Lifted Generalized Dual Decomposition

Nicholas Gallo
University of California Irvine
Irvine, CA 92637-3435
ngallo1@uci.edu

Alexander Ihler
University of California Irvine
Irvine, CA 92637-3435
ihler@ics.uci.edu

Abstract

Many real-world problems, such as Markov Logic Networks (MLNs) with evidence, can be represented as a highly symmetric graphical model perturbed by additional potentials. In these models, variational inference approaches that exploit exact model symmetries are often forced to ground the entire problem, while methods that exploit approximate symmetries (such as by constructing an over-symmetric approximate model) offer no guarantees on solution quality. In this paper, we present a method based on a lifted variant of the generalized dual decomposition (GenDD) for marginal MAP inference which provides a principled way to exploit symmetric sub-structures in a graphical model. We develop a coarse-to-fine inference procedure that provides any-time upper bounds on the objective. The upper bound property of GenDD provides a principled way to guide the refinement process, providing good any-time performance and eventually arriving at the ground optimal solution.

Introduction

A central task in many application domains (Wainwright and Jordan 2008) is computing likelihoods and marginal probabilities over distributions defined by a graphical model. These tasks are, in general, intractable, and this has motivated the development of many approximate inference techniques. Of significant theoretical and practical interest are variational techniques which formulate the inference problem as a constrained optimization problem. Within this class, methods that provide upper bounds on the partition function and give rise to convex optimization problems are particularly attractive since convergence can be guaranteed and the quality of two bounds can be easily compared. This class contains the tree-reweighted bounds (Wainwright, Jaakkola, and Willsky 2005) and primal variants such as GenDD (Ping, Liu, and Ihler 2015), which we study here.

Recently, there has been a growing interest in the development of *lifted* inference techniques, both exact (Bui, Huynh, and de Salvo Braz 2012; Poole 2003) and approximate (Bui, Huynh, and Riedel 2012; Mladenov, Ahmadi, and Kersting 2012; Mladenov and Kersting 2015; Bui, Huynh, and Sontag 2014; Mladenov, Globerson, and Kersting 2014), which exploit model symmetries. Most of these works take a well established variational approximation and identify exact problem symmetries either algorithmically (e.g., redundant message computations of loopy belief propagation (LBP) (Singla and Domingos 2008; Kersting, Ahmadi, and Natarajan 2009)) or via graph theoretic notions of symmetry (Bui, Huynh, and Riedel 2012; Mladenov, Ahmadi, and Kersting 2012; Mladenov and Kersting 2015; Bui, Huynh, and Sontag 2014; Mladenov, Globerson, and Kersting 2014).

These methods, however, often have significant trouble dealing with models possessing highly symmetric substructures, but few exact symmetries. Some works (Venugopal and Gogate 2014; Van den Broeck and Darwiche 2013) address this by generating a symmetric approximation to an MLN (Richardson and Domingos 2006) with evidence, but offer no guarantees on the quality of the approximation. (Broeck and Niepert 2014) correct this "biasedness" problem by using an over-symmetric approximation to construct a proposal distribution for a Metropolis-Hastings chain, which guarantees convergence to the true marginals, but gives no guarantees on mixing time. Other works approximate lifted BP by grouping messages with similar values (Kersting et al. 2010) or satisfying a desired structural property (Singla, Nath, and Domingos 2014). (Singla, Nath, and Domingos 2014) analyzes errors the approximation induces on the true BP messages, but like BP, offers no guarantees on solution quality or convergence.

Similar in spirit to our work is (Habeeb et al. 2017) which sequentially solves a set of relaxed MAP inference problems for computer vision tasks. Their method operates in a coarse to fine manner solving, at each level, a MAP problem where groups of pixels are restricted to have the same value. This pixel grouping is relaxed and the finer MAP problem is initialized with the solution to the coarser problem, guaranteeing monotonic improvement of the solution.

Our paper provides a framework for exploiting approximate model symmetries, based on a lifted variant of GenDD (Ping, Liu, and Ihler 2015). In contrast to works that create an over-symmetric model approximation, our method imposes a set of over-symmetric restrictions on the variational cost-shifting updates (messages) associated with the GenDD problem. Guided by a measure of solution quality, these constraints are gradually relaxed, producing a sequence of problems of increasing accuracy, but increasing cost. Our experimental results show the superiority of the objective based refinement criteria, and good anytime performance compared to methods that exploit exact symmetries.

Background

A Markov random field (MRF) over n discrete random variables $X = [X_1 \dots X_n]$ taking values $x = [x_1 \dots x_n] \in (\mathcal{X}^1 \times \dots \times \mathcal{X}^n)$ has probability density function

$$p(X = x; \theta) = \exp\big[\sum_{\alpha \in \mathcal{F}} \theta_\alpha(x_\alpha) - \Phi(\theta)\big],$$
$$\Phi(\theta) = \log \sum_{x \in \mathcal{X}^n} \exp\big[\sum_{\alpha \in \mathcal{F}} \theta_\alpha(x_\alpha)\big]$$

where $\mathcal{F}$ represents the set of cliques of the distribution, with $\alpha \in \mathcal{F}$ being a subset of the variables, associated with a potential table θ_α. $\Phi(\theta)$ is the log partition function which normalizes the distribution. Furthermore, denote the set of variable indices as $\mathcal{V} = \{1 \dots n\}$.

Generalized dual decomposition

Motivated by the intractability of computing $\Phi(\theta)$, (Ping, Liu, and Ihler 2015) develops efficient upper bounds on $\Phi(\theta)$ via a decomposition based on Hölder's inequality. Their bound generalizes the dual decomposition bound (Globerson and Jaakkola 2008) for the MAP problem, maintaining two key properties: (1) it decomposes into a sum of terms defined over cliques of the graphical model and (2) a family of bounds can be indexed by a set of variational parameters where finding the parameter setting yielding the tightest bound can be formulated as a convex optimization problem. We review their main results.

For any non-negative function $f(x)$ and $w \in \mathbb{R}_{>0}$, the power sum operator is defined as

$$\sum_x^w f(x) = \big[\sum_x f(x)^{1/w}\big]^w.$$

The power sum reduces to standard sum when $w = 1$ and approaches $\max_x f(x)$ as $w \to 0^+$. We also define the vector power sum $\sum_x^{\mathbf{w}} F(x) = \sum_{x_k}^{w_k} \cdots \sum_{x_1}^{w_1} F(x)$, where $\mathbf{w}$ is a vector of k weights, x is a vector of k random variables, and F is a non-negative function with k inputs. The set of variational parameters $\boldsymbol{\delta} = \{\delta_\alpha^r \mid \alpha \in \mathcal{F}, r \in \mathbb{N}^+_{|\alpha|}\}$ and $\boldsymbol{w} = \{w_\alpha^r \mid \alpha \in \mathcal{F}, r \in \mathbb{N}^+_{|\alpha|}\}$, where $\mathbb{N}^+_i = \{1 \dots i\}$ for any positive integer i, reparameterize the distribution to provide a tighter bound. The bound of (Ping, Liu, and Ihler 2015) can be written as

$$\Phi(\theta) \le \sum_{v \in \mathcal{V}} u_v(\boldsymbol{\delta}, \boldsymbol{w}) + \sum_{\alpha \in \mathcal{F}} c_\alpha(\boldsymbol{\delta}, \boldsymbol{w}) \overset{\text{def}}{=} L(\boldsymbol{\delta}, \boldsymbol{w}). \quad (1)$$

where, letting $\boldsymbol{w}_\alpha = [w_\alpha^1 \dots w_\alpha^{|\alpha|}]$, the clique terms are

$$c_\alpha(\boldsymbol{\delta}, \boldsymbol{w}) = \log \sum_{x_\alpha}^{\boldsymbol{w}_\alpha} \exp\left[\theta_\alpha(x_\alpha) - \sum_{r=1}^{|\alpha|} \delta_\alpha^r(x_{\alpha_r})\right] \quad (2)$$

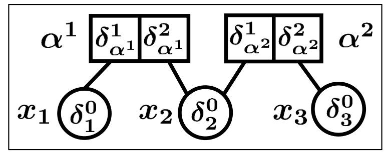


Figure 1: Position explicit factor graph associated with $\alpha^1 = (x_1, x_2)$, $\alpha^2 = (x_2, x_3)$. Factor cells (α^i, r) are associated with parameters $(\delta^r_{\alpha^i}, w^r_{\alpha^i})$ and unary terms are sums of neighboring cost shifting variables: $\delta_1^0 = \delta^1_{\alpha^1}$, $\delta_2^0 = \delta^2_{\alpha^1} + \delta^1_{\alpha^2}$, $\delta_3^0 = \delta^2_{\alpha^2}$ (similarly for the weights).

and the unary terms are

$$u_v(\boldsymbol{\delta}, \boldsymbol{w}) = \log \sum_{x_v}^{w_v^0} \exp(\delta_v^0(x_v)), \quad (3)$$

with $\delta_v^0(x_v) = \sum_{(\alpha,r) \in N_v} \delta_\alpha^r(x_v)$ and $w_v^\Delta = 1 - \sum_{(\alpha,r) \in N_v} w_\alpha^r$ where $N_v = \{(\alpha, r) \mid \alpha_r = v\}$ are the neighbors of v. Furthermore, we define $w_v^0 = \max(0, w_v^\Delta)$ and require $w_\alpha^r \ge 0$ for all (α, r).[1] We also require that the local elimination order of each clique power sum term (2) be consistent with a global elimination order $\mathbf{o}$, meaning that for all $i < j$, $\mathbf{o}(\alpha_i) < \mathbf{o}(\alpha_j)$. If this is violated for an α in the input, that α and its associated potential are permuted to obey this order.

Ground factor graphs In order to facilitate development of our lifted inference procedure, we modify the standard factor graph depiction of graphical models to make each variable's position in its participating factors explicit. To this end, we draw a circular node associated with each variable v and a rectangle partitioned into $|\alpha|$ cells associated with each factor α. Position cell (α, r) is associated with terms $(\delta_\alpha^r, w_\alpha^r)$ (labeled with just δ_α^r for compactness) and has an edge with its variable neighbor $v = \alpha_r$, which is associated with terms δ_v^0,w_v^Δ,u_v (labeled with just δ_v^0). A simple example is illustrated in figure 1.[2]

Exploiting symmetries

Lifted inference techniques identify a set of symmetries in the ground variational problem that can be exploited computationally. Exact lifted variational inference (Singla and Domingos 2008; Mladenov, Globerson, and Kersting 2014; Mladenov and Kersting 2015; Mladenov, Ahmadi, and Kersting 2012) identifies a stable partition (Berkholz, Bonsma, and Grohe 2013) of the graph as sufficient to characterize symmetries in the solution to the ground variational problem. In this paper, we impose a set of over-symmetric constraints on the ground variational parameters $(\boldsymbol{\delta}, \boldsymbol{w})$, inducing symmetries in the objective (1) that will, in general, be

[1] In (Ping, Liu, and Ihler 2015), w_v^0 is a free variable with restrictions $w_v^0 \ge 0$ and $w_v^0 + \sum_{(\alpha,r) \in N_v} w_\alpha^r = 1$. In our formulation, this sum is always larger than or equal to 1 (providing a valid bound), with equality holding at the optimum.

[2] (Mladenov and Kersting 2015) used a similar representation where each position cell is its own node, connected to its factor node. For compactness, we prefer the notation presented here.

coarser than those implied by a stable partition, yielding looser but computationally cheaper bounds.

Parameter and objective symmetries

Consider a disjoint partition $\mathcal{P}_{\mathcal{F}} = \{\mathcal{F}_1 \ldots \mathcal{F}_F\}$ of the ground factors $\mathcal{F}$ ($\mathcal{F}_i \cap \mathcal{F}_j = \emptyset$ for $i \neq j$, $\cup_f \mathcal{F}_f = \mathcal{F}$) where all factors in the same partition are required to have the same potential function. That is, for each f there exists a representative factor $\bar{\theta}_f$ (with associated scope size $\bar{r}_f$ and domains $\bar{\mathcal{X}}_f^r$ for $r \in \mathbb{N}^+_{\bar{r}_f}$) such that for all $\alpha \in \mathcal{F}_f$, $\theta_\alpha = \bar{\theta}_f$ (and $|\alpha| = \bar{r}_f$ and $\mathcal{X}_{\alpha_r} = \bar{\mathcal{X}}_f^r$ for $r \in \mathbb{N}^+_{\bar{r}_r}$).

In this paper, we consider over-symmetric constraints on the variational parameters which forces terms associated with factors in the same partition to be identical. That is, we consider parameters in the set

$$\mathcal{S}(\mathcal{P}_{\mathcal{F}}) = \left\{ (\boldsymbol{\delta}, \boldsymbol{w}) \mid {}^{(\exists\, \bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})\ \forall \mathcal{F}_f \in \mathcal{P}_{\mathcal{F}},\ \alpha \in \mathcal{F}_f,\ r \in \mathbb{N}^+_{|\alpha|}}_{\delta^r_\alpha(\bar{x}^r_f) = \bar{\delta}^r_f(\bar{x}^r_f),\ w^r_\alpha = \bar{w}^r_f} \right\}$$

where $\bar{\boldsymbol{\delta}} = \{\bar{\delta}^r_f \mid \mathcal{F}_f \in \mathcal{P}_{\mathcal{F}}, r \in \mathbb{N}^+_{\bar{r}_f}\}$ and $\bar{\boldsymbol{w}} = \{\bar{w}^r_f \mid \mathcal{F}_f \in \mathcal{P}_{\mathcal{F}}, r \in \mathbb{N}^+_{\bar{r}_f}\}$ are restricted sets of parameters. For any $(\boldsymbol{\delta}, \boldsymbol{w}) \in \mathcal{S}(\mathcal{P}_{\mathcal{F}})$, we identify symmetries in the clique terms (2) as $c_\alpha(\boldsymbol{\delta}, \boldsymbol{w}) = \bar{c}_f(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$ for all $\alpha \in \mathcal{F}_f$ where

$$\bar{c}_f(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}) \overset{\text{def}}{=} \log \sum_{\bar{x}}^{\bar{\boldsymbol{w}}_{\boldsymbol{f}}} \exp\left[\bar{\theta}_f(\bar{x}) - \sum_{r=1}^{\bar{r}_f} \bar{\delta}^r_f(\bar{x}^r_f)\right] \quad (4)$$

and $\bar{\boldsymbol{w}}_{\boldsymbol{f}} = [\bar{w}^1_f \ldots \bar{w}^{\bar{r}_f}_f]$ and $\bar{x} = [\bar{x}^1_f \ldots \bar{x}^{\bar{r}_f}_f]$.

To recognize the symmetries induced in the aggregated cost shifting δ^0_v, w^0_v (and hence u_v (3)) terms, first let $N^{(f,r)}_v = \{\alpha \mid \alpha \in \mathcal{F}_f, (\alpha, r) \in N_v\}$ represent the neighbors of v which contribute $\delta^r_\alpha(\bar{x}^r_f) = \bar{\delta}^r_f(\bar{x}^r_f)$ to the sum. There are $M^{(f,r)}_v = |N^{(f,r)}_v|$ such neighbors. Furthermore, define a disjoint partition $\mathcal{P}_{\mathcal{V}} = \{\mathcal{V}_1 \ldots \mathcal{V}_K\}$ of $\mathcal{V}$ where all variables in the same partition have the same domain ($\mathcal{X}_v = \bar{\mathcal{X}}_k$ for all $v \in \mathcal{V}_k$) and the same neighbor counts given $\mathcal{P}_F$. That is, there exist representative counts $\bar{M}^{(f,r)}_k$ such that for $M^{(f,r)}_v = \bar{M}^{(f,r)}_k$ for all $v \in \mathcal{V}_k$ and all (f, r). Now, for all $v \in \mathcal{V}_k$ we have

$$\delta^0_v(\bar{x}_k) = \bar{\delta}^0_k(\bar{x}_k) \overset{\text{def}}{=} \sum_{(f,r) \in \bar{N}_k} \bar{M}^{(f,r)}_k \cdot \bar{\delta}^r_f(\bar{x}_k) \quad (5)$$

for all values $\bar{x}_k$ and $\bar{N}_k = \{(f, r) \mid \bar{M}^{(f,r)}_k > 0\}$ (and similarly, $w^\Delta_v = \bar{w}^\Delta_k \overset{\text{def}}{=} 1 - \sum_{(f,r) \in \bar{N}_k} \bar{M}^{(f,r)}_k \cdot \bar{w}^r_f$). These symmetries imply that $u_v(\boldsymbol{\delta}, \boldsymbol{w}) = \bar{u}_k(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$ for all $v \in \mathcal{V}_k$.

Our objective can now be evaluated over the smaller set of representative parameters $(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$, with symmetries fully specified by $\mathcal{P} = (\mathcal{P}_{\mathcal{F}}, \mathcal{P}_{\mathcal{V}})$. We say that such $\mathcal{P}$ is *valid* and we have $L(\boldsymbol{\delta}, \boldsymbol{w}) = \bar{L}(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$ where

$$\bar{L}(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}) \overset{\text{def}}{=} \sum_{\mathcal{F}_f \in \mathcal{P}_{\mathcal{F}}} |\mathcal{F}_f| \cdot \bar{c}_f(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}) + \sum_{\mathcal{V}_k \in \mathcal{P}_{\mathcal{V}}} |\mathcal{V}_k| \cdot \bar{u}_k(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}). \quad (6)$$

Lifted factor graphs Associated with any valid partition $\mathcal{P}$ is a lifted graph which compactly illustrates the symmetries in the variational parameters and problem terms (1). A lifted factor graph has a super-node associated with each $\mathcal{V}_k$ and a super-factor with $\bar{r}_f$ position cells associated with each $\mathcal{F}_f$. Position cell (f, r) is associated with $\bar{\delta}^r_f, \bar{w}^r_f$ (labeled with just $\bar{\delta}^r_f$) and has a super-edge labeled $\bar{M}^{(f,r)}_k$ with super-node k if $\bar{M}^{(f,r)}_k > 0$. The term $\bar{N}_k$ defined after (5) can be seen as the neighborhood of k in the lifted graph. We similarly define the neighborhood of (f, r) as $\bar{N}^{(f,r)} = \{k \mid \bar{M}^{(f,r)}_k > 0\}$.

Example 1. Consider a model with factor graph depicted in figure 2a (with elimination order $\boldsymbol{o} = (R_1, R_2, T_1, T_2)$) and clique symmetries $\theta_{\alpha^2} = \theta_{\alpha^3} = \theta_{\alpha^4} = \theta_{\alpha^5} = \bar{\theta}$. Associated with the graph is the partition $\mathcal{P}_{\mathcal{F}} = \{\mathcal{F}_g, \mathcal{F}_y\}$ where $\mathcal{F}_g = \{\alpha^1\}$ (green), $\mathcal{F}_y = \{\alpha^2, \alpha^3, \alpha^4, \alpha^5\}$ (yellow) and $\mathcal{P}_{\mathcal{V}} = \{\mathcal{V}_p, \mathcal{V}_r, \mathcal{V}_b\}$ where $\mathcal{V}_p = \{R_1\}$ (purple), $\mathcal{V}_r = \{R_2\}$ (red), and $\mathcal{V}_b = \{T_1, T_2\}$ (blue).[3]

The top half of the middle panel indicates the over-symmetric parameter constrains specified by $\mathcal{P}_{\mathcal{F}}$. The bottom half of that panel indicates the induced symmetries in the unary terms. For example, by examining the neighborhood of R_1 in the factor graph, we see that $\delta^0_{R_1} = \delta^1_{\alpha^2} + \delta^1_{\alpha^3} + \delta^1_{\alpha^1}$. Via the parameter symmetries $\delta^1_{\alpha^2} = \delta^1_{\alpha^3} = \bar{\delta}^1_y$ and $\delta^1_{\alpha^1} = \bar{\delta}^1_g$, we see that this is equivalent to $\delta^0_{R_1} = 2 \cdot \bar{\delta}^1_y + \bar{\delta}^1_g$.

Gradient symmetries

Just as the over-symmetric parameter constraints induce symmetries in the ground objective function, they also induce (a different set of) symmetries in the gradient of the ground objective. These gradient symmetries will be used to derive the gradient of the lifted problem, to identify exact symmetries in the ground problem, and to guide our coarse-to-fine inference procedure.

First, notice that the parameter symmetries imply not only symmetries in the values of the clique (4) and unary (5) terms, but also in their gradients. That is, for any $(\boldsymbol{\delta}, \boldsymbol{w}) \in \mathcal{S}(\mathcal{P}_{\mathcal{F}})$, we have

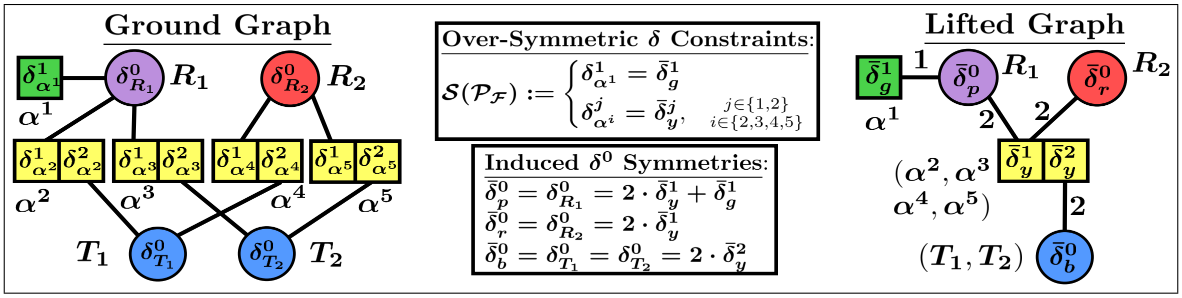
$$\frac{\partial c_\alpha}{\partial \delta^r_\alpha(\bar{x}^r_f)} = \frac{\partial \bar{c}_f}{\partial \bar{\delta}^r_f(\bar{x}^r_f)} \quad \text{and} \quad \frac{\partial u_v}{\partial \delta^0_v(\bar{x}_k)} = \frac{\partial \bar{u}_k}{\partial \bar{\delta}^0_k(\bar{x}_k)}$$

for any $\alpha \in \mathcal{F}_f$, $v \in \mathcal{V}_k$, and for all values of $\bar{x}^r_f$ and $\bar{x}_k$.

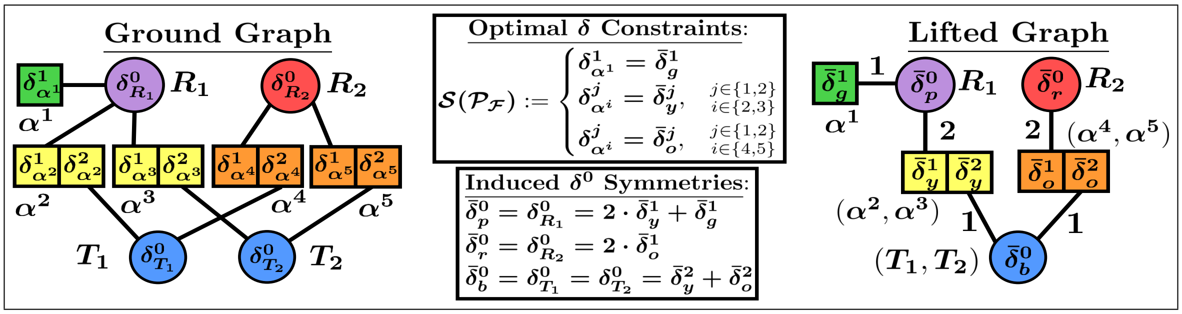
Now, consider the set $\mathcal{E}^{(f,r)}_k = \cup_{v \in \mathcal{V}_k} N^{(f,r)}_v$ of ground factors $\alpha \in \mathcal{F}_f$ whose r^{th} position cell (α, r) has neighbor $v = \alpha_r \in \mathcal{V}_k$. For any $\alpha \in \mathcal{E}^{(f,r)}_k$, the parameter δ^r_α appears only in clique term c_α in position r (multiplied by -1) and unary term u_v (multiplied by +1). We can therefore identify symmetries in the objective gradient as

$$\begin{aligned} \frac{\partial L}{\partial \delta^r_\alpha(\bar{x}_k)} &= \frac{\partial \bar{u}_v}{\partial \bar{\delta}^0_v(\bar{x}_k)} - \frac{\partial \bar{c}_f}{\partial \bar{\delta}^r_\alpha(\bar{x}_k)} \\ &= \frac{\partial \bar{u}_k}{\partial \bar{\delta}^0_k(\bar{x}_k)} - \frac{\partial \bar{c}_f}{\partial \bar{\delta}^r_f(\bar{x}_k)} \overset{\text{def}}{=} \bar{g}^{(f,r)}_{k;\delta}(\bar{x}_k) \end{aligned} \quad (7)$$

[3]The main text used integers to index factor and variable partitions; here we use letters, representing figure colors, for clarity.



(a) Valid, non-stable $\mathcal{P}$, specifies over-symmetric constraints on the variational parameters of the ground problem.



(b) Stable $\mathcal{P}$, specifies symmetries in variational parameters that yield an exact solution to the ground problem.

Figure 2: (a) valid, non-stable partition imposes over-symmetric constraints, yielding a looser bound. (b) Stable partition which characterizes exact symmetries of ground problem. More details in examples 1 and 2 of the text.

for all values of $\bar{x}_k$ (and similarly for the $\boldsymbol{w}$ gradients).

Lifted gradients and ground optimality The gradients of the lifted objective (6) can be written using the gradient symmetries of the ground problem. Specifying the lifted gradients in this way allows us to identify partitions whose symmetric parameter constraints specify an exact solution to the ground problem.

Since, via (7), $\delta_\alpha^r(\bar{x}_f^r) = \bar{\delta}_f^r(\bar{x}_f^r)$ for all $\alpha \in \mathcal{F}_f$ (and all values of $\bar{x}_f^r$), we can write

$$\frac{\partial \bar{L}}{\partial \bar{\delta}_f^r(\bar{x}_f^r)} = \sum_{\alpha \in \mathcal{F}_f} \frac{\partial L}{\partial \delta_\alpha^r(\bar{x}_f^r)} = \sum_{k \in \bar{N}^{(f,r)}} |\mathcal{E}_k^{(f,r)}| \cdot \bar{g}_{k;\delta}^{(f,r)}(\bar{x}_f^r) \tag{8}$$

where $|\mathcal{E}_k^{(f,r)}| = \bar{M}_k^{(f,r)} \cdot |\mathcal{V}_k|$ (and similarly for the $\bar{\boldsymbol{w}}$ gradients). In the special case where (f,r) has one lifted neighbor $\bar{N}^{(f,r)} = \{k\}$, $\mathcal{F}_f = \mathcal{E}_k^{(f,r)}$ and (8) simplifies to

$$\frac{\partial \bar{L}}{\partial \bar{\delta}_f^r(\bar{x}_f^r)} = |\mathcal{F}_f| \cdot \frac{\partial L}{\partial \delta_\alpha^r(\bar{x}_f^r)},$$

meaning that $\partial L / \partial \delta_\alpha^r(\bar{x}_f^r) = 0$ when $\partial \bar{L} / \partial \bar{\delta}_f^r(\bar{x}_f^r) = 0$. Now, if all of the gradients of the lifted problem are zero, and its lifted graph has $|\bar{N}^{(f,r)}| = 1$ for all (f,r), then all of the ground gradients are also zero. This means that a solution of the lifted problem provides a solution to the ground problem. A partition $\mathcal{P}$ whose lifted graph satisfies the property that $|\bar{N}^{(f,r)}| = 1$ for all (f,r) is called *stable*.

Example 2. If we replace $\mathcal{P}_\mathcal{F}$ from example 1 with $\mathcal{F} = \{\mathcal{F}_g, \mathcal{F}_y, \mathcal{F}_o\}$ where $\mathcal{F}_g = \{\alpha^1\}$ (green), $\mathcal{F}_y = \{\alpha^2, \alpha^3\}$ (yellow), $\mathcal{F}_o = \{\alpha^4, \alpha^5\}$ (orange), then the partition is stable and $\mathcal{S}(\mathcal{P}_\mathcal{F})$ contains a solution to the ground variational problem. This is illustrated in figure 2b.

Optimizing the lifted objective

In this paper, we use a black-box procedure to perform global optimization utilizing the gradient of $\bar{L}$. To ensure positivity of the weights, we define the parameters $\bar{\boldsymbol{\eta}} = \{\bar{\eta}_f^r \mid \mathcal{F}_f \in \mathcal{P}_\mathcal{F}, r \in \mathbb{N}_{\bar{r}_f}^+\}$, substituting $\bar{w}_f^r$ with $\epsilon + \exp(\bar{\eta}_f^r)$ (and $\partial \bar{L}/\partial \bar{\eta}_f^r = (\partial \bar{L}/\partial \bar{w}_f^r) \cdot \exp(\bar{\eta}_f^r)$). To ensure that the function is differentiable, we set $\bar{w}_k^0 = a(\bar{w}_k^\Delta)$ where $a(x) = \epsilon + t \cdot \log(1 + \exp(x/t))$.[4] This enables unconstrained optimization over the $\bar{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\delta}}$ parameters. The gradients are straightforward to derive from the gradients in (Ping, Liu, and Ihler 2015) and are omitted for space.

Coarse to fine approximation

We now develop a coarse to fine procedure that interleaves relaxations of the over-symmetric parameter constraints $\mathcal{S}(\mathcal{P}_\mathcal{F})$ with optimization, eventually arriving at the coarsest stable partition $\mathcal{P}$ and a solution to the ground problem. While optimization touches only the lifted graph, relaxation steps must touch a subset of the ground factors and variables. To control this cost, we adapt the key ideas from (Berkholz, Bonsma, and Grohe 2013), which presented an asymptotically optimal algorithm for generating the coarsest stable partition of a graph, to our work. This procedure is

[4] Any $t > 0$, $\epsilon > 0$ bound the objective since it generates larger unary weights: $a_t(w_v^\Delta) > \epsilon + \max(0, w_v^\Delta) > \max(0, w_v^\Delta)$.

Algorithm 1 Syntactic coarsest stable partition

1: Set $\mathcal{P}_\mathcal{V}$ to group variables with identical $\mathcal{X}_v$
2: Set $\mathcal{F}_f^0$ for $f \in \mathbb{N}_{F^0}^+$ to group factors with identical θ
3: $Q \leftarrow \emptyset$;
4: **for** $F = 1$ **to** F^0 **do**
5: $\quad \mathcal{F}_F \leftarrow \mathcal{F}_f^0$; **refinePV**$(0)$;
6: **end for**
7: **while** $Q \neq \emptyset$ **do**
8: $\quad$ Choose any $(f, r) \in Q$, set $Q \leftarrow Q \setminus (f, r)$
9: $\quad$ Choose a $k \in \bar{N}^{(f,r)}$ where $|\mathcal{E}_k^{(f,r)}| \leq \frac{1}{2}|\mathcal{F}_f|$
10: $\quad F \leftarrow F + 1$;
11: $\quad \mathcal{F}_F \leftarrow \mathcal{E}_k^{(f,r)}$; $\mathcal{F}_f \leftarrow \mathcal{F}_f \setminus \mathcal{F}_F$;
12: $\quad$ **refinePV**(f);
13: **end while**

used as a pre-processing step in works on exact lifted variational inference such as (Mladenov, Globerson, and Kersting 2014).

Syntactic coarsest stable partition

Throughout our description, we maintain a valid partition and its associated lifted graph. Recall from the subsection on lifted gradients (and Figure 2b) that a valid partition is stable if and only if each position cell in its lifted graph has exactly one neighbor. Algorithm 1 maintains the set $Q = \{(f, r) \mid |\bar{N}^{(f,r)}| \geq 2\}$ of position cells that violate this property. The algorithm iteratively selects an $(f, r) \in Q$ and one of its neighbors $k \in \bar{N}^{(f,r)}$. Ground factors in $\mathcal{F}_f$ are then split into two groups (line 11), based on whether their r^{th} position neighbor is a variable in $\mathcal{V}_k$ (recall from the section on gradient symmetries that this is $\mathcal{E}_k^{(f,r)}$), or not.

In the appendix, we define the function **refinePV** that updates the partition $\mathcal{P}_\mathcal{V}$ to be consistent with the new neighbor counts M_v^f and M_v^F for all v, updating the lifted graph and Q as needed. The restriction of choosing k such that $|\mathcal{E}_k^{(f,r)}| \leq \frac{1}{2} \cdot |\mathcal{F}_f|$ (line 9) is necessary to maintain the complexity results of (Berkholz, Bonsma, and Grohe 2013), as discussed in the appendix.

Coarse-to-fine algorithm

One simple way to construct a coarse to fine inference procedure is to interrupt the main loop of Algorithm 1 to perform lifted inference with relaxations and symmetries specified by the current partition $\mathcal{P}$. This would yield any-time bounds and ensure that we never produce a model finer than the coarsest stable partition.

In practice, however, superior results are obtained by choosing refinement operations based on their predicted decrease of the objective value. Algorithm 2 illustrates this idea, where S_f^r (detailed in next section) judges the quality of a split of (f, r), $cost(\mathcal{P})$ is the estimated cost of one inference iteration (we count the number of operations used to compute the lifted score (6)), and β controls how much refinement we allow between inference calls.

Algorithm 2 Coarse to fine inference

Execute lines 1 to 6 of Algorithm 1 $\quad \triangleright$ Initialize $\mathcal{P}$
$\Psi \leftarrow 0$;
repeat
$\quad$ **if** $cost(\mathcal{P}) > \Psi$ **then**
$\qquad [\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}] =$ **inference**$(\mathcal{P}, \bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$
$\qquad$ Set S_f^r via (9) $\forall (f, r) \in Q$
$\qquad \Psi \leftarrow cost(\mathcal{P}) \cdot \beta$;
$\quad$ **end if**
$\quad (f^*, r^*) \leftarrow arg\,min_{(f,r) \in Q}\, S_f^r$
$\quad$ Set $\bar{N}_1' \subset \bar{N}^{(f^*,r^*)}$ as $arg\,min$ of $S_{f^*}^{r^*}$ via (9)
$\quad F \leftarrow F + 1$;
$\quad \mathcal{F}_F \leftarrow \cup_{k \in \bar{N}_1'} \mathcal{E}_k^{(f^*,r^*)}$; $\mathcal{F}_{f^*} \leftarrow \mathcal{F}_{f^*} \setminus \mathcal{F}_F$;
$\quad$ **refinePV**(f^*)
$\quad$ Set $S_{f'}^r$ via (9) for $f' \in \{f^*, F\}, r \in \mathbb{N}_{\bar{r}_f}^+$
until $\mathcal{P}$ is stable
$[\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}}] =$ **inference**$(\mathcal{P}, \bar{\boldsymbol{\delta}}, \bar{\boldsymbol{w}})$ $\quad \triangleright$ Stable $\mathcal{P}$ inference

Objective based scoring function We first define a metric to measure the quality of a proposed parameter partition, then we will show how to optimize this metric efficiently, using only information contained in the lifted graph.

For the set of ground parameters associated with an (f, r), we measure the error induced by the current parameter tying restrictions as the distance of the ground gradients to their projected gradient. Letting g_α^r be the set of ground gradients associated with an (α, r), this metric is $d(\mathcal{F}_f, r) = \sum_{\alpha \in \mathcal{F}_f} ||g_\alpha^r - \gamma||_2^2$, with $\gamma = (\sum_{\alpha \in \mathcal{F}_f} g_\alpha^r)/|\mathcal{F}_f|$.

We would like to find a split of $\mathcal{F}_f$ into $\mathcal{F}_{f^1}$ and $\mathcal{F}_{f^2}$ that minimizes $d(\mathcal{F}_{f^1}, r) + d(\mathcal{F}_{f^2}, r) - d(\mathcal{F}_f, r)$, which is the change in the projected gradient distance. By symmetry, this split groups together factors with the same g_α^r. Recalling from (7) that $g_\alpha^r = \bar{g}_k^{(f,r)}$ for all $\alpha \in \mathcal{E}_k^{(f,r)}$, we write this grouping as $\mathcal{F}_{f^i} = \cup_{k \in \bar{N}_i'} \mathcal{E}_k^{(f,r)}$ where $\bar{N}_1' \subset \bar{N}^{(f,r)}$ and $\bar{N}_2' = \bar{N}^{(f,r)} \setminus \bar{N}_1'$ (adopting the convention $|\mathcal{F}_{f^1}| \leq |\mathcal{F}_{f^2}|$).

We can now form a compact optimization problem to find the best split. For any $\bar{N}' \subseteq \bar{N}^{(f,r)}$ we have $d(\cup_{k \in \bar{N}'} \mathcal{E}_k^{(f,r)}, r) = \bar{d}_f^r(\bar{N}')$ where

$$\bar{d}_f^r(\bar{N}') = \sum_{k \in \bar{N}'} |\mathcal{E}_k^{(f,r)}| \cdot ||\bar{g}_k^{(f,r)} - \gamma||_2^2$$

and $\gamma = (\sum_{k \in \bar{N}'} |\mathcal{E}_k^{(f,r)}| \cdot \bar{g}_k^{(f,r)})/(\sum_{k \in \bar{N}'} |\mathcal{E}_k^{(f,r)}|)$. Letting $\mathbb{P}(\bar{N}^{(f,r)})$ be the power set (all subsets) of $\bar{N}^{(f,r)}$, we solve

$$S_f^r = \min_{\bar{N}_1' \in \mathbb{P}(\bar{N}^{(f,r)})} \sum_{i=1}^{2} \bar{d}_f^r(\bar{N}_i') - \bar{d}_f^r(\bar{N}^{(f,r)}). \qquad (9)$$

This is a weighted K-Means (K=2) problem with "data" $\bar{g}_k^{(f,r)}$ and weights $|\mathcal{E}_k^{(f,r)}|$. If we approximate this by using only one of the elements of $\bar{g}_k^{(f,r)}$ (for example only the gradient wrt $\bar{\delta}_k^{(f,r)}(1)$, as we use on binary problems in this paper's experiments), the optimum can be found exactly by sorting $\bar{g}_k^{(f,r)}$ and looping over all splits.

Weight	Formula
b	$(\forall x \neq y)\ V(x) \Leftrightarrow V(y)$
u_x	$(\forall x)\ V(x)$

(a) Complete Graph

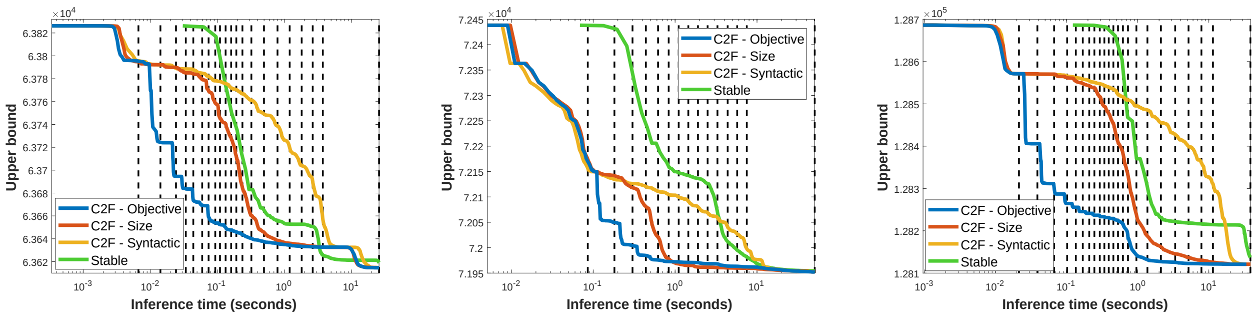
Weight	Formula
b	$(\forall\ x \neq y)\ L(x,y) \Rightarrow (C(y) \Leftrightarrow C(x))$
u_x	$(\forall x)\ C(x)$
l_{xy}	$(\forall x \neq y)\ L(x,y)$

(b) Binary Collective Classification

Weight	Formula
b	$(\forall\ x \neq y)\ Q_1(x) \Leftrightarrow Q_2(y)$
b	$(\forall\ x \neq y)\ Q_2(x) \Leftrightarrow Q_3(y)$
b	$(\forall\ x \neq y)\ Q_3(x) \Leftrightarrow Q_1(y)$
u_x^i	$(\forall x)\ Q_i(x), \quad i \in \{1,2,3\}$

(c) Clique-Cycle

Figure 3: Test models with distinct soft evidence (u and l terms).



(a) Complete (b) Collective Classification (c) Clique Cycle

Figure 4: Comparison of optimization at stable coloring (exact model symmetries) with our coarse to fine inference framework for three different splitting methods, on each of the test models with $d = 160$ domain size for the logical variables. Vertical black dashed lines indicate coarse-to-fine transitions for objective based splitting ("C2F - Objective" curve). Transitions for other C2F curves occur at approximately the same positions.

Experiments

This section provides an empirical illustration of our lifted GenDD algorithm. We demonstrate the superiority of our objective based approach over purely syntactic based approaches for generating a coarse to fine approximation sequence. We also demonstrate excellent any-time performance on MLNs with distinct soft evidence on every node compared to exact lifted variational approaches which are forced to ground the entire problem.

Datasets and methodology

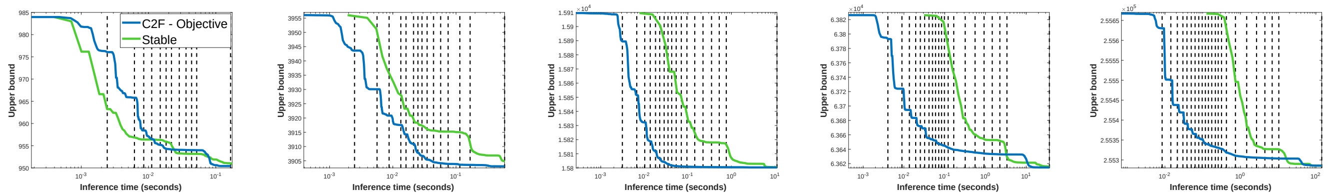
Datasets. Figure 3 shows the models we use (similar models without evidence were used in (Mladenov and Kersting 2015; Bui, Huynh, and Sontag 2014) to evaluate performance on marginalization tasks). In all models b was set to 2.5 and the u terms specify random evidence uniformly distributed on $[-\zeta, \zeta]$ where $\zeta = \log((1-10^{-2})/10^{-2})$, so that the u terms act like "probabilities" between 0.01 and 0.99. For the collective classification problem, we also choose a random 10% of the x, y values associated with links $L(x,y)$ and set l_{xy} to a random value on [0, 5.0]. This can be interpreted as the strength of a link between two web pages.

Parameter settings. In our experiments, we set $\epsilon = 10^{-3}$, $t = 10^{-2}$ (in the definition of $a(x)$) and perform an LBFGS black box optimization with rank 20 Hessian correction. For the coarse to fine procedure, it is important to balance performing inference with model refinement. We found it worked well to perform a small number of inference iterations (30), followed by a small amount of model refinement (setting $\beta = 1.25$ in Algorithm 2). This worked better than trying to diagnose inference convergence to determine the transition point. Future work will look at more principled methods of balancing inference work with refinement.

Timing methodology. In our experiments, we measure only the cost of performing inference, ignoring the cost of identifying symmetries. This is common in experiments on approximate lifted inference in other works (Mladenov, Globerson, and Kersting 2014; Broeck and Niepert 2014; Venugopal and Gogate 2014) as well as exact lifted variational inference approaches (Bui, Huynh, and Riedel 2012; Bui, Huynh, and Sontag 2014; Mladenov and Kersting 2015) which assume problem symmetries are available before inference. Works such as (Singla, Nath, and Domingos 2014; Kersting et al. 2010) note that identifying (approximate) symmetries can be a bottleneck. (Singla, Nath, and Domingos 2014) used a sparse hyper-cube based representation to specify approximate symmetries in MLNs; adapting lifted GenDD to leverage a similar representation is an interesting direction for future work.

Results

Figure 4 compares our coarse to fine procedure vs. performing inference on the stable coloring (ground model in our setup) on each of our three test models. The blue curve ("C2F - objective") performs objective-based splitting as per Algorithm 2 using the metric (9) to guide splitting choice. We also show syntactic splitting ("C2F - syntactic", orange) as per Algorithm 1, interleaving inference when the predicted inference cost exceeds a cost bound (then multiplying that bound by β as in Algorithm 2). "C2F - Size" (red) splits the largest super-factor in Q and chooses a random split that is as even as possible (adding random edges from $\bar{N}^{(f,r)}$ to



(a) $d = 20$ (b) $d = 40$ (c) $d = 80$ (d) $d = 160$ (e) $d = 320$

Figure 5: Scalability comparison. Panels (a)-(e) show bound vs time on instances of complete graph with evidence, varying logical variable domain size d (in MLN specification). C2F exhibits increasing advantage as model grows larger.

$\bar{N}'_2$, stopping when $\sum_{k \in \bar{N}'_1} |\mathcal{E}_k^{(f,r)}| > \frac{1}{2} \cdot |\mathcal{F}_f|$ and setting $\bar{N}'_1 = \bar{N}^{(f,r)} \setminus \bar{N}'_2$ in line 10 of Algorithm 2), interleaving inference with refinement as in the other C2F methods. The stable (here, ground) inference process is shown in green.

We see that objective-based refinement significantly outperforms the others, with size-based refinement worse, but still better than syntactic splitting. Objective-based splitting provides orders of magnitude speed-up over the stable coloring for similar quality results, and all methods provide better early performance than the stable coloring, which must touch the entire model to even provide an initial bound.

Figure 5 demonstrates the scalability of our approach. Plots (a)-(e) show the running time for our coarse-to-fine ("C2F - Objective" (blue)) method and the stable coloring method on instances of the complete graph, varying its size.

Conclusions

We presented a framework for lifting GenDD for approximate inference by imposing over-symmetric constraints on the optimization parameters to induce symmetry in the variational bound. We developed a coarse-to-fine procedure that, guided by the objective, provides high-quality any-time approximations even in models with no exact symmetry. Our method provides orders of magnitude speed-up on our benchmarks, with increasing advantage for larger models.

Acknowledgements

This work is sponsored in part by NSF grants IIS-1526842, IIS-1254071, and by the United States Air Force under Contract No. FA8750-14-C-0011 and FA9453-16-C-0508.

Appendix

Description of refinePV.

After creating a new partition $\mathcal{F}_F$ (possibly via splitting it from $\mathcal{F}_f$), algorithms 1 and 2 call **refinePV** to update $\mathcal{P}_\mathcal{V}$ such that variables having identical neighbor counts $M_v = [M_v^1, \ldots, M_v^F]$ are grouped together (where $M_v^{f'} \stackrel{\text{def}}{=} [M_v^{(f',1)} \ldots M_v^{(f',\bar{r}_{f'})}]$ for all f'). We assume that at entry, the neighbor counts $M_v^{f'}$ are correct for all v and $f' = 1 \ldots F-1$, and that $\mathcal{P}_\mathcal{V}$ correctly groups variables by these counts, i.e., $M_v^{f'} = \bar{M}_k^{f'}$ for all $v \in \mathcal{V}_k$.

The first loop updates $N_v^{(f,r)}$ and $N_v^{(F,r)}$ and their counts. The second loop groups variables by their signature $[\mathcal{K}_v, M_v^{(F,r)}]$, creating the sets $\mathcal{V}_k^m = \{v \mid \mathcal{K}_v = k, m = M_v^{(F,r)}\}$, where $\mathcal{K}_v$ is the partition holding variable v ($v \in \mathcal{V}_k$ with $k = \mathcal{K}_v$). This produces the same grouping as would be obtained by grouping the full signature $[M_v^1 \ldots M_v^{(F-1)}, M_v^F]$ because only $M_v^{(f,r)}$ and $M_v^{(F,r)}$ counts change during the call. Note that the change to $M_v^{(f,r)}$ does not affect our grouping since for all $v \in \mathcal{V}_k^m$, $M_v^{(f,r)}$ is equal to its value before the call minus $M_v^{(F,r)}$ (the updates to $\bar{M}_{K'}^{(f,r)}$ on line 19 reflect this fact).

Lastly, the third loop adds $\mathcal{V}_k^m$ sets to $\mathcal{P}_\mathcal{V}$ and updates quantities associated with the lifted graph ($\bar{M}$, $\bar{N}$ terms). $\mathcal{V}_k$ will be over-written with $\mathcal{V}_k^{m'}$ where m' is the last (k, m) encountered in the loop. In all other cases, k splits into a new super-node and its neighbors are copied (line 16).

Algorithm 3 refinePV

Input: f: if $f \neq 0$, $\mathcal{F}_F$ was deleted from $\mathcal{F}_f$ before call.

1: **for** $r = 1$ **to** $\bar{r}_F$ **do**
2: **for** $\alpha \in \mathcal{F}_F$ **do** ▷ Update ground neighbor lists
3: $v = \alpha_r$;
4: $N_v^{(f,r)} \leftarrow N_v^{(f,r)} \setminus \alpha$; $N_v^{(F,r)} \leftarrow N_v^{(F,r)} \cup \alpha$;
5: $M_v^{(f,r)} \leftarrow M_v^{(f,r)} + 1$; $M_v^{(F,r)} \leftarrow M_v^{(F,r)} - 1$;
6: **end for**
7: $\mathcal{V}_k^m \leftarrow \emptyset\ (\forall k, m)$ ▷ New variable groupings
8: **for** $\alpha \in \mathcal{F}_F$ **do**
9: $v = \alpha_r$; $k \leftarrow \mathcal{K}_v$; $m = M_v^{(F,r)}$;
10: $\mathcal{V}_k \leftarrow \mathcal{V}_k \setminus v$; $\mathcal{V}_k^m \leftarrow \mathcal{V}_k^m \cup v$;
11: **end for**
12: **for** $(k, m) : \mathcal{V}_k^m \neq \emptyset$ **do** ▷ Update lifted graph
13: $K' \leftarrow k$;
14: **if** $m \neq$ last_m(k) **then** ▷ New super-node
15: $K \leftarrow K + 1$; $K' \leftarrow K$
16: $\bar{M}_K^{(f',r')} \leftarrow \bar{M}_k^{(f',r')}\ \forall (f', r') \in \bar{N}_k$
17: **end if**
18: $\mathcal{V}_{K'} \leftarrow \mathcal{V}_k^m$; $\mathcal{K}_v \leftarrow K'\ (\forall v \in \mathcal{V}_{K'})$
19: $\bar{M}_{K'}^{(F,r)} \leftarrow m$; $\bar{M}_{K'}^{(f,r)} \leftarrow \bar{M}_{K'}^{(f,r)} - m$;
20: **end for**
21: **end for**

Note: Updating $\bar{M}_{k'}^{(f',r')}$ (for any k', f', r') changes $\bar{N}_{k'}$, $\bar{N}^{(f',r')}$, Q. For compactness, these changes are not shown.

Complexity analysis Each set is implemented as a hash table supporting $O(1)$ addition and deletion ($N_v^{(f',r')}$ stores pointers to factor scope vectors). Hence, each $\alpha \in \mathcal{F}_F$ contributes $O(|\alpha|)$ time to the first two loops of **refinePV**. Throughout algorithms 1 and 2, an α is moved to an $\mathcal{F}_F$ that is at most half the size of its current set. Thus, each α participates in at most $\log_2 |\mathcal{F}|$ sets throughout either algorithm. Equivalent to the result of (Berkholz, Bonsma, and Grohe 2013), the total time of the these loops is $O(R \cdot |\mathcal{F}| \cdot \log |\mathcal{F}|)$ where $R = \max_{\alpha \in \mathcal{F}} |\alpha|$.

Throughout algorithms 1 and 2, the total time spent in the third loop of **refinePV** is proportional to the size of the lifted graph. This is because the loop adds nodes and edges to the lifted graph in $O(1)$ and each addition is accompanied by at most one deletion (if $\bar{M}_K^{(f,r)} = 0$ in line 19). Since the size of the lifted graph is at most the size of the ground graph ($R \cdot |\mathcal{F}|$), the total time is dominated by the first two loops.

References

Berkholz, C.; Bonsma, P.; and Grohe, M. 2013. Tight lower and upper bounds for the complexity of canonical colour refinement. In *European Symposium on Algorithms*, 145–156. Springer.

Broeck, G. V. d., and Niepert, M. 2014. Lifted probabilistic inference for asymmetric graphical models. *arXiv preprint arXiv:1412.0315*.

Bui, H. B.; Huynh, T. N.; and de Salvo Braz, R. 2012. Exact lifted inference with distinct soft evidence on every object. In *AAAI*.

Bui, H. H.; Huynh, T. N.; and Riedel, S. 2012. Automorphism groups of graphical models and lifted variational inference. *arXiv preprint arXiv:1207.4814*.

Bui, H. H.; Huynh, T. N.; and Sontag, D. 2014. Lifted tree-reweighted variational inference. *arXiv preprint arXiv:1406.4200*.

Globerson, A., and Jaakkola, T. S. 2008. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In *Advances in neural information processing systems*, 553–560.

Habeeb, H.; Anand, A.; Singla, P.; et al. 2017. Coarse-to-fine lifted map inference in computer vision. *arXiv preprint arXiv:1707.07165*.

Kersting, K.; Ahmadi, B.; and Natarajan, S. 2009. Counting belief propagation. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 277–284. AUAI Press.

Kersting, K.; El Massaoudi, Y.; Hadiji, F.; and Ahmadi, B. 2010. Informed lifting for message-passing. In *AAAI*.

Mladenov, M.; Ahmadi, B.; and Kersting, K. 2012. Lifted linear programming. In *AISTATS*, 788–797.

Mladenov, M., and Kersting, K. 2015. Equitable partitions of concave free energies. In *UAI*, 602–611.

Mladenov, M.; Globerson, A.; and Kersting, K. 2014. Lifted message passing as reparametrization of graphical models. In *UAI*, 603–612.

Ping, W.; Liu, Q.; and Ihler, A. T. 2015. Decomposition bounds for marginal map. In *Advances in Neural Information Processing Systems*, 3267–3275.

Poole, D. 2003. First-order probabilistic inference. In *IJCAI*, volume 3, 985–991.

Richardson, M., and Domingos, P. 2006. Markov logic networks. *Machine learning* 62(1):107–136.

Singla, P., and Domingos, P. M. 2008. Lifted first-order belief propagation. In *AAAI*, volume 8, 1094–1099.

Singla, P.; Nath, A.; and Domingos, P. M. 2014. Approximate lifting techniques for belief propagation. In *AAAI*, 2497–2504.

Van den Broeck, G., and Darwiche, A. 2013. On the complexity and approximation of binary evidence in lifted inference. In *Advances in Neural Information Processing Systems*, 2868–2876.

Venugopal, D., and Gogate, V. 2014. Evidence-based clustering for scalable inference in markov logic. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 258–273. Springer.

Wainwright, M. J., and Jordan, M. I. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning* 1(1–2):1–305.

Wainwright, M. J.; Jaakkola, T. S.; and Willsky, A. S. 2005. A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory* 51(7):2313–2335.