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 $\mathsf{MI}(\Delta_{\mathbb{B}}) = \mathsf{MI}(\Delta)$. 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.

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

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 $\mathsf{MI}(\Delta) = \mathsf{MI}(\Delta_{\mathbb{B}})$ where $\mathsf{MI}(\Delta_{\mathbb{B}})$ 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. □

A.2. THEOREM 12

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

We prove the other way by contradiction. Suppose that there exists a WMI problem $\nu = \mathsf{WMI}(\Delta, w) \in \mathsf{treeWMI}(\mathbf{\Omega})$ with a per-literal weight function $w_\ell \notin \mathbf{\Omega}^{\mathsf{SMI}}$ such that $\rho(\nu) \in \mathsf{treeMI}$. Since the per-literal weight function $w_\ell \notin \mathbf{\Omega}^{\mathsf{SMI}}$, from the definition of $\mathbf{\Omega}^{\mathsf{SMI}}$, 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 $\ell \neq \ell_i, \forall i = 1, \cdots, k$. During the reduction, a clause $\Gamma' = \ell \Rightarrow \wedge_j^m \theta_j$ 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 $\rho(\nu) \in \mathsf{treeMI}$. Therefore, if $\forall \nu \in \mathsf{treeWMI}(\mathbf{\Omega})$, $\rho(\nu) \in \mathsf{treeMI}$, then $\mathbf{\Omega} \subseteq \mathbf{\Omega}^{\mathsf{SMI}}$. □

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.

$$\mathsf{WMI}(\Delta, w) = \int_{\boldsymbol{x} \models \Delta} \prod_{\ell \in LITS(\Delta)} w_\ell(\boldsymbol{x})^{[\![\boldsymbol{x} \models \ell]\!]} \, d\boldsymbol{x}$$

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

$$[\![\boldsymbol{x} \models \Delta]\!] = \prod_{\mathcal{S}} [\![\boldsymbol{x}_{\mathcal{S}} \models \Delta_{\mathcal{S}}]\!] = \prod_{\mathcal{S}} \prod_{\Gamma \in CLS(\Delta_{\mathcal{S}})} [\![\boldsymbol{x}_{\mathcal{S}} \models \Gamma]\!].$$

Moreover, it holds that

$$\prod_{\ell \in LITS(\Delta)} w_\ell(\boldsymbol{x})^{[\![\boldsymbol{x} \models \ell]\!]} = \prod_{\mathcal{S}} \prod_{\Gamma \in CLS(\Delta)} \prod_{\ell \in LITS(\Gamma)} w_\ell(\boldsymbol{x}_{\mathcal{S}})^{[\![\boldsymbol{x}_{\mathcal{S}} \models \ell]\!]}.$$

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

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 $f_{\mathcal{S}}$ is $\mathsf{m}_{X_i \to f_{\mathcal{S}}}(X_i) = 1$; when the leaf node is a factor node f_i, by definition the messages it sends to the variable node X_i is $\mathsf{m}_{f_i \to X_i}(X_i) = f_i(X_i)$. 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 $\mathsf{m}_{f_{\mathcal{S}} \to X_i}$ sent from factor nodes $f_{\mathcal{S}}$ to variable nodes X_i, the domain of variable X_i is divided into different pieces by constraints in formula $\Delta_{\mathcal{S}}$ that correspond to different integration bounds and thus the resulting messages from integration is again univariate piecewise integration. This concludes the proof. □

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 $f_{\mathcal{S}}$ that is a neighbour of the variable node X_i. Then the unnormalized joint distribution can be rewritten as follows.

$$p(\boldsymbol{x}) = \prod_{f_{\mathcal{S}} \in \mathsf{neigh}(X_i)} F_{\mathcal{S}}(x_i, \boldsymbol{x}_{\mathcal{S}})$$

where $\boldsymbol{x}_{\mathcal{S}}$ denotes the set of all variables in the subtree connected to the variable X_i via the factor node $f_{\mathcal{S}}$, and $F_{\mathcal{S}}(x_i, \boldsymbol{x}_{\mathcal{S}})$ denotes the product of all the factors in the group associated with factor $f_{\mathcal{S}}$. Then we have that

$$\begin{aligned} p(x_i) &= \prod_{f_{\mathcal{S}} \in \mathsf{neigh}(X_i)} \mathsf{m}_{f_{\mathcal{S}} \to X_i}(X_i) \\ &= \prod_{f_{\mathcal{S}} \in \mathsf{neigh}(X_i)} \int F_{\mathcal{S}}(x_i, \boldsymbol{x}_{\mathcal{S}})\, d\boldsymbol{x}_{\mathcal{S}} = \int p(\boldsymbol{x})\, d\boldsymbol{x} \backslash x_i \end{aligned}$$

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. □

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. □

A.7. PROPOSITION 22

Proof. (Proposition 22) The statement (i) holds since the message $\mathsf{m}_{X_i \to f_{ij}}$ 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 $\mathsf{m}_{f_{ij} \to X_j}$ 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 $\mathsf{m}_{X_i \to f_{ij}}$ 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 $2mc + c^2$. □

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 $\overrightarrow{f_s}$ bounds the number of pieces in the message $\mathsf{m}_{X_i \to f_s}$ 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 $\mathsf{m}_{X_j \to f_{s'}}$ for some variable X_j, the number of message pieces in each message $\mathsf{m}_{f_{s'} \to X_i}$ 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 $\mathsf{m}_{X_i \to f_s}$ 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. □

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

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

From the recurrence above, it can be obtained that

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

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