
Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing

Zhe Zeng^{*1} Paolo Morettin^{*2} Fanqi Yan^{*3} Antonio Vergari¹ Guy Van den Broeck¹

Abstract

Weighted model integration (WMI) is an appealing framework for probabilistic inference: it allows for expressing the complex dependencies in real-world problems, where variables are both continuous and discrete, via the language of Satisfiability Modulo Theories (SMT), as well as to compute probabilistic queries with complex logical and arithmetic constraints. Yet, existing WMI solvers are not ready to scale to these problems. They either ignore the intrinsic dependency structure of the problem entirely, or they are limited to overly restrictive structures. To narrow this gap, we derive a factorized WMI computation enabling us to devise a scalable WMI solver based on message passing, called MP-WMI. Namely, MP-WMI is the first WMI solver that can (i) perform exact inference on the full class of tree-structured WMI problems, and (ii) perform inter-query amortization, e.g., to compute all marginal densities simultaneously. Experimental results show that our solver dramatically outperforms the existing WMI solvers on a large set of benchmarks.

1. Introduction

In many real-world scenarios, performing probabilistic inference requires reasoning over domains with complex logical and arithmetic constraints while dealing with variables that are heterogeneous in nature, i.e., both continuous and discrete. Consider for example the task of matching players in a game by their skills. Performing probabilistic inference for this task has been popularized by Minka et al. (2018) and is at the core of several online gaming services. A probabilistic

model for this task has to deal with continuous variables, such as the player and team performance, and reason over discrete attributes such as membership in a squad and the achieved scores. Moreover, such a model would need to take into account constraints such as the team performance being bounded by that of the players in it, and that forming a squad boosts performance. Ultimately, this translates into performing probabilistic inference in the presence of logical and arithmetic constraints and dependencies.

These hybrid scenarios are beyond the reach of probabilistic models including variational autoencoders (Kingma & Welling, 2013) and generative adversarial networks (Goodfellow et al., 2014), whose inference capabilities, despite their recent success, are limited. Classical probabilistic graphical models (Koller & Friedman, 2009), while providing more flexible inference routines, are generally incapacitated when dealing with continuous and discrete variables at once (Shenoy & West, 2011), or they tend to make simplistic (Heckerman & Geiger, 1995; Lauritzen & Wermuth, 1989) or overly strong assumptions about their parametric forms (Yang et al., 2014). Even recent efforts in modeling these hybrid scenarios while delivering tractable inference (Molina et al., 2018; Vergari et al., 2019) can not perform inference in the presence of complex constraints.

Weighted Model Integration (WMI) (Belle et al., 2015; Morettin et al., 2017) is a recent framework for probabilistic inference that offers all the aforementioned “ingredients” needed for hybrid probabilistic reasoning with logical constraints, *by design*. WMI leverages the expressive language of Satisfiability Modulo Theories (SMT) (Barrett et al., 2010) for describing problems over continuous and discrete variables. Moreover, WMI provides a principled way to perform hybrid probabilistic inference: asking for the probability of a complex query with logical and arithmetic constraints can be done by integrating weight functions over the regions that satisfy the constraints and query at hand.

Despite these appealing features, current state-of-the-art WMI solvers are far from being applicable to high-dimensional real-world scenarios. This is due to the fact that most solvers ignore the dependency structure of the problem, here expressible through the notion of a primal or factor graph of an SMT formula (Dechter & Mateescu,

^{*}Equal contribution ¹Computer Science Department University of California, Los Angeles, USA ²DISI, University of Trento, Italy, Italy ³Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China. Correspondence to: Zhe Zeng <zhezeng@cs.ucla.edu>.

2007). Thus, their practical utility is limited by their inability to scale up the WMI inference. In contrast, SMI (Zeng & Van den Broeck, 2019), is a recently proposed solver that directly exploits the problem structure encoded in primal graphs while reducing a WMI problem to an unweighted one. However, in order to perform a tractable reduction, SMI is limited to a restricted set of weights, and hence a very narrow set of WMI problems.

The contribution we make in this work is twofold. First, we theoretically trace the boundaries for the classes of tractable WMI problems known in the literature. Second, we expand these boundaries by devising a polytime algorithm for exact WMI inference on a class that is strictly larger than the class previously known to be tractable. Our proposed WMI solver, called MP-WMI, adopts a novel message-passing scheme for WMI problems. It is able to exactly compute all the variable marginal densities at once. By doing so, we are able to scale inference beyond the capabilities of all current exact WMI solvers. Moreover, we can amortize inference *inter-queries* for rich SMT queries that conform to the problem structure.

The paper is organized as follows. We start by reviewing the necessary SMT and WMI background. Then we trace the boundaries between hard and tractable WMI problem classes in Section 3. Next, we present our exact message-passing WMI solver in Section 4 together with its complexity analysis in Section 5. Before comparing our solver to the existing WMI solvers on a set of benchmarks, we discuss related work in Section 6.

2. Background

Notation. We use uppercase letters for random variables (e.g., X, B) and lowercase letters for their assignments (e.g., x, b). Bold uppercase letters denote sets of variables (e.g., \mathbf{X}, \mathbf{B}) and their lowercase denote their assignments (e.g., \mathbf{x}, \mathbf{b}). We represent logical formulas by capital Greek letters, (e.g., Λ, Φ, Δ), and literals (i.e., atomic formulas or their negation) by lowercase ones (e.g., ϕ, δ) or ℓ . We denote satisfaction of a formula Φ by one assignment \mathbf{x} by $\mathbf{x} \models \Phi$ and we denote its corresponding indicator function as $\llbracket \mathbf{x} \models \Phi \rrbracket$. For undirected graphs, neigh denotes the set of neighboring nodes; for directed ones, pa and ch denote the parent node and the set of child nodes respectively.

Satisfiability Modulo Theories (SMT). SMT (Barrett & Tinelli, 2018) generalizes the well-known SAT problem (Biere et al., 2009) to determining the satisfiability of a logical formula w.r.t. a decidable theory. Rich mixed logical/arithmetic constraints can be expressed in SMT for hybrid domains. In particular, we consider quantifier-free SMT formulas in the theory of linear arithmetic over the reals, or $\text{SMT}(\mathcal{LRA})$. Here, formulas are propositional com-

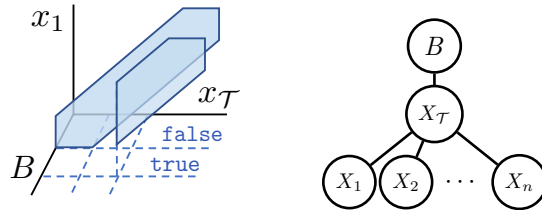


Figure 1: **Feasible region** (left) of formula Γ with one player and **primal graph** (right) of formula Γ with n players from Example 1.

binations of atomic Boolean literals and of atomic \mathcal{LRA} literals over real variables, for which satisfaction is defined in a natural way. W.l.o.g. we assume SMT formulas to be in conjunctive normal form (CNF). In the following, we will use the shorthand SMT to denote $\text{SMT}(\mathcal{LRA})$.

Example 1 (SMT representation of a skill matching system). *In a skill rating system for online games, the team performance $X_{\mathcal{T}}$ of each team \mathcal{T} is defined by the performance X_i of each player i in team \mathcal{T} , both of which are real variables. The team performance $X_{\mathcal{T}}$ is also related to a Boolean variable B indicating whether players in the team form a squad, i.e., a group of friends, which offsets (boosts) the team performance. We can build an SMT formula Γ of the relationship among these variables as follows. For brevity, we omit the domains for real variables in the formula.*

$$\Gamma := \bigwedge_{i \in \mathcal{T}} |X_{\mathcal{T}} - X_i| < 1 \bigwedge (B \Rightarrow X_{\mathcal{T}} > 2)$$

We show in Figure 1 the feasible regions of formula Γ i.e., the volumes for which the constraints are satisfied.

Weighted Model Integration (WMI). Weighted Model Integration (WMI) (Belle et al., 2015; Morettin et al., 2017) provides a framework for probabilistic inference with models defined over the logical constraints given by SMT formulas.

Definition 2. (WMI) *Let \mathbf{X} be a set of continuous random variables defined over \mathbb{R} , and \mathbf{B} a set of Boolean random variables defined over $\mathbb{B} = \{\text{true}, \text{false}\}$. Given an SMT formula Δ over \mathbf{X} and \mathbf{B} , and a weight function $w : (\mathbf{x}, \mathbf{b}) \mapsto \mathbb{R}^+$ belonging to some parametric weight function family Ω , the weighted model integration (WMI) task computes*

$$\text{WMI}(\Delta, w; \mathbf{X}, \mathbf{B}) \triangleq \sum_{\mathbf{b} \in \mathbb{B}^{|\mathbf{B}|}} \int_{(\mathbf{x}, \mathbf{b}) \models \Delta} w(\mathbf{x}, \mathbf{b}) d\mathbf{x}. \quad (1)$$

That is, summing over all possible Boolean assignments $\mathbf{b} \in \mathbb{B}^{|\mathbf{B}|}$ while integrating over the weighted assignments of \mathbf{X} making the evaluation of the formula SAT: $(\mathbf{x}, \mathbf{b}) \models \Delta$.

Weight functions w are usually defined as products of literal weights (Belle et al., 2015; Chavira & Darwiche, 2008; Zeng & Van den Broeck, 2019). That is, for a set of literals \mathcal{L} , a set of per-literal weight functions $\mathcal{W} = \{w_\ell(\mathbf{x})\}_{\ell \in \mathcal{L}}$ is given, with weight functions w_ℓ defined over variables in literal ℓ . Then, the weight of assignment (\mathbf{x}, \mathbf{b}) is:

$$w(\mathbf{x}, \mathbf{b}) = \prod_{\ell \in \mathcal{L}} w_\ell(\mathbf{x})^{\llbracket \mathbf{x}, \mathbf{b} \rrbracket \ell}.$$

When all variables are Boolean (i.e., $\mathbf{X} = \emptyset$), the per-literal weights $w_\ell(\mathbf{x})$ are constants and we retrieve the original definition of the well-known weighted model counting (WMC) task (Chavira & Darwiche, 2008) as a special case of WMI. In this paper, we assume that all per-literal weights are from some certain weight function family, and for literals not in the set \mathcal{L} , their weights are the constant function one. This setting is expressive enough to approximate many continuous distributions (Belle et al., 2015).

Example 3 (WMI formulation of a skill matching system). Consider the team performance SMT model Γ in Example 1. Assume that a set of per-literal weights $w_{\ell_i}(X_{\mathcal{T}}, X_i) = 0.1 \cdot (X_{\mathcal{T}} + X_i - 6)^2$ is associated to literals $\ell_i = X_{\mathcal{T}} - X_i < 1$, quantifying how likely the team performance is upper bounded by player performances. Then the WMI of the formula Γ with two players is $\text{WMI}(\Gamma, w; \mathbf{X}, \mathbf{B}) \approx 170.69$.

Intuitively, $\text{WMI}(\Delta, w; \mathbf{X}, \mathbf{B})$ equals the partition function of the unnormalized probability distribution induced by weights w on formula Δ . In the following, we will adopt the shorthand $\text{WMI}(\Delta, w)$ for computing the WMI with all the variables in Δ in scope. The set of weight functions w together act as an unnormalized probability density while the formula Δ represents logical constraints defining its structure. Therefore, it is possible to compute the (now normalized) probability of any logical query Φ expressible as an SMT formula involving complex constraints as

$$\Pr_{\Delta}(\Phi) = \text{WMI}(\Delta \wedge \Phi, w) / \text{WMI}(\Delta, w).$$

Example 4 (WMI inference for skill rating). Suppose we want to quantify the squad effect in a 2v2 game. Specifically, given two teams \mathcal{T}_1 and \mathcal{T}_2 whose players have the same performance, but team \mathcal{T}_1 is a squad while \mathcal{T}_2 is not, that is, $\Phi_c = (B_1 = \text{true} \wedge B_2 = \text{false})$. We wonder what is the probability of query $\Phi = X_{\mathcal{T}_1} > X_{\mathcal{T}_2}$, that is team \mathcal{T}_1 wins and \mathcal{T}_2 loses. The probability of query Φ can be computed by two WMI tasks as follows.

$$\Pr_{\Delta}(\Phi | \Phi_c) = \frac{\text{WMI}(\Delta \wedge \Phi_c \wedge \Phi, w)}{\text{WMI}(\Delta \wedge \Phi_c, w)} = \frac{4,206}{7,225} \approx 58.22\%$$

with the SMT formula $\Delta := \Gamma_1 \wedge \Gamma_2$ where the two sub-formulas Γ_1 and Γ_2 model the two teams as in Example 1.

W.l.o.g, from here on we will focus on WMI problems on continuous variables only. We can safely do this since a

WMI problem on continuous and Boolean variables of the form $\text{WMI}(\Delta, w; \mathbf{X}, \mathbf{B})$ can always be reduced in polytime to a new WMI problem $\text{WMI}(\Delta', w'; \mathbf{X}')$ on continuous variables only, by properly introducing auxiliary variables in \mathbf{X}' to account for Boolean variables \mathbf{B} without increasing the problem size (Zeng & Van den Broeck, 2019).

From WMI to MI. Recently, model integration (MI) (Luu et al., 2014) has been proposed as an alternative way to perform WMI inference in Zeng & Van den Broeck (2019). MI is the task of computing the volumes corresponding to the models of an SMT formula. As such, MI is a special case of WMI in which the weights equate to one everywhere.

Definition 5. (Model Integration) Let \mathbf{X} consist of continuous random variables over \mathbb{R} , and let Δ be an SMT formula. The model integration (MI) of \mathbf{X} over Δ is:

$$\text{MI}(\Delta; \mathbf{X}) \triangleq \int_{\mathbf{x} \models \Delta} 1 \, d\mathbf{x} = \int_{\mathbb{R}^{|\mathbf{x}|}} \llbracket \mathbf{x} \models \Delta \rrbracket \, d\mathbf{x}$$

Zeng & Van den Broeck (2019) propose a polytime reduction of a WMI problem with polynomial weights to an MI one such that their proposed MI solver is amenable to a certain class of WMI problems. This reduction provides the basis for the largest class of tractable WMI problems known before our work. We will review it in the next section, before considerably expanding upon the class of WMI problems that can be solved tractably in the prior work.

3. Tractable WMI inference

The major efforts in advancing WMI inference have been so far concentrated on devising sophisticated WMI solvers to deliver exact inference routines for general scenarios without investigating the effect of the structure of a WMI problems on its complexity. Little to no attention has gone to formally understand which classes of WMI problems can be guaranteed to be solved exactly and in polynomial time, that is, *tractably*.

One notable exception can be found in Zeng & Van den Broeck (2019) where the search-based MI (SMI) solver is introduced. WMI problems for which SMI guarantees polytime exact inference constitute the first class of tractable WMI. Intuitively, SMI solves MI problems by using search to leverage the conditional independence among variables.

As in Zeng & Van den Broeck (2019) we characterize the structure of an SMT formula via its *primal graph*.

Definition 6. (Primal graph of SMT) The primal graph of an SMT formula Δ is an undirected graph \mathcal{G}_{Δ} whose vertices are variables in formula Δ and whose edges connect any two variables that appear in a same clause in the formula Δ .

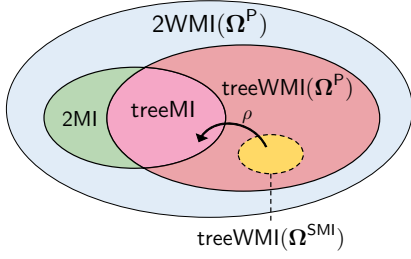


Figure 2: **The current landscape of classes of WMI problems.** We enlarge the boundaries of tractable WMI inference from treeMI to treeWMI and prove the hardness of 2MI and 2WMI.

An example primal graph of the SMT formula in Example 1 is shown in Figure 1. The SMI solver guarantees polynomial time execution for the class of MI problems with certain tree-shaped primal graphs, which we denote as treeMI.

Definition 7. (*treeMI Problem Class*) Let treeMI be the set of all MI problems over real variables whose SMT formula Δ induces a primal graph \mathcal{G}_Δ with treewidth one and with bounded diameter d . Problems in treeMI can be solved in polytime via SMI (Zeng & Van den Broeck, 2019).

Note that in Definition 7 the primal graph diameter here plays the role of a constant since, otherwise, SMI complexity can be worst-case exponential in diameter d . In the following we will try to answer if larger classes than treeMI are still amenable to tractable inference. We start by demonstrating a novel result that states the hardness of a larger class of MI problems, still focusing on dependencies between two variables, but allowing for non-tree-shaped primal graphs.

Definition 8. (*2MI Problem Class*) Let 2MI be the set of all MI problems over real variables whose SMT formula Δ is a conjunction of clauses comprising at most two variables.

Note that a clause comprising at most two variables can be a conjunction of arbitrarily many literals. Moreover, when there are more than two variables in a clause, in the primal graph there must be a loop and thus the treewidth of the primal graph is larger than one. Hence all MI problems with tree-shaped primal graph must be in the class 2MI.

Theorem 9. (*Hardness of 2MI*) Given an MI problem in 2MI with an SMT formula Δ , computing $\text{MI}(\Delta)$ is #P-hard.

Sketch of Proof. The proof is done by reducing the #P-complete problem #2SAT to an MI problem in 2MI with an SMT formula Δ such that counting the number of satisfying assignments to the #2SAT problem equates to the MI of formula Δ . See Appendix for a detailed proof. \square

From Theorem 9 it follows that the problem class $2\text{WMI}(\Omega)$, i.e., the WMI problems with SMT formulas

being a conjunction of clauses comprising at most two variables, and with per-literal weights in weight function family Ω , is also hard since class 2MI is a sub-class of $2\text{WMI}(\Omega)$. We revert our attention to WMI problems exhibiting a dependency tree structure. Notice that for WMI problems, the tractability not only depends on the logical structure defined by the SMT formulas, but also the statistical structure defined by weight functions. Next in our analysis, we take into consideration the weight function families. Analogously to what Definition 7 states, we introduce the notion of $\text{treeWMI}(\Omega)$ with the associated weight function family specified as follows.

Definition 10. (*treeWMI(Ω) Problem Class*) Let $\text{treeWMI}(\Omega)$ be the set of all WMI problems over real variables whose SMT formula Δ induces a primal graph \mathcal{G}_Δ with treewidth one and with bounded diameter d , and whose per-literal weights are in a function family Ω .

Zeng & Van den Broeck (2019) propose a WMI-to-MI reduction such that some $\text{treeWMI}(\Omega)$ problems with polynomial weights are reduced in polynomial time to treeMI problems amenable to tractable inference by the SMI solver. Intuitively, the reduction process introduces auxiliary continuous variables and SMT formulas over these variables to encode the polynomial weight functions. We refer the readers to Zeng & Van den Broeck (2019) for a detailed description of the reduction. However, as shown next, the set of treeWMI problems that can be reduced to treeMI is rather limited.

Definition 11. (*Ω^{SMI} Weight Function Family*) Let Ω^{SMI} be the family of per-literal weight functions that are monomials associated with either (i) univariate literals or (ii) a literal that appears exclusively in a unit clause, i.e., a clause consisting of a single literal.

Theorem 12. Let ρ be the polytime WMI-to-MI reduction for $\text{treeWMI}(\Omega)$ problems as defined in Zeng & Van den Broeck (2019). Then the image $\{\rho(\nu) \mid \nu \in \text{treeWMI}(\Omega)\} \subset \text{treeMI}$ if-and-only-if $\Omega \subset \Omega^{\text{SMI}}$.

Sketch of Proof. The necessary condition can be proved by the reduction process and the sufficient one can be proved by contradiction. See Appendix for a detailed proof. \square

Therefore, the SMI solver is limited to a rather restricted subset of $\text{treeWMI}(\Omega)$ since from the definition of Ω^{SMI} we can tell that it is a strict subset of monomial per-literal weights. In order to enlarge the tractable class of WMI problems, next we will define a rich family of weight functions.

Definition 13. (*Tractable Weight Conditions*) Let Ω be a family of per-literal weight functions. We say that the tractable weight conditions (TWC) hold for Ω if we have:

- (i) *closedness under product:* $\forall f, g \in \Omega, f \cdot g \in \Omega$;

- (ii) *tractable symbolic integration*: $\forall f \in \Omega$, the symbolic antiderivative of function f can be tractably computed by symbolic integration;
- (iii) *closedness under definite integration*: $\forall f \in \Omega$ with its antiderivative denoted by F , given integration bounds $l(x), u(x)$ in \mathcal{LRA} with $x \in \mathbf{X}$, $F(u(x)) - F(l(x)) \in \Omega$.

Some example weight function families that satisfy TWC include the polynomial family, exponentiated linear function family and the function family resulting from their product. Moreover note that piecewise function families, when pieces belong to the above families, also satisfy TWC. It turns out that the weight function families that satisfy TWC subsume and extend all the parametric weight functions adopted in the WMI literature so far. The following proposition is a direct result from the fact that the piecewise polynomial weight family Ω^P is a strict superset of the family Ω^{SMI} .

Proposition 14. *Let Ω^P be the piecewise polynomial weight function family. The WMI problem class $\text{treeWMI}(\Omega^P)$ is a strict superset of problem class $\text{treeWMI}(\Omega^{\text{SMI}})$.*

Theorem 15. *If a weight function family Ω satisfies TWC as in Definition 13, WMI problems in class $\text{treeWMI}(\Omega)$ are tractable, i.e., they can be solved in polynomial time.*

The proof to the above theorem is provided in the next two sections by construction where in Section 4 we proposed our WMI solver, called MP-WMI, operating on WMI problems in $\text{treeWMI}(\Omega)$ with its complexity analysis in Section 5. A summary of the WMI problem classes is shown in Figure 2.

4. Message-Passing WMI

Message passing on tree-structured graphs has achieved remarkable attention in the PGM literature (Pearl, 1988; Kschischang et al., 2001). Its classical formulation and efficiency relies on compact factor representations allowing easy computations. However, adapting existing message-passing algorithms to WMI inference is non-trivial. This is due to the fact that inference is computed in a hybrid structured space with logical and arithmetic constraints. We present our message-passing scheme by first deriving a factorized representation of WMI problems.

4.1. Factor Graph Representation of WMI

In the literature of WMC, *incidence graphs* are proposed to characterize the structure of problems defined by Boolean CNF formulas (Samer & Szeider, 2010). Incidence graphs are bipartite graphs with clause nodes and variable nodes, where a clause and a variable node are joined by an edge if the variable occurs in the clause. We derive the analogous representation for the more general SMT formulas, which we then turn into a factor graph of WMI problems.

Recall that for the joint distribution represented by a WMI problem, the support is defined by the logical constraints and the unnormalized density is defined by weight functions. In the following, we first factorize the SMT formula Δ of a WMI problem $\text{WMI}(\Delta, w)$ in the class treeWMI :

$$\Delta = \bigwedge_{i \in \mathcal{V}} \Delta_i \wedge \bigwedge_{i,j \in \mathcal{E}} \Delta_{ij} \quad (2)$$

where the set \mathcal{V} is the index set of variables and the set \mathcal{E} is the index pairs of variables in the same clause. Then a WMI problem can be conveniently represented as a bipartite graph, known as factor graph, that includes two sets of nodes: variable nodes X_i , and factor nodes $f_{\mathcal{S}}$, where \mathcal{S} denotes a factor scope, i.e., the set of indices of the variables appearing in it. A variable node X_i is connected to a factor node $f_{\mathcal{S}}$ if and only if $i \in \mathcal{S}$. Specifically, the factors are defined as follows:

$$f_{\mathcal{S}}(\mathbf{x}_{\mathcal{S}}) = \prod_{\Gamma \in \text{CLS}(\Delta_{\mathcal{S}})} \llbracket \mathbf{x}_{\mathcal{S}} \models \Gamma \rrbracket \prod_{\ell \in \text{LITS}(\Gamma)} w_{\ell}(\mathbf{x}_{\mathcal{S}})^{\llbracket \mathbf{x}_{\mathcal{S}} \models \ell \rrbracket} \quad (3)$$

where $\mathbf{x}_{\mathcal{S}}$ denotes the restriction of \mathbf{x} to the variables in factor $f_{\mathcal{S}}$ and analogously $\Delta_{\mathcal{S}}$ is the restriction of formula Δ to the clauses over the variables in \mathcal{S} . Here, the set of clauses in the SMT formula Δ is denoted by $\text{CLS}(\Delta)$, and the set of literals in a clause Γ is denoted by $\text{LITS}(\Gamma)$. Intuitively, the factors include the parameterized densities as in the classic PGM literature, here represented by the per-literal weights, but also the structure enforced by the logical constraints in the SMT formula, via the indicator functions. Figure 3 shown an example of a factor graph.

As in every tree-shaped factor graphs, we define an unnormalized joint distribution corresponding to the WMI problem in the form of a product of factors as follows.

$$p(\mathbf{x}) = \prod_{\mathcal{S}} f_{\mathcal{S}}(\mathbf{x}_{\mathcal{S}}) = \prod_{i \in \mathcal{V}} f_i(X_i) \prod_{i,j \in \mathcal{E}} f_{ij}(X_i, X_j) \quad (4)$$

By construction, it is easy to see that the normalization constant of such a distribution equals computing the corresponding weighted model integral.

Proposition 16. *Given a problem $\text{WMI}(\Delta, w)$ in treeWMI , let $p(\mathbf{x})$ being the unnormalized joint distribution as defined in Equation 4. Then the partition function of distribution $p(\mathbf{x})$ is equal to $\text{WMI}(\Delta, w)$.*

4.2. Message-Passing Scheme

Deriving a message-passing scheme for WMI poses unique and considerable challenges. First, different from discrete domains, on continuous or hybrid domains one generally does not have universal and compact representations for messages, and logical constraints in WMI make it even

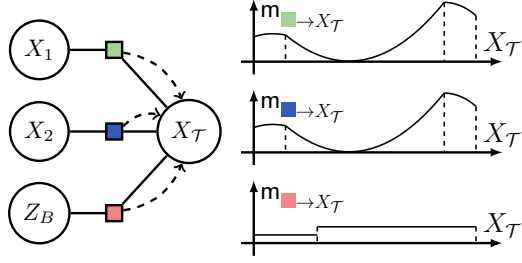


Figure 3: **Factor graph** (left) of formula γ with two players and **piecewise polynomial messages** (right) sent from the three factor nodes to variable node $X_{\mathcal{T}}$ when solving the WMI in Example 3 by MP-WMI.

harder to derive such representations. Moreover, marginalization over real variables requires integration over polytopes, which is already #P-hard (Dyer & Frieze, 1988). The integration poses the problem of whether the messages defined are integrable and how hard it is to perform the integration. In the following part, we will present our solutions to these challenges by first describing a general message-passing scheme for WMI and then investigating of which form the messages are, given certain weight families.

Given the factorized representation of WMI in Section 4.1, our message-passing scheme, called MP-WMI and summarized in Algorithm 1, comprises exchanging messages between nodes in the factor graph. Before the message passing starts, we choose an arbitrary node in the factor graph as root and orient all edges away from the root to define the message sending order. MP-WMI operates in two phases: an upward pass and a downward one. First, we send messages up from the leaves to the root (upward pass) such that each node has all information from its children and then we incorporate messages from the root down to the leaves (downward pass) such that each node also has information from its parent. The messages are formulated as follows.

Proposition 17. *Both messages $m_{f_{ij} \rightarrow X_i}$ from factor node to variable node and messages $m_{X_i \rightarrow f_{ij}}$ from variable node to factor node have iterative formulations as follows.*

- (i) $m_{f_{ij} \rightarrow X_i}(x_i) = \int f_{ij}(x_i, x_j) \cdot m_{X_j \rightarrow f_{ij}}(x_j) dx_j$;
- (ii) $m_{X_i \rightarrow f_S}(x_i) = \prod_{f_{S'} \in \text{neigh}(X_i) \setminus f_S} m_{f_{S'} \rightarrow X_i}(x_i)$.

For the start of sending messages, when a leaf node is a variable node X_i , the message that it sends along its one and only edge to a factor f_S is $m_{X_i \rightarrow f_S}(c_i) = 1$; in the case when a leaf node is a factor node f_i , the message from the factor node f_i to a variable node X_i is $m_{f_i \rightarrow X_i}(x_i) = f_i(x_i)$. Even though the weight function family is not specified here, it can be shown that when the integration in Proposition 17 is well-defined, i.e., the integrands are integrable, then the messages are univariate piecewise functions, which is a striking difference with clas-

Algorithm 1 MP-WMI(Δ)

- 1: $V_{\text{up}} \leftarrow$ sort variable nodes in factor graph
 - 2: **for each** $X_i \in V_{\text{up}}$ **do** {upward pass}
 - 3: send-message($X_i, f_{i, \text{pa}(i)}$)
 - 4: send-message($f_{i, \text{pa}(i)}, X_{\text{pa}(i)}$)
 - 5: **end for**
 - 6: $V_{\text{down}} \leftarrow$ sort nodes in set V_{up} in reverse order
 - 7: **for each** $X_i \in V_{\text{down}}$ **do** {downward pass}
 - 8: **for each** $X_c \in \text{ch}(X_i)$ **do**
 - 9: send-message(X_i, f_{ic})
 - 10: send-message(f_{ic}, X_c)
 - 11: **end for**
 - 12: **end for**
 - 13: **return** $\{m_{X_i \rightarrow f_S}, m_{f_S \rightarrow X_i}\}_{(x_i, f_S) \in \mathcal{E}}$
-

sical message-passing schemes.

Proposition 18. *For any problem in treeWMI, the messages as in Proposition 17 are univariate piecewise functions.*

The specific form of messages also depends on the chosen weight function family as mentioned in Section 3. For example, when the weight functions are chosen to be polynomials, the messages are piecewise polynomials, as in the example in Figure 3. We show how to compute the piecewise polynomial messages in Algorithm 2 with functions *critical-points* and *get-msg-pieces* as subroutines to compute the numeric and symbolic integration bounds for the message pieces. Both of them can be efficiently implemented, see Zeng & Van den Broeck (2019) for details. The actual integration of the polynomial pieces can be efficiently performed symbolically, as supported by many scientific computing packages.

When MP-WMI terminates, the information stored in the obtained messages is sufficient to compute the unnormalized marginals for each variable and it is independent of the choice of root. Moreover, the integration of unnormalized marginals equals to WMI(Δ, w).

Proposition 19. *Let Δ be an SMT formula with a tree factor graph and with per-literal weights w . For any variable X_i , the unnormalized marginal $p(x_i)$ is*

$$p(x_i) = \prod_{f_S \in \text{neigh}(X_i)} m_{f_S \rightarrow X_i}(x_i).$$

Moreover, the partition function for any x_i is the WMI of SMT formula Δ , i.e., $\text{WMI}(\Delta, w) = \int_{x_i} p(x_i) dx_i$.

4.3. Amortization

We will show that by leveraging the messages pre-computed in MP-WMI, we are able to speed up (amortize) inference time over multiple queries on formula Δ . More specifically, when answering queries that do not change the tree structure

Algorithm 2 send-message(s, t)

```

1: if  $s = X_i$  and  $t = f_{ij}$  then
2:   Return  $\prod_{f_{s'} \in \text{neigh}(X_i) \setminus f_{ij}} m_{f_{s'} \rightarrow X_i}$ 
3: else if  $s = f_{ij}$  and  $t = X_i$  then
4:    $\mathcal{P} \leftarrow \text{critical-points}(m_{X_j \rightarrow f_{ij}}, \Delta_{ij})$ 
5:    $\mathcal{I} \leftarrow \text{intervals-from-points}(\mathcal{P})$ 
6:   for interval  $I \in \mathcal{I}$  consistent with formula  $\Delta_{ij}$  do
7:      $\langle l_s, u_s, p \rangle \leftarrow \text{get-msg-pieces}(m_{X_j \rightarrow f_{ij}}, I, w)$ 
8:      $p'(x_i) \leftarrow \int_{l_s}^{u_s} p(x_i, x_j) dx_j$ 
9:      $m_{f_{ij} \rightarrow X_i} \leftarrow m_{f_{ij} \rightarrow X_i} + \llbracket x_i \in I \rrbracket \cdot p'(x_i)$ 
10:  end for
11: end if
12: return  $m_{s \rightarrow t}$ 

```

in the factor graph of formula Δ , we only need to update messages that are related to the queries while other messages are pre-computed. Some examples are SMT queries on a node variable or queries over a pair of variables that are connected by an edge in the factor graph, since these queries either add leaf nodes or do not change existing nodes. Thus we can reuse the local information encoded in messages.

Proposition 20. Let $\text{WMI}(\Delta, w)$ be a problem in treeWMI, and Φ be an SMT query over a factor f_s^* involving a variable $X_i \in \mathbf{X}$. Given pre-computed messages $\{m_{f_s \rightarrow X_i}\}_{f_s \in \text{neigh}(X_i)}$,

$$\text{WMI}(\Delta \wedge \Phi) = \int_{\mathbb{R}} m_{f_s^* \rightarrow X_i}^*(x_i) \cdot \prod_{f_s \in \text{neigh}(X_i) \setminus f_s^*} m_{f_s \rightarrow X_i}(x_i) dx_i$$

with message $m_{f_s^* \rightarrow X_i}^*$ computed over factor $f_s^*(\mathbf{x}_s) := f_s(\mathbf{x}_s) \cdot \llbracket \mathbf{x}_s \models \Phi \rrbracket$ as in Proposition 17.

Pre-computing messages can dramatically speed up inference by amortization, as we will show in our experiments, especially when traversing the factor graph is expensive or the number of queries is large.

5. Complexity Analysis

This section provides a complexity analysis of our proposed WMI solver MP-WMI. Given the SMT formula Δ with a tree factor graph with a chosen root node, each factor node would be traversed exactly once in each phase of the message-passing scheme. We denote the set of directed factor nodes by $\mathcal{F} := \{\vec{f}_s\} = \{f_s^+, f_s^- \mid f_s \in \mathcal{V}\}$ where f_s^+ denotes the factor node f_s visited in the upward pass and f_s^- denotes the one visited in downward pass respectively.

To characterize the message-passing scheme, we define a nilpotent matrix A as follows. The matrix $A \in \mathbb{N}^{|\mathcal{F}| \times |\mathcal{F}|}$ has both its columns and rows denoted by the factor nodes in

set \mathcal{F} . At each column denoted by \vec{f}_s , only entries at rows denoted by factor nodes visited right after \vec{f}_s are non-zero.

Proposition 21. The nilpotent matrix A as described above has its order at most the diameter of the factor graph.

Next we show how to define the non-zero entries in matrix A with parameters about the SMT formulas in WMI problems.

Proposition 22. Suppose that the two variables X_i and X_j are connected in the factor graph by a factor f_{ij} associated with a sub-formula Δ_{ij} of size c , then in MP-WMI:

- (i) the number of pieces in message $m_{X_i \rightarrow f_{ij}}$ is bounded by $\sum m_s$, where m_s is the number of pieces in message $m_{f_s \rightarrow X_i}$ with $f_s \in \text{neigh}(X_i) \setminus f_{ij}$;
- (ii) the number of pieces in message $m_{f_{ij} \rightarrow X_j}$ is bounded by $2mc + c^2$ with m being the number of pieces in message $m_{X_i \rightarrow f_{ij}}$.

Now we show how to use the matrix A to bound the number of pieces in messages. We define the non-zero entries in the nilpotent matrix A to be $2c$ with c being a constant that bounds the size of sub-formulas associated to factors. Define a vector $v^{(t)} \in \mathbb{N}^{|\mathcal{E}|}$ for the state of the message-passing scheme at step t – by state it means that each entry in vector $v^{(t)}$ is denoted by a factor node in set \mathcal{F} and the entry denoted by \vec{f}_s bounds the number of pieces in the message sent to f_s in the MP-WMI. For the initial state vector $v^{(0)}$, it has all non-zero entries to be c , the constant bounding the sub-formula size, and these entries are those denoted by $\vec{f}_s = f_s^+$ with factor node f_s connected with a leaf.

Proposition 23. Let A be the nilpotent matrix and v the initial state vector as described above. Also let $v^{(t)} := Av^{(t-1)} + c^2 \cdot \text{sgn}(Av^{(t-1)})$ with sgn being the sign function. Then each entry in vector $v^{(t)}$ denoted by \vec{f}_s bounds the number of pieces in the message $m_{X_i \rightarrow f_s}$ received by factor f_s from some variable node X_i at step t in MP-WMI.

Proposition 24. Let A be the nilpotent matrix and $v^{(t)}$ the state vectors as described above. The total number of pieces in all the messages is bounded by $\|\sum_{t=0}^d v^{(t)}\|_1$ with d being the diameter of the factor graph. Moreover, it holds that $\|\sum_{t=0}^d v^{(t)}\|_1$ is of $\mathcal{O}((4nc)^{2d+2})$.

This gives the worst-case total number of message pieces in MP-WMI. From Proposition 24, it holds that the problems in class treeWMI(Ω) with the weight function family Ω satisfying TWC are tractable to MP-WMI, since the complexity of MP-WMI is the total number of message pieces multiplied by the symbolic integration cost of each piece, which is tractable for functions in family Ω by definition. This finishes the constructive proof for Theorem 15 in Section 3. Notice the complexity of WMI problems depends on the graph structures. In our experiments, we will compare

solvers on WMI problems with three representative problem classes with different factor graph diameters.

6. Related Work

WMI generalizes weighted model counting (WMC) (Sang et al., 2005) to hybrid domains (Belle et al., 2015). WMC is one of the state-of-the-art approaches for inference in many discrete probabilistic models. Existing exact WMI solvers for arbitrarily structured problems include DPLL-based search with numerical (Belle et al., 2015; Morettin et al., 2017; 2019) or symbolic integration (de Salvo Braz et al., 2016) and compilation-based algorithms (Kolb et al., 2018; Zuidberg Dos Martires et al., 2019a).

Motivated by their success in WMC, Belle et al. (2016) present a caching scheme for WMI that allows reusing computations at the cost of not supporting algebraic constraints between variables. Different from usual, Merrell et al. (2017) adopt Gaussian distributions, while Zuidberg Dos Martires et al. (2019a) fixed univariate parametric assumptions for weight functions. Closest to our MP-WMI, SMI (Zeng & Van den Broeck, 2019) is an exact solver which leverages context-specific independence to perform efficient search and operates on tree-shaped primal graphs. Many recent efforts in WMI converged in the *pywmi* (Kolb et al., 2019) python framework.

Tree-shaped dependency structures, as the ones characterizing our treeWMI(Ω) class, naturally arise in many fields, such genetics (Nei & Kumar, 2000), system analysis (Vesely et al., 1981), linguistics (Petrov et al., 2006), and telecommunications (Leon-Garcia & Widjaja, 2003). Moreover, thanks to their appealing mathematical properties, trees serve as practical approximations of non tree-shaped problems (Rubinstein et al., 1983; Robins & Zelikovsky, 2000; Binev & DeVore, 2004).

Message-passing schemes have been widely used for developing exact and approximate inference algorithms for probabilistic graphical models on discrete (Kschischang et al., 2001), continuous (Guo et al., 2019; Wang et al., 2018) and hybrid domains (Gogate & Dechter, 2012). Our amortization scheme is closely related to the reuse of local computation in the join tree algorithm (Huang & Darwiche, 1996; Lepar & Shenoy, 2013), which has never been explored in hybrid domains for WMI inference, however. Similarly to us, Gamarnik et al. (2012) adopts piecewise polynomial messages, specifically piecewise-linear convex functions, in a belief propagation scheme for non-probabilistic min-cost network flow problems.

Research on learning WMI distributions from data is at its early stages. Parameter learning for piecewise constant densities has been addressed in (Belle et al., 2015). Recently, an approach for jointly learning the structure and parameters

of a WMI problem has been proposed in (Morettin et al., 2020). Developing faster inference algorithms is thus beneficial in learning scenarios as, typically, learning a full model requires numerous calls to an inference procedure. WMI inference is closely related to probabilistic program inference, where complex arithmetic and logical constraints are induced by the program structure or its abstraction (Holtzen et al., 2017; 2018).

7. Experiments

In this Section, we aim to answer the following research questions:¹ **Q1**) Can we effectively scale WMI inference with MP-WMI? **Q2**) How beneficial is inter-query amortization with MP-WMI?

To answer **Q1**, we generated a benchmark of WMI problems with tree-shaped primal graphs of different diameters: star-shaped graphs (STAR), complete ternary trees (SNOW) and linear chains (PATH). These structures were originally investigated by the authors of SMI and are prototypical of tree shapes that can be encountered in real-world scenarios such as phylogenetic trees (Nei & Kumar, 2000), hierarchies in file and networks systems (Vesely et al., 1981), and natural language grammars (Petrov et al., 2006).

We sampled random SMT formulas with N variables with the tree structures described above and polynomial weights mapping a subset of literals to a random non-negative polynomials. We generated problems with N ranging from 2 to 20 with step size 2, and from 20 to 100 with step size 10. We compared our MP-WMI python implementation against the following baselines: WMI-PA (Morettin et al., 2019), a solid general-purpose WMI solver exploiting SMT-based predicate abstraction techniques that is less sensitive to the problem structure; and F-XSDD(BR) (Zuidberg Dos Martires et al., 2019b), a compilation-based solver achieving state-of-the-art results in several WMI benchmarks.

Fig. 4 shows that, with timeout being an hour, our proposed solver MP-WMI is able to scale up to 60 variables for STAR problems and up to 90 variables for SNOW and PATH problems, while the other two solvers stop at problem size 20 for all three classes. Note that the results are in line with those reported in (Zuidberg Dos Martires et al., 2019b). This answers **Q1** affirmatively, raising the bar of the size of WMI problems that can be solved exactly up to 100 variables.

We tackle **Q2** by comparing MP-WMI with SMI (Zeng & Van den Broeck, 2019) on tree-structured MI problems. SMI is a search-based MI solver that has been shown to be efficient for such problems. WMI-PA, F-XSDD and the SGDPLL(T) (de Salvo Braz et al., 2016) solver are not

¹Our implementation of MP-WMI and the code for reproducing our empirical evaluation can be found at <https://github.com/UCLA-StarAI/mpwmi>.

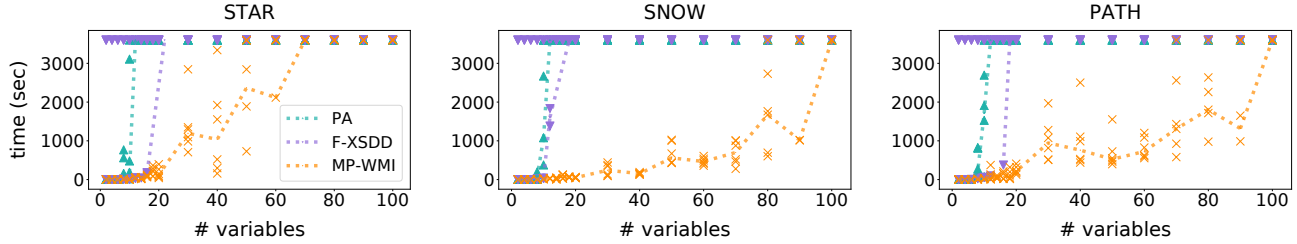


Figure 4: Results of the comparison between MP-WMI, WMI-PA and F-XSDD on WMI problems with tree dependencies. In this setting, MP-WMI remarkably scales to problems having up to 60 variables on STAR, while solving SNOW and PATH problems having up to 90 variables, considerably “raising the bar” for the size of tractable WMI inference problems.

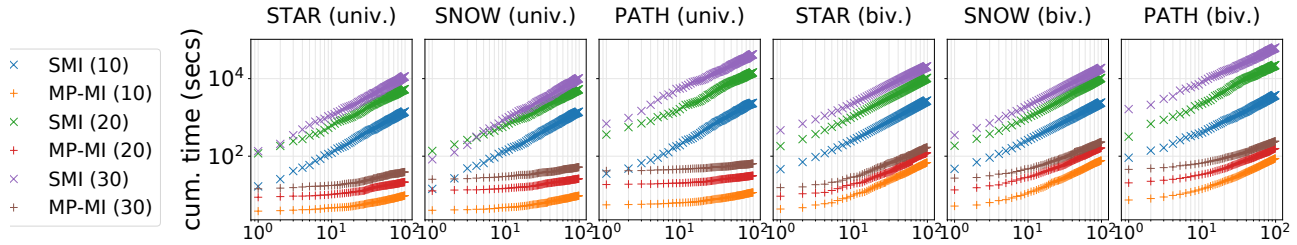


Figure 5: Log-log plot of cumulative time (seconds, y-axis) for MP-WMI (orange, red, brown) and SMI (blue, green, purple) over STAR, SNOW and PATH primal graphs (see text) with 10, 20 and 30 variables for increasing numbers of univariate and bivariate queries (x-axis). For every class, MP-WMI takes up to two order of magnitude less time when amortizing 100 queries, while being faster than SMI on a single query.

included in the comparison since they were already shown in Zeng & Van den Broeck (2019) to not be competitive on such problems. The synthetic SMT formulas range over $n \in \{10, 20, 30\}$ variables with tree factor graphs being STAR, SNOW and PATH. We generate 100 univariate or bivariate random queries for each MI problem.

Figure 5 shows the cumulative runtime of answering random queries by both solvers. As expected, MP-WMI takes a fraction of the time of SMI (up to two order of magnitudes) to answer 100 univariate or bivariate queries in all experimental scenarios, since it is able to amortize inference *inter-query*. More surprisingly, by looking at the first point of each curve, we can tell that MP-WMI is even faster than SMI to compute a single query. This is because SMI solves polynomial integration numerically, by reconstructing the univariate polynomials before the numeric integration via interpolation, e.g., Lagrange interpolation; while in MP-WMI we adopt symbolic integration. Hence the complexity of the former is always quadratic in the degree of the polynomial, while for the latter the average case is linear in the number of monomials in the polynomial to integrate, which in practice might be much less than the degree of the polynomial.

8. Conclusions

In this paper, we theoretically traced the boundaries of tractable WMI inference and proposed a novel exact WMI solver based on message-passing, MP-WMI, which is efficient on a rich class of tractable WMI problems with tree-shaped factor graphs, the largest known so far. Furthermore, MPWMI dramatically reduces the answering time of multiple queries by amortizing local computations and allows to compute all marginals and moments simultaneously.

We believe this provides a theoretical and algorithmic stepping stone needed to devise principled approximate WMI inference schemes that can scale even further to larger and non tree-shaped problem structures.

Acknowledgements

This work is partially supported by NSF grants #IIS-1943641, #IIS-1633857, #CCF-1837129, DARPA XAI grant #N66001-17-2-4032, a Sloan fellowship, and gifts from Intel and Facebook Research.

The authors would like to thank Arthur Choi for insightful discussions about message passing routines and generalizations in the PGM literature.

References

- Barrett, C. and Tinelli, C. Satisfiability modulo theories. In *Handbook of Model Checking*, pp. 305–343. Springer, 2018.
- Barrett, C., de Moura, L., Ranise, S., Stump, A., and Tinelli, C. The SMT-LIB initiative and the rise of SMT (HVC 2010 award talk). In *Proceedings of the 6th international conference on Hardware and software: verification and testing*, pp. 3–3. Springer-Verlag, 2010.
- Belle, V., Passerini, A., and Van den Broeck, G. Probabilistic inference in hybrid domains by weighted model integration. In *Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2770–2776, 2015.
- Belle, V., Van den Broeck, G., and Passerini, A. Component caching in hybrid domains with piecewise polynomial densities. In *AAAI*, pp. 3369–3375, 2016.
- Biere, A., Heule, M., and van Maaren, H. *Handbook of satisfiability*, volume 185. IOS press, 2009.
- Binev, P. and DeVore, R. Fast computation in adaptive tree approximation. *Numerische Mathematik*, 97(2):193–217, 2004.
- Chavira, M. and Darwiche, A. On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6-7):772–799, 2008.
- de Salvo Braz, R., O’Reilly, C., Gogate, V., and Dechter, R. Probabilistic inference modulo theories. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 3591–3599. AAAI Press, 2016.
- Dechter, R. and Mateescu, R. And/or search spaces for graphical models. *Artificial intelligence*, 171(2-3):73–106, 2007.
- Dyer, M. E. and Frieze, A. M. On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, 17(5):967–974, 1988.
- Gamarnik, D., Shah, D., and Wei, Y. Belief propagation for min-cost network flow: Convergence and correctness. *Operations research*, 60(2):410–428, 2012.
- Gogate, V. and Dechter, R. Approximate inference algorithms for hybrid bayesian networks with discrete constraints. *arXiv preprint arXiv:1207.1385*, 2012.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- Guo, Y., Xiong, H., and Ruoizzi, N. Marginal inference in continuous Markov random fields using mixtures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 7834–7841, 2019.
- Heckerman, D. and Geiger, D. Learning Bayesian networks: a unification for discrete and gaussian domains. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 274–284. Morgan Kaufmann Publishers Inc., 1995.
- Holtzen, S., Millstein, T., and Van den Broeck, G. Probabilistic program abstractions. In *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.
- Holtzen, S., Van den Broeck, G., and Millstein, T. Sound abstraction and decomposition of probabilistic programs. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.
- Huang, C. and Darwiche, A. Inference in belief networks: A procedural guide. *International journal of approximate reasoning*, 15(3):225–263, 1996.
- Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Kolb, S., Mladenov, M., Sanner, S., Belle, V., and Kersting, K. Efficient symbolic integration for probabilistic inference. In *IJCAI*, pp. 5031–5037, 2018.
- Kolb, S., Morettin, P., Zuidberg Dos Martires, P., Somavilla, F., Passerini, A., Sebastiani, R., and De Raedt, L. The pywmi framework and toolbox for probabilistic inference using weighted model integration. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 6530–6532, 2019.
- Koller, D. and Friedman, N. Probabilistic graphical models. 2009.
- Kschischang, F. R., Frey, B. J., and Loeliger, H.-A. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.
- Lauritzen, S. L. and Wermuth, N. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The annals of Statistics*, pp. 31–57, 1989.
- Leon-Garcia, A. and Widjaja, I. *Communication networks*. McGraw-Hill, Inc., 2003.
- Lepar, V. and Shenoy, P. P. A comparison of Lauritzen-Spiegelhalter, Hugin, and Shenoy-Shafer architectures for computing marginals of probability distributions. *arXiv preprint arXiv:1301.7394*, 2013.

- Luu, L., Shinde, S., Saxena, P., and Demsky, B. A model counter for constraints over unbounded strings. In *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*, pp. 565–576, 2014.
- Merrell, D., Albarghouthi, A., and D’Antoni, L. Weighted model integration with orthogonal transformations. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017. doi: 10.24963/ijcai.2017/643.
- Minka, T., Cleven, R., and Zaykov, Y. Trueskill 2: An improved bayesian skill rating system. 2018.
- Molina, A., Vergari, A., Di Mauro, N., Natarajan, S., Esposito, F., and Kersting, K. Mixed sum-product networks: A deep architecture for hybrid domains. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
- Morettin, P., Passerini, A., and Sebastiani, R. Efficient weighted model integration via smt-based predicate abstraction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 720–728. AAAI Press, 2017.
- Morettin, P., Passerini, A., and Sebastiani, R. Advanced smt techniques for weighted model integration. *Artificial Intelligence*, 275:1–27, 2019.
- Morettin, P., Kolb, S., Teso, S., and Passerini, A. Learning weighted model integration distributions. In *AAAI*, pp. 5224–5231, 2020.
- Nei, M. and Kumar, S. *Molecular evolution and phylogenetics*. Oxford university press, 2000.
- Pearl, J. Probabilistic reasoning in intelligent systems: Networks of plausible inference, 1988.
- Petrov, S., Barrett, L., Thibaux, R., and Klein, D. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 433–440, 2006.
- Robins, G. and Zelikovsky, A. Improved steiner tree approximation in graphs. In *SODA*, pp. 770–779. Citeseer, 2000.
- Rubinstein, J., Penfield, P., and Horowitz, M. A. Signal delay in rc tree networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2(3): 202–211, 1983.
- Samer, M. and Szeider, S. Algorithms for propositional model counting. *Journal of Discrete Algorithms*, 8(1): 50–64, 2010.
- Sang, T., Beame, P., and Kautz, H. A. Performing Bayesian inference by weighted model counting. In *AAAI*, volume 5, pp. 475–481, 2005.
- Shenoy, P. P. and West, J. C. Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52(5):641–657, 2011.
- Vergari, A., Molina, A., Peharz, R., Ghahramani, Z., Kersting, K., and Valera, I. Automatic Bayesian density analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 5207–5215, 2019.
- Vesely, W. E., Goldberg, F. F., Roberts, N. H., and Haasl, D. F. Fault tree handbook. Technical report, Nuclear Regulatory Commission Washington DC, 1981.
- Wang, D., Zeng, Z., and Liu, Q. Stein variational message passing for continuous graphical models. In *Proceedings of the International Conference of Machine Learning*, 2018.
- Yang, E., Baker, Y., Ravikumar, P., Allen, G., and Liu, Z. Mixed graphical models via exponential families. In *Artificial Intelligence and Statistics*, pp. 1042–1050, 2014.
- Zeng, Z. and Van den Broeck, G. Efficient search-based weighted model integration. *Proceedings of UAI*, 2019.
- Zuidberg Dos Martires, P. M., Dries, A., and De Raedt, L. Exact and approximate weighted model integration with probability density functions using knowledge compilation. In *Proceedings of the 30th Conference on Artificial Intelligence*. AAAI Press, 2019a.
- Zuidberg Dos Martires, P. M., Kolb, S., and De Raedt, L. How to exploit structure while solving weighted model integration problems. *UAI 2019 Proceedings*, 2019b.

A. Proofs

A.1. THEOREM 9

Proof. (Theorem 9) The proof is done by reducing the #P-complete problem #2SAT over a 2SAT formula $\Delta_{\mathbb{B}}$ to an MI problem on a 2-Clause SMT(\mathcal{LRA}) formula Δ .

By the Boolean-to-real reduction from (Zeng & Van den Broeck, 2019), there exists an SMT(\mathcal{LRA}) formula Δ over real variables only such that $\text{MI}(\Delta_{\mathbb{B}}) = \text{MI}(\Delta)$. The formula Δ can be obtained in the following way. Any Boolean literal B or $\neg B$ in propositional formula $\Delta_{\mathbb{B}}$ is substituted by \mathcal{LRA} literals $Z_B > 0$ and $Z_B < 0$ respectively where the real variable Z_B is an auxiliary real variable with bounding box $(Z_B \geq -1) \wedge (Z_B \leq 1)$. Denote the formula after replacement by Δ' . Then we have formula Δ as follows.

$$\Delta = \Delta' \wedge \bigwedge_{B \in \text{vars}(\Delta_{\mathbb{B}})} (Z_B \geq -1) \wedge (Z_B \leq 1)$$

For each clause in formula Δ , since it contains at most two Boolean variables before substitution, it also contains at most two real variables now. Therefore formula Δ is a 2-Clause SMT(\mathcal{LRA}) formula over real variables only. Moreover, the reduction guarantees that $\text{MI}(\Delta) = \text{MI}(\Delta_{\mathbb{B}})$ where $\text{MI}(\Delta_{\mathbb{B}})$ is the number of satisfying assignments to $\Delta_{\mathbb{B}}$ by the definition of WMI. Thus, computing MI of a 2-Clause SMT(\mathcal{LRA}) formula over real variables is #P-hard. \square

A.2. THEOREM 12

Proof. (Theorem 12) When the weight function family $\Omega = \Omega^{\text{SMI}}$, by the WMI-to-MI reduction process in Zeng & Van den Broeck (2019), any WMI problem in $\text{treeWMI}(\Omega)$ can be reduced to an MI problem in class treeMI .

We prove the other way by contradiction. Suppose that there exists a WMI problem $\nu = \text{WMI}(\Delta, w) \in \text{treeWMI}(\Omega)$ with a per-literal weight function $w_{\ell} \notin \Omega^{\text{SMI}}$ such that $\rho(\nu) \in \text{treeMI}$. Since the per-literal weight function $w_{\ell} \notin \Omega^{\text{SMI}}$, from the definition of Ω^{SMI} , it holds that ℓ is a bivariate literal defined in a clause Γ which is a conjunction of more than one distinct literals, i.e., $\Gamma = \ell \vee \bigvee_{i=1}^k \ell_i$, $k \geq 1$ with $\ell \neq \ell_i, \forall i = 1, \dots, k$. During the reduction, a clause $\Gamma' = \ell \Rightarrow \bigwedge_j^n \theta_j$ is conjoined to the formula Δ to encode the weight function w_{ℓ} with at least one auxiliary variable in formula θ_j . Then there are at least three distinct variables in clause Γ' since given the form of clause Γ , clause Γ' can not be further simplified by resolution. This causes a loop in the primal graph of the reduced MI problem $\rho(\nu)$, which contradicts the assumption that $\rho(\nu) \in \text{treeMI}$. Therefore, if $\forall \nu \in \text{treeWMI}(\Omega), \rho(\nu) \in \text{treeMI}$, then $\Omega \subseteq \Omega^{\text{SMI}}$. \square

A.3. PROPOSITION 16

Proof. (Proposition 16) Recall that given a WMI problem with SMT formula Δ over real variables only, the WMI can be computed as follows by the definition of WMI in Equation 1.

$$\text{WMI}(\Delta, w) = \int_{\mathbf{x} \models \Delta} \prod_{\ell \in \text{LITS}(\Delta)} w_{\ell}(\mathbf{x})^{\llbracket \mathbf{x} \models \ell \rrbracket} d\mathbf{x}$$

Notice that this is equivalent to integrating on domain $\mathbb{R}^{|\mathbf{X}|}$ over the integrand $f(\mathbf{x}) = \llbracket \mathbf{x} \models \Delta \rrbracket \prod_{\ell \in \text{LITS}(\Delta)} w_{\ell}(\mathbf{x})^{\llbracket \mathbf{x} \models \ell \rrbracket}$. Next, we show how to factorize over the integrand $f(\mathbf{x})$ based on the factorization on formula Δ in Equation 2. First, for the indicator function, we have that

$$\llbracket \mathbf{x} \models \Delta \rrbracket = \prod_S \llbracket \mathbf{x}_S \models \Delta_S \rrbracket = \prod_S \prod_{\Gamma \in \text{CLS}(\Delta_S)} \llbracket \mathbf{x}_S \models \Gamma \rrbracket.$$

Moreover, it holds that

$$\prod_{\ell \in \text{LITS}(\Delta)} w_{\ell}(\mathbf{x})^{\llbracket \mathbf{x} \models \ell \rrbracket} = \prod_S \prod_{\Gamma \in \text{CLS}(\Delta)} \prod_{\ell \in \text{LITS}(\Gamma)} w_{\ell}(\mathbf{x}_S)^{\llbracket \mathbf{x}_S \models \ell \rrbracket}.$$

Together they complete the proof that the integrand $f(\mathbf{x})$ here equals to the unnormalized joint distribution $p(\mathbf{x})$ defined in Equation 4 and therefore the partition function of distribution $p(\mathbf{x})$ equals to the WMI of formula Δ . \square

A.4. PROPOSITION 18

Proof. (Proposition 18) This follows by induction on the message-passing scheme. Consider the base case of the messages sent by leaf nodes. When the leaf node is a variable node X_i , by definition the messages it sends to a factor node f_S is $m_{X_i \rightarrow f_S}(X_i) = 1$; when the leaf node is a factor node f_i , by definition the messages it sends to the variable node X_i is $m_{f_i \rightarrow X_i}(X_i) = f_i(X_i)$. By the definition of factor functions in Equation 3, the function f_i is a univariate piecewise function in variable X_i with pieces defined by the logical constraints in formula Δ_i as in Equation 2. Then it holds that messages sent from the leaf nodes in the message-passing scheme are piecewise function.

Further, by the recursive formulation of messages in Proposition 17, since the piecewise functions are close under product, messages sent from variable nodes to factor nodes are again univariate piecewise functions; for messages $m_{f_S \rightarrow X_i}$ sent from factor nodes f_S to variable nodes X_i , the domain of variable X_i is divided into different pieces by constraints in formula Δ_S that correspond to different integration bounds and thus the resulting messages from integration is again univariate piecewise integration. This concludes the proof. \square

A.5. PROPOSITION 19

Proof. (Proposition 19) Given the tree structure of the factor graph as well as the factorization of WMI as in Equation 4, the factors functions can be partitioned into groups, with each group associated with each factor nodes f_S that is a neighbour of the variable node X_i . Then the unnormalized joint distribution can be rewritten as follows.

$$p(\mathbf{x}) = \prod_{f_S \in \text{neigh}(X_i)} F_S(x_i, \mathbf{x}_S)$$

where \mathbf{x}_S denotes the set of all variables in the subtree connected to the variable X_i via the factor node f_S , and $F_S(x_i, \mathbf{x}_S)$ denotes the product of all the factors in the group associated with factor f_S . Then we have that

$$\begin{aligned} p(x_i) &= \prod_{f_S \in \text{neigh}(X_i)} m_{f_S \rightarrow X_i}(X_i) \\ &= \prod_{f_S \in \text{neigh}(X_i)} \int F_S(x_i, \mathbf{x}_S) d\mathbf{x}_S = \int p(\mathbf{x}) d\mathbf{x} \setminus x_i \end{aligned}$$

where the last equality is obtained by interchanging the integration and product. Thus it holds that $p(x_i)$ obtained from the product of messages to variable node X_i is the unnormalized marginal. The fact that the partition function of marginal $p(x_i)$ is the WMI of formula Δ follows Proposition 16. \square

A.6. PROPOSITION 21

Proof. (Proposition 21) W.l.o.g, assume that both the chosen root node and leaf nodes are variable nodes. Recall that the tree-height h is the longest path from root node to any leaf node. Let n_f be the number of factor nodes in the longest path in the factor graph from root node to a leaf node that defines the tree-height h . Then it holds that $h = 2n_f$ since the factor graph is a bipartite graph.

For another, consider a directed graph \mathcal{G} whose nodes are the directed factor nodes in \mathcal{F} and whose directed edges go from one factor node to factor nodes if they are visited right after in the MP-WMI. By definition, we have that $A = 2c \cdot M$ where M is the adjacency matrix of \mathcal{G} , and c is the constant that bounds the size of sub-formulas associated to factors.

For adjacency matrix M , since the power matrix M^k has non-zero entries only when there exists at least one path in graph \mathcal{G} with length k , the order of matrix M is the length of longest path in graph \mathcal{G} plus one which is two times the number of number of factor nodes in the longest path in the factor graph, i.e., $2n_f$. Therefore the adjacency matrix M is a nilpotent matrix with order being at most $2n_f$, i.e., the tree-height of the factor graph, which is at most the diameter of the factor graph. So is matrix A . \square

A.7. PROPOSITION 22

Proof. (Proposition 22) The statement (i) holds since the message $m_{X_i \rightarrow f_{ij}}$ is the product of messages hence intersection of corresponding pieces by definition in Proposition 17.

For the statement (ii), the end points of the message pieces in message $m_{f_{ij} \rightarrow X_j}$ are obtained by the solving linear equations with respect to variable x_j as described in Zeng & Van den Broeck (2019) where they define them as critical points. For these equations, each side can be either an endpoint in message $m_{X_i \rightarrow f_{ij}}$ or an \mathcal{LRA} atom from a literal in sub-formula Δ_{ij} . Then there are at most $2mc$ equations with one side as an endpoint and the other size as an \mathcal{LRA} atom, and at most c^2 equations with both sides as \mathcal{LRA} atoms. Thus the total number of critical points from solving the equations is $2mc + c^2$, which indicates that the number of pieces, whose domains are bounded intervals with critical points being their endpoints, is at most $2mc + c^2$. \square

A.8. PROPOSITION 23

Proof. (Proposition 23) The proof is done by mathematical induction at steps in MP-WMI. Given a directed factor node $f_s \in \mathcal{F}$, denote the set $\mathcal{S}(f_s) := \{f_{s'} \mid A_{f_s, f_{s'}} \neq 0\}$.

For step 0, the statement holds by the definition of $v^{(0)}$. Suppose that for step t , each entry in vector $v^{(t-1)}$ denoted by $(v^{(t-1)})_{f_s}$ bounds the number of pieces in the message $m_{X_i \rightarrow f_s}$ received by factor f_s from some variable node X_i at step $t-1$. For step t , it holds for $v^{(t)}$ by its definition that $(v^{(t)})_{f_s} = \sum_{f_{s'} \in \mathcal{S}(f_s)} (A_{f_s, f_{s'}} (v^{(t-1)})_{f_{s'}} + c^2)$.

Moreover, for a factor node $f_s \in \mathcal{F}$, there exists an variable X_i such that nodes in $\mathcal{S}(f_s)$ are connected to f_s by the variable node X_i in the factor graph. Since the entry $(v^{(t-1)})_{f_{s'}}$ bounds the number of message pieces in $m_{X_j \rightarrow f_{s'}}$ for some variable X_j , the number of message pieces in each message $m_{f_{s'} \rightarrow X_i}$ is bounded by $2c \cdot (v^{(t-1)})_{f_{s'}} + c^2$ by Proposition 22. It further indicates that the number of message pieces in $m_{X_i \rightarrow f_s}$ is bounded by $\sum_{f_{s'} \in \mathcal{S}(f_s)} (2c \cdot (v^{(t-1)})_{f_{s'}} + c^2) = (v^{(t)})_{f_s}$ since the non-zero entries in A are defined as $2c$. Thus the statement holds for step t , which finishes the induction and the proof. \square

A.9. PROPOSITION 24

Proof. (Proposition 24) For brevity, we denote the $L1$ -norm by $\|\cdot\|$. Denote the cardinality of set \mathcal{F} to be s . From the definition of matrix A , it holds that $\|A\| \leq 2cs$. Then for all t , it holds that

$$\|v^{(t)}\| \leq \|Av^{(t-1)} + c^2 \cdot \text{sgn}(Av^{(t-1)})\| \leq 2cs \|v^{(t-1)}\| + c^2 s$$

From the recurrence above, it can be obtained that

$$\begin{aligned} \left\| \sum_{t=0}^d v^{(t)} \right\| &\leq \sum_{t=0}^d \|v^{(t)}\| \\ &\leq \sum_{t=0}^d [(2cs)^t \|v^{(0)}\| + \sum_{i=0}^{t-1} (2cs)^i cs] \leq 2(2cs)^{2d+2} \end{aligned}$$

Moreover, since the cardinality $s \leq 2n$, we have that $\left\| \sum_{t=0}^d v^{(t)} \right\|$ is of $\mathcal{O}((4nc)^{2d+2})$. \square