Polynomial-Time Relational Probabilistic Inference in Open Universes

Luise Ge[1] , **Brendan Juba**[1] , **Kris Nilsson**[1]

[1]Washington University in St. Louis

{g.luise,bjuba,k.b.nilsson}@wustl.edu,

Abstract

Reasoning under uncertainty is a fundamental challenge in Artificial Intelligence. As with most of these challenges, there is a harsh dilemma between the expressive power of the language used, and the tractability of the computational problem posed by reasoning. Inspired by human reasoning, we introduce a method of first-order relational probabilistic inference that satisfies both criteria, and can handle hybrid (discrete and continuous) variables. Specifically, we extend sum-of-squares logic of expectation to relational settings, demonstrating that lifted reasoning in the bounded-degree fragment for knowledge bases of bounded quantifier rank can be performed in polynomial time, even with an *a priori* unknown and/or countably infinite set of objects. Crucially, our notion of tractability is framed in *proof-theoretic terms*, which extends beyond the syntactic properties of the language or queries. We are able to derive the tightest bounds provable by proofs of a given degree and size and establish completeness in our sum-of-squares refutations for fixed degrees.

1 Introduction

Intelligent agents must cope with limits to their knowledge of the world. One of the most widely studied and mature approaches treats the state of the world as probabilistic, using inference from observed evidence to draw conclusions about what is likely—and what is not—regarding unobserved parts of the world, future states, and so on.

At the same time, it is profitable to use relational representations for our knowledge of the world. In this way, a single, short statement can assert a common collection of relationships about a large collection of similar objects. In a vast world, such relational knowledge enables an agent to draw on some basic understanding of situations, locations, or objects that it has never before encountered. There is substantial evidence from cognitive science showing that a core aspect of human intelligence arises from relational thinking and reasoning [Hummel and Holyoak, 2003; Krawczyk, 2012]. And one can argue that relational representations are highly desir-

able and perhaps even essential for many areas of AI such as planning and natural language processing.

Unfortunately, power rarely comes for free. Naïve approaches to both probabilistic inference and relational inference, that might explicitly consider the various possible states of the world, are not computationally feasible to perform at scale. Since our focus is on addressing a vast world, we need approaches that do not explicitly consider the possible world states, but instead operate at a more abstract "lifted" level, reasoning generically about entire classes of objects. In general, it is still challenging to carry out such a strategy efficiently, and this is an area of active research [Van den Broeck *et al.*, 2021]. In this work, we consider a new logic with a tractable fragment that can capture knowledge that is out of reach of the existing tractable methods.

Our approach builds on and extends three distinct threads in the literature. The first is work on tractable probability logics. Our work presents a relational generalization of the sum-of-squares probability logic [Lasserre, 2001; Juba, 2019], while retaining a powerful tractable fragment given by bounded-degree expressions.

The second thread is relational optimization. Sum-of-squares, in the primal formulation, is a family of semidefinite optimization problems. Kersting *et al.* [2017] and Mladenov *et al.* [2017] respectively proposed relational generalizations of linear and convex quadratic programs, and developed tools for efficiently solving such problems. Here, we show that the techniques may be extended to countably infinite domains, and present a logic that characterizes the power of the programs we obtain via these techniques.

Finally, the third thread concerns the existing technique for inference in such infinite universes: Belle [2017] showed how to perform weighted first-order model counting in such open universes, which enables certain kinds of probabilistic inference. We extend Belle's approach, and overcome two main limitations: first, that first-order model counting is only known to be tractable for various two-variable fragments, and is postulated to be intractable for many three-variable formulas and beyond [Beame *et al.*, 2015]. Second, Belle's technique to extend to infinite universes relies on compactness for Boolean propositional logic, which clearly does not hold even for linear arithmetic—consider the example of the formulas $e(x) \geq n$ for all integer n (where $e(x)$ represents the expected value for x, which is included in our language), the

infinite set of constraints $\{e(x) \geq n : n$ is an integer$\}$ is unsatisfiable, whereas any finite subset is satisfiable.

Our generalization, in contrast to others, enables us to reason about the expected values of random variables that take numeric values as well as bounded-degree moments.

2 Related Work

Probability logics Most directly, our work proposes a logic of expectation, as discussed by Halpern and Pucella [2007], extended to a first-order language with a powerful tractable fragment. It directly generalizes the tractable fragment considered by Juba [2019]. Indeed, it can also serve as a logic of probability, of the kind that Halpern [1990] calls a "type 2" first-order logic of probability – meaning, the probabilities are taken over valuations of the formulas, as opposed to bindings from the domain of quantification. (See Appendix C for an example)

Alternatively, our knowledge is expressed as a combination of logical constraints and bounds on the marginals of certain expressions; as shown by Kuželka *et al.* [2018], if we consider a maximum-entropy objective, marginal equality constraints alone capture Markov Logic Networks [Richardson and Domingos, 2006], so we may obtain a (strict) generalization of such models. But rather than assuming interest in a maximum-entropy model, we will be primarily interested in what values may possibly be taken by the expected value of other expressions.

Markov Logic Indeed, Markov Logic is a prominent example of the vast family of different models that have been considered as targets for probabilistic modeling and inference with relational representations; Van den Broeck et al. [2021] give a broad overview of this area. Typically, inference in these models is reduced to Weighted (First-Order) Model Counting (WFOMC) [Chavira and Darwiche, 2008; Van den Broeck *et al.*, 2011; Van den Broeck *et al.*, 2014], in which possible worlds are assigned a (pseudo-) likelihood given by the product of the (explicitly defined) weights associated with the specified formulas satisfied in each such possible world. Indeed, the likelihoods assigned to possible worlds by Markov Logic (for example again) are usually defined in such a way. Again, in contrast to these works, our focus is not specifically on inference in models of such form.

Moreover, we consider expectation more generally than probability per se – i.e., expressions that may take numeric values. In the literature, such models are referred to as "hybrid" probabilistic models, and reduced to Weighted (First-Order) Model Integration [Belle *et al.*, 2015; Feldstein and Belle, 2021]. This problem is generally more challenging and is only tractable in limited cases [Zeng *et al.*, 2020; Feldstein and Belle, 2021]. Indeed, even WFOMC, which only concerns probability, is only known to be tractable in various two-variable fragments [Van den Broeck *et al.*, 2011; Van den Broeck *et al.*, 2014; Beame *et al.*, 2015; Kazemi *et al.*, 2016; Kazemi *et al.*, 2017; Kuusisto and Lutz, 2018; Kuzelka, 2021; Van Bremen and Kuželka, 2023; Tóth and Kuželka, 2023], and is believed to be intractable for some three-variable fragments [Beame *et al.*, 2015].

Although our tractable fragment similarly requires some

kind of bound on the quantifier rank, each finite bound corresponds to a polynomial-time fragment, and there is no such hard barrier against the many-variable fragment.

As mentioned previously, Belle's work [2017] extends these WFOMC-based approaches to infinite and open universe settings; but, being based on WFOMC, it also inherits the above barriers to tractability.

Other approaches In addition, a number of other languages have been proposed for reasoning about probability in open universes, but without promises of tractability or generalization to reasoning about expectation [Poole, 2003; Milch *et al.*, 2005; Carbonetto *et al.*, 2005; Poole, 2008].

3 Background and Notation

By reasoning, we refer to the ability to resolve a query based on a knowledge base, both of which are encoded using the logic fragment we have defined. In this section, we first introduce the first-order probabilistic relational logic that underpins our inference method. We then discuss how sum-of-squares relaxation techniques can be applied to facilitate efficient reasoning at the propositional level, leaving the discussion of lifted reasoning for the following section.

3.1 (First-Order) Probability Relational Logic

Language: We start by defining a first-order language $\mathscr{L}$ that includes relational symbols $\{P(x), Q(x, y), R(x, y, z), ..., P'(x), ...\}$ of every arity, variables $\mathscr{V} = \{x, y, z, ...\}$, and a countably infinite set of names $\mathscr{N}$ serving as the domain of quantification. These names can be thought of as the integers in $\mathbb{N}$, but we use proper names $\{john, james, jane, ...\}$ for readability. We will also, in general, consider a finite set of constants $\mathscr{C} \subseteq \mathscr{N}$. and refer to $\mathscr{G} = \mathscr{N} \setminus \mathscr{C}$ as generic names. A renaming substitution is given by a permutation on $\mathscr{G}$, that is extended to $\mathscr{N}$ by the identity map on $\mathscr{C}$. The logical terms of the logic $\mathscr{T}$ are given by a relation symbol together with a tuple of the corresponding arity from $\mathscr{V} \cup \mathscr{C}$. A *ground* term is given by a relation symbol together with a tuple of the corresponding arity from $\mathscr{N}$.

In general, we consider expressions of the form of polynomial inequalities, where the indeterminates are given by terms: a monomial μ is given by a finite subset of $\mathscr{T}$, $\tau_1 \ldots, \tau_k$ with corresponding positive integer exponents $d_1 \ldots, d_k$: $\mu = \tau_1^{d_1} \cdots \tau_k^{d_k}$. We refer to $\deg(\mu) = \sum_{i=1}^{k} d_i$ as the degree of the monomial. For each monomial μ, we have a moment term $e(\mu)$.

A polynomial inequality is now given by a finite set of monomials M, together with a real-valued coefficient c_μ for each monomial $\mu \in M$, and a numeric relation symbol from $\{\geq, =\}$. We will write these expressions as $\sum_{\mu \in M} c_\mu \mu \geq 0$ and $\sum_{\mu \in M} c_\mu \mu = 0$, respectively. We now refer to $\max_{\mu \in M} \deg(\mu)$ as the degree of the polynomial inequality expression. Now, our polynomial inequalities will in general be bound by universal quantifiers, and analogous to the languages introduced by Lakemeyer and Levesque [2002], we will allow the domain of quantification to be specified by the following kind of equality expressions: these have atomic formulas consisting of a pair from $\mathscr{V} \cup \mathscr{C}$ (variables or constants),

which we write e.g. as $x = y$, $jane = x$, etc., that are given by any Boolean expression on these atomic formulas (using the usual De Morgan connectives, $\wedge, \vee, \neg$).

Logical Constraints Now, for each pair of an equality expression Ξ and polynomial inequality Λ, we have a logical constraint formula $\Phi = \forall \Xi \supset \Lambda$. When Ξ is a trivial tautology, we simply write $\forall \Lambda$. We refer to the number of distinct variables occurring in Ξ and Λ together as the quantifier rank of Φ. The semantics of Φ is that for all substitutions of names for variables θ satisfying Ξ, the same substitution into the polynomial inequality $\Lambda\theta$ is also satisfied.

Example 1. *Although in general the relation symbols are interpreted as numeric indeterminates, we can assert that relations P and Q take Boolean values by the formulas $\forall P(x)^2 - P(x) = 0$ and $\forall Q(x, y)^2 - Q(x, y) = 0$. We can moreover define a negative literal for each relation symbol by considering another relation $\bar{P}$ and relating its value to P like so: $\forall P(x) + \bar{P}(x) - 1 = 0$.*

There are two ways of encoding clauses. The first approach uses a monomial of Boolean terms to represent a conjunction, and uses an equality formula to assert that the conjunction is false; by De Morgan's law, this is a clause. Concretely, the clause $\forall x, y\, P(x) \vee \neg Q(x, y)$ is represented by $\forall \bar{P}(x) Q(x, y) = 0$. This method produces monomials with degree equal to the width of the clause. The second approach uses a linear inequality, asserting at least one term is true: we represent $\forall x, y\, P(x) \vee \neg Q(x, y)$ by $\forall P(x) + \bar{Q}(x, y) - 1 \geq 0$. These expressions have degree one. The two different encodings can be used to simulate two different tractable logics.

Finally, the equality formulas allow us to write expressions such as $\forall x \neq y \wedge x \neq jane \supset P(x) + \bar{Q}(x, y) - 1 \geq 0$, which is the aforementioned clause with additional restrictions on the possible bindings. As with Belle [2017], we can thereby represent proper+ knowledge bases [Lakemeyer and Levesque, 2002], which consist of universal clauses with such equality expressions defining the domain of quantification.

Expectation Constraints Next, we use these relational symbols, and their associated variables, as the random variables in our expectation logic in the sense of Halpern and Pucella [2007]. The term $e(P(james, jane))$ represents the expected value of the relation P on the names of $james$ and $jane$. We now consider expectation bound expressions that are defined by a finite set of moment terms M, together with real-valued coefficients $c_{e(\mu)}$ for each moment $e(\mu) \in M$, corresponding to the linear inequality $\sum_{e(\mu) \in M} c_{e(\mu)} e(\mu) \geq 0$. The degree of the expectation bound is similarly defined to be $\max_{e(\mu) \in M} \deg(\mu)$, where we also refer to the degree of the monomial μ as the degree of the moment $e(\mu)$. Similarly to the logical constraint formulas, for each pair of an equality expression Ξ and expectation bound B, we have an expectation constraint formula $\Psi = \forall \Xi \supset B$. The quantifier rank of an expectation constraint Ψ is likewise equal to the number of distinct variables occurring in Ξ and B together. Again, the semantics is that for any substitution of names for variables θ satisfying Ξ, the substitution into the expectation bound $B\theta$ holds.

Knowledge Base and Query: In general, we define a knowledge base Δ as a finite non-empty set of logical and expectation constraint formulas about various relations. A query q is another constraint formula pending verification.

Thus, assuming the knowledge base is satisfiable, checking whether the query is compatible with it is equivalent to checking the consistency of the expanded knowledge base $\Delta \cup q$. From now on, we will refer to $\Delta \cup q$ simply as the knowledge base for brevity.

Semantics: A model $\mathscr{M}$ for the knowledge base Δ is given by a probability measure space, together with an assignment of a measurable function for each binding of a relational variable to a tuple of names of the appropriate arity, such that the logical constraints in Δ are satisfied with probability 1, and the expectation bounds are also satisfied for all bindings.

Now we are ready to resolve queries using knowledge bases in the described form. For instance, as the expectation of a Boolean variable coincides with the probability of the variable taking the value one, we can directly reason about various conditional probabilities, as illustrated in the following example.

Example 2. *In the following example, we would like to speculate the likelihood of a war between Antony and Octavian. We will consider a binary Boolean relation **War** and an ternary Boolean relation **LoveTriangle**, with constants Antony and Cleopatra. We can encode an assertion like "$\Pr[War(x, y)|LoveTriangle(x, y, z)] \geq .75$" by using the definition of conditional probability to rewrite it as "$\Pr[War(x, y) \wedge LoveTriangle(x, y, z)] \geq .75 \Pr[LoveTriangle(x, y, z)]$." Thus we obtain a knowledge base*

$$\forall LoveTriangle(x, y, z)^2 - LoveTriangle(x, y, z) = 0$$
$$\forall War(x, y)^2 - War(x, y) = 0$$
$$\forall e(War(x, y) LoveTriangle(x, y, z))$$
$$- .75 e(LoveTriangle(x, y, z)) \geq 0.$$

Given also:

$$\forall x \neq Antony \wedge x \neq Cleopatra$$
$$\supset e(LoveTriangle(x, Antony, Cleopatra) \geq 1$$

this would imply that Antony has a greater than 75% chance of being at war with everyone in the universe (apart from himself and Cleopatra), including Octavian.

3.2 Propositional Sum-of-squares Refutations

At the propositional level, a knowledge base consists solely of a collection of *ground* logical constraints and *ground* expectation bounds. A sum-of-squares refutation demonstrates that it is impossible for a joint distribution on random variables to be consistent with the knowledge base, reducing reasoning to a moment problem.

A convenient form for our purposes is due to Putinar [1993]: suppose our knowledge base consists of logical inequalities $\{g_i \geq 0\}_{i \in I}$, equalities $\{h_j = 0\}_{j \in J}$, and expectation bounds $\{b_k \geq 0\}_{k \in K}$, where the various g_i, h_j, and b_k are written as polynomials in the relations. A sum-of-squares polynomial σ is, as the name suggests, equal to a sum of squares of arbitrary polynomials $\sum_{\ell \in L}(p_\ell)^2$. Now, for sum-of-squares polynomials σ_0 and σ_i for $i \in I$, arbitrary

polynomials q_j for $j \in J$, and nonnegative real numbers r_k for $k \in K$, suppose we have an expression

$$\sigma_0 + \sum_{i \in I} \sigma_i g_i + \sum_{j \in J} q_j h_j + \sum_{k \in K} r_k b_k = -1. \qquad (1)$$

This is called a sum-of-squares refutation. It is sound, since for any joint distribution on the relational variables the sum-of-squares polynomials must be nonnegative; assuming that the logical constraints hold, and hence that each g_i is nonnegative and each h_j is identically 0 over the support of the distribution, we see that by linearity of the expectation operator, the sum-of-squares expression must be nonnegative. Since, it is formally equal to -1, which certainly has negative expectation.

In order to establish a sum-of-squares refutation, we need to find σ_0 and σ_i for $i \in I$, arbitrary polynomials q_j for $j \in J$, and nonnegative real numbers r_k for $k \in K$ satisfying the above equality 1. A manually established sum-of-squares refutation is provided below.

Example 3. *We demonstrate a proof of Chebyshev's inequality in sum-of-squares. This is a standard tool in probability with a simple, low-degree proof. We suppose X is a mean-0 random variable with variance $\lambda > 0$. We define a tail event T to be a Boolean (indicator) random variable corresponding to $|X|$ exceeding $\sqrt{k\lambda}$ – or equivalently, $X^2 \geq k\lambda$. For any given $\delta > 0$, we'll prove that the probability of T is at most $\frac{1}{k} + \delta$. Formally, now, we have the system:*

$$T(X^2 - k\lambda) \geq 0$$
$$(1 - T)(k\lambda - X^2) \geq 0$$
$$T^2 - T = 0$$
$$e(X^2) - \lambda = 0$$
$$e(X) = 0$$
$$e(T) - (\frac{1}{k} + \delta) \geq 0$$

which we wish to refute. The sum-of-squares expression

$$((1 - T)X)^2 + T(X^2 - k\lambda) - X^2(T^2 - T)$$
$$- (X^2 - \lambda) + k\lambda \left(T - \left(\frac{1}{k} + \delta \right) \right)$$

is formally equal to $-k\lambda\delta$, so by rescaling by $\frac{1}{k\lambda\delta}$, we obtain a sum-of-squares refutation. Observe that the total degree of the expression is 4.

3.3 Tractability of Constant-Degree Sum-of-Squares (SOS)

Deriving a sum-of-squares refutation is a non-trivial task, as demonstrated by Example 3. Luckily, as noted independently by many authors [Shor, 1987; Nesterov, 2000; Parrilo, 2000; Lasserre, 2001], for any fixed $d \in \mathbb{N}$, the fragment of ground sum-of-squares consisting of expressions of degree at most d is tractable: suppose $\vec{R}$ is the vector of all ground monomials of relational random variables up to degree $d/2$ (assume d is even), ordered by increasing total degree, including the degree-0 constant 1. We refer to the expectation of the outer-product $\vec{R}\vec{R}^\top$ as a moment matrix $[M]$. For a polynomial

p, we define the degree-d localizing matrix $[pM]$ to be, for the largest index s^* such that the degree of R_{s^*} is at most $d/2 - \deg(p)$, the expectation of $p \cdot \vec{R}_{1:s^*} \vec{R}_{1:s^*}^\top$. Observe that this matrix is constructed so that all expressions appearing in it have total degree at most d. Finally, the degree-d sum-of-squares semidefinite program corresponding to logical inequalities $\{g_i \geq 0\}_{i \in I}$, equalities $\{h_j = 0\}_{j \in J}$, and expectation bounds $\{b_k \geq 0\}_{k \in K}$ is

$$[M] \succeq 0$$
$$[g_i M] \succeq 0 \quad i \in I$$
$$[h_j M] = 0 \quad j \in J$$
$$b_k \geq 0 \quad k \in K$$

Assuming that the logical inequalities assert that each term only obtains bounded values – specifically, for all ground terms τ, $\tau^2 \leq U$ for some common U – this program is infeasible iff there is a degree-d sum-of-squares refutation. (Such systems are said to be explicitly compact.) Indeed, degree-d refutations are described by the dual semidefinite program. Hence, we can test for the existence of refutations with coefficients up to a given size using a polynomial-time algorithm for deciding the feasibility of this semidefinite program:

Theorem 1. (Soundness [Shor, 1987; Nesterov, 2000; Parrilo, 2000; Lasserre, 2001]). *Let $\{g_i \geq 0\}_{i \in I}, \{h_j = 0\}_{j \in J}, \{b_k \geq 0\}_{k \in K}$ be a system of constraints that is explicitly compact. Then either there is a degree-d sum-of-squares refutation or there is a solution to the degree-d sum-of-squares semidefinite program.*

Lasserre [2001] showed that ground sum-of-squares is complete for explicitly compact knowledge bases: for every inconsistent system, there is a sum-of-squares refutation that, in particular, has some finite degree.

Theorem 2. (Completeness, a corollary of [Putinar, 1993]). *There exists a probability distribution with expected values $\{e(x^{\vec{\alpha}})\}_{\vec{\alpha} \in \mathbb{N}^n}$ supported on a set given by an explicitly compact system $\{g_i \geq 0\}_{i \in I}, \{h_j = 0\}_{j \in J}, \{b_k \geq 0\}_{k \in K}$ iff every moment matrix is positive semidefinite, every localizing matrix for each g_i is positive semidefinite, every localizing matrix for each h_j is zero, and the bounds b_k are nonnegative (i.e., solutions exist for all degrees d).*

Naturally, there is no reasonable bound on the degree, and even for systems in which all of the relational variables are Boolean, the degree may need to be linear in the number of such variables; observe that in general, the number of monomials of a given degree is exponential in the degree. Hence, this corresponds to an exponential-size semidefinite program formulation.

But, the constant-degree fragments are already quite expressive. Many theorems of probabilistic analysis are captured by these fragments [Barak *et al.*, 2012; O'Donnell and Zhou, 2013; De *et al.*, 2013; Kauers *et al.*, 2014]. More generally, Berkholz [2018] showed that constant-degree fragments of ground sum-of-squares simulate the constant degree fragments of polynomial calculus [Clegg *et al.*, 1996], and Juba [2019] showed that they also simulate bounded-space treelike resolution [Esteban and Torán, 2001], which naturally generalizes unit propagation and other tractable fragments of resolution [Ansótegui *et al.*, 2008].

We remark briefly that extensions of Theorem 2 to countably-infinite dimension have been obtained; see [Ghasemi *et al.*, 2016] and [Curto *et al.*, 2023]. While conceptually pleasing, as with the finite-dimensional (ground) case, these cannot provide practically usable bounds on the size of the refutation. For the sake of efficient algorithms, we will have to focus on a limited fragment.

4 Efficient Reasoning in Open Universes

From this point onward, we enter the first-order world with an *open-universe* (OU) setting, where we do not assume a complete knowledge of all the objects in the universe. Thus, the ability to reason in an open universe demonstrates robustness and adaptability in the face of uncertainty. This assumption determines how we need to ground our knowledge base.

4.1 Grounding in Probability Logic

A ground theory is obtained from Δ by substituting variables with names. Let the rank of Δ be the maximum quantifier rank of any formula in Δ. Given a knowledge base of this form, we are going to define a grounding of Δ as a substitution of all variables in Δ with names.

Definition 1.

$$GND(\Delta) = \{\phi\theta | [\forall \Xi \supset \phi] \in \Delta, \vDash \Xi\theta\}$$

Typically, to limit the domain of variables used in a grounding we will use:

$$GND(\Delta, k) = \{\phi\theta | [\forall \Xi \supset \phi] \in \Delta, \vDash \Xi\theta, \theta \in K\}$$

where K consists of all constants present in Δ plus k additional generic names, for $k \geq 0$.

The generic names capture the values the formula may take when its variables are bound to elements outside the set of constants. As opposed to the *domain closure assumption*, where we do not consider any objects outside of the knowledge base (i.e., $DC(\Delta) = GND(\Delta, 0)$), we will be capturing an open universe via $OU(\Delta) = GND(\Delta, k)$, where k is exactly the rank of Δ. Intuitively, we need at least k distinct names since Ξ can require that all variables are bound to distinct elements, but moreover we will see (cf. Theorem 4) that k generic names are also sufficient to capture the behavior of all possible groundings.

Example 4. *Continuing our previous example, we can get a clearer picture of Antony's ongoing wars. For the knowledge base Δ as defined in Example 2, we get GND(Δ, 3) with 3 generic names $Octavian, Caesar, Cicero$ (ignoring the Boolean axioms and equality constraints):*

$$e(War(Oct., Ant.)L.T.(Oct., Ant., Cleo.))$$
$$-.75e(L.T.(Oct., Ant., Cleo.)) \geq 0$$
$$e(War(Cae., Ant.)L.T.(Cae., Ant., Cleo.))$$
$$-.75e(L.T.(Cae., Ant., Cleo.)) \geq 0$$
$$e(War(Cic., Ant.)L.T.(Cic., Ant., Cleo.))$$
$$-.75e(L.T.(Cic., Ant., Cleo.)) \geq 0$$
$$e(L.T.(Oct., Ant., Cleo.)) - 1 \geq 0$$
$$e(L.T.(Cae., Ant., Cleo.)) - 1 \geq 0$$
$$e(L.T.(Cic., Ant., Cleo.)) - 1 \geq 0$$

Observe, the more generic names we add, the more wars Antony likely ends up in. Imagine how bad his issues would be in GND(Δ) (The full grounding including the Boolean axioms is included in the Appendix A for completeness).

4.2 Satisfiability

In this section we cast the usual notion of logical satisfiability into the algebraic language of sum-of-squares feasibility. Concretely, we say that a knowledge base Δ is *satisfiable* if and only if its associated sum-of-squares semidefinite program admits a feasible solution, which we call a pseudo-model:

Definition 2. *A degree-d pseudomodel for Δ is given by an assignment of a real number to each $e(\mu)$ for each monomial μ up to degree d, such that the assignments satisfy the infinite sum-of-squares program.*

For a finite set of names, used as the domain of quantification for GND(Δ), the corresponding program is a (finite) ground sum-of-squares program Theorem 2 asserts that for explicitly compact Δ and sufficiently high degree, satisfiability w.r.t. models and pseudomodels coincide. This equivalence lets us reduce logical satisfiability questions to the feasibility of polynomial-size semidefinite programs, paving the way for tractable relational inference in open universes.

4.3 Equivalence Classes

Definition 3. *Two ground monomials μ and μ' are said to be in the same renaming equivalence class if there is a renaming substitution θ such that $\mu\theta = \mu'$.*

Example 5. *For Δ containing:*

$$\forall e(Q(x, y)) - 3 \geq 0$$
$$\forall e(Q(x, james)) \geq 0$$
$$e(P(james)) - 1 \geq 0$$

and the subsequent grounding GND(Δ,2):

$$e(Q(jack, jill)) - 3 \geq 0$$
$$e(Q(jill, jack)) - 3 \geq 0$$
$$e(Q(james, jack)) - 3 \geq 0$$
$$e(Q(james, jill)) - 3 \geq 0$$
$$e(Q(jack, james)) \geq 0$$
$$e(Q(jill, james)) \geq 0$$
$$e(P(james)) - 1 \geq 0$$

We get that $Q(jack, jill)$ and $Q(jill, jack)$ are in the same equivalence class, but $Q(james, jack)$ is not because james is not a generic name, and thus can't be freely renamed.

Each name will fall into a single equivalence class, and it will be used when lifting our grounded logic.

4.4 Lifted Sum-of-Squares

We now present our lifted sum-of-squares system. Let Δ be a knowledge base consisting of logical constraint formulas and expectation bounds, and let k be the quantifier rank of Δ. We obtain our *lifted* sum-of-squares system by adding equality constraints for the generic names to the knowledge base.

Definition 4 (Lifted Sum-of-Squares). *For any given degree bound d,* degree-d lifted sum-of-squares *for a first-order knowledge base Δ uses the language of propositional degree-d sum-of-squares with the following propositional knowledge base:* $\mathrm{GND}(\Delta, k)$ *union with the set of equality constraints $e(\mu) - e(\mu') = 0$ for each pair of ground monomials μ, μ' in the names used by $\mathrm{GND}(\Delta, k)$ of degree up to d such that for a renaming substitution θ, $\mu = \mu'\theta$. We denote this propositional sum-of-squares knowledge base by* $\mathrm{GND}_{\mathrm{liftedSOS}}(\Delta, k)$. *Thus, a degree-$d$ lifted sum-of-squares refutation of Δ means a propositional degree-d sum-of-squares refutation of* $\mathrm{GND}_{\mathrm{liftedSOS}}(\Delta, k)$.

Observe that since degree-d refutations of Δ are simply degree-d ground sum-of-squares refutations, it follows immediately from Theorem 1 that refutations exist iff the sum-of-squares semidefinite program is infeasible. This program naïvely has dimension equal to the number of monomials that can be constructed from relational variables bound to names from $\mathscr{C}$ or our k generic names, which is polynomial in $|\mathscr{C}|$ and the number of relation symbols as long as both k and the arity of all relations is bounded by a constant. But, instead of including the equality constraints $e(\mu) - e(\mu') = 0$ for each pair of equivalent ground monomials μ, μ', we can use a single variable to represent the value for the entire equivalence class. Then the number of constraints is similarly bounded, yielding a polynomial time guarantee for our lifted inference:

Theorem 3. *Given a Δ with c constants, n $\forall$-clauses, each mentioning at most m predicates, and with rank k, the running time for a lifted degree-d sum-of-squares inference in a open universe is polynomial in $c, n,$ and m for fixed k and d.*

Proof. For each $\forall$-clause we will have at most $(c + k)^k$ substitutions for each predicate. Then the total number of atoms in $\mathrm{GND}(\Delta, k)$ is $O(nm(c + k)^k)$.

As there are algorithms solving a semidefinite program up to arbitrary precision in polynomial time with respect to the size of the program (e.g. the ellipsoid method [Bland *et al.*, 1981]), we thus can decide whether or not a degree-d refutation exists in polynomial time in $c, n,$ and m for fixed k and d. $\qquad\square$

For example, we can carry out the following probabilistic inference in polynomial time.

Example 6. *We're going to assume a grounding that includes the 3 generic names Oct., Cic., and Cae., (As partially shown in Example 4) and infer a bound on the value of e(War(Ant.,Oct.)). Given our boolean constraints on both relations War and L.T.. We can then infer that the expression $e(War(Ant., Oct.)) \geq .75$ via the sum-of-squares expression (denoting e(War(Ant.,Oct.)) by W and e(LoveTriangle(Ant.,Oct.,Cleo.)) by LT),*

$$W^2(1 - LT)^2 + W^2(LT - LT^2) + (1 - LT)(W - W^2)$$
$$+ (W \cdot LT - .75LT) + (.75LT - .75) = W - .75$$

(See Appendix B for details.)

4.5   Soundness and Completeness

We will now show that our lifted sum-of-squares logic is both sound and complete for a $\mathrm{GND}(\Delta)$ that uses an open-universe. To begin, recall the results on ground sum-of-squares, Theorems 1 and 2. We will leverage these two theorems to show that our lifted sum-of-squares logic is sound and complete when finding a satisfying model of the equivalence classes, assuming explicit compactness.

We first argue that pseudomodels can be taken to agree within an equivalence class without loss of generality.

Lemma 1. *If $GND(\Delta, k)$ is satisfiable then there exists a pseudomodel $\mathcal{M}$ in which all members of each equivalence class take the same value.*

Proof. Assume $\mathrm{GND}(\Delta, k)$ is satisfiable. Then for its associated semidefinite program there is some feasible "generic" assignment $e_{\mathscr{G}}$ of moment-variables. Now, consider any renaming substitution θ that takes the k generic names in $\mathrm{GND}(\Delta, k)$ to the same set of k generic names. For any constraint $\forall \Xi \supset \phi$ in Δ, for any grounding θ' such that $\vDash \Xi\theta'$, observe that $\vDash \Xi\theta'\theta$ as well: indeed, for any atom of Ξ, if θ' grounds a variable to a constant, θ does not rename the constant, so equality w.r.t. any constant is unchanged; and θ' grounds both variables to the same name iff θ maps both to the same name, so equality between generic names is also preserved. Therefore, $\phi\theta'\theta$ is also in $\mathrm{GND}(\Delta, k)$. Now consider the assignment $\mathcal{M}$ that assigns each $e_{\mathcal{M}}(\mu)$ the value $\frac{1}{k!} \sum_{\text{renaming } \theta} e_{\mathscr{G}}(\mu\theta)$. This $e_{\mathcal{M}}(\mu)$ is then going to be our proposed pseudomodel. Observe that for any renaming θ'',

$$e_{\mathcal{M}}(\mu\theta'') = \frac{1}{k!}\sum_{\theta} e_{\mathscr{G}}(\mu\theta''\theta) = \frac{1}{k!}\sum_{\theta} e_{\mathscr{G}}(\mu\theta) = e_{\mathcal{M}}(\mu)$$

so indeed, the equivalence classes share a common value. Likewise, since $\varphi\theta''$ is in $\mathrm{GND}(\Delta, k)$, the assignments $e_{\mathscr{G}\theta''}(\mu) = e_{\mathscr{G}}(\mu\theta'')$ are also solutions. Moreover, since $e_{\mathcal{M}}$ is a convex combination of these $e_{\mathscr{G}\theta''}$, and the feasible region of any semidefinite program is convex, $e_{\mathcal{M}}$ must also be feasible. Therefore, $e_{\mathcal{M}}$ is indeed a pseudomodel. $\qquad\square$

Next, we observe that a set of generic names equal to the quantifier rank suffices to obtain representatives of all of the equivalence classes of monomials that appear in $\mathrm{GND}(\Delta)$; we will use this property to extend to a pseudomodel for the entire $\mathrm{GND}(\Delta)$ next.

Lemma 2. *For a given Δ a pseudomodel for $GND(\Delta, k)$, where k is the quantifier rank of Δ, assigns a value for some grounding from every equivalence class occurring in $GND(\Delta)$.*

Proof. Consider any grounding θ such that for $\forall \Xi \supset \phi$ in Δ, $\phi\theta$ is in $\mathrm{GND}(\Delta)$. Since $\forall \Xi \supset \phi$ has quantifier rank k, at most k names from $\mathscr{G}$ are assigned by θ in $\phi\theta$. So, consider the renaming substitution θ'' that takes these k names to the k names used in $\mathrm{GND}(\Delta, k)$ (arbitrarily). Observe that $\phi\theta\theta''$ is in $\mathrm{GND}(\Delta, k)$, so the pseudomodel for $\mathrm{GND}(\Delta, k)$ assigns a value to each monomial in $\phi\theta\theta''$. For each monomial occurring in $\phi\theta$, there is an equivalent monomial in $\phi\theta\theta''$ that has therefore indeed been assigned a value in the pseudomodel. $\qquad\square$

Theorem 4. *For all d, GND(△) has a degree-d pseudomodel iff GND(△, k) has a degree-d pseudomodel, where k is equal to the rank of △.*

Proof. The direction of most interest is showing that given GND(△, k) has a pseudomodel then GND(△) also has a pseudomodel. Starting with a pseudomodel $\mathscr{M}$ that satisfies GND(△, k) we can apply Lemma 1 to get $\mathscr{M}'$ which will be our pseudomodel where all members of an equivalence class take the same value. By Lemma 2 we have that each monomial in GND(△) belongs to an equivalence class with a representative appearing in GND(△,k). This guarantees a pseudomodel exists for GND(△), because for every $\phi\theta$ in GND(△), we assign the same values to its monomials as in $\phi\theta\theta''$ where θ'' is the renaming substitution that assigns the k names in $\phi\theta$ to the names in GND(△, k). Since the constraints of the semidefinite program associated with $\phi\theta\theta''$ are satisfied by these values, so are the constraints associated with $\phi\theta$. $\qquad\square$

Theorem 5. *Given an explicitly compact knowledge base △ and a grounding GND(△, k), a lifted sum-of-squares program using GND(△, k) is sound and complete for degree-d refutations of △.*

Proof. For soundness, we first observe that if △ is satisfiable, then the degree-d lifted sum-of-squares program for GND(△, k) is feasible by Lemma 1, and hence by Theorem 1 there is no degree-d lifted sum-of-squares refutation of △.

For completeness with respect to degree-d, suppose that there is no degree-d pseudomodel of GND(△). Then by Theorem 4, there is no degree-d pseudomodel of GND(△, k), either. By Theorem 1, there is a degree-d sum-of-squares refutation of GND(△, k); by renaming, this is a degree-d lifted sum-of-squares refutation of △. $\qquad\square$

4.6 Discussion of Tractability

We stress that some probabilistic inference problems are inherently intractable and we do not claim to solve these problems in polynomial time. We know that even approximate inference in Bayesian belief networks is NP-hard [Dagum and Luby, 1993]; but this was due to the fact that any language capable of expressing all 3-CNFs can encode NP-hard problems. Thus, restricting the language seems to be an unpromising approach to obtaining tractable inference.

Indeed, recall that the same issue arises in even in classical Boolean inference. There, the approach broadly pursued in "SAT solvers" in practice does not attempt to restrict the language, but rather the algorithm is only feasible to run on a subset of the instances. Following [Beame *et al.*, 2004], we know that the subset of tractable instances is captured by a fragment of resolution in which the proof size is small.

Here, we pursue an analogous approach to probabilistic inference. We have opted for a language that can also represent all 3-CNFs and so in general (in the absence of a degree bound) we encounter an intractable inference problem. As with Boolean reasoning, we can guarantee that our algorithm is tractable for clearly defined fragments (of sum-of-squares), but we do not know in advance what moment degrees will

suffice to solve the problem. Unfortunately, even for Boolean inference, SAT solvers face the analogous difficulty.

5 Comparison to Other Models

We will use three existing models for probabilistic inference and show where lifted sum-of-squares logic improves in either tractability, expressiveness, or both. A short summary of these comparisons can be seen in Table 1.

Model	Ternary Relations	Full Arithmetic
Lifted SOS	✓	✓
WFOMC-OU	✗	✗
PSL	✓	✗
TML	✓	✓

Table 1: Comparisons of tractabilities among existing statistical relational learning methods

5.1 Tractable Markov Logic (TML)

Tractable Markov Logic (TML), as defined in [Domingos and Webb, 2012] has two very distinct characteristics. First of all, they require the objects all be inside a single hierarchy, which is highly restrictive. Secondly, they adopt weights together with those sub-class structures to maintain tractability. The explicitly strong structural assumption is not obvious for many cases including our example. Along with these drawbacks there is a fundamental difference between TML and Lifted Sum-of-Squares. Markov Logics in general focus only on bounding the maximum entropy distribution, whereas we focus on bounding all distributions. This can be particularly useful when using these models to reject a null hypothesis. By having certifiable bounds, our model can reject a hypothesis by itself without further experimentation being necessary. An example of this would be the biological benchmarks used in [Ribeiro *et al.*, 2022].

5.2 Weighted First-Order Model Counting with an Open Universe (WFOMC-OU)

As mentioned above, Belle's approach of performing weighted first-order model counting in an open universe [Belle, 2017] is believed to be intractable for many ternary relations and beyond. This means that Example 6 is likely to be intractable using this approach. However, our runtime guarantees are polynomial in the rank of our knowledge base meaning this example is well within our tractable fragment. Along with this, many model counting algorithms rely on approximation to increase their tractability. While this may be useful for some domains it runs into a similar issue as TML. If an exact answer is necessary these models may be unable to provide an answer.

5.3 Probabilisitc Soft Logic (PSL)

This model, introduced under the name Probabilistic Soft Logic, is another commonly applied method to solving probabilistic reasoning in relational domains [Kimmig *et al.*, 2012]. Much similar to Markov Logic Networks, Probabilisitc Soft Logic combines graphical models and first-order logic, but allows the truth values to be soft as in any time between

zero and one. These methods have had success in practice but currently lack the theoretical guarantees provided in our work. In addition to this, PSL is limited to arithmetic rules of linear combinations of relations. This is in contrast to our Example 2, where we are able to leverage the term $War(x, y)L.T.(x, y, z)$ to create a rule that models conditional probability. PSL models are also an example of a model that utilizes weighted counting to handle the probabilitistic nature. This leads to many of the same drawbacks that were mentioned for WFOMC-OU.

6 Future Work and Conclusion

In summary our work is able to extend previous work on probabilistic inference using sum-of-squares to a first order logic. This allows for a substantially larger tractable fragment than previous work. While approach manages to employ first-order reasoning, but there may still be more simplification available for the matrices resulting from our semidefinite program. By utilizing an approach similar to that of [Kersting *et al.*, 2017] for relational linear programming, one may be able to simplify the semi-definite program even more. Further, if sparsity exists in the formulation, we may also exploit it using methods mentioned in [Lasserre, 2006]. One downside of our approach is that it inherits the limitation that sum-of-squares cannot represent or reason about independence of random variables. Although the most straightforward approach to representing independence leads immediately to an intractable polynomial optimization problem, it is natural to ask if this is truly inherent.

Acknowledgments

References

[Ansótegui *et al.*, 2008] Carlos Ansótegui, María Luisa Bonet, Jordi Levy, and Felip Manyá. Measuring the hardness of SAT instances. In *Proc. AAAI'08*, pages 222–228, 2008.

[Barak *et al.*, 2012] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

[Beame *et al.*, 2004] Paul Beame, Henry Kautz, and Ashish Sabharwal. Towards understanding and harnessing the potential of clause learning. *Journal of artificial intelligence research*, 22:319–351, 2004.

[Beame *et al.*, 2015] Paul Beame, Guy Van den Broeck, Eric Gribkoff, and Dan Suciu. Symmetric weighted first-order model counting. In *Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 313–328, 2015.

[Belle *et al.*, 2015] Vaishak Belle, Andrea Passerini, and Guy Van den Broeck. Probabilistic inference in hybrid domains by weighted model integration. In *Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2770–2776. AAAI Press/International Joint Conferences on Artificial Intelligence, 2015.

[Belle, 2017] Vaishak Belle. Open-universe weighted model counting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Berkholz, 2018] Christoph Berkholz. The relation between polynomial calculus, Sherali-Adams, and sum-of-squares proofs. In *Proc. 35th STACS*, LIPIcs, pages 11:1–11:14, 2018.

[Bland *et al.*, 1981] Robert G Bland, Donald Goldfarb, and Michael J Todd. The ellipsoid method: A survey. *Operations research*, 29(6):1039–1091, 1981.

[Carbonetto *et al.*, 2005] Peter Carbonetto, Jacek Kisyński, Nando de Freitas, and David Poole. Nonparametric bayesian logic. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 85–93, 2005.

[Chavira and Darwiche, 2008] Mark Chavira and Adnan Darwiche. On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6–7):772–799, 2008.

[Clegg *et al.*, 1996] Matthew Clegg, Jeff Edmonds, and Russell Impagliazzo. Using the Gröbner basis algorithm to find proofs of unsatisfiability. In *Proc. 28th STOC*, pages 174–183, 1996.

[Curto *et al.*, 2023] Raul E Curto, Mehdi Ghasemi, Maria Infusino, and Salma Kuhlmann. The truncated moment problem for unital commutative $\mathbb{R}$-algebras. *Journal of Operator Theory*, 90(1):223–261, 2023.

[Dagum and Luby, 1993] Paul Dagum and Michael Luby. Approximating probabilistic inference in bayesian belief networks is np-hard. *Artificial intelligence*, 60(1):141–153, 1993.

[De *et al.*, 2013] Anindya De, Elchanan Mossel, and Joe Neeman. Majority is stablest: discrete and SoS. In *Proc. 45th STOC*, pages 477–486, 2013.

[Domingos and Webb, 2012] Pedro Domingos and William Webb. A tractable first-order probabilistic logic. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 1902–1909, 2012.

[Esteban and Torán, 2001] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Inf. Comp.*, 171(1):84–97, 2001.

[Feldstein and Belle, 2021] Jonathan Feldstein and Vaishak Belle. Lifted reasoning meets weighted model integration. In *Uncertainty in Artificial Intelligence*, pages 322–332. PMLR, 2021.

[Ghasemi *et al.*, 2016] Mehdi Ghasemi, Salma Kuhlmann, and Murray Marshall. Moment problem in infinitely many variables. *Israel Journal of Mathematics*, 212:1012–1012, 2016.

[Halpern and Pucella, 2007] Joseph Y Halpern and Riccardo Pucella. Characterizing and reasoning about probabilistic and non-probabilistic expectation. *Journal of the ACM (JACM)*, 54(3):15–es, 2007.

[Halpern, 1990] Joseph Y Halpern. An analysis of first-order logics of probability. *Artificial intelligence*, 46(3):311–350, 1990.

[Hummel and Holyoak, 2003] John E Hummel and Keith J Holyoak. A symbolic-connectionist theory of relational inference and generalization. *Psychological review*, 110(2):220, 2003.

[Juba, 2019] Brendan Juba. Polynomial-time probabilistic reasoning with partial observations via implicit learning in probability logics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7866–7875, 2019.

[Kauers et al., 2014] Manuel Kauers, Ryan O'Donnell, Li-Yang Tan, and Yuan Zhou. Hypercontractive inequalities via SOS, and the Frankl-Rödl graph. In *Proc. 25th SODA*, pages 1644–1658, 2014.

[Kazemi et al., 2016] Seyed Mehran Kazemi, Angelika Kimmig, Guy Van den Broeck, and David Poole. New liftable classes for first-order probabilistic inference. *Advances in Neural Information Processing Systems*, 29, 2016.

[Kazemi et al., 2017] Seyed Mehran Kazemi, Angelika Kimmig, Guy Van den Broeck, and David Poole. Domain recursion for lifted inference with existential quantifiers. *arXiv preprint arXiv:1707.07763*, 2017.

[Kersting et al., 2017] Kristian Kersting, Martin Mladenov, and Pavel Tokmakov. Relational linear programming. *Artificial Intelligence*, 244:188–216, 2017.

[Kimmig et al., 2012] Angelika Kimmig, Stephen Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. A short introduction to probabilistic soft logic. In *NIPS Workshop on probabilistic programming: Foundations and applications*, volume 1, page 3, 2012.

[Krawczyk, 2012] Daniel C Krawczyk. The cognition and neuroscience of relational reasoning. *Brain research*, 1428:13–23, 2012.

[Kuusisto and Lutz, 2018] Antti Kuusisto and Carsten Lutz. Weighted model counting beyond two-variable logic. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 619–628, 2018.

[Kuželka et al., 2018] Ondřej Kuželka, Yuyi Wang, Jesse Davis, and Steven Schockaert. Relational marginal problems: Theory and estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Kuzelka, 2021] Ondrej Kuzelka. Weighted first-order model counting in the two-variable fragment with counting quantifiers. *Journal of Artificial Intelligence Research*, 70:1281–1307, 2021.

[Lakemeyer and Levesque, 2002] Gerhard Lakemeyer and Hector J Levesque. Evaluation-based reasoning with disjunctive information in first-order knowledge bases. In *KR*, pages 73–81, 2002.

[Lasserre, 2001] Jean Bernard Lasserre. Global optimzation with polynomials and the problem of moments. *SIAM J. Optimization*, 11(3):796–817, 2001.

[Lasserre, 2006] Jean B Lasserre. Convergent sdp-relaxations in polynomial optimization with sparsity. *SIAM Journal on optimization*, 17(3):822–843, 2006.

[Milch et al., 2005] Brian Milch, Bhaskara Marthi, Stuart Russell, David Sontag, Daniel L Ong, and Andrey Kolobov. Blog: Probabilistic models with unknown objects. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 1352–1359, 2005.

[Mladenov et al., 2017] Martin Mladenov, Leonard Kleinhans, and Kristian Kersting. Lifted inference for convex quadratic programs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Nesterov, 2000] Y. Nesterov. Squared functional systems and optimization problems. *High performance optimization*, 13:405–440, 2000.

[O'Donnell and Zhou, 2013] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proc. 24th SODA*, pages 1537–1556, 2013.

[Parrilo, 2000] Pablo A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[Poole, 2003] David Poole. First-order probabilistic inference. In *Proceedings of the 18th international joint conference on Artificial intelligence*, pages 985–991, 2003.

[Poole, 2008] David Poole. The independent choice logic and beyond. In *Probabilistic inductive logic programming: theory and applications*, pages 222–243. Springer, 2008.

[Putinar, 1993] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana U. Math. J.*, 42:969–984, 1993.

[Ribeiro et al., 2022] Tony Ribeiro, Maxime Folschette, Morgan Magnin, and Katsumi Inoue. Learning any memory-less discrete semantics for dynamical systems represented by logic programs. *Machine Learning*, pages 1–78, 2022.

[Richardson and Domingos, 2006] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62:107–136, 2006.

[Shor, 1987] N. Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics and Systems Analysis*, 23(5):695–700, 1987.

[Tóth and Kuželka, 2023] Jan Tóth and Ondřej Kuželka. Lifted inference with linear order axiom. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12295–12304, 2023.

[Van Bremen and Kuželka, 2023] Timothy Van Bremen and Ondřej Kuželka. Lifted inference with tree axioms. *Artificial Intelligence*, 324:103997, 2023.

[Van den Broeck *et al.*, 2011] Guy Van den Broeck, Nima Taghipour, Wannes Meert, Jesse Davis, and Luc De Raedt. Lifted probabilistic inference by first-order knowledge compilation. In *IJCAI*, pages 2178–2185, 2011.

[Van den Broeck *et al.*, 2014] Guy Van den Broeck, Wannes Meert, and Adnan Darwiche. Skolemization for weighted first-order model counting. In *Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2014.

[Van den Broeck *et al.*, 2021] Guy Van den Broeck, Kristian Kersting, Sriraam Natarajan, and David Poole. *An Introduction to Lifted Probabilistic Inference*. MIT Press, 2021.

[Zeng *et al.*, 2020] Zhe Zeng, Paolo Morettin, Fanqi Yan, Antonio Vergari, and Guy Van den Broeck. Probabilistic inference with algebraic constraints: Theoretical limits and practical approximations. *Advances in Neural Information Processing Systems*, 33:11564–11575, 2020.

A Full Grounded Knowledge Base

$$War(Ant., Oct.) - War(Ant., Oct.)^2 = 0$$
$$War(Ant., Cleo.) - War(Ant., Cleo.)^2 = 0$$
$$War(Ant., Cic.) - War(Ant., Cic.)^2 = 0$$
$$War(Ant., Cae.) - War(Ant., Cae.)^2 = 0$$
$$War(Cleo., Oct.) - War(Cleo., Oct.)^2 = 0$$
$$War(Oct., Cic.) - War(Oct., Cic.)^2 = 0$$
$$War(Oct., Cae.) - War(Oct., Cae.)^2 = 0$$
$$War(Cleo., Cic.) - War(Cleo., Cic.)^2 = 0$$
$$War(Cleo., Cae.) - War(Cleo., Cae.)^2 = 0$$
$$War(Cic., Cae.) - War(Cic., Cae.)^2 = 0$$
$$L.T(Ant., Oct., Cleo) - L.T(Ant., Oct., Cleo)^2 = 0$$
$$L.T(Ant., Oct., Cic.) - L.T(Ant., Oct., Cic.)^2 = 0$$
$$L.T(Ant., Oct., Cae.) - L.T(Ant., Oct., Cae.)^2 = 0$$
$$L.T(Ant., Cleo, Cic.) - L.T(Ant., Cleo., Cic.)^2 = 0$$
$$L.T(Ant., Cleo, Cae.) - L.T(Ant., Cleo., Cae.)^2 = 0$$
$$L.T(Ant., Cic, Cae.) - L.T(Ant., Cic., Cae.)^2 = 0$$
$$L.T(Oct., Cleo., Cic.) - L.T(Oct., Cleo., Cic.)^2 = 0$$
$$L.T(Oct., Cleo., Cae.) - L.T(Oct., Cleo., Cae.)^2 = 0$$
$$L.T(Cleo., Cae., Cic.) - L.T(Cleo., Cae., Cic.)^2 = 0$$
$$L.T(Oct., Cae., Cic.) - L.T(Oct., Cae., Cic.)^2 = 0$$
$$e(L.T.(Oct., Ant., Cleo.)) - 1 \geq 0$$
$$e(L.T.(Cae., Ant., Cleo.)) - 1 \geq 0$$
$$e(L.T.(Cic., Ant., Cleo.)) - 1 \geq 0$$
$$e(War(Oct., Ant.)L.T.(Oct., Ant., Cleo.))$$
$$- .75e(L.T.(Oct., Ant., Cleo.)) \geq 0$$
$$e(War(Oct., Ant.)L.T.(Oct., Ant., Cic.))$$
$$- .75e(L.T.(Oct., Ant., Cic.)) \geq 0$$
$$e(War(Oct., Ant.)L.T.(Oct., Ant., Cae.))$$
$$- .75e(L.T.(Oct., Ant., Cae.)) \geq 0$$
$$e(War(Cae., Ant.)L.T.(Cae., Ant., Cleo.))$$
$$- .75e(L.T.(Cae., Ant., Cleo.)) \geq 0$$
$$e(War(Cae., Ant.)L.T.(Cae., Ant., Cic.))$$
$$- .75e(L.T.(Cae., Ant., Cic.)) \geq 0$$
$$e(War(Cae., Ant.)L.T.(Cae., Ant., Oct.))$$
$$- .75e(L.T.(Cae., Ant., Oct.)) \geq 0$$
$$e(War(Cic., Ant.)L.T.(Cic., Ant., Cleo.))$$
$$- .75e(L.T.(Cic., Ant., Cleo.)) \geq 0$$
$$e(War(Cic., Ant.)L.T.(Cic., Ant., Oct.))$$
$$- .75e(L.T.(Cic., Ant., Oct.)) \geq 0$$
$$e(War(Cic., Ant.)L.T.(Cic., Ant., Cae.))$$
$$- .75e(L.T.(Cic., Ant., Cae.)) \geq 0$$

$$e(War(Cleo., Ant.)L.T.(Cleo., Ant., Oct.))$$
$$- .75e(L.T.(Cleo., Ant., Oct.)) \geq 0$$
$$e(War(Cleo., Ant.)L.T.(Cleo., Ant., Cic.))$$
$$- .75e(L.T.(Cleo., Ant., Cic.)) \geq 0$$
$$e(War(Cleo., Ant.)L.T.(Cleo., Ant., Cae.))$$
$$- .75e(L.T.(Cleo., Ant., Cae.)) \geq 0$$
$$e(War(Cleo., Oct.)L.T.(Cleo., Oct., Ant.))$$
$$- .75e(L.T.(Cleo., Oct., Ant.)) \geq 0$$
$$e(War(Cleo., Oct.)L.T.(Cleo., Oct., Cic.))$$
$$- .75e(L.T.(Cleo., Oct., Cic.)) \geq 0$$
$$e(War(Cleo., Oct.)L.T.(Cleo., Oct., Cae.))$$
$$- .75e(L.T.(Cleo., Oct., Cae.)) \geq 0$$
$$e(War(Cleo., Cic.)L.T.(Cleo., Cic., Ant.))$$
$$- .75e(L.T.(Cleo., Cic., Ant.)) \geq 0$$
$$e(War(Cleo., Cic.)L.T.(Cleo., Cic., Oct.))$$
$$- .75e(L.T.(Cleo., Cic., Oct.)) \geq 0$$
$$e(War(Cleo., Cic.)L.T.(Cleo., Cic., Cae.))$$
$$- .75e(L.T.(Cleo., Cic., Cae.)) \geq 0$$
$$e(War(Cleo., Cae.)L.T.(Cleo., Cae., Ant.))$$
$$- .75e(L.T.(Cleo., Cae., Ant.)) \geq 0$$
$$e(War(Cleo., Cae.)L.T.(Cleo., Cae., Oct.))$$
$$- .75e(L.T.(Cleo., Cae., Oct.)) \geq 0$$
$$e(War(Cleo., Cae.)L.T.(Cleo., Cae., Cic.))$$
$$- .75e(L.T.(Cleo., Cae., Cic.)) \geq 0$$
$$e(War(Cic., Cae.)L.T.(Cic., Cae., Ant.))$$
$$- .75e(L.T.(Cic., Cae., Ant.)) \geq 0$$
$$e(War(Cic., Cae.)L.T.(Cic., Cae., Oct.))$$
$$- .75e(L.T.(Cic., Cae., Oct.)) \geq 0$$
$$e(War(Cic., Cae.)L.T.(Cic., Cae., Cleo.))$$
$$- .75e(L.T.(Cic., Cae., Cleo.)) \geq 0$$
$$e(War(Oct., Cae.)L.T.(Oct., Cae., Ant.))$$
$$- .75e(L.T.(Oct., Cae., Ant.)) \geq 0$$
$$e(War(Oct., Cae.)L.T.(Oct., Cae., Cic.))$$
$$- .75e(L.T.(Oct., Cae., Cic.)) \geq 0$$
$$e(War(Oct., Cae.)L.T.(Oct., Cae., Cleo.))$$
$$- .75e(L.T.(Oct., Cae., Cleo.)) \geq 0$$
$$e(War(Oct., Cic.)L.T.(Oct., Cic., Ant.))$$
$$- .75e(L.T.(Oct., Cic., Ant.)) \geq 0$$
$$e(War(Oct., Cic.)L.T.(Oct., Cic., Cae.))$$
$$- .75e(L.T.(Oct., Cic., Cae.)) \geq 0$$
$$e(War(Oct., Cic.)L.T.(Oct., Cic., Cleo.))$$
$$- .75e(L.T.(Oct., Cic., Cleo.)) \geq 0$$

Permutations are omitted.

B Omitted Steps of Example 6

Starting with an expression that takes the form of the law of total probability $\forall x \neq Ant., x \neq$

$Oct.e(War(Ant., Oct.)L.T.(Ant., Oct., x))...$ We can then make substitutions using the laws of expectation and probability. These substitutions can be continued until giving a final sum-of-squares expression, which can be simplified to $e(War(Ant., Oct.)) - .75$, of:

$$W^2(1-LT)^2+(1-LT)^2(W-W^2)+W(LT-LT^2)+WLT-.75LT+.75(LT-1)$$

Where $W = e(War(Ant., Oct.))$, and $e(LT = L.T(Ant., Oct., Cleo.))$.

This is an expression of degree 4, meaning that by invoking Theorem 1 we have a degree-4 sum-of-squares program that will detect infeasibility with the constraint $e(War(Ant., Oct.) \leq .75 - \delta$ for any $\delta > 0$.

First of all, we verify that indeed our expression is equivalent to the query $War(Antony, Octavian)$:
$W^2(1-LT)^2+(1-LT)^2(W-W^2)+W^2(LT-LT^2)+W \cdot LT - .75LT + .75(LT-1)$
$= W^2((1-LT)^2 - (1-LT)^2) + W((1-LT)^2 + (LT-LT^2) + LT) - .75$
$= W - .75,$
where $W = War(Antony, Octavian)$, and $LT = LoveTriangle(Antony, Octavian, Cleopatra)$.
Since every term is indeed nonnegative by either the knowledge base or because it is a square, we can use this as our sum-of-squares equation to bound the likelihood of a war between Antony and Octavian ($War(Antony, Octavian)$).

Next, we will show how this equation was constructed. Keeping the same values for W and LT we start with the square polynomial: $W^2(1-LT)^2 = W^2(1-LT)^2$ which, recall, represents $W^2 \geq W^2LT^2$. Then using the Boolean axioms we can reduce the degree of W and LT. First LT:

$$W^2(1-LT)^2 + W^2(LT-LT^2) = W^2(1-LT)$$

and next, we will make a substitution of W in for W^2:
$W^2(1-L)^2 + W^2(LT-LT^2) + (1-LT)(W-W^2) = W(1-LT)$
Now that we have $W \geq W \cdot LT$, we can add the expression $W \cdot LT - .75LT$ to obtain $W \geq .75LT$:
$W^2(1-L)^2 + W^2(LT-LT^2) + (1-LT)(W-W^2) + (W \cdot LT - .75LT) = W - .75LT$
For the final step we will use our given bound for $LT \geq 1$ which can be re-scaled to $.75LT \geq .75$, giving us the expression $.75LT - .75$ for us to add:
$W^2(1-L)^2 + W^2(LT-LT^2) + (1-LT)(W-W^2) + (W \cdot LT - .75LT) + (.75LT - .75) = W - .75$.

C Expectation Logics Example

Given a population where 20% experience an elevated heart rate (100BPM or above) and every person has a heart rate of at least 60, what can we say about the average heart rate of that population:

$$e(HR) = e(HR * HighHR) + e(HR * (1 - HighHR))$$

Now, using the distributivity axiom, and the fact that every person with a high HR has a HR over 100, we can infer:

$$e(HR * HighHR) \geq e(100 * HighHR)$$

We can then do the same with the standard population, using the fact that a heart rate must be above 60, and infer:

$$e(HR * (1 - HighHR)) \geq e(60 * (1 - HighHR))$$

Substituting into the original equation we get:

$$e(HR) \geq e(100 * HighHR) + e(60 * (1 - HighHR))$$
$$e(HR) \geq 100e(HighHR) + 60e(1 - HighHR)$$
$$e(HR) \geq 100 * .2 + 60 * .8$$
$$e(HR) \geq 68$$

This means that in the overall population the average heart rate is at least 68.