Dissociation-Based Oblivious Bounds for Weighted Model Counting

Li Chou
The University of Texas at Dallas
Richardson, TX 75080

Wolfgang Gatterbauer
Northeastern University
Boston, MA 02115

Vibhav Gogate
The University of Texas at Dallas
Richardson, TX 75080

Abstract

We consider the weighted model counting task which includes important tasks in graphical models, such as computing the partition function and probability of evidence as special cases. We propose a novel partition-based bounding algorithm that exploits logical structure and gives rise to a set of inequalities from which upper (or lower) bounds can be derived efficiently. The bounds come with optimality guarantees under certain conditions and are oblivious in that they require only limited observations of the structure and parameters of the problem. We experimentally compare our bounds with the mini-bucket scheme (which is also oblivious) and show that our new bounds are often superior and never worse on a wide variety of benchmark networks.

1 INTRODUCTION

Logic and probability theory are formalisms employed for the task of automated reasoning. Logic facilitates deterministic representations and decisions, while probability theory accommodates situations where uncertainty arises. Propositional logic (Boolean satisfiability) is a prominent construct for performing deductive reasoning, particularly within a combinatorial setting. Extensive research efforts have resulted in state-of-the-art satisfiability solvers that have been successfully deployed in fields such as software/hardware model checking, planning and cybersecurity (Zhang and Malik, 2002). Graphical models (GM) have emerged as an effective scheme for modeling uncertainty. For example, Bayesian networks (Pearl, 1988) have been used in medical domains, while Markov networks are widely used in areas such as computer vision and natural language processing. However, in order to effectively model problems in real-world domains, it is of great practical interest to solve the harder problem of developing models with the capacity to account for knowledge that is both deterministic and uncertain in an unified manner.

Propositional model counting is the generalization of the Boolean satisfiability problem. Extending the task of determining satisfiability, the objective is to count the number of distinct instances that result in satisfiability. This is also referred to as solution counting. Counting is a fundamental aspect to probabilistic computations (sum inference) and thus propositional model counting provides an intuitive connection between logic and uncertainty. In this paper, we address a further extension, namely the problem of *weighted model counting* (WMC). WMC allows for additional probabilistic interpretations of the variables in the model by associating a weight function either at the variable level or the clause level (Chavira and Darwiche, 2008; Gogate and Domingos, 2010; Sang et al., 2005).

It is well known that probabilistic inference in GM can be reduced to WMC (Chavira and Darwiche, 2008). The reduction has two main components: (1) encode the GM as a propositional knowledge base; and (2) leverage state-of-the-art propositional model counters to develop a WMC-based algorithm for solving the desired inference task. However, a major drawback of the aforementioned methods is that they are computationally intractable for most real-world problems. Therefore, developing fast, scalable approximate schemes is a subject of fundamental interest.

While there exists several approaches to propositional *approximate counting*, most of those are intrinsically stochastic (Ermon et al., 2013; Gogate and Dechter, 2007, 2011), and little attention has been given to deterministic methods that can bound estimates with correctness guarantees. In this paper we propose a *deterministic bounding scheme for WMC*. Our approach is partition-based (Dechter and Rish, 2003) and gives rise to a novel class of inequalities from which upper (or lower) bounds can be derived efficiently. In addition, the bounds are oblivious, i.e. they require only limited observations of the structure and parameters of the problem, which yields fast methods.

Specifically, we extend the work of Gatterbauer and Suciu (Gatterbauer and Suciu, 2014, 2017), which is applicable to only monotone SAT formulas, to the task of WMC for arbitrary (*non-monotone*) formulas. Our method is related to the class of bounded complexity inference schemes such as mini-buckets (MB) (Dechter and Rish, 2003) and their extensions (Choi et al., 2007; Liu, 2014). MB relaxes the original problem by decomposing it into local sub-problems (by splitting/dissociating nodes) that are then solved exactly. The result is an approximate scheme that generates bounds for various inference tasks.

We make the following contributions. (1) We analyze the idea of dissociation based oblivious bounds (Gatterbauer and Suciu, 2014) using the framework of weighted model counting and extend it to the general non-monotone case; (2) we take advantage of logical structure and derive a novel set of inequalities for bounding methods that dissociate until the formula has a tree structure (namely the i-bound in MB is equal to 1); (3) we theoretically compare the idea of dissociation with MB and show that MB bounds are a special case of our bounds and can be quite inferior; and (4) we empirically demonstrate that dissociation based bounds are more accurate than MB on several synthetic and real-world datasets.

2 BACKGROUND

Let $\mathbf{X}, \mathbf{Y}$, etc. be sets of propositional variables that take values (i.e., truth assignments) from the set $\{$*false*, *true*$\}$ (or $\{0, 1\}$). Given $\mathbf{X} = \{X_1, \ldots, X_n\}$, let Ω be the set of the 2^n truth assignments to $\mathbf{X}$. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \Omega$ be a truth assignment to all variables in $\mathbf{X}$ s.t. $X_i = x_i$. We use the symbol '$*$' to denote the case when X_i can take either values, namely $(0 \vee 1)$ or otherwise known as the *don't care* condition. Let F be a propositional formula in conjunctive normal form (CNF) over $\mathbf{X}$, i.e. F is a conjunction of clauses, where a clause is a disjunction of literals, and each literal is defined as a variable X_i (positive literal, $+$) or its negation $\overline{X_i}$ (negative literal, $-$). Let $\mathbf{C}$ be the set of clauses of F. In this paper, we will focus on arbitrary (non-monotone) CNF.

Definition 1. (Monotonicity). *A formula F is "monotone in variable X_i" iff X_i appears in F as either positive or negative (but not both). A formula F is "monotone" iff it is monotone in all variables. Otherwise F is non-monotone.*

2.1 WEIGHTED MODEL COUNTING

Given a propositional formula F, a satisfying assignment or *model* of F is a truth assignment to all variables in F such that F evaluates to *true* ($\mathbf{x} \models F$). The problem of determining if there exists a satisfying assignment $\mathbf{x}$ for F is called the *Boolean satisfiability problem* or SAT. *Propositional model counting* or #SAT is the task of computing the number of models of F. This is the canonical #P-complete problem that generalizes SAT.

Weighted model counting (WMC) (Chavira and Darwiche, 2008; Sang et al., 2005) extends model counting by associating the following probability distribution (weight function) ϕ_i to each propositional variable X_i:

$$\phi_i(X_i) = \begin{cases} p_i & \text{if } X_i \text{ evaluates to } 1 \\ \overline{p_i} & \text{otherwise} \end{cases},$$

where $p_i \in [0, 1]$ and $\overline{p_i} \triangleq 1 - p_i$.[1] The functions ϕ_i yield a *weighted representation* $\mathcal{F}$ of the CNF F and is called WCNF. Formally, $\mathcal{F}$ is a triple $\langle \mathbf{X}, \Phi, \mathbf{C} \rangle$, where $\mathbf{X}$ is a set of n Boolean variables in F, Φ is a set of weight functions ϕ_i associated with each Boolean variable $X_i \in \mathbf{X}$ and $\mathbf{C}$ is a set of clauses of F. $\mathcal{F}$ represents the following probability distribution

$$P_{\mathcal{F}}(\mathbf{x}) = \begin{cases} \frac{1}{Z_{\mathcal{F}}} \prod_{i=1}^{n} \phi_i(X_i = x_i) & \text{if } \mathbf{x} \models F \\ 0 & \text{otherwise} \end{cases},$$

where $Z_{\mathcal{F}}$ is the *partition function*, also referred to as the *weighted model count* (WMC) of $\mathcal{F}$, and is given by

$$Z_{\mathcal{F}} = \sum_{(\mathbf{x} \in \Omega \,\wedge\, \mathbf{x} \models F)} \prod_{i=1}^{n} \phi_i(X_i = x_i).$$

When $p_i = 1/2$ for all variables, the product $2^n Z_{\mathcal{F}}$ equals the special case of (unweighted) model count of F.

2.2 GRAPHICAL MODELS

Graphical models (GM) provide a compact representation of joint probability distributions over a set of variables $\mathbf{X}$. For simplicity, we will focus on pairwise binary Markov networks since every GM can be converted to this form (cf. (Koller and Friedman, 2009)). Let $\mathcal{I} \subseteq \mathcal{A}$ where $\mathcal{A}$ denotes the set of all pairs (i, j) such that $i < j$ and $1 \leq i, j \leq n$. In a pairwise graphical model, we associate a potential function $\psi_{i,j}$ over each pair $(i, j) \in \mathcal{I}$. The probability distribution is given by

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{(i,j) \in \mathcal{I}} \psi_{i,j}(x_i, x_j),$$

where Z is the normalization constant (partition function) and (x_i, x_j) is the projection of $\mathbf{x}$ on $\{X_i, X_j\}$.

[1]WMC is typically defined by attaching weights to literals, and the corresponding potential function over each variable is constructed by exponentiating the weights. We consider an equivalent representation in which the potential function is normalized to yield a probability distribution.

Table 1: Clauses for W2CNF Encoding of a GM

$(\overline{X_i} \vee Y_{i,j,1})$	$(\overline{X_j} \vee Y_{i,j,1})$
$(\overline{X_i} \vee Y_{i,j,2})$	$(X_j \vee Y_{i,j,2})$
$(X_i \vee Y_{i,j,3})$	$(\overline{X_j} \vee Y_{i,j,3})$
$(X_i \vee Y_{i,j,4})$	$(X_j \vee Y_{i,j,4})$

2.3 WCNF ENCODING OF A GM

We describe here a possible translation from GM to WCNF. For more details see (Chavira and Darwiche, 2008; Gogate and Domingos, 2010, 2016). Since we focus on pairwise binary GMs, we can convert them to WCNFs in which each clause has at most two literals. We will refer to such WCNFs as W2CNF.

Given a GM, we can construct an equivalent W2CNF as follows. We start with a W2CNF $\mathcal{F}$ defined over the variables $\mathbf{X}$ of the GM such that the set of clauses $\mathbf{C}$ of $\mathcal{F}$ is empty and $p_{X_i} = 0.5$ for each variable $X_i \in \mathbf{X}$. Then, for each pairwise binary potential $\psi_{i,j}$ in the GM such that $\psi_{ij} : X_i = 0, X_j = 0 \rightarrow w_{i,j,1}$, $\psi_{ij} : X_i = 0, X_j = 1 \rightarrow w_{i,j,2}$, $\psi_{ij} : X_i = 1, X_j = 0 \rightarrow w_{i,j,3}$, $\psi_{ij} : X_i = 1, X_j = 1 \rightarrow w_{i,j,4}$, we add a variable for each weight to $\mathcal{F}$. We will denote the variables associated with $w_{i,j,1}$, $w_{i,j,2}$, $w_{i,j,3}$ and $w_{i,j,4}$ by $Y_{i,j,1}$, $Y_{i,j,2}$, $Y_{i,j,3}$ and $Y_{i,j,4}$ respectively. Utilizing these weight variables, we add the the clauses given in Table 1 to $\mathbf{C}$ for $k = 1, \ldots, 4$. We also add the following probability distribution for each variable $Y_{i,j,k}$

$$\phi(y_{i,j,k}) = \begin{cases} \frac{w_{i,j,k}-1}{w_{i,j,k}} & \text{if } y_{i,j,k} \text{ is } \textit{false} \text{ or } 0 \\ \frac{1}{w_{i,j,k}} & \text{otherwise} \end{cases} .$$

Note that when $w_{i,j,k} < 1$, $\phi(y_{i,j,k})$ will be negative. To avoid this condition, we can easily rescale the potentials of the GM by multiplying them with an appropriate constant. Also, zero weights can be handled by adding the corresponding negated assignment as a clause to $\mathbf{C}$. For example, if $w_{i,j,1} = 0$, we add the clause $X_i \vee X_j$ to $\mathbf{C}$. Using previous work (Chavira and Darwiche, 2008; Gogate and Domingos, 2010), it is straight-forward to show that:

Proposition 2. W2CNF *output by Encoding 1 represents the same probability distribution over* $\mathbf{X}$ *as the input GM.*

2.4 MINI-BUCKET ELIMINATION

We can utilize inference algorithms such as bucket or variable elimination (Dechter, 1996; Zhang and Poole, 1994) to compute the weighted model count of a W2CNF. However, since the complexity of using such algorithms is in general exponential in the treewidth, a more practical approach is to approximate the task by introducing relaxations techniques that control model complexity (i.e., the induced width given a fixed elimination order). Mini-bucket (MB) (Dechter and Rish, 2003) is one such approximate scheme that builds on bucket elimination (BE) for generating upper (or lower) bounds on the partition function or weighted model count. We will use the following running example to illustrate BE and MB for WMC.

Example 3. *Consider the* W2CNF $\mathcal{F}$ *such that* $\mathbf{X} = \{X_1, Y_2, Y_3\}$, $\mathbf{C} = \{(X_1 \vee Y_2), (X_1 \vee Y_3)\}$ *and* $\Phi = \{\phi_1, \phi_2, \phi_3\}$. *For simplicity we denote* ϕ_1 *for* $\phi_1(X_1)$, *etc. We can convert the clauses and potentials of* $\mathcal{F}$ *to the following two potentials yielding a more convenient form for BE.*

X_1	Y_2	$\psi_{12}(X_1, Y_2)$
0	0	0
0	1	$\overline{p_1}p_2$
1	0	$p_1\overline{p_2}$
1	1	p_1p_2

X_1	Y_3	$\psi_{13}(X_1, Y_3)$
0	0	0
0	1	p_3
1	0	$\overline{p_3}$
1	1	p_3

Without loss of generality, we assume the elimination ordering as $[X_1, Y_2, Y_3]$ *(although it is clearly not optimal, it will help us illustrate the main ideas). BE begins by creating* $|\mathbf{X}|$ *number of buckets and groups the functions by placing each function involving some variable* X_i *(or* Y_i *in our example) in a bucket* $\mathbf{B}_{X_i}$ *according to the position of* X_i *in the ordering. The resulting computation is* $Z_{\mathcal{F}}^{\text{BE}} = \sum_{Y_3}\sum_{Y_2}\sum_{X_1}\psi_{12}\psi_{13}$ *where* $\mathbf{B}_{X_1} = \{\psi_{12}, \psi_{13}\}$ *is first processed by taking the product of the two potentials and summing out variable* X_1. *The resulting new potential* ψ_{23}' *is placed in bucket* $\mathbf{B}_{Y_2}$ *in which variable* Y_2 *is summed out. Summing out* Y_3 *from the subsequent function* ψ_3' *yields* $Z_{\mathcal{F}}^{\text{BE}}$.

MB follows similarly. However, MB partitions each bucket into two or more so called mini-buckets according to an input parameter called the i-bound, which defines the maximum number (i-bound + 1) of variables in each mini-bucket. The mini-buckets are then processed independently. To obtain an upper bound, the sum-product operation is performed on one of the mini-buckets and the max-product for the remaining (min-product for lower bound). Using i-bound = 1, $\mathbf{B}_{X_1}$ *is split into two mini-buckets* $\mathbf{B}'_{X_1} = \{\psi_{12}\}$ *and* $\mathbf{B}''_{X_1} = \{\psi_{13}\}$. *One possible resulting computation is*

$$\underbrace{\sum_{Y_3 Y_2}\left(\sum_{X_1}\psi_{12}\right)\left(\min_{X_1}\psi_{13}\right)}_{Z_{\mathcal{F}}^{\mathrm{MB}(L)}} \leq \sum_{Y_3}\sum_{Y_2}\sum_{X_1}\psi_{12}\psi_{13}$$

$$\leq \underbrace{\sum_{Y_3 Y_2}\left(\sum_{X_1}\psi_{12}\right)\left(\max_{X_1}\psi_{13}\right)}_{Z_{\mathcal{F}}^{\mathrm{MB}(U)}},$$

where the MB upper bound on the partition function,

$Z_{\mathcal{F}}^{\text{MB}(U)}$*, is computed by maxing out* X_1 *from* ψ_{13} *independently from summing out* X_1 *from* ψ_{12}*. Summing out* Y_2 *and* Y_3 *from the resulting two new potentials,* ψ'_2, ψ'_3*, and taking their product gives the upper bound. The lower bound,* $Z_{\mathcal{F}}^{\text{MB}(L)}$*, is computed similarly using min instead of max.*

MB is a fast and simple algorithm for computing upper (or lower) bounds. The resulting complexity of inference is exponential in the *i-bound*. Lower *i-bound* values translates to simpler models and provides the trade-off between complexity and accuracy.

Next, we present the idea of dissociation based oblivious bounds for the case of monotone W2CNF and extend it to the non-monotone case by exploiting logical structure in Section 4. As mentioned earlier, in this paper, we focus on the case where variables are dissociated until the resulting formula is a tree. In other words, our scheme is comparable to the case when the *i-bound* in MB equals 1.

3 DISSOCIATION

Our task is to compute the WMC $Z_{\mathcal{F}}$ of a given WCNF $\mathcal{F}$. Since the problem is computationally intractable in general (e.g., high treewidth), approximate methods are required. In this paper we use a *bounded inference* approach, where we approximate the original $\mathcal{F}$ with $\mathcal{F}'$ from which the upper (or lower) bounds on $Z_{\mathcal{F}}$ can be computed efficiently. We build upon (Gatterbauer and Suciu, 2014, 2017) which presents a bounding scheme called *dissociation* that can be applied to WMC. The derived bounds are *oblivious* to the set of weight functions ϕ_i, i.e. they can be calculated by only observing a limited subset of clauses. However, these bounds only apply to monotone formulas, whereas we are interested in extending the underlying ideas to more general non-monotone formulas (Section 4). Here, we first give a general intuition of prior results followed by the formal definition and then present optimal oblivious bounds for monotone formulas.

At a high level, dissociation is the process of replacing an existing variable X_i in $\mathcal{F}$ with new variables $X_{i;1}, \ldots, X_{i;d}$ and assigning them new probability distributions. The technique is closely related to *variable or node splitting* (Choi et al., 2007) in which the new variables are referred to as clones. The partitioning of mini-buckets can also be classified under the general notion of variable splitting.

By creating new variables, we are implicitly ignoring (or relaxing) a set of equality constraints (Choi and Darwiche, 2009). However, we can recover the set by defining and incorporating the function $\varphi(X_{i;1} = x_{i;1}, \ldots, X_{i;n} = x_{i;d})$ which evaluates to 1 iff $x_{i;1} = \ldots = x_{i;d}$, and 0 otherwise, for the d copies $X_{i;j}, j \in [d]$ of variable X_i, and $x_{i;j}$ being the corresponding truth assignment. We can also incorporate equivalence clauses for each new pair of variables into a formula with the new clauses. For example, consider the formula $F = (X_1 \vee Y_1)(X_1 \vee Y_2)$. We can create the equivalent formula $F' = (X_{1;1} \vee Y_1)(X_{1;2} \vee Y_2)(X_{1;1} \Leftrightarrow X_{1;2})$ using copies of X_1 for the unweighted model counting case. We see two issues arising. First, for general 2-CNF formulas, we will require $d-1$ equality constraints (equivalence ($\Leftrightarrow$) formulas). Second, it is not immediately clear on how to integrate the weight functions so that weighted model counts can be computed using this scheme.

Dissociation expands on the notion of variable duplication and provides an *algebraic framework* to analyze and approximate the aforementioned set of equality constraints. The result is a novel class of inequalities to construct upper (or lower) bounds on the WMC. We first give the formal definition of dissociation for W2CNF.

Definition 4. (Dissociation). *Let* $\mathcal{F} = \langle \mathbf{X}, \Phi, \mathbf{C} \rangle$*. Select a variable* $X_i \in \mathbf{X}$ *and let* $C(X_i) \subseteq \mathbf{C}$ *be the subset of all clauses that involve variable* X_i*. We say* $\mathcal{F}' = \langle \mathbf{X}', \Phi', \mathbf{C}' \rangle$ *is a dissociation of* $\mathcal{F}$ *on* X_i *iff*

- $\mathbf{X}' = \mathbf{X} \setminus X_i \cup X_{i;1} \cup \cdots \cup X_{i;d}$ *with* $d \leq |C(X_i)|$*,*
- $\Phi' = \Phi \setminus \phi_i \cup \phi_{i;1} \cup \cdots \cup \phi_{i;d}$*, and*
- $\mathbf{C}'[\theta_{X_i}(\mathbf{X}')] = \mathbf{C}[\mathbf{X}]$ *with* θ_{X_i} *being the substitution* $\theta_{X_i}[\{(X_{i;j}/X_i), j \in [d]\}]$.

We say a dissociation is full if $d = |C(X_i)|$.

Example 5. (Dissociation). *Consider* $\mathcal{F}$ *from example 3. Dissociating* X_1 *results in adding two new variables,* $\mathbf{X}' = \mathbf{X} \setminus X_1 \cup X_{1;1} \cup X_{1;2}$*, and two new associated weight functions* $\Phi' = \Phi \setminus \phi_1 \cup \phi_{1;1} \cup \phi_{1;2}$*. Applying the substitution* $\theta_{X_1}[(X_{1;1}/X_1), (X_{1;2}/X_1)]$ *on* $C(X_i)$ *results in* $\mathbf{C}' = \mathbf{C} \setminus C(X_i) \cup (X_{1;1} \vee Y_2) \cup (X_{1;2} \vee Y_3)$*.*

Once we have defined the new weight functions (for dissociated variables), the question we are interested in is how to parameterize the new functions in order to obtain guaranteed upper (or lower) bounds. In particular, we are interested in *oblivious bounds*, i.e. when the *new probabilities are chosen independently of the probabilities of all other variables*. We achieve that by considering all possible valuations (or truth assignments) of the non-dissociated variables, The assignments give rise to a set of inequalities which are then evaluated to develop necessary and sufficient conditions for upper (or lower) bounds. We next illustrate with an example.

Example 6. (Oblivious bounds). *Consider the two sets of clauses,* $\{(X_1 \vee Y_2), (X_1 \vee Y_3)\}$ *and* $\{(X_{1;1} \vee Y_2), (X_{1;2} \vee Y_3)\}$ *from examples 3 and 5. We analyze the* $2^2 = 4$ *possible truth assignments to the non-dissociated variables*

Table 2: Dissociation valuation analysis (example 6).

Y_2	Y_3	X_1	$X_{1;1}$	$X_{1;2}$	ϕ_1	$\phi_{1;1}, \phi_{1;2}$
0	0	1	1	1	p_1	$p_{1;1}p_{1;2}$
0	1	1	1	$*$	p_1	$p_{1;1}$
1	0	1	$*$	1	p_1	$p_{1;2}$
1	1	$*$	$*$	$*$	1	1

Y_2 and Y_3. Table 2 shows each possible valuation of Y_2 and Y_3 and the corresponding assignments to $X_1, X_{1;1}$ and $X_{1;2}$ required to satisfy the clauses. We also show the weights (probabilities) of the original (column ϕ_1) and dissociated formulas (column $\phi_{1;1}\phi_{1;2}$).

As example, consider the assignment, $Y_2 = 0 \wedge Y_3 = 1$: The assignment $X_1 = 1$ is required to satisfy $\mathcal{F}$, resulting in the term $p_1\overline{p_2}p_3$. The assignments $(X_{1;1} = 1 \wedge X_{1;2} = 0)$ or $(X_{1;1} = 1 \wedge X_{1;2} = 1)$ are required to satisfy $\mathcal{F}'$, i.e. $X_{1;2}$ can take any assignment $(*)$, resulting in the term $p_{1;1}\overline{p_2}p_3$. Utilizing the two terms, simplifying by removing the common terms $(\overline{p_2}p_3)$ and assuming that we are interested in computing lower bounds, we create the inequality $p_1 \geq p_{1;1}$. Repeating the same analysis for the three remaining cases results in the inequalities $p_1 \geq p_{1;1}p_{1;2}$ and $p_1 \geq p_{1;2}$. The last case $1 \geq 1$ is trivially satisfied. Combining the resulting inequalities, and doing a similar analysis for computing upper bounds (where we replace $\geq$ by $\leq$) gives rise to the following conditions for oblivious (U)pper and (L)ower bounds:

- *U*: $(p_1 \leq p_{1;1}p_{1;2}) \wedge (p_1 \leq p_{1;1}) \wedge (p_1 \leq p_{1;2})$.
- *L*: $(p_1 \geq p_{1;1}p_{1;2}) \wedge (p_1 \geq p_{1;1}) \wedge (p_1 \geq p_{1;2})$.

Notice the valuation process creates $2^{|C(X_i)|}$ inequalities, one for each truth assignment. However, we can simplify the conditions by removing subsumed inequalities.

Definition 7. (Subsumed inequality). *We say an inequality I_i subsumes inequality I_j $(i \neq j)$ iff $I_i \Rightarrow I_j$, i.e. satisfying I_i also satisfies I_j.*

Example 8. *Consider the upper and lower bound conditions in example 6. For the upper bound, clearly $p_1 \leq p_{1;1}p_{1;2}$ subsumes the remaining inequalities since $\forall p_1, p_{1;1}, p_{1;2} \in [0,1] \; : \; (p_1 \leq p_{1;1}p_{1;2}) \Rightarrow (p_1 \leq p_{1;1}) \wedge (p_1 \leq p_{1;2})$. For the lower bound, clearly $(p_1 \geq p_{1;1}) \wedge (p_1 \geq p_{1;2})$ subsumes the remaining inequality since $\forall p_1, p_{1;1}, p_{1;2} \in [0,1] : ((p_1 \geq p_{1;1}) \wedge (p_1 \geq p_{1;2})) \Rightarrow (p_1 \geq p_{1;1}p_{1;2})$. Therefore, we can reduce the required conditions for the oblivious bounds to:*

- *U*: $p_1 \leq p_{1;1}p_{1;2}$.
- *L*: $(p_1 \geq p_{1;1}) \wedge (p_1 \geq p_{1;2})$.

Following the preceding analysis, we can now state the conditions for oblivious bounds for monotone W2CNF.

Theorem 9. (Gatterbauer and Suciu, 2014) (Oblivious bounds for monotone W2CNF). *Let $\mathcal{F}$ be a monotone W2CNF. Let $\mathcal{F}'$ be the result of applying a series of dissociation steps on $\mathcal{F}$. For every set of weight functions defined for a dissociate variable, namely $X_{i;1}, \dots, X_{i;d}$ and $\{\phi_{i;1}, \dots, \phi_{i;d}\}$ with $d > 1$, we have the following* oblivious bounds*:*

- *U*: $\prod_{j=1}^{d} p_{i;j} \geq p_i$.
- *L*: $\forall j : p_{i;j} \leq p_i$.

Optimal oblivious bounds are defined as those that are not dominated, i.e. they cannot be improved without knowledge of the probabilities of all other variables. They are obtained by replacing inequality with equality. Notice that optimal oblivious lower bounds are uniquely defined, $\forall j : p_{i;j} = p_i$, whereas there are infinitely many optimal oblivious upper bounds, e.g. symmetric ones: $\forall j : p_{i;j} = \sqrt[d]{p_i}$, and finding the best one requires access to all other probabilities (den Heuvel et al., 2018).

Note that optimal oblivious bounds are different from augmented mini-buckets (AMB) (Liu, 2014). For example, in AMB for computing upper bounds, the potential over each dissociated variable is initialized to $\phi_{i;j}(X_{i;j} = 1) = \phi_i(X_i = 1)^{1/d}$ and $\phi_{i;j}(X_{i;j} = 0) = \phi_i(X_i = 0)^{1/d}$ where we have d dissociations. A better initialization would be $\phi_{i;j}(X_{i;j} = 1) = \phi_i(X_i = 1)^{1/d}$, and $\phi_{i;j}(X_{i;j} = 0) = 1 - \phi_i(X_{i;j} = 1)^{1/d}$.

3.1 COMPARISON WITH MINI-BUCKET

We use example 3 to analyze the base case bounds for monotone dissociation (X_1 to $X_{1;1}$ and $X_{1;2}$) and compare it with MB (*i-bound* = 1).

Lower bound. Dissociation results in the partition function $Z_{\mathcal{F}'}^{\text{DIS}(U)} = p_2p_3 + p_{1;1}\overline{p_2}p_3 + p_{1;2}p_2\overline{p_3} + p_{1;1}p_{1;2}\overline{p_2}\overline{p_3}$. The two possible partition functions according to MB are (1) $\sum_{X_1}\psi_1 \min_{X_1}\psi_2 \Rightarrow Z_{\mathcal{F}}^{\text{MB}(L1)} = p_2p_3 + p_1\overline{p_2}p_3$; (2) $\min_{X_1}\psi_1 \sum_{X_1}\psi_2 \Rightarrow Z_{\mathcal{F}}^{\text{MB}(L2)} = p_2\min(\overline{p_1}, p_1)(1 + p_3)$. Clearly, $Z_{\mathcal{F}'}^{\text{DIS}(L)} \geq Z_{\mathcal{F}}^{\text{MB}(L)}$ $\forall p_1, p_{1;1}, p_{1;2}, p_2, p_3 \in [0,1]$.

Upper bound. Notice there exist an infinite number of settings to $p_{1;1}$ and $p_{1;2}$ that satisfy $p_{1;1}p_{1;2} = p_1$ under dissociation. We analyze two possible cases. (1) $(p_{1;1} = p_1) \wedge (p_{1;2} = 1) \Rightarrow Z_{\mathcal{F}'}^{\text{DIS}(U1)} = p_{1;1} + \overline{p_{1;1}}p_2$; (2) $(p_{1;2} = p_1) \wedge (p_{1;1} = 1) \Rightarrow Z_{\mathcal{F}'}^{\text{DIS}(U2)} = p_{1;2} + \overline{p_{1;2}}p_3$. The two possible partition functions according to MB are (1) $\sum_{X_1}\psi_1 \max_{X_1}\psi_2 \Rightarrow Z_{\mathcal{F}}^{\text{MB}(U1)} = p_1 + \overline{p_1}p_1$; (2) $\max_{X_1}\psi_1 \sum_{X_1}\psi_2 \Rightarrow Z_{\mathcal{F}}^{\text{MB}(U2)} = (1 + p_3)(p_2\max(\overline{p_1}, p_1) + p_1\overline{p_1})$. We first observe the bounds are equivalent between dissociation and MB in setting

(1) and also for (2) if the functions are unweighted (e.g., $\forall i\ p_i = 1/2$). However, note there exist more degrees of freedom (solutions) for dissociation, and this example simply demonstrates one such setting for which we observe equivalency under certain conditions.

4 DISSOCIATION FOR NON-MONOTONE FORMULAS

In this section, we extend dissociation bounds from the monotone case to arbitrary non-monotone W2CNFs. Unlike monotone W2CNFs, we can apply logical inference techniques such as resolution and unit propagation to reduce non-monotone W2CNFs which in turn may improve our dissociation-based bounds. Moreover, logical propagation can be applied as a pre-processing step before dissociating a variable X_i.

4.1 PREPROCESSING

We say that a W2CNF $\mathcal{F}$ is *minimal* if the following steps are applied to its set of clauses $\mathbf{C}$ until convergence.

1. **(Binary) Resolution**: If $\mathbf{C}$ contains two clauses of the form $L_i \vee L_j$ and $\overline{L}_i \vee L_k$, where L_i, L_j and L_k are literals of variables X_i, X_j and X_k respectively, we add the clause $L_j \vee L_k$ to $\mathbf{C}$.
2. **Unit Resolution**: If $\mathbf{C}$ contains two clauses of the form $L_i \vee L_j$ and $\overline{L}_i \vee L_j$, where L_i and L_j are literals of variables X_i and X_j respectively, we add the unit clause L_j to $\mathbf{C}$.
3. **Clause Deletion and Reduction**: If $\mathbf{C}$ contains a unit clause L_i where L_i is a literal of X_i then we delete all clauses of the form $L_i \vee L_j$ and remove $\overline{L}_i$ from all clauses that mention $\overline{L}_i$. If $\mathbf{C}$ contains both unit clauses L_i and $\overline{L_i}$, $\mathbf{C}$ is inconsistent and we return a lower/upper bound of 0.[2]

Example 10 (Minimal formula)**.** *Consider* $\mathbf{C} = \{(X_1 \vee \overline{X_2}), (X_1 \vee X_2), (\overline{X_2} \vee Y_4), (X_1 \vee X_3), (X_3 \vee X_5)\}$. $\mathbf{C}$ *is not minimal and we can make it minimal using the aforementioned steps. After applying Unit Resolution on the first two clauses, we get* $\mathbf{C} = \{(X_1), (X_1 \vee \overline{X_2}), (X_1 \vee X_2), (\overline{X_2} \vee Y_4), (X_1 \vee X_3), (X_3 \vee X_5)\}$. *After applying Clause deletion and Reduction, we get* $\mathbf{C} = \{(X_1), (\overline{X_2} \vee Y_4), (X_3 \vee X_5)\}$, *which is minimal.*

4.2 TYPES OF NON-MONOTONE FORMULAS

In the sequel, we assume that the input W2CNF $\mathcal{F}$ to our algorithm is minimal. To formulate oblivious bounds for non-monotone W2CNF, we first establish a *canonical representation* that helps us take advantage of symmetry and reduces the number of cases (inequalities) we need to consider for our proposed oblivious bounds. Specifically, given a candidate *dissociation variable* X_i, we convert the set of clauses $\mathbf{C}$ into a canonical representation:

Definition 11 (Canonical representation)**.** *We say that $\mathcal{F}$ is canonical w.r.t. a variable X_i if $\mathcal{F}$ is minimal and all clauses in $C(X_i)$ satisfy the following two properties:*

1. *If a variable Y_j appears only once in $C(X_i)$ then it only appears positively, i.e. it appears in clauses of the form $X_i \vee Y_j$ or $\overline{X}_i \vee Y_j$ (but not of the form $X_i \vee \overline{Y}_j$ or $\overline{X}_i \vee \overline{Y}_j$).*
2. *If a variable Y_j appears twice in $C(X_i)$, then it appears in the following two clauses $X_i \vee Y_j$ and $\overline{X}_i \vee \overline{Y}_j$ (but not in the clauses $\overline{X}_i \vee Y_j$ and $X_i \vee \overline{Y}_j$).*

Note that since $\mathcal{F}$ is minimal, Y_j cannot appear more than twice in $C(X_i)$, nor twice with the same sign. If $C(X_i)$ is not in canonical form, we can easily make it canonical by using the following procedure:

- If Y_j violates either condition (1) or (2) in definition 11, then replace Y_j by a new variable Y_k in all clauses of $\mathcal{F}$ (where Y_j appears) such that $Y_k = \overline{Y}_j$, and set $\phi(Y_k) = \phi(\overline{Y_j})$ and $\phi(\overline{Y_k}) = \phi(Y_j)$.

Example 12 (Canonical representation)**.** *Consider* $\mathbf{C} = \{(X_1 \vee \overline{Y_2}), (\overline{X_1} \vee Y_2), (X_1 \vee \overline{Y_3})\}$. $\mathbf{C}$ *is not in canonical form w.r.t. X_1 because Y_2 and Y_3 violate the second and first property respectively in definition 11. To convert it to canonical form, set $Y_4 = \overline{Y_2}$, $Y_5 = \overline{Y_3}$, $\phi(Y_4) = \phi(\overline{Y_2})$, $\phi(\overline{Y_4}) = \phi(Y_2)$, $\phi(Y_5) = \phi(\overline{Y_3})$ and $\phi(\overline{Y_5}) = \phi(Y_3)$. Thus, the canonical representation of* $\mathbf{C}$ *is the set* $\{(X_1 \vee Y_4), (\overline{X_1} \vee \overline{Y_4}), (X_1 \vee Y_5)\}$.

We call variables Y_j which appear only once in $C(X_i)$ *single-occurrence neighbors* of X_i and those which appear twice *two-occurrence neighbors*.

4.3 CHARACTERIZING OBLIVIOUS BOUNDS

We now derive oblivious bounds based on whether $C(X_i)$ has two-occurrence neighbors or not. In the following, let $\mathcal{F}$ denote a W2CNF that is canonical w.r.t. X_i and let $\mathcal{F}'$ be the result of applying a series of dissociation steps on $\mathcal{F}$. Let Y_j be a single-occurrence neighbor of X_i. Let S^+ and S^- denote the set of indices of the dissociated variables that appear in clauses $(X_i \vee Y_j)$ and $(\overline{X}_i \vee Y_j)$ respectively in $C(X_i)$. Let Y_k be a two-occurrence neighbor of X_i. Let T^+ and T^- denote the set of indices of the dissociated variables in clauses $X_i \vee Y_k$ and $\overline{X}_i \vee \overline{Y}_k$ respectively in $C(X_i)$. (We use S and T to refer to "single-occurrence" and "two-occurrence" variables, respectively.)

Example 13 (Indices)**.** *Consider* $\mathbf{C} = \{(X_1 \vee Y_5), (X_1 \vee Y_8), (\overline{X_1} \vee Y_6), (X_1 \vee Y_7), (\overline{X_1} \vee \overline{Y_7}), (X_1 \vee$

[2] Note that our scheme will return an upper bound of 0 only when $\mathbf{C}$ is inconsistent.

$Y_9), (\overline{X_1} \vee \overline{Y_9})\}$. *After applying dissociation on* X_1, *we get* $C(X'_1) = \{(X_{1;1} \vee Y_5), (X_{1;2} \vee Y_8), (\overline{X_{1;3}} \vee Y_6), (X_{1;4} \vee Y_7), (\overline{X_{1;5}} \vee \overline{Y_7}), (X_{1;6} \vee Y_9), (\overline{X_{1;7}} \vee \overline{Y_9})\}$. *Then* $S^+ = \{1, 2\}, S^- = \{3\}, T^+ = \{4, 6\}$, *and* $T^- = \{5, 7\}$.

We next analyze the two possible non-monotone cases in Theorems 14 and 16. The proofs are presented in an extended version of the paper.

The first case is when $C(X_i)$ has only single-occurrence neighbors (but no two-occurrence neighbors). This generalizes the monotone case, in which only one type of single-occurrence variables are present. In particular, in the monotone case either clauses of the form $(X_i \vee Y_j)$ or $(\overline{X_i} \vee Y_j)$ are present but not both while in the non-monotone case both clauses can be present in $C(X_i)$. Note that bounds given in Theorem 9 are a special case of the bounds in Theorem 14 presented next.

Theorem 14. (Oblivious bounds for W2CNFs having only single-occurrence neighbors w.r.t. X_i). *For a given variable* X_i, *if* $\mathcal{F}$ *contains only single-occurrence neighbors but no two-occurrence neighbors then we have the following oblivious bounds for* X_i:

- U: $\Big(\prod_{j \in S^+} p_{i;j} \geq p_i\Big) \wedge \Big(\prod_{j \in S^-} \overline{p_{i;j}} \geq \overline{p_i}\Big)$
- L: *Either of following two conditions hold:*
  1. $\Big(\forall j \in S^+ : p_{i;j} \leq p_i\Big) \wedge \Big(\forall j \in S^- : \overline{p_{i;j}} = 0\Big)$
  2. $\Big(\forall j \in S^- : \overline{p_{i;j}} \leq \overline{p_i}\Big) \wedge \Big(\forall j \in S^+ : p_{i;j} = 0\Big)$

Optimal oblivious bounds are obtained by replacing inequality with equality in the bound conditions.

Example 15. *Consider* $C(X'_1) = \{(X_{1;1} \vee Y_2), (\overline{X_{1;2}} \vee Y_3), (X_{1;3} \vee Y_4), (\overline{X_{1;4}} \vee Y_5)\}$. *Theorem 14 gives the conditions for upper and lower oblivious bounds as:*

- U: $(p_{1;1} p_{1;3} \geq p_1) \wedge (\overline{p_{1;2}}\,\overline{p_{1;4}} \geq \overline{p_1})$.
- L: *Either of following two conditions hold:*
  1. $(p_{1;1} \leq p_1) \wedge (p_{1;3} \leq p_1) \wedge (\overline{p_{1;2}} = \overline{p_{1;4}} = 0)$
  2. $(\overline{p_{1;2}} \leq \overline{p_1}) \wedge (\overline{p_{1;4}} \leq \overline{p_1}) \wedge (p_{1;1} = p_{1;3} = 0)$

Our second non-monotone case is when $\mathcal{F}$ has at least one two-occurrence neighbor. Intuitively, dissociated variables which form clauses with two-occurrence neighbors are more constrained than those that appear with single-occurrence neighbors. Thus, there are more constraints on probabilities associated with two-occurrence neighbors (indexed by T^+ and T^-) than those associated with single-occurrence neighbors (indexed by S^+ and S^-); see conditions 1. and 2. in Theorem 16.

Theorem 16. (Oblivious bounds for W2CNFs having two-occurrence neighbors w.r.t. X_i). *For a given variable* X_i, *if* $\mathcal{F}$ *contains at least one two-occurrence neighbor then we have the following oblivious bounds for* X_i:

- U: $\Big(\prod_{j \in (S^+ \cup T^+)} p_{i;j} \geq p_i\Big) \wedge \Big(\prod_{j \in (S^- \cup T^-)} \overline{p_{i;j}} \geq \overline{p_i}\Big)$
- L: *Either of following three conditions hold:*
  1. $\Big(\prod_{j \in T^+} p_{i;j} \leq p_i\Big) \wedge \Big(\forall j \in (S^- \cup T^-) : \overline{p_{i;j}} = 0\Big)$
  2. $\Big(\prod_{j \in T^-} \overline{p_{i;j}} \leq \overline{p_i}\Big) \wedge \Big(\forall j \in (S^+ \cup T^+) : p_{i;j} = 0\Big)$
  3. *If* $|T^+| = |T^-| = 1$ *and* $T^+ = \{a\} \wedge T^- = \{b\}$: $\Big(p_{i;a} \leq p_i\Big) \wedge \Big(\forall j \in S^- : \overline{p_{i;j}} = 0\Big) \wedge$ $\Big(\overline{p_{i;b}} \leq \overline{p_i}\Big) \wedge \Big(\forall j \in S^+ : p_{i;j} = 0\Big)$

Optimal oblivious bounds are obtained by replacing inequality with equality in the bound conditions.

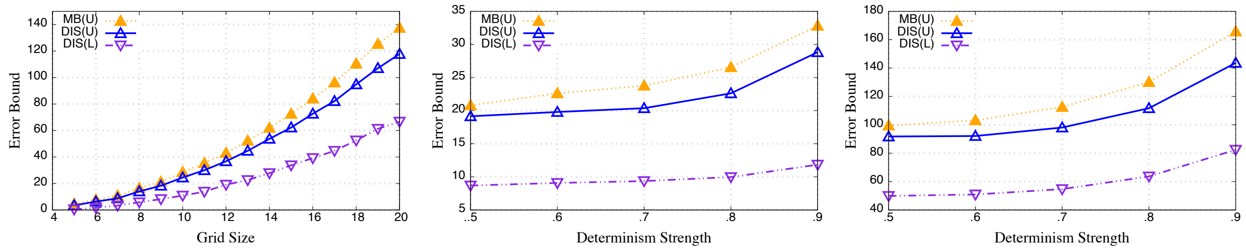
Algorithm 1: (DIS) Dissociation Bounds for WMC

Input: W2CNF $\mathcal{F} = \langle \mathbf{X}, \Phi, \mathbf{C} \rangle$,
Variable ordering $o = [X_1, X_2, \ldots, X_{|\mathbf{X}|}]$
Output: Lower (or upper) bound on the WMC
1. Initialize: $Z_B = 1$ (Bound on the partition function)
2. for i = 1 *to* $|\mathbf{X}|$ **do**
  2a. Convert $\mathcal{F}$ to a minimal $\mathcal{F}$
  2b. Convert $C(X_i)$ to canonical form
  2c. if C *is inconsistent* **then**
    return 0
  else if $C(X_i) = \{X_i\}$ **then**
    $Z_B = Z_B \times p_i$
  else if $C(X_i) = \{\overline{X_i}\}$ **then**
    $Z_B = Z_B \times \overline{p_i}$
  else if $C(X_i)$ *has two-occurrence neighbors* **then**
    Update Z_B using Theorem 16
  else if $C(X_i)$ *has single-occurrence neighbors* **then**
    Update Z_B using Theorem 14
return Z_B

Example 17. *Consider* $C(X'_1) = \{(X_{1;1} \vee Y_3), (X_{1;2} \vee Y_4), (\overline{X_{1;3}} \vee \overline{Y_4}), (X_{1;4} \vee Y_5)(\overline{X_{1;5}} \vee Y_6)\}$. *Theorem 16 gives the following conditions for upper and lower oblivious bounds:*

- U: $\Big(p_{1;1} p_{1;2} p_{1;4} \geq p_1\Big) \wedge \Big(\overline{p_{1;3}}\,\overline{p_{1;5}} \geq \overline{p_1}\Big)$
- L: *Either of the following three conditions hold:*
  1. $(p_{1;2} \leq p_1) \wedge (\overline{p_{1;3}} = \overline{p_{1;5}} = 0)$
  2. $(\overline{p_{1;3}} \leq \overline{p_1}) \wedge (p_{1;1} = p_{1;2} = p_{1;4} = 0)$
  3. $(p_{1;2} \leq p_1) \wedge (\overline{p_{1;3}} \leq \overline{p_1}) \wedge (p_{1;1} = p_{1;4} = \overline{p_{1;5}} = 0)$

Table 3 summarizes the oblivious bound conditions. Theorems 14 and 16 yield the algorithm given in Algorithm 1 for bounding the partition function of a given W2CNF.



(a) (b) (c)

Figure 1: Upper bound estimates for dissociation DIS(U) and mini-bucket MB(U), and lower bound estimates for dissociation DIS(L). Error bound by varying (a) grid size (b) level of determinism for 10×10 grid (c) 20×20 grid. Lower value is better.

Table 3: Summary of oblivious bound conditions. T: whether $C(X_i)$ has two-occurrence neighbors, S^+ and S^-: whether $C(X_i)$ has single-occurrence neighbors which appear in clauses $(X_i \vee Y_j)$ and $(\overline{X_i} \vee Y_j)$ respectively. An entry in a cell means that neighbors of the respective types are either present ($\surd$), absent ($\times$), or either present or absent ($*$). Bold text in Case and Solution columns denote novel contributions of this paper while normal font text indicates previous work.

S^+	S^-	T	Case	Solution
$\surd$	$\times$	$\times$	Monotone	Theorems 9 & **14**
$\times$	$\surd$	$\times$		
$\surd$	$\surd$	$\times$	**Single-occurrence**	**Theorem 14**
$*$	$*$	$\surd$	**Two-occurrence**	**Theorem 16**

5 EXPERIMENTS

We evaluated the performance of DIS (see Algorithm 1) and compared it with MB on generated synthetic datasets and benchmark datasets from the UAI 2008 probabilistic inference competition repository (http://graphmod.ics.uci.edu/uai08) for the task of computing upper and lower bounds on the weighted model count (or partition function). All experiment were conducted on quad-core Intel i7 based machines with 24GB RAM running Ubuntu.

5.1 SYNTHETIC DATASETS

We generated non-monotone W2CNF formulas encoded as $m \times m$ grid structure graphical models parameterized by univariate and pairwise binary potentials. We then compared error bound performance of DIS and MB (*i-bound* $= 1$) from the aspects of (1) varying grid sizes under random weight function settings; and (2) varying weight function settings according to determinism strength under fixed grid sizes. For each model, we computed the true weighted model count Z^*. We then compared each algorithm's approximated bound Z^{algo} and calculated the error bound as $\log(Z^*/Z^{\text{algo}})$ for the lower bound and the same negated for the upper bound. A lower error bound value is better. For each setting, we generated 50 random problem instances and ran DIS and MB 100 times for each instance. From the 100 solutions, we selected the best, namely either the lowest upper bound or the highest lower bound. We then computed the average error bound across the 50 problem instances.

Grid size. We generated $m \times m$ grids using values of $m = \{5, 6, 7, \ldots, 20\}$. For the weight function values, we sampled from an uniform $U(0, 1)$ distribution. We also uniformly generated the clauses. The results are shown in Figure 1a. For the upper bound, DIS noticeably begins to outperform MB starting at around grid size 10×10 and the performance gap widens as the grid size increases. Since MB utilizes the max function, it has a higher tendency to overestimate the upper bound. This was accomplished only by setting the weight function values to the k-th root (e.g., $p_{X_{1;1}} = p_{X_{1;2}} = \sqrt{p_{X_1}}$ for $|C(X_1)| = 2$). We would expect the performance gap to be wider, favoring DIS, by optimizing the inequalities. For the lower bound, MB produced 0 for all problems and thus was not plotted. MB has a high tendency to converge to the so called degenerate solution (i.e., 0) due to the min function. The lower bound for DIS is tighter, as compared to the upper bounds, since the settings to the lower bound inequalities do not need to be optimized.

Determinism strength. We analyzed the performance of DIS and MB according to various levels of determinism, namely the distance from uniform .5 (unweighted) towards 0 and 1. To accomplish this, we set all weight functions to the same value $p_X \in \{.5, .6, .7, .8, .9\}$. The results are shown in Figures 1b and 1c. For the lower bound, MB produced 0 for all problems and thus was not plotted. The overall relative performance comparison is similar to that of varying grid size. Again, the lower bound performance for DIS is tighter and all bounds had higher bound error as the determinism strength increased. Intuitively, as the gap between p_{X_i} and $\overline{p_{X_i}}$ widens, the tendency to overestimate (underestimate) the upper (lower) bound increases.

Table 4: The log relative upper bound between dissociation DIS(U) and mini-bucket MB(U) on UAI 2008 repository problem instances. Lower value is better for DIS.

Instance	$\log \frac{Z^{\text{DIS}(U)}}{Z^{\text{MB}(U)}}$	Instance	$\log \frac{Z^{\text{DIS}(U)}}{Z^{\text{MB}(U)}}$
sg2-17	-277.8	orc111	-87.6
sg7-11	-293.4	orc175	-96.3
sg8-18	-281.9	orc180	-124.4
sg9-24	-292.8	orc203	-111.0
sg17-4	-303.3	orc218	-4.4
smk10	-50.9	orc62	-393.4
smk20	-165.9	orc154	-97.0
orc42	-119.6	orc225	-137.3
orc45	-261.1	orc139	-155.0

Table 5: The log relative lower bound between ground truth and Dissociation DIS(L) on UAI 2008 repository problem instances. Lower value is better for DIS.

Instance	$\log \frac{Z^*}{Z^{\text{DIS}(L)}}$	Instance	$\log \frac{Z^*}{Z^{\text{DIS}(L)}}$
sg2-17	732.4	orc111	209.8
sg7-11	759.4	orc175	342.6
sg8-18	727.3	orc180	375.0
sg9-24	774.5	orc203	346.8
sg17-4	752.1	orc218	18.2
smk10	191.3	orc62	$-$
smk20	799.8	orc154	354.7
orc42	407.9	orc225	499.7
orc45	747.8	orc139	576.6

5.2 UAI INFERENCE DATASETS

We also compared DIS to MB on the segmentation (sg), promedas (orc) and smokers (smk) dataset from the UAI 2008 repository. The variables in the models are binary and the number of variables range from $\sim$100 to 1000. We converted the non-pairwise models to pairwise models and then encoded them as W2CNF. We used *i-bound* $= 1$ for MB. We ran DIS and MB 100 times and similarly, we selected the best. For the upper bound, we evaluated using the log relative upper bound, namely $\log(Z^{DIS(U)}/Z^{MB(U)})$. Lower value is better for DIS. The results are shown in Table 4. DIS outperforms MB by a wide margin on the majority of the datasets. The solution quality of DIS for sg was quite consistent while for orc it had higher variance. For the lower bound, we evaluated dissociation's lower bound against the ground truth, namely $\log(Z^*/Z^{DIS(L)})$. MB produced 0 for all problems and thus was not shown. The results are shown in Table 5 (orc62 was not tractable).

In summary, DIS performs consistently better than MB on harder WMC problems. In particular, the lower bounds output by DIS are always better than MB.

6 CONCLUSION AND FUTURE WORK

We proposed an approximate, oblivious bounding scheme for WMC, extending the idea of dissociation to non-monotone formulas and exploiting logical structure. Dissociation yields a novel set of inequalities for which upper and lower bounds can be derived efficiently. Empirically, we showed that our method outperforms mini-buckets—a popular oblivious bounding scheme—on various datasets. The lower bounds are robust since they do not require optimization (in the monotone case). For upper bounds, we utilized naïve settings, namely the k-th root applied to the parameter of a dissociated variable.

For future work, we are interested in obtaining better (tighter) upper and lower bounds. To do so, we can leverage four powerful complementary techniques described in literature (cf. (Gogate and Domingos, 2011, 2013; Ihler et al., 2012; Lam et al., 2014; Liu and Ihler, 2011; Ping et al., 2015)): cost-shifting (or re-parameterization), higher *ibound*, quantization and Hölder's inequality. For instance, applying Hölder's inequality to our running example (see Example 3) gives the optimization problem $\min_\omega (p_{X_1}^{\omega} + (\overline{p_{X_1}} p_{Y_2})^{\omega})^{1/\omega} (1 + p_{Y_3}^{(1-\omega)})^{(1/1-\omega)}$ such that $0 \leq \omega \leq 1$. We can also apply Hölder's inequality to dissociation which alternatively gives us the optimization problem $\min_{p_{X_{1;1}}, p_{X_{1;2}}, \omega} (p_{X_{1;1}}^{\omega} + (\overline{p_{X_{1;1}}} p_{Y_2})^{\omega})^{1/\omega} (p_{X_{1;2}}^{(1-\omega)} + (\overline{p_{X_{1;2}}} p_{Y_3})^{(1-\omega)})^{(1/1-\omega)}$ such that $p_{X_{1;1}} p_{X_{1;2}} = p_{X_1}$ and $0 \leq \omega \leq 1$. We are particularly interested in developing algorithms to optimize the latter problem and to determine which formulation will consistently yield tighter upper and lower bounds. Another line of future work is investigating the utility of our approach when applied to other inference tasks such as maximum a posteriori (MAP) estimation and marginal maximum a posteriori (MMAP) estimation.

Acknowledgments

This work was supported by the DARPA Explainable Artificial Intelligence (XAI) Program under contract number N66001-17-2-4032, and by the National Science Foundation grants IIS-1652835, IIS-1528037, and IIS-1762268.

References

Chavira, M. and Darwiche, A. (2008). On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6-7):772–799.

Choi, A., Chavira, M., and Darwiche, A. (2007). Node splitting: A scheme for generating upper bounds in

Bayesian networks. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pages 57–66.

Choi, A. and Darwiche, A. (2009). Relax then compensate: On max-product belief propagation and more. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems*, pages 351–359.

Dechter, R. (1996). Bucket elimination, a unifying framework for probabilistic inference. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence*, pages 211–219.

Dechter, R. and Rish, I. (2003). Mini-buckets: A general scheme for bounded inference. *Journal of the ACM*, 50(2):107–153.

den Heuvel, M. V., Gatterbauer, W., Theobald, M., and Geerts, F. (2018). A general framework for anytime approximation in probabilistic databases. In *8th International Workshop on Statistical Relational AI*. (to appear).

Ermon, S., Gomes, C. P., Sabharwal, A., and Selman, B. (2013). Embed and project: Discrete sampling with universal hashing. In *Advances in Neural Information Processing Systems 26*.

Gatterbauer, W. and Suciu, D. (2014). Oblivious bounds on the probability of boolean functions. *ACM Trans. Database Syst.*, 39(1):5:1–5:34.

Gatterbauer, W. and Suciu, D. (2017). Dissociation and propagation for approximate lifted inference with standard relational database management systems. *VLDB J.*, 26(1):5–30.

Gogate, V. and Dechter, R. (2007). Approximate counting by sampling the backtrack-free search space. In *Proceedings of the 22nd Conference on Artificial Intelligence*, pages 198–203.

Gogate, V. and Dechter, R. (2011). SampleSearch: Importance sampling in presence of determinism. *Artificial Intelligence*, 175(2):694–729.

Gogate, V. and Domingos, P. (2010). Formula-based probabilistic inference. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 210–219.

Gogate, V. and Domingos, P. (2011). Approximation by Quantization. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 247–255.

Gogate, V. and Domingos, P. (2013). Structured Message Passing. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 252–261.

Gogate, V. and Domingos, P. M. (2016). Probabilistic theorem proving. *Communications of the ACM*, 59(7):107–115.

Ihler, A., Flerova, N., Dechter, R., and Otten, L. (2012). Join-graph based cost-shifting schemes. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 397–406.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Lam, W., Kask, K., Dechter, R., and Ihler, A. (2014). Beyond static mini-bucket: Towards integrating with iterative cost-shifting based dynamic heuristics. In *Proceedings of the 7th Annual Symposium of Combinatorial Search*, pages 105–113.

Liu, Q. (2014). *Reasoning and Decisions in Probabilistic Graphical Models–A Unified Framework*. University of California, Irvine.

Liu, Q. and Ihler, A. (2011). Bounding the partition function using Hölder's inequality. In *Proceedings of the 28th International Conference on Machine Learning*, pages 849–856.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Ping, W., Liu, Q., and Ihler, A. (2015). Decomposition bounds for marginal map. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pages 3267–3275.

Sang, T., Bearne, P., and Kautz, H. (2005). Performing Bayesian inference by weighted model counting. In *Proceedings of the 20th National Conference on Artificial Intelligence*, pages 475–481.

Zhang, L. and Malik, S. (2002). The quest for efficient Boolean satisfiability solvers. In *Proceedings of the 14th International Conference on Computer Aided Verification*, pages 17–36.

Zhang, N. L. and Poole, D. (1994). A simple approach to Bayesian network computations. In *Proceedings of the 10th Canadian Conference on Artificial Intelligence*, pages 171–178.