = \frac{q_{ij}}{r_i}(t+1)(r_i) = \frac{q_{ij}}{r_i}(m+1).$$

which implies that  $\mu(\{z < m+1 : \mathcal{B}_{m+1} \models \phi(\bar{a}, z)\}) \leq q_{ij}/r_i$ , that is,  $\mathcal{B}_{m+1} \models (P(z) \leq q_{ij}/r_i)\phi(\bar{a}, z)$ , which give us case 2. for this formula.

The proofs for both type of probability quantifiers under the assumption that case 2. holds for  $\phi$  are just the counterpositive versions of the two cases just proved.  $\square$

The above lemma can be used to prove separation of different fragments of  $\mathcal{LP}_{MOD}$ .

**Theorem 2.** *Let  $r, r_1, r_2, \dots, r_k$  be distinct non zero natural numbers, and such that  $r$  is relatively prime with each  $r_1, \dots, r_k$ . Then  $\mathcal{LP}_{MOD}[r_1, \dots, r_k]$  is properly contained in  $\mathcal{LP}_{MOD}[r_1 \dots r_k, r]$ .*

*Proof.* It is obvious that  $\mathcal{LP}_{MOD}[r_1, \dots, r_k]$  is contained in  $\mathcal{LP}_{MOD}[r_1, \dots, r_k, r]$ . Furthermore, we saw (Example 4) that the query: “the size of the model is a multiple of  $r$ ” is expressible in  $\mathcal{LP}_{MOD}(\Gamma)[r]$ . We will show that this query is not expressible in  $\mathcal{LP}[r_1, \dots, r_k]_{MOD}(\Gamma)$ . More specifically, we will show that there is no sentence  $\phi$  in  $\mathcal{LP}_{g_n}[r_1, \dots, r_k](\Gamma)$  that defines the above query, where  $g_n$  is the sublinear function defined in Example 3, for all  $n > (\prod_{i=1}^k r_i)r$ .

Recall that the collection of all arithmetic models  $\mathbf{C} = \{\mathcal{A}_m\}_{m=1}^\infty$  forms a chain. It follows that for every  $n$ , the collection  $\mathbf{C}^{g_n} = \{\mathcal{A}_m^{g_n}\}_{m=1}^\infty$  forms a  $g_n$ -chain. Suppose now that there exists a sentence  $\phi$  in  $\mathcal{LP}_{g_n}[r_1, \dots, r_k](\Gamma)$  that captures the query “the size of the model is a multiple of  $r$ ” for all (except finitely many) structures  $\mathcal{A}_m^{g_n}$ . Then we can apply Lemma 3 and get the following:

For every two models  $\mathcal{A}_m^{g_n}$  and  $\mathcal{A}_{m+1}^{g_n}$  in  $\mathbf{C}^{g_n}$  such that  $m+1 > r_i$ ,  $m \equiv_{r_i} -1$  for every  $i$ , and  $g_n(m) = g_n(m+1)$ , we have that at least one of the following two cases hold  
 (1)  $\mathcal{A}_m^{g_n} \models \phi$  implies  $\mathcal{A}_{m+1}^{g_n} \models \phi$ , or (2)  $\mathcal{A}_{m+1}^{g_n} \models \phi$  implies  $\mathcal{A}_m^{g_n} \models \phi$ .

Suppose it is case 1. that is true. Using that  $r$  is relatively prime with the  $r_i$ ’s together with the Generalized Chinese Remainder Theorem we can get a natural number  $b \leq (\prod_{i=1}^k r_i)r$  such that  $b \equiv_{r_i} -1$  for every  $i$  and  $b \equiv_r 0$ . Let  $D$  be the collection of naturals  $m$  such that  $m = r(\prod_{i=1}^k r_i)tn + b$  for some natural  $t > 0$ . Clearly  $m+1 > r_i$ ,  $m \equiv_{r_i} -1$  for every  $i$ , and  $g_n(m) = g_n(m+1)$ . Furthermore,  $D$  is infinite and for every  $m \in D$ ,  $m \equiv_r 0$ . It follows that for almost all the  $m \in D$ ,  $\mathcal{A}_m^{g_n} \models \phi$  and, in consequence, for almost all the  $m \in D$ ,  $\mathcal{A}_{m+1}^{g_n} \models \phi$ , i.e. for almost all elements  $m$  of  $D$ ,  $m+1$  is a multiple of  $r$ , which is impossible.

Suppose it is case 2. that is true. Then by a similar argument as above we prove the existence of  $b \leq (\prod_{i=1}^k r_i)r$  such that  $b \equiv_{r_i} -1$  for every  $i$  and  $b \equiv_r -1$ . Let  $D$  be the same as above. Then  $D$  is infinite and for every  $m \in D$ ,  $m \equiv_r -1$ . It follows that for almost all the  $m \in D$ ,  $\mathcal{A}_{m+1}^{g_n} \models \phi$  and, in consequence, for almost all the  $m \in D$ ,  $\mathcal{A}_m^{g_n} \models \phi$ , i.e. for almost all elements  $m$  of  $D$ ,  $m$  is a multiple of  $r$ , which is impossible.

We conclude that such sentence  $\phi$  can not exists in  $\mathcal{LP}_{g_n}[r_1, \dots, r_k](\Gamma)$ .  $\square$

**Corollary 2.** *The expressive power of  $FO_{MOD}$  is strictly weaker than the expressive power of  $\mathcal{LP}_{MOD}[2]$ .  $\square$*

This last result, for modular logics, corresponds to the separation of FO and  $FO + M$  in the context of arithmetic models, which in turn is equivalent to the separation of  $AC^0$  from  $TC^0$  shown by Ajtai and independently by Furst, Saxe and Sipser (see [5] for a nice exposition of this result and references).

## 5 Approximating $\mathcal{LP}_A$ with $\mathcal{LP}_{MOD}$

We introduce the notion of approximate formulas. This concept will provide a link between satisfaction in arithmetic structures and satisfaction in modular approximations of these arithmetic structures.

**Definition 8 (Approximate Formulas).** For every formula in prenex normal form  $\theta(\bar{x}) \in \mathcal{LP}(\Gamma^+)$ , for every  $0 \leq \epsilon < 1$ , we define the  $\epsilon$ -approximation of  $\theta(\bar{x})$  as follows:

**Atomic formulas** If  $\theta(\bar{x}) := R_s(\bar{x}, \bar{c})$  then  $\theta_\epsilon(\bar{x}) := R_s(\bar{x}, \bar{c})$ .

**Negation of atomic formulas** If  $\theta(\bar{x}) := \neg R_s(\bar{x}, \bar{c})$  then  $\theta_\epsilon(\bar{x}) := \neg R_s(\bar{x}, \bar{c})$ .

**Conjunction** If  $\theta(\bar{x}) := \phi(\bar{x}) \wedge \psi(\bar{x})$  then  $\theta_\epsilon(\bar{x}) := \phi_\epsilon(\bar{x}) \wedge \psi_\epsilon(\bar{x})$ .

**Disjunction** If  $\theta(\bar{x}) := \phi(\bar{x}) \vee \psi(\bar{x})$  then  $\theta_\epsilon(\bar{x}) := \phi_\epsilon(\bar{x}) \vee \psi_\epsilon(\bar{x})$ .

**Existential quantification** If  $\theta(\bar{x}) := \exists z \phi(\bar{x}, z)$  then  $\theta_\epsilon(\bar{x}) := \exists z \phi_\epsilon(\bar{x}, z)$ .

**Universal quantification** If  $\theta(\bar{x}) := \forall z \phi(\bar{x}, z)$  then

$$\theta_\epsilon(\bar{x}) := (P(z) > 1 - \epsilon) \phi_\epsilon(\bar{x}, z).$$

**Probability quantifiers** If  $\theta(\bar{x}) := (P(z) > r) \phi(\bar{x}, z)$  then

$$\theta_\epsilon(\bar{x}) := (P(z) > r - \min(\epsilon, r)) \phi_\epsilon(\bar{x}, z).$$

$$\text{If } \theta(\bar{x}) := (P(z) \geq r) \phi(\bar{x}, z) \text{ then } \theta_\epsilon(\bar{x}) := (P(z) \geq r - \min(\epsilon, r)) \phi_\epsilon(\bar{x}, z).$$

The next lemma provides the basic operational properties of the approximate formulas.

**Lemma 4.** For every formula (in prenex normal form)  $\theta(\bar{x}) \in \mathcal{LP}(\Gamma^+)$ , for every  $0 < \epsilon < 1$ , for every finite structure  $\mathcal{B}_m$  and every vector  $\bar{a} < m$  the following holds:

- If  $0 < \epsilon < \delta < 1$  then  $\mathcal{B}_m \models \theta(\bar{a}) \rightarrow \theta_\epsilon(\bar{a}) \rightarrow \theta_\delta(\bar{a})$ .
- If  $\{\epsilon_i\}_{i=1}^\infty$  is a sequence of real numbers less than 1 and converging to 0, then

$$\text{If } (\forall i \in \mathbb{N}, \mathcal{B}_m \models \theta_{\epsilon_i}(\bar{a})) \text{ then } \mathcal{B}_m \models \theta(\bar{a}).$$

The purpose of the next theorem is to establish an “approximation” relationship between satisfaction in the modular logic  $\mathcal{LP}_{MOD}$  and satisfaction in  $\mathcal{LP}_A$  via the approximate formulas.

**Theorem 3. (Bridge Theorem).** Fix a natural  $n$ . For every formula in prenex normal form  $\theta(\bar{x}) \in \mathcal{LP}(\Gamma^+)$ , for every arithmetic model  $\mathcal{A}_m$  with  $m > n^2$ , for every  $\bar{a} < g_n(m)$ , the following holds:  $\mathcal{A}_m^{g_n} \models \theta(\bar{a})$  implies  $\mathcal{A}_m \models \theta_{1/n}(\bar{a})$ .

*Proof.* By induction in the complexity of the formula.

**Atomic formulas and negation of atomic formulas** (Hint: for atomic formulas and their negation  $\theta_{1/n}$  is the same as  $\theta$ .)

**Conjunction, disjunction** Direct.

**Existential quantifier** (Hint: Suppose  $\mathcal{A}_m^{g_n} \models \exists z \theta(\bar{a}, z)$ . Then use Lemma 1 and that  $\mathcal{A}_m^{g_n}$  is  $g_n$ -modular to conclude  $\theta(\bar{x}, z)$  is  $g_n$ -modular for  $\mathcal{A}_m^{g_n}$  and, hence,  $\mathcal{A}_m^{g_n} \models \theta(\bar{a}, [c]_{g_n(m)})$  for some  $c < m$ .)

**Universal quantifier** Suppose that  $\mathcal{A}_m^{g_n}$  satisfies the formula  $\forall z \theta(\bar{a}, z)$ . Then for every  $c < g_n(m)$  we have that  $\mathcal{A}_m^{g_n} \models \theta(\bar{a}, c)$ . We can apply now the induction hypothesis to obtain that for every  $c < g_n(m)$  we have that  $\mathcal{A}_m \models \theta_{1/n}(\bar{a}, c)$ . Since  $\frac{m - g_n(m)}{m} \leq \frac{n}{m}$  and  $m > n^2$  we get that  $\frac{g_n(m)}{m} > 1 - \frac{1}{n}$ , which implies  $\mathcal{A}_m \models (P(z) > 1 - \frac{1}{n}) \theta_{1/n}(\bar{a}, c)$ .

**Probability quantification** Suppose that  $\mathcal{A}_m^{g_n}$  satisfies the formula  $(P(z) > r)\theta(\bar{a}, z)$  for  $0 < r < 1$ . It follows that  $|\{c < m : \mathcal{A}_m^{g_n} \models \theta(\bar{a}, c)\}| > rm$ . Then we get that

$$|\{c < g_n(m) : \mathcal{A}_m^{g_n} \models \theta(\bar{a}, c)\}| > rm - (m - g_n(m)).$$

Applying the induction hypothesis we obtain that

$$|\{c < m : \mathcal{A}_m \models \theta_{1/n}(\bar{a}, c)\}| > rm - (m - g_n(m)).$$

It follows that

$$\begin{aligned} \mu_m(\{c < m : \mathcal{A}_m \models \theta_{1/n}(\bar{a}, c)\}) &> \frac{rm - (m - g_n(m))}{m} = \\ r - \frac{(m - g_n(m))}{m} &= r - \frac{1}{n} \text{ since } m > n^2. \end{aligned}$$

But this last statement is just  $\mathcal{A}_m \models (P(z) > r - \frac{1}{n})\theta_{1/n}(\bar{a}, z)$ .  $\square$

The gist of the above result is to give a quantifiable relationship between satisfaction of a formula in the structures  $\mathcal{A}_m^{g_n}$  and satisfaction of its approximation in  $\mathcal{A}_m$ . It implies the following relationship between boolean queries captured by  $\mathcal{LP}_A$  and the boolean queries captured in  $\mathcal{LP}_{MOD}$ . (We will abbreviate by  $(\neg\theta)_\epsilon$ , for  $\theta \in \mathcal{LP}(\Gamma^+)$ , the  $\epsilon$ -approximation of the formula equivalent to  $\neg\theta$ .)

**Corollary 3.** *Assume there is a boolean query  $B$ , a natural  $n$  and a formula  $\theta \in \mathcal{LP}(\Gamma^+)$  such that for every arithmetic model  $\mathcal{A}_m$ , with  $m > n^2$ , if  $\mathcal{A}_m \models \theta_{1/n}$  then  $\mathcal{A}_m \in B$ , and if  $\mathcal{A}_m \models (\neg\theta)_{1/n}$  then  $\mathcal{A}_m \notin B$ . Then for every  $m > n^2$ ,  $\mathcal{A}_m \in B$  iff  $\mathcal{A}_m^{g_n} \models \theta$ .  $\square$*

## 6 P and the logic LP extended

The first problem shown to be complete for the class **P**, deterministic polynomial time, was *Path System Accessibility* due to Cook [2]. An instance of the Path System Accessibility problem, which we abbreviate from now on as PS, is a finite structure  $\mathcal{A} = \langle A, R, T, s \rangle$ , 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 set of *targets*  $T \subseteq A$  such that  $s \notin T$ . A positive instance of PS is a path system  $\mathcal{A}$  where some target in  $T$  is *accessible* from the source  $s$ , 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 [8], Stewart shows that PS is complete for **P** via quantifier free first order reductions; in fact, via *projections* (see [8] for definitions and also [5] Section 11.2), and we will use that result to show that an approximation version of PS which we present in Example 6 below is also complete for **P** via reductions that are projections, and that would help us to show that a certain extension of our  $\mathcal{LP}$  logic captures **P** on finite ordered structures. (We remark that Stewart considers the path systems in [8] as having only one target, and not a set of targets as we do here. However one can see that his results on completeness of PS via first order reductions holds also for our version of this problem.)

**Definition 9.** Let  $X$  be a second order variable of arity 1, and  $\alpha(\bar{x}, X)$  a first order formula over some (finite) vocabulary  $\tau$  with first order variables  $\bar{x} = (x_1, \dots, x_m)$  and second order variable  $X$ . Let  $r \in [0, 1]$ . Then

$$(P(X) > r)\alpha(\bar{x}, X) \quad \text{and} \quad (P(X) \geq r)\alpha(\bar{x}, X)$$

are new formulas with the following semantic. For an appropriate finite  $\tau$ -model  $\mathcal{A}_n$ , and elements  $\bar{a} = (a_1, \dots, a_m)$  from  $\{0, \dots, n-1\}$ , the universe of  $\mathcal{A}_n$ ,

$$\begin{aligned} \mathcal{A}_n &\models (P(X) > r)\alpha(\bar{a}, X) \\ \iff &\text{the least subset } A \subseteq \{0, \dots, n-1\} \text{ such that} \\ &\mathcal{A}_n \models \alpha(\bar{a}, A) \text{ has } |A|/n > r \end{aligned}$$

Similarly for  $(P(X) \geq r)\alpha(\bar{a}, X)$ .

*Example 6.* Let  $\tau = \{R, T, s\}$  where  $R$  is a ternary relation symbol,  $T$  is a unary relation symbol and  $s$  is a constant symbol. We think of  $\tau$ -structures as path systems with source  $s$ , a target set  $T$  and set of rules  $R$ . Let  $r$  be a rational with  $0 < r < 1$ . We define

$\text{NPS}_{\geq r} := \{\mathcal{A} = \langle A, R, T, s \rangle : \mathcal{A} \text{ is a path system and at least a fraction } r \text{ of the elements accessible from } s \text{ are not in } T\}$

Let  $\alpha_{nps}(X)$  be the following formula (the constant symbol  $\perp$  stands for false),

$$\begin{aligned} \alpha_{nps}(X) &:= \forall x(x = s \longrightarrow X(x)) \\ &\quad \wedge \forall x \forall y \forall z (X(x) \wedge X(y) \wedge R(x, y, z) \longrightarrow X(z)) \\ &\quad \wedge \forall x (X(x) \wedge T(x) \longrightarrow \perp) \end{aligned}$$

Then

$$\mathcal{A}_n \in \text{NPS}_{\geq r} \iff \mathcal{A}_n \models (P(X) \geq r)\alpha_{nps}(X)$$

$\text{NPS}_{\geq r}$  is an approximation version of the problem PS, definable by our probability quantifiers over unary second order variables acting on formulas with a particular form to which we give a name below.

**Definition 10.** Let  $\tau = \{R_1, \dots, R_m, C_1, \dots, C_k\}$  be some vocabulary with relation symbols  $R_1, \dots, R_m$ , and constant symbols  $C_1, \dots, C_k$ , and let  $X$  be a unary second order variable. A first order formula  $\alpha$  over  $\tau \cup \{X\}$ , and extra symbols  $=$  (equality) and the constant  $\perp$  (standing for false), is a universal Horn formula, if  $\alpha$  is the conjunction of universally quantified formulas over  $\tau \cup \{X\}$  of the form

$$\psi_1 \wedge \psi_2 \wedge \dots \wedge \psi_s \longrightarrow \varphi$$

where  $\varphi$  is either  $X(\bar{u})$  or  $\perp$ , and  $\psi_1, \dots, \psi_s$  are atomic  $(\tau \cup \{X\})$ -formulas with any occurrence of the variable  $X$  being positive (there are no restrictions on the predicates in  $\tau$  or  $=$ ).

The logic  $\mathcal{LP}_{\text{Horn}}$  is the set of formulas

$$\text{FO} + \{(P(X) > r)\alpha_1(\bar{x}, X), (P(X) \geq r)\alpha_2(\bar{x}, X) : \alpha_i(\bar{x}, X) \text{ is universal Horn (first order) formula with second order variable } X\}$$



Example 6 shows that the problem  $\text{NPS}_{\geq r}$  is definable in  $\mathcal{LP}_{Horn}$ . We shall see that this is true of all problems in **P**

**Lemma 5.** *The set of finite structures that satisfy a sentence  $\theta$  in  $\mathcal{LP}_{Horn}$  is in **P**.*

*Proof.* Let  $\theta \in \mathcal{LP}_{Horn}$  be of the form

$$(P(X) > r) \left[ \bigwedge_{i=1}^m \forall \bar{x}_i (\psi_{i1} \wedge \dots \wedge \psi_{is} \longrightarrow \varphi_i) \right],$$

and let  $\mathcal{A}_n$  be a model of the appropriate vocabulary of size  $n$ . Then it's not difficult to describe a polynomial time procedure that decides whether  $\mathcal{A}_n$  satisfies the above sentence.  $\square$

Thus, according to this lemma, our problem  $\text{NPS}_{\geq r}$  is in **P**. We show next that it is hard for **P**.

**Lemma 6.** *The problem  $\text{NPS}_{\geq r}$  is complete for **P** via projections.*

*Proof.* We exhibit a (successor free) projection from the complement of the problem PS to  $\text{NPS}_{\geq r}$ . Let  $\mathcal{A} = \langle A, R, T, s \rangle$  be an instance of PS. Define  $\mathcal{A}' = \langle A', R', T', s' \rangle$  as follows: its universe  $A' = A^2$ , and

$$\begin{aligned} T' &= T \times s = \{(x, s) : x \in T\} \\ R' &= \{((x, s), (y, s), (z, s)) : (x, y, z) \in R\} \cup \\ &\quad \{((x, s), (y, s), (z, s)) : x \in T \wedge x \neq s \wedge y \in T \wedge y \neq s \wedge z \neq s\} \\ s' &= (s, s) \end{aligned}$$

Then,  $\mathcal{A} \in \text{PS} \iff \mathcal{A}' \notin \text{NPS}_{\geq r}$ .  $\square$

**Corollary 4.** *Every problem in **P** is a set of finite ordered structures that satisfy a sentence in  $\mathcal{LP}_{Horn}$*

*Proof.* Every problem in **P** is reducible to  $\text{NPS}_{\geq r}$  via projections;  $\text{NPS}_{\geq r}$  is definable in  $\mathcal{LP}_{Horn}$  and this logic is closed via projections.  $\square$

**Corollary 5.** *Over finite ordered structures, the logic  $\mathcal{LP}_{Horn}$  captures **P**.*  $\square$

The logic  $\mathcal{LP}_{Horn}$  verifies Lemma 1; namely, for a sublinear function  $F$ ,  $F$ -modularity is preserved. Indeed, we need only to check for formulas of the form  $(P(X) > r)\alpha(\bar{z}, X)$ : Suppose  $\bar{a}, \bar{b} < m$ ,  $\bar{a} \equiv_{F(m)} \bar{b}$  and  $\mathcal{B}_m \models (P(X) > r)\alpha(\bar{a}, X)$ . Then there exists a  $B \subseteq \{0, 1, \dots, m-1\}$ , such that  $\mathcal{B}_m \models \alpha(\bar{a}, B)$  and  $|B| > rm$ . The parameters in  $\bar{a}$  do not occur in  $B$ ; hence, by inductive hypothesis  $\mathcal{B}_m \models \alpha(\bar{b}, B)$ . Thus,  $\mathcal{B}_m \models (P(X) > r)\alpha(\bar{b}, X)$ .  $\square$

## References

1. Barrington, D., Immerman, N., Straubing, H.: On uniformity within  $NC^1$ . *J. Computer and Syst. Sci.* **41** (1990) 274–306.
2. Cook, S. A.: An observation on time-storage trade off, *J. Comput. System Sci.* **9** (1974) 308–316.
3. Etessami, K., Immerman, N.: Reachability and the power of local ordering, *Theo. Comp. Sci.* **148**, 2 (1995) 261–279.
4. Gurevich, Y.: Logic and the challenge of computer science. In: “Current trends in theoretical computer science“ (E. Börger, Ed.) Computer Science Press. (1988) 1-57.
5. Immerman, N.: *Descriptive Complexity*. Springer (1998).
6. Keisler, H. J.: Hyperfinite model theory. In: “Logic Colloquium 76” R.C. Gandy and J.M. E. Hyland, Eds.), North-Holland (1977) .
7. Libkin, L., Wong, L.: Lower bounds for invariant queries in logics with counting. *Theoretical Comp. Sci.* **288** (2002), 153-180.
8. Stewart, I.: Logical description of monotone NP problems, *J. Logic Computat.* **4**, 4 (1994) 337-357.