Lifted Message Passing as Reparametrization of Graphical Models

Martin Mladenov
TU Dortmund University
{fn.ln}@cs.tu-dortmund.de

Amir Globerson
Hebrew University
gamir@cs.huji.ac.il

Kristian Kersting
TU Dortmund University
{fn.ln}@cs.tu-dortmund.de

Abstract

Lifted inference approaches can considerably speed up probabilistic inference in Markov random fields (MRFs) with symmetries. Given evidence, they essentially form a lifted, i.e., reduced factor graph by grouping together indistinguishable variables and factors. Typically, however, lifted factor graphs are not amenable to off-the-shelf message passing (MP) approaches, and hence requires one to use either generic optimization tools, which would be slow for these problems, or design modified MP algorithms. Here, we demonstrate that the reliance on modified MP can be eliminated for the class of MP algorithms arising from MAP-LP relaxations of pairwise MRFs. Specifically, we show that a given MRF induces a whole family of MRFs of different sizes sharing essentially the same MAP-LP solution. In turn, we give an efficient algorithm to compute from them the smallest one that can be solved using off-the-shelf MP. This incurs no major overhead: the selected MRF is at most twice as large as the fully lifted factor graph. This has several implications for lifted inference. For instance, running MPLP results in the first convergent lifted MP approach for MAP-LP relaxations. Doing so can be faster than solving the MAP-LP using lifted linear programming. Most importantly, it suggests a novel view on lifted inference: it can be viewed as standard inference in a reparametrized model.

1 INTRODUCTION

Probabilistic logical languages [5] provide powerful formalisms for knowledge representation and inference. They allow one to compactly represent complex relational and uncertain knowledge. For instance, in the friends-and-smokers Markov logic network (MLN) [17], the weighted formula $1.1\ :\ \mathtt{fr}(\mathtt{X},\mathtt{Y}) \Rightarrow (\mathtt{sm}(\mathtt{X}) \Leftrightarrow \mathtt{sm}(\mathtt{Y}))$ encodes that friends in a social network tend to have similar smoking habits. Yet, performing inference in these languages is extremely costly, especially if it is done at the propositional level. Instantiating all atoms from the formulae in such a model induces a standard graphical model (potentially) with symmetries, i.e., with repeated factor structures for all grounding combinations. Recent advances in lifted probabilistic inference [16] such as [3, 15, 1, 14, 18] (see [9] for an overview that also covers exact inference approaches), have rendered many of these large, previously intractable models quickly solvable by exploiting the induced symmetries. For instance, lifted message-passing (MP) approaches such as [19, 10, 22, 8, 1] have been proven successful in several important AI applications such as link prediction, social network analysis, satisfiability and boolean model counting problems. Lifted MP approaches such as lifted Belief Propagation (BP) first automatically group together variables and factors of the graphical model into supervariables and superfactors if they have identical computation trees (i.e., the tree-structured "unrolling" of the graphical model computations rooted at the nodes). Then, they run modified MP algorithms on this lifted network. These modified MP algorithms, however, can also be considered a downside of today's lifted MP approaches. They require more information than is actually captured by a standard factor graph. More precisely, lifted MP will typically exponentiate a message from a supervariable to a superfactor by the count of ground instances of this superfactor, which are neighbors to a ground instance of the supervariable. Since these multi-dimensional counts have to be stored in the network, the lifted factor graph becomes a multigraph (i.e., a factor graph with edge counts and self-loops), in contrast to a standard factor graph where no multiedges or loops are allowed. Hence, lifted factor graphs are not amenable to off-the-shelf MP approaches. Instead, lifted MP has its own ecosystem of lifted data structures and lifted algorithms. In this ecosystem, considerable effort is required to keep up with the state of the art in propositional inference.

In this paper we demonstrate that the reliance on modified MP can be eliminated for the class of MP algorithms arising from linear programming (LP) relaxations of MAP inference (MAP-LPs) of pairwise MRFs. MAP-LPs approximate the MAP problem as an LP with polynomially many constraints [25], which is therefore tractable, and have several nice properties. First, they yield an upper bound on the MAP value, and can thus be used within branch and bound methods. Second, they provide certificates of optimality, so that one knows if the problem has been solved exactly. Third, the LP can be solved using simple algorithms such as coordinate descent, many of which have a nice message passing structure. [12] Fourth, the LP relaxations can be progressively tightened by adding gradually constraints of a higher order. This has been shown to solve challenging MAP problems [20].

Indeed, it is already known that MAP-LP relaxations of MRFs can be lifted efficiently [13, 3, 14], and the resulting lifted LPs can be solved using any off-the-shelf LP solver[1]. Unfortunately, however, the liftings employed there may not preserve the MRF structure of the underlying LP. That is, if we lift a MAP-LP, we end up with constraints that do not conform to the MAP-LP template as already observed by Bui et al. (see Section 7 in [3]). In turn, existing MP solvers for MAP-LPs such as MPLP and TRW-BP — that have been reported to often solve the MAP-LP significantly faster than generic LP solvers — will not work without modifying them. Doing so, however, takes a lot effort (if it is at all possible): it has do be done for each existing MP approach separately; there is no general methodology for doing this, and the extra coding itself is error prone. Hence this "upgrading methodology" may significantly delay the development of lifted MP approaches. Fortunately, as we demonstrate here, the theory of lifted LPs provides us with a way around these issues. The main insight is that a given MRF induces actually a whole family of MRFs of different sizes sharing essentially the same MAP-LP solution. From these, one can select the smallest one where MAP beliefs can be computed using off-the-shelf MP approaches. These beliefs then are also valid (after a simple transformation) for the original problem. Moreover, this incurs no major overhead: the selected MRF is at most twice as large than the fully lifted factor graph. In this way we eliminate the need for modified MP algorithms.

To summarize, our contributions are two-fold. **(1)** By making use of lifted linear programming, we show that LP-based lifted inference in MRFs can be formulated as ground inference on a reparametrized MRF. **(2)** We give an efficient algorithm that given a ground MRF finds the smallest reparametrized MRF and show that its size is not more than twice the size of the fully lifted model.

[1] A similar approach has been proposed for exact MAP by Noesner *et al.* [15]. Moreover, Sarkhel *et al.* [18] have recently shown that MAP over MLNs can be reduced to MAP over Markov networks if the MLN has very restrictive properties. In contrast our approach is generic for MAP-LP relaxations.

This has several implications for lifted inference. For instance, using MPLP [6] results in the first convergent MP approach for MAP-LP relaxations, and using other MP approaches such as TRW-BP [24] actually spans a whole family of lifted MP approaches. This suggests a novel view on lifted probabilistic inference: *it can be viewed as standard inference in a reparametrized model.*

We proceed as follows. We start off with reviewing MAP-LP basics. Then, we touch upon equitable partitions and lifted LPs, and use them to develop the reparametrization approach. Before concluding we provide empirical illustrations, which support our theoretical results.

2 BACKGROUND

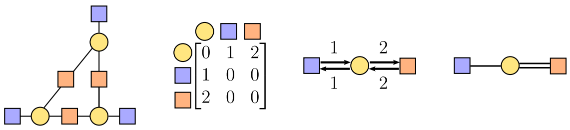
We start off by introducing MAP inference and its LP relaxation. Then we will touch upon equitable partitions and recall how they can be used in lifted linear programming.

MAP Inference in MRFs. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a set of n discrete-valued random variables and let x_i represent the possible realizations of random variable X_i. Markov random fields (MRFs) compactly represent a joint distribution over $\mathbf{X}$ by assuming that it is obtained as a product of functions defined on small subsets of variables [11]. For simplicity, we will restrict our discussion to a specific subset of MRFs, namely Ising models with arbitrary topology[2]. In an Ising model $I = (G, \boldsymbol{\theta})$ on a graph $G = (V, E)$, all variables are binary, i.e., $\mathbf{X}_i \in \{0, 1\}$. Moreover, in an Ising model G must be a *simple* graph, i.e. G must have no self-loops or multiple edges between vertices. The model is then given by: $p(x) \propto \exp\left[\sum_{ij \in E} \theta_{ij} x_i x_j + \sum_i \theta_i x_i\right]$. In the following we will find it convenient to represent Ising models as factor graphs. The factor graph of an Ising model combines the structure and parameters of the model into a single bipartite graph. In this graph we have a variable vertex v_i for each probabilistic variable $\mathbf{X}_i$ and a factor vertex ϕ_i for each θ_i and ϕ_{ij} for θ_{ij}. Moreover, ϕ_i is connected to v_i and ϕ_{ij} to v_i and v_j. While for ground Ising models, the factor graph does not capture any additional information, it makes the presentation of lifted structures simpler and reveals the essence of the conflict between lifting and message-passing. Hence, from now on when we refer to an Ising model $I = (G, \boldsymbol{\theta})$, by G we will mean the corresponding factor graph.

The Maximum a-posteriori (MAP) inference problem is defined as finding an assignment maximizing $p(x)$. This can equivalently be formulated as the following LP

$$\boldsymbol{\mu}^* = \operatorname*{argmax}_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum\nolimits_{ij \in E} \mu_{ij}\theta_{ij} + \sum\nolimits_i \mu_i \theta_i = \boldsymbol{\theta} \cdot \boldsymbol{\mu} \quad (1)$$

[2] The shorter description of MAP-LP for Ising models makes the presentation easier. Our approach, however, can be applied to any pairwise MRF with only minor modifications.



(a) $G, \mathcal{P}$ (b) $\mathrm{DM}(G, \mathcal{P})$ (c) $G/\mathcal{P}$ (d) $G \wr \mathcal{P}$

Figure 1: Lifted Structures: Example of a factor graph and its partitions and quotients. (a) A factor graph G (variables in circles, factors in squares) and its coarsest EP $\mathcal{P}$ represented by the colors. (b) The degree DM matrix of G according to $\mathcal{P}$. (c) The corresponding quotient graph $G/\mathcal{P}$. (d) The factor quotient $G \wr \mathcal{P}$.

where the set $\mathcal{M}(G)$ is known as the marginal polytope [25]. Even though Eq. 1 is an LP, the polytope $\mathcal{M}(G)$ generally requires an exponential number of inequalities to describe [25], and is NP-complete to maximize over. Hence one typically considers tractable relaxations (outer bounds) of $\mathcal{M}(G)$. The outer bounds we consider are equivalent to the standard local consistency bounds typically considered in the literature (e.g., see [25] Eq. 8.32). However, we present them in a slightly different manner, which simplifies our presentation. Define the following set in $[0, 1]^{|V|+|E|}$:

$$L(G) = \left\{ \begin{array}{l} \boldsymbol{\mu} \geq 0 \ , \forall \phi_{ij} \in G : \\ \alpha(ij) \equiv \mu_{ij} \leq \mu_j \ , \ \beta(ij) \equiv \mu_{ij} \leq \mu_i ; \\ \gamma(ij) \equiv \mu_i + \mu_j - \mu_{ij} \leq 1 \end{array} \right\} . \quad (2)$$

The polytope $L(G)$ is sometimes referred to as the local marginal polytope [21]. The vectors with $\{0, 1\}$ coordinates in $L(G)$ are the vertices of the $\mathcal{M}(G)$. In other words $\mathcal{M}(G)$ is the convex hull of $L(G) \cap \{0, 1\}^{|V|+|E|}$. We call the relaxed inference problem over $L(G)$ MAP-LP. Note that whenever $\mathcal{M}(G)$ and $L(G)$ do not coincide, $L(G)$ (which is an outer bound on $\mathcal{M}(G)$) has fractional vertices and the resulting LP may have optima which are not valid assignments. However, all integral points in $L(G)$ correspond to valid assignments, thus if the solution $\boldsymbol{\mu}^*$ happens to be integral, then this $\boldsymbol{\mu}^*$ solves the MAP problem.

Equitable Partitions (EPs) of Graphs and Matrices. Lifted inference approaches essentially work with reduced models by grouping together indistinguishable variables and factors. In other words, they exploit symmetries. For linear programs, Mladenov *et al.* [13, 14] have shown that such symmetries can be formally captured by equitable partitions of weighted graphs and matrices. Since these partitions also play an important part in our argument we will next review the most relevant concepts and results.[3] For an illustration, we refer to Fig. 1.

[3] Note, however, that the definitions we present here are tailored towards bipartite structures (e.g. factor graphs with variables and factors, matrices with rows and columns) for the sake of clarity. They are not the most general ones found in literature.

Let $U = V \cup F$ be a set consisting of two kinds of objects, e.g. the variables and factors of a factor graph as in Fig. 1(a), or the row and column indices of a matrix. A partition $\mathcal{P} = \{P_1, \ldots, P_p\} \cup \{Q_1, \ldots Q_q\}$ is a family of disjoint subsets of U, such that $\bigcup_{i=1}^{p} P_i = V$ and $\bigcup_{i=1}^{q} Q_i = F$. In Fig. 1 the partition is indicated by the colors of the nodes. A convenient data structure for performing algebraic operations using partitions is the incidence matrix $B \in \{0, 1\}^{|U| \times |\mathcal{P}|}$. The incidence matrix shows the assignment of the elements of U to the classes of $\mathcal{P}$ – it has one row for every object and one column for every class. The entry in the row of object u and the column of class P_p is

$$B_{up} = 1 \text{ if } u \in P_p \text{ and } 0 \text{ if } u \notin P_p \ .$$

We shall also make use of the normalized transpose of B, which we denote by $\widehat{B} \in \mathbb{Q}^{|\mathcal{P}| \times |U|}$ and define as

$$\widehat{B}_{pu} = 1/|P_p| \text{ if } u \in P_p \text{ and } 0 \text{ if } u \notin P_p \ .$$

Algebraically, B and $\widehat{B}$ are related as $\widehat{B} = (B^T B)^{-1} B^T$, i.e., $\widehat{B}$ is the left pseudoinverse of B: $\widehat{B} B = I_{|\mathcal{P}|}$.

The partitions we consider will never group elements of V with elements in F. Thus, the matrix B will always be of the form $B = \left(\begin{smallmatrix} B_P & 0 \\ 0 & B_Q \end{smallmatrix}\right)$, where B_P and B_Q correspond to the partitions of V and F respectively. We shall also use the notation $B = (B_P, B_Q)$ to refer to this block diagonal matrix, and similarly $\widehat{B} = (\widehat{B}_P, \widehat{B}_Q)$.

Let $u \in \mathbb{R}^{|U|}$ be a real vector composed as $\mathbf{u} = [\mathbf{c}, \mathbf{b}]^T$, $\mathbf{c} \in \mathbb{R}^{|V|}$, $\mathbf{b} \in \mathbb{R}^{|F|}$. The values of $\mathbf{u}$ can be thought of as labels for the elements of U. We say that a partition $\mathcal{P}$ respects $\mathbf{u}$ if for every $x, y \in U$ that are in the same class of $\mathcal{P}$, we have $u_x = u_y$. Note that if $\mathcal{P}$ respects $\mathbf{u}$, then $(\mathbf{c}^T B_P)_i = |P_i| c_x$ where x is any member of P_i (and similarly for $B_Q, \mathbf{b}$). Moreover, $(\widehat{B}_P \mathbf{c})_i = c_x$ where x is any member of P_i (and similarly for $\hat{B}_Q, \mathbf{b}$).

We next define a special class of partitions of graphs and matrices, which play a central role in our argument. Let us first consider a bipartite graph $G = (V \cup F, E)$. Here V and F are the two sides of the graph, and E are the edges connecting them. The neighbors of a node v in this graph are denoted by $\mathrm{nb}(v)$.

Definition 1 (Equitable partition of a bipartite graph). *An equitable partition of a bipartite graph $G = (V \cup F, E)$ given a vector $\mathbf{u} \in R^{|V|+|F|}$ is a partition $\mathcal{P} = \{P_1, \ldots, P_p, Q_1, \ldots Q_q\}$ of the vertex set V and F such that* **(a)** *for every pair $v, v' \in V$ in some P_m, and for every class Q_n, $|\mathrm{nb}(v) \cap Q_n| = |\mathrm{nb}(v') \cap Q_n|$;* **(b)** *for every pair $f, f' \in F$ in some Q_m, and for every class P_n, $|\mathrm{nb}(f) \cap P_n| = |\mathrm{nb}(f') \cap P_n|$. Furthermore, $\mathcal{P}$ must respect the vector* $\mathbf{u}$.

If we are dealing with matrices, the above definition can be extended. Essentially, we view a matrix $A \in \mathbb{R}^{m \times n}$ as a weighted graph over the set $\{\mathrm{row}[1], \ldots, \mathrm{row}[m]\} \cup$

$\{\mathrm{col}[1], \ldots, \mathrm{col}[n]\}$, where A_{ij} is the weight of edge between $\mathrm{row}[i]$ and $\mathrm{col}[j]$. More precisely:

Definition 2 (Equitable partition of a matrix). *An equitable partition of a matrix $A \in \mathbb{R}^{m \times n}$ given a vector $\mathbf{u}$ is a partition $\mathcal{P} = \{P_1, \ldots, P_p, Q_1, \ldots Q_q\}$ of the sets $V = \{\mathrm{row}[1], \ldots, \mathrm{row}[m]\}$ and $F = \{\mathrm{col}[1], \ldots, \mathrm{col}[n]\}$ s.t.* **(a)** *for every pair $v, v' \in V$ in some P_m, and for every class Q_n, $\sum_{f \in Q_n} A_{vf} = \sum_{f \in Q_n} A_{v'f}$ and* **(b)** *for every pair $f, f' \in F$ in some Q_m, and for every class P_n, $\sum_{v \in P_n} A_{vf} = \sum_{v \in P_n} A_{vf'}$. In addition, $\mathcal{P}$ must respect $\mathbf{u}$.*

Note that Def. 1 is an instance of Def. 2 when we take as A the (biparite) adjacency matrix of a graph G. An illustration of an equitable partition of a graph is given Fig. 1(a).

One notable kind of equitable partitions (EPs) are orbit partitions (OPs) – the partitions that arise under the action of the automorphism group of a graph or matrix. Their role in MAP inference has been studied in [3]. Although OP-based lifting is indeed practical in a number of cases or even the only applicable one, in particular for exact inference approaches, computing them is a GI-complete problem. Because of this we will stick to EPs which are more efficiently computable and yield more reduction (to be discussed shortly). Still, we would like to stress that our result applies to any EP, in particular to OPs.

Using an EP of a graph or a matrix, we can derive condensed representations of that graph or matrix using the partition. This is the essence of lifting: *the reduced representation is as good as the original representation for some computational task at hand, while (potentially) having a significantly smaller size.* A key insight that we exploit here is that there is a one-to-one relationship between EPs of the factor graph of an Ising model (as in Def. 1) and the EPs of its MAP-LP matrix (as in Def. 2).

One useful representation of a graph and its equitable partition is via a degree matrix, as illustrated in Fig. 1(b). The degree matrix, $\mathrm{DM}(G, \mathcal{P})$, has $|\mathcal{P}| \times |\mathcal{P}|$ entries, where each entry represents how members of different classes interact. More precisely, $\mathrm{DM}(G, \mathcal{P})_{ij} = |\,\mathrm{nb}(u) \cap P_j|$, where u is any element of P_i. Due to the bipartiteness of G, this matrix will have the block form $\mathrm{DM}(G, \mathcal{P}) = \left(\begin{smallmatrix} 0 & \mathrm{DV} \\ \mathrm{DF} & 0 \end{smallmatrix}\right)$, where DV represents the relationship of the P-classes to the Q-classes and DF vice-versa. As a shorthand, we use the notation $\mathrm{DM} = (\mathrm{DF}, \mathrm{DV})$. Graphically (see Fig. 1(c)), a degree matrix can be visualized as a quotient graph $G/\mathcal{P}$, which is a directed multi-graph. In $G/\mathcal{P}$ there is a node for every class of $\mathcal{P}$. Given two nodes u, v we have $|\,\mathrm{nb}(u) \cap P_j|, u \in P_i$ many edges going from u to v. $\mathrm{DM}(G, \mathcal{P})$ is essentially the weighted adjacency matrix of $G/\mathcal{P}$.

Later on, we will be interested in the interaction of the factors with variables rather than the other way around. Therefore, we introduce the factor quotient graph $G \wr \mathcal{P}$ of G, which corresponds only to the DF-block of $\mathrm{DM}(G, \mathcal{P})$ as shown in Fig. 1(d). That is, we draw only edges going from factor classes to variable classes, but not the other way around. Moreover, as our MRFs are pairwise, a factor class can have a degree of at most two to any variable class. We will thus not write numbers on top of the arcs, but draw double or single edges. To stay consistent with existing terminology, we call the nodes of $G \wr \mathcal{P}$ corresponding to variable classes of G "supervariables", and factor-class nodes "superfactors". Note that if $\mathcal{P}$ is the OP of G, the resulting factor quotient is the "lifted graph" considered in [3].

Finally, to compute EPs one can use color-passing (also known as "color refinement" or "1-dimensional Weisfeiler-Lehman"). It is a basic algorithmic routine for graph isomorphism testing. It iteratively partitions, or colors, vertices of a graph according to an iterated degree sequence in the following fashion: initially, all vertices get their label in G as color, and then at each step two vertices that so far have the same color get different colors if for some color c they have a different number of neighbors of color c. The iteration stops if the partition remains unchanged. For matrices, a suitable extension was introduced in [7]. The resulting partition is called the *coarsest equitable partition* (CEP) of the graph, and can be computed asynchronously in quasi-linear time $O((n + m) \log n)$ (e.g., see [2]).

Lifted Linear Programming and Lifted MAP-LPs. By computing an EP of the matrix of a linear program (LP) one can derive a smaller but equivalent LP – the "lifted" LP – whose optimal solutions can easily be translated back to a solution of the original LP [13, 7]. This will be key in what follows, and is reviewed below.

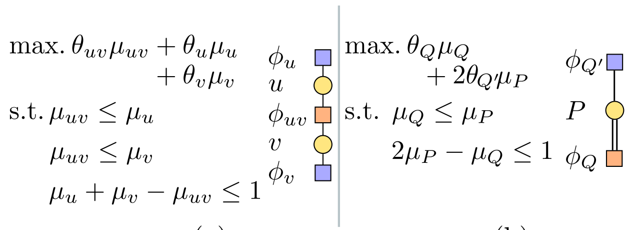
Let $L = (A, \mathbf{b}, \mathbf{c})$ be a linear program, corresponding to the optimization problem $\mathbf{x}^* = \mathrm{argmax}_{A\mathbf{x} \leq \mathbf{b}}\, \mathbf{c}^T \mathbf{x}$.

Theorem 3 (Lifted Linear Programs [7][4]). *Let $\mathcal{P} = \{P_1, \ldots, P_p\} \cup \{Q_1, \ldots, Q_q\}$ be an equitable partition with incidence matrix $B = (B_P, B_Q)$ of the rows and columns of A, which respects the vector $\mathbf{u} = [\mathbf{c}, \mathbf{b}]^T$. Then, $L' = (\widehat{B}_Q A B_P, \widehat{B}_Q \mathbf{b}, B_P^T \mathbf{c})$ is an LP with fewer variables and constraints. The following relates L and L':*

(a) *If $\mathbf{x}'$ is a feasible point in L', then $B_P \mathbf{x}'$ is feasible in L. If in addition $\mathbf{x}'$ is optimal in L', $B_P \mathbf{x}'$ is optimal in L with the same objective value.*
(b) *if $\mathbf{x}$ is a feasible point in L, then $\widehat{B}_P \mathbf{x}$ is feasible in L'. If in addition $\mathbf{x}$ is optimal in L, $\widehat{B}_P \mathbf{x}$ is optimal in L' with the same objective value.*

In previous works, only part **(a)** has been exploited. That is, as illustrated in Fig. 2, given any LP we construct L' using equitable partitions, solve it (often faster than the original one), and finally transfer the solution to the larger problem by virtue of **(a)** above. This "standard" way applies to MAP-LPs as follows (see [13, 7] for more details).

[4]We restrict the theorem to what is relevant for the discussion.



$$\begin{aligned} \max.\ & \theta_{uv}\mu_{uv} + \theta_u\mu_u + \theta_v\mu_v \\ \text{s.t. } & \mu_{uv} \le \mu_u \\ & \mu_{uv} \le \mu_v \\ & \mu_u + \mu_v - \mu_{uv} \le 1 \end{aligned}$$

(a)

$$\begin{aligned} \max.\ & \theta_Q\mu_Q + 2\theta_Q\mu_P \\ \text{s.t. } & \mu_Q \le \mu_P \\ & 2\mu_P - \mu_Q \le 1 \end{aligned}$$

(b)

Figure 2: Illustration of lifting MAP-LPs. (a) A factor graph and the corresponding MAP-LP; (b) A factor quotient and the corresponding lifted MAP-LP. This example also illustrates that the lifted MAP-LP in (b) is not amenable to standard MP anymore, since the constraint $2\mu_P - \mu_Q \le 1$ does not appear in standard MAP-LPs. However, note that the lifted structure is identical to the one in Fig. 1(d), yet the original factor graph in (a) above is smaller. That is, factor graphs of different sizes may share the (structurally) same lifted MAP-LP.

$$\begin{tikzcd} \text{MAP-LP}(J) \arrow[d, "\boldsymbol{\tau}' = \widehat{B}_P^J \boldsymbol{\tau}"', "\text{lift}"] & \text{MAP-LP}(I) \\ \text{LMAP-LP}(J) \arrow[r, "\text{identical}", "\boldsymbol{\tau}' = \boldsymbol{\mu}'"'] & \text{LMAP-LP}(I) \arrow[u, "\text{unlift}", "\boldsymbol{\mu} = B_P^I \boldsymbol{\mu}'"'] \end{tikzcd}$$

Figure 3: Commutative diagram established by Thm. 6 underlying our reparametrization approach to lifted inference.

Given an MRF with graph G and parameter $\boldsymbol{\theta}$, which we denote by $I = (G, \boldsymbol{\theta})$, denote its standard MAP-LP relaxation by the LP defined via $(A, \mathbf{b}, \mathbf{c})$. To obtain a potentially smaller LP, we calculate the equitable partition of the LP, and its corresponding $B, \widehat{B}$ matrices. The LP defined via $(\widehat{B}_Q A B_P, \widehat{B}_Q \mathbf{b}, B_P^T \mathbf{c})$ is then equivalent to the original MAP-LP in the sense of Thm. 3. We refer to this LP as LMAP-LP(I).

Unfortunately, as we will show in the next section, LMAP-LP(I) is not a standard MAP-LP. In turn, it is not amenable to standard message passing based MAP-LP solvers. Fortunately, by making heavily use of part **(b)** of Thm. 3 as well, we will show how to produce LPs that have this special structure. This results "lifting by reparametrization".

3 LIFTING BY REPARAMETRIZATION

To introduce the "lifting by reparametrization" approach we proceed in two steps. First, we introduce the class of all MRFs whose MAP-LPs are equivalent, in the sense that solving one such LP results in a solution to all LPs in the class. Then, we will show how to construct the smallest such equivalent MRFs for a given MRF in the class.

LP-equivalence of MRFs. We start off by discussing the structure of the lifted MAP-LP (LMAP-LP). The main purpose is to illustrate why lifting generally does not preserve the message-passing structure of the LP (see also Fig. 2). Then, as an alternative, we introduce an equivalence theorem in the spirit of Thm. 3. As illustrated in Fig. 3, instead of relating ground (original) and lifted LPs, it relates ground LPs of different sizes that have the same lifting. This is the main insight underlying our "lifting by reparametrization" approach: *instead of lifting the ground MAP-LP of an MRF at hand, we replace it by the ground LP of another equivalent MRF, hopefully of much smaller size.* In the following we will now formally define what "equivalence" means here.

Our starting point is to note that LMAP-LP(I) only depends on I via the structure of $G \wr \mathcal{P}$. To provide the formal result, we need a few more notations. First, the graph $G \wr \mathcal{P}$ has nodes corresponding to groups of variables in the original factor graph, and nodes corresponding to groups of factors in the original factor graph. We denote those by $V(G \wr \mathcal{P})$ and $F(G \wr \mathcal{P})$ respectively. Second, $G \wr \mathcal{P}$ is a multigraph and may have multiple edges between its nodes. We define $\mathrm{nb}(Q)$ to be the neighbors of $Q \in F(G \wr \mathcal{P})$ in $G \wr \mathcal{P}$, with repetitions. In other words, if there are two edges between $P, Q \in G \wr \mathcal{P}$, then $\mathrm{nb}(Q)$ will contain P twice.

Proposition 4. *Let $\mathcal{P} = \{P_1, \ldots, P_p\} \cup \{Q_1, \ldots, Q_q\}$ be an equitable partition of the variables and factors of the MRF specified by $I = (G, \boldsymbol{\theta})$. Let $G \wr \mathcal{P}$ be the factor quotient of I. Then, LMAP-LP(I) can be written as:*

$$\boldsymbol{\mu}'^* = \operatorname*{argmax}_{\boldsymbol{\mu}' \in L'(G)} \sum_{P \in V(G \wr \mathcal{P})} \theta_P |P| \mu_P + \sum_{Q \in F(G \wr \mathcal{P})} \theta_Q |Q| \mu_Q \,.$$

Where θ_P, θ_Q are the parameters of I for the corresponding partition elements. The constraints $L'(G)$ are defined as the set of $\boldsymbol{\mu} \ge 0$ such that:

$$\left\{ \begin{array}{l} \forall P, P' \in V(G \wr \mathcal{P}), Q \in F(G \wr \mathcal{P}) \\ \qquad \textit{s.t. } P, P' \in \mathrm{nb}(Q): \\ \alpha'(Q) \equiv \mu_Q \le \mu_P \ ; \ \beta'(Q) \equiv \mu_Q \le \mu_{P'}; \\ \gamma'(Q) \equiv \mu_P + \mu_{P'} - \mu_Q \le 1 \end{array} \right\} \tag{3}$$

Proof. We omit a detailed proof of this proposition due to space restrictions. Essentially, the argument is that the reformulation of Sec. 7 in [3] holds for any equitable partition, not just the orbit partition of an MRF. In this case, the Q-classes generalize edge orbits while the P-classes generalize variable orbits. □

This proposition tells us that we can construct $L'(G)$ by the following procedure: (1) we instantiate an LP variable μ_P for every variable class $P \in V(G \wr \mathcal{P})$ (i.e., every supervariable) (2) we instantiate an LP variable μ_Q for every factor class $Q \in F(G \wr \mathcal{P})$ (i.e. superfactor); (3) for every pair

Algorithm 1: Solving MRF I using an equivalent MRF J

Input: Ground MRF I and LP-equivalent MRF J
Output: MAP-LP(I) solution $\boldsymbol{\mu}$

Solve MRF J, i.e., compute $\boldsymbol{\tau} = \operatorname{argmax}_{\boldsymbol{\mu}} \text{MAP-LP}(J)$;
Lift the solution $\boldsymbol{\tau}$ to LMAP-LP(J). That is, compute $\boldsymbol{\tau}' = \widehat{B}_P^J \boldsymbol{\tau}$ (Thm. 3(b));
Recover solution of I, i.e., compute $\boldsymbol{\mu} = B_P^I \boldsymbol{\tau}'$ (Thm. 3(a));

of classes P, Q, if some ground variable $x_i \in P$ is adjacent to some ground factor $\phi_{ij} \in Q$, we add the constraint $\mu_Q \leq \mu_P$. For every triplet P, P', Q such that there exist $x_i \in P, x_j \in P'$ adjacent to $\phi_{ij} \in Q$, we add the constraint $\mu_P + \mu_{P'} - \mu_Q \leq 1$.

Observe that the factor quotient graph $G \wr \mathcal{P}$ actually contains exactly the necessary and sufficient information to construct $L'(G)$: it gives us the number of classes and the relations between them. Hence, it would seem that $L'(G)$ is just $L(G \wr \mathcal{P})$, and we are done. Unfortunately this is not exactly the case, and we have to be a little bit more careful.

Recall our running example from Fig. 2. There can be a factor ϕ_{ij} in some Q, whose adjacent variables x_i, x_j fall into the same class, $P = P'$. In terms of constraints, the corresponding triplet P, P', Q, with $P = P'$ yields the constraint $2\mu_P - \mu_Q \leq 1$. Graphically, this situation occurs whenever $G \wr \mathcal{P}$ contains a double edge. This also happens in our running examples (see Fig. 2(b)). Unfortunately, such constraints have no analogue in MAP-LP(I)!

How can we deal with this? Assume for the moment that for any ground factor ϕ_{ij}, $\mathcal{P}(i) \neq \mathcal{P}(j)$, in other words $G \wr \mathcal{P}$ happens to be a simple graph (no edge connects at both ends to the same vertex, and there is no more than one edge between any two different vertices). Then we can compute a new weight vector $\boldsymbol{\theta}' \in \mathbb{R}^q$ as $\theta'_Q = |Q|\theta_Q,\ \theta'_P = |P|\theta_P$ (*cf.* Eq. 3).In this case, the MRF $I' = (G \wr \mathcal{P}, \boldsymbol{\theta}')$ would indeed be a smaller MRF, whose MAP-LP is identical to the LMAP-LP of I. This enables us to view lifting as reparametrization: **(1)** we compute $G \wr \mathcal{P}$ from G; **(2)** instead of solving LMAP-LP(I), we solve MAP-LP(I') using any solver we want, including message-passing algorithms such as MPLP, TRWBP, among others; **(3)** because of the equivalence, we treat the solution of MAP-LP(I') as a solution of LMAP-LP(I') and unlift it using Thm. 3(a).

While our assumption does not hold in general (see e.g. Fig. 1) — and we will indeed account for it below — the procedure just outlined above is the main idea underlying "lifting by reparametrization" method. Since the LMAP-LP of I will potentially contain constraints such as $2\mu_P - \mu_Q \leq 1$, it will not be the MAP-LP of any simple graph. So instead, we will look for something else, namely a proper (potentially much smaller) MRF J, where instead of LMAP-LP(I) =MAP-LP(J) we ask that LMAP-LP(I) = LMAP-LP(J) . We call any pair of MRF where this holds LP-equivalent.

Definition 5 (LP equivalent MRFs). *Two MRFs $I = (G, \boldsymbol{\theta}^I)$ and $J = (H, \boldsymbol{\theta}^J)$ having simple graphs are LP-equivalent if we can find an equitable partition $\mathcal{P}$ of G with incidence matrix $B = (B_P, B_Q)$ and an equitable partition $\mathcal{P}'$ of H with incidence matrix $B' = (B'_P, B'_Q)$ such that LMAP-LP(I) $:= (\widehat{B}_Q^T A B_P, \widehat{B}_Q^T \mathbf{b}, B_P^T \mathbf{c}) = ((\widehat{B}'_Q)^T A' B'_P, (\widehat{B}'_Q)^T \mathbf{b}', (B'_P)^T \mathbf{c}') =:$ LMAP-LP(J).*

Then, we apply the lifted equivalence of Thm. 3(b) and are done. As summarized in Alg. 1, we solve the smaller MAP-LP(J) using any MRF-structure-aware LP solver. We obtain an optimal solution of LMAP-LP(J) using $B_{P'}^T$ as prescribed by Thm. 3(b). Due to the lifted equivalence, this solution is also a solution of LMAP-LP(I), hence we recover (or "unlift") the solution with respect to I using B_P. In doing so, we end up with an optimal solution of MAP-LP(I). This procedure is outlined in Fig. 3. We will shortly prove its soundness.

Theorem 6. *Let I and J be two LP-equivalent MRFs of possibly different sizes. Then,* **(A)** *if $\boldsymbol{\tau}$ is feasible in MAP-LP(J), $\boldsymbol{\mu} = B_P \widehat{B}'_P \boldsymbol{\tau}$ is feasible in MAP-LP(I). Moreover, if $\boldsymbol{\tau}$ is optimal, $\boldsymbol{\mu}$ is optimal as well.* **(B)** *if $\boldsymbol{\mu}$ is feasible in MAP-LP(I), $\boldsymbol{\tau} = B'_P \widehat{B}_P \boldsymbol{\mu}$ is feasible in MAP-LP(J). Moreover, if $\boldsymbol{\mu}$ is optimal, $\boldsymbol{\tau}$ is optimal as well.*

Proof. We prove only **(A)** due to the symmetry of the statement. Let $\boldsymbol{\tau}$ be feasible in MAP-LP(J). By Thm. 3(b), $\boldsymbol{\tau}' = B'_P \boldsymbol{\tau}$ is feasible in LMAP-LP(J). Due to LP-equivalence, LMAP-LP(J) = LMAP-LP(I), $\boldsymbol{\tau}'$ is also a solution to LMAP-LP(I). Now, we unlift $\boldsymbol{\tau}'$ with respect to LMAP-LP(I). Due to Thm. 3(b), $\boldsymbol{\mu} = B_P(\widehat{B}'_P)\boldsymbol{\tau}$ is feasible in MAP-LP(I). Moreover, if $\boldsymbol{\tau}$ is optimal in MAP-LP(J), Thm. 3 tells us that optimallity will hold throughout the entire chain of LPs. □

To summarize our argument so far, Thm. 6 provides us with a way to exploit the MAP-LP equivalence between MRFs of different sizes. What is still missing is a way to efficiently construct such smaller LP-equivalent MRFs as input to Alg. 1. We will now address this issue.

Finding equivalent MRFs. So far we discussed the equivalence of MRFs of different sizes in terms of their (lifted) MAP-LPs. Making use of our result, however, requires efficient algorithm to find LP-equivalent MRFs of considerably smaller size. Given an MRF I and its EP, Alg. 2 finds the smallest LP-equivalent MRF I' in linear time. Next to illustrating Alg. 2 and proving that it is sound, we will also show that the size of I' is at most $2|G \wr \mathcal{P}|$.

Let $I = (G, \boldsymbol{\theta})$ be an MRF and $\mathcal{P}$ be an EP of its variables and factors. We will introduce the algorithm in two

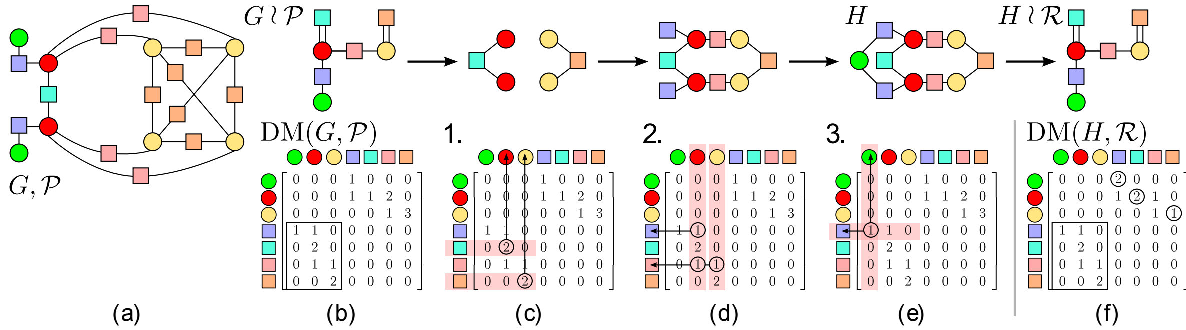


Figure 4: Illustration of "lifted inference by reparametrization" as summarized in Alg. 2 and used as input to Alg. 1. (a) input: a graph G and its coarsest equitable partition $\mathcal{P}$; (b) the factor quotient $G \wr \mathcal{P}$ and the corresponding degree matrix; (c) **Step (A)**: representing all degree 2 factors and the corresponding degree 2 variables. (d) **Step (B)**: instantiating degree 1 factors adjacent to degree 2 variables; (e) **Step (C)**: instantiating the remaining classes. After this step we terminate and output the result H; (f) soundness: $H \wr \mathcal{R}$ is identical to $G \wr \mathcal{P}$. Observe that the upper-right corner of $\mathrm{DM}(G, \mathcal{P})$ is different from that of $\mathrm{DM}(H, \mathcal{R})$. (Best viewed in color.)

steps: first, we discuss how to obtain a simple graph G' which gives the structure of an LP-equivalent MRF, i.e. $L'(G) = L'(G')$. Then we show how to find weights $\boldsymbol{\theta}'$ for this graph, such that the MRF $I' = (G', \boldsymbol{\theta}')$ is LP-equivalent to I. Finally, we will show correctness and minimality of our approach.

Recall that in Eq. 3 the lifted Ising polytope of LMAP-LP(I) is fully defined by the factor quotient $G \wr \mathcal{P}$. Hence, a necessary and sufficient condition for LP-equivalence (regarding the constraints of the LP, we will deal with the objective shortly) in MRFs is that the corresponding graphs exhibit the same factor quotients for some equitable partitions. Thus, the problem of finding an LP-equivalent structure boils down to finding G' such that $G \wr \mathcal{P} = G' \wr \mathcal{P}'$ for some partition $\mathcal{P}'$. Moreover, to maximize the compression, we want $\mathcal{P}$ to be the coarsest EP of G (resulting in the smallest possible $G \wr \mathcal{P}$) and that G' is the smallest possible LP-equivalent MRF. Let us now see how to find G'. As a running example, we will use the factor graph in Fig 4(a).

Suppose $G \wr \mathcal{P}$ is given, e.g. computed using color-refinement. For our running example, $G \wr \mathcal{P}$ is shown in Fig. 4(b). Let us divide the superfactors and supervariables of $G \wr \mathcal{P}$ into classes based on their connectivity. A superfactor connected to a supervariable via a double edge is called a $(\mathbf{2})$**-superfactor**. In Fig. 4(b), these are the cyan and orange superfactors. Correspondingly, we call a variable connected to a superfactor via a double edge a $(\mathbf{2})$**-supervariable** (red and yellow in Fig. 4(b)). Next, a superfactor connected to at least one (2)-supervariable via a *single* edge is called a $(\mathbf{1}, \mathbf{2})$**-superfactor** (violet and pink in Fig. 4(b)). Finally, all other superfactors and supervariables are $(\mathbf{1})$**-superfactors** and $(\mathbf{1})$**-supervariables** respectively (e.g. the green supervariable).

We then compute G' in the following way as also illustrated in Fig. 4(c)-(e). We start with an empty graph. Then, **Step (A)** as illustrated in Fig. 4(c) consists of adding for every (2)-superfactor in $G \wr \mathcal{P}$ exactly one representative factor to G'. Furthermore, for every (2)-supervariable, we add two representatives in G' and connect them to the corresponding (2)-superfactor representatives whenever the supernodes they represent are connected in $G \wr \mathcal{P}$. In **Step (B)**, see Fig. 4(d), for every $(1, 2)$-superfactor, we instantiate two representatives. Moreover, for every (2)-supervariable (all of them are already represented in G'), we match the two $(1, 2)$-superfactor representatives to the two (2)-supervariable representatives whenever the represented supernodes are connected in $G \wr \mathcal{P}$. Finally, **Step (C)** as shown in Fig. 4(e) introduces one representative for every other supernode and connects it to other representatives based on $G \wr \mathcal{P}$. If it happens that the represented supernode is connected to a (2)-supervariable or $(1, 2)$-superfactor in $G \wr \mathcal{P}$, we connect the representative to both representatives of the corresponding neighbor.

This is summarized in Alg. 2 and provably computes a minimal structure of an LP-equivalent MRF. Finally, we must compute a parameter vector for I' to facilitate LP-equivalence. Suppose $\mathcal{P}'$ is the EP of G' induced by Alg. 2 (the partition which groups nodes in G' together if they represent the same supernode of $G \wr \mathcal{P}$). Let Q be any factor class in $\mathcal{P}$ and Q' be the corresponding class in $\mathcal{P}'$. We then compute the weight $\theta_{Q'}$ of the factors $\phi' \in Q'$ of I' as

$$\theta'_{Q'} = (|Q|/|Q'|)\theta_Q \; , \tag{4}$$

where θ_Q is the weight associated with the class Q in $\mathcal{P}$ (recall Prop. 4). We now argue that the resulting Ising model $I' = (G', \boldsymbol{\theta}')$ is LP-equivalent to $I = (G, \boldsymbol{\theta})$.

Theorem 7 (Soundness). *$I' = (G', \boldsymbol{\theta}')$ as computed above is LP-equivalent to $I = (G, \boldsymbol{\theta})$.*

Proof. Following Def. 5 we must show that given I and its EP $\mathcal{P}$, there is a partition $\mathcal{P}'$ of I' such that the lifted

Algorithm 2: Computing the smallest LP-equivalent MRF.

Input: Fully lifted factor graph $G \wr \mathcal{P}$ of G

Output: G' such that $\exists \mathcal{P}' : G' \wr \mathcal{P}' = G \wr \mathcal{P}$.

Initialize $G' \leftarrow \emptyset$, i.e., the empty graph;

/* **Step (A)** Treat double edges */

for *every (2)-superfactor Q in $G \wr \mathcal{P}$ with neighboring (2)-supervariable P* **do**

  Add a factor q representing Q to G';

  Add two variables p, p' representing P in G' and connect them to the factor q;

end

/* **Step (B)** To preserve degrees, treat now single edges in G' incident to double edges */

for *every (1,2)-superfactor Q in $G \wr \mathcal{P}$* **do**

  Add two factors q, q' representing Q to G';

  Connect q to p and q' to p' where p, p' are the representatives of a (2)-supervariable P in $G \wr \mathcal{P}$ that is neighboring Q;

end

/* **Step (C)** Add remaining nodes and edges to G' */

for *all supervariables variables P and superfactor Q in $G \wr \mathcal{P}$ not represented in G' so far* **do**

  Add a single variable p resp. factor q to G';

  Connect p to all representatives of superfactor Q neighboring to P in $G \wr \mathcal{P}$;

end

LPs are equal. We take the partition $\mathcal{P}'$ to be the one induced by Alg. 2. $\mathcal{P}'$ is equitable on G' by construction: we can go through Alg. 2 to verify that every two nodes in G' representing the same supernode of $G \wr \mathcal{P}$ are connected to the same number of representatives of every other supernode of $G \wr \mathcal{P}$ (we omit this due to space restrictions). Now, to show that LMAP-LP(I) has the same constraints as LMAP-LP(I'), we need $G \wr \mathcal{P} = G' \wr \mathcal{P}'$. To see that this holds, observe that Alg. 2 connects p to q in G' if only if P is connected to Q in $G \wr \mathcal{P}$: if Q is a (2)-superfactor, P is a (2)-supervariable – q will be connected to p in **Step (A)**. If P is a (2)-supervariable and Q is $(1,2)$-superfactor, p and q will be connected in **Step (B)**. If Q is $(1,2)$- of a (1)-superfactor and P is a (1)-supervariable, p and q will be connected in **Step (C)**. There are no other possible combinations. Hence, as P' consists of all representatives of P and Q' consists of all representatives of Q, P' and Q' are connected in $G' \wr \mathcal{P}'$ iff P is connected to Q. Moreover, representatives of (2)-superfactors are the only ones connected to two representatives of the same supervariable in G', hence Q' is connected to P' via a double edge in $G' \wr \mathcal{P}'$ if and only if Q is connected to P via a double edge in $G \wr \mathcal{P}$.

Next, we argue that the objectives of the lifted LPs are the same. Using the parameters calculated with Eq. 4, the objective of LMAP-LP(I') is $\sum_{Q' \in \mathcal{P}'} |Q'| \theta_{Q'} \mu_{Q'} = \sum_{Q' \in \mathcal{P}'} |Q'|(|Q|/|Q'|) \theta_Q \mu_{Q'} = \sum_{Q' \in \mathcal{P}'} |Q| \theta_Q \mu_{Q'} = \sum_{Q \in \mathcal{P}} |Q| \theta_Q \mu_Q$. Observe that the final term is exactly the objective of LMAP-LP(I) as given by Prop. 4. We conclude LMAP-LP(I) $=$ LMAP-LP(I'). $\square$

We have thus shown that Alg. 2 and Eq. 4 together produce an LP-equivalent MRF. We will now show that this MRF is the smallest LP-equivalent MRF to the original.

Theorem 8 (Minimality). *Let $I = (G, \boldsymbol{\theta})$ be an Ising model and an $I' = (G', \boldsymbol{\theta}')$ be computed as above. Then there is no other LP-Equivalent MRF with less factors or less vertices than G'. Moreover, $|V(G')| \leq 2|V(G \wr \mathcal{P})|$ and G' and $|E(G')| \leq 2|E(G \wr \mathcal{P})|$, i.e., the size of I' is at most twice the size of the fully lifted model.*

Proof. Let H be any graph with the same factor quotient as G. Then, let Q be a (2)-superfactor in $G \wr \mathcal{P}$ adjacent to some (2)-supervariable P. Due to equivalence, Q' is a (2)-superfactor in $H \wr \mathcal{P}'$ as well and P' is a (2)-supervariable. Hence, the class $P' \in \mathcal{P}'$ must have at least two ground variables from H. Next, let Q be a $(1,2)$-factor in $G \wr \mathcal{P}$ adjacent to a (2)-supervariable. Analogously, Q' is a $(1,2)$-factor in $H \wr \mathcal{P}'$ and P' is a (2)-supervariable. As we have established P' must have at least two ground elements in H. Since P' is connected to Q' via a single edge, the same holds on the ground level: any $p \in P'$ is connected to $q \in Q'$ via a single edge. This means that there are at least as many $q \in Q'$ as there are $p \in P'$, that is, at least two. All other supernodes must have at least one representative. These conditions are necessary for any LP-equivalent H.

Now, let G' be computed from Alg. 2 and $\mathcal{P}'$ be the corresponding partition. To see why G' is minimal, observe that G' has *exactly* two representatives of any (2)-supervariable in $G \wr \mathcal{P}$ (step 1) and *exactly* two representatives of any $(1,2)$-superfactor (step 2). All other supernodes have *exactly* one representative (steps 1 and 3). Therefore, G' meets the conditions with equality and is thus minimal. Finally, since we represent any supernode of $G \wr \mathcal{P}$ by at most 2 nodes in G', G' can have at most twice as many factors and variables as $G \wr \mathcal{P}$. $\square$

Since Alg. 2 makes only one pass over the lifted factor graph, the overall time to compute the LP-equivalent MRF (which is then input to Alg. 1) is dominated by color-refinement, which is quasi-linear in the size of G.

4 EMPIRICAL ILLUSTRATION

The empirical illustration of our theoretical results investigates two questions. **(Q1)** Is reparametrization comparable to lifted MAP-LPs in terms of time and quality when using LP solvers? **(Q2)** If so, can lifted MPLP and TRW by reparametrization pay-off when solving MAP-LPs? And finally, **(Q3)** how does reparametrization behave in the presence of approximate evidence?

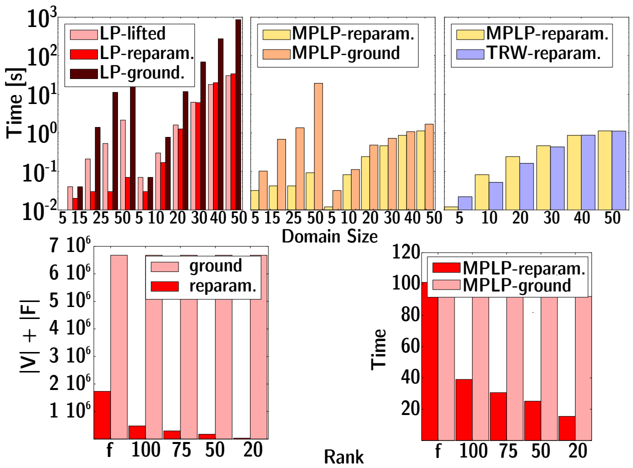


Figure 5: Experimental illustration (for ease of comparison the first three plots have the same scales). From top to bottom, from left to right: (a) Comparison (domain size vs. end-to-end running time) of solving LPs using GLPK (ground, fully lifted, and reparametrization). The first block of domain sizes $(5, 15, 25, 50)$ are from the friends-smoker MLN; the second block $(5, 10, \ldots, 50)$ is on the CORA MLN. (b) Performance of MPLP (ground vs. reparametrization) on the same MLNs (same block structure). (c) Comparison (domain size vs. end-to-end running time) of TRW and MPLP by reparametrization on CORA. (d,e) Model sizes for exact evidence ("f") and approximations of ranks 100 to 20 and running times.

To this aim we implemented the reparametrization approach on a single Linux machine (4×3.4 GHz cores, 32 GB main memory) using Python and C/C++. For evaluation we considered three sets of MRFs. One was generated from grounding a modified version of a Markov Logic Network (MLN) used for entity resolution on the CORA dataset. Five different MRFs were generated by grounding the model for 5, 10, 20, 30, 40 and 50 entities, having 960, 4081, 13933, 27850, 4699 and 76274 factors respectively. The second set was generated from a pairwise version of the friends-smokers MLN [4] for 5, 15, 25 and 50 people, having 190, 1620, 4450 and 17650 factors. The third set considers a simple $\mathtt{fr}(\mathtt{X}, \mathtt{Y}) \Rightarrow (\mathtt{sm}(\mathtt{X}) \Leftrightarrow \mathtt{sm}(\mathtt{Y}))$ rule (converted to a pairwise MLN) where we used the `link_common` observations from the "Cornell" dataset as evidence for `fr`. Then we computed different low-rank approximations of the evidence using [23] .

In all cases, there were only few additional factors due to treating double edges. What is more interesting are the running times and overall performances. Fig. 5(a) shows the end-to-end running time for solving the corresponding ground, (fully) lifted, and reparametrized LPs using GLPK. As one can see, reparametrization is competitive to lifted linear programming (LLP) in time. Actually, it can even save time since it runs directly on the factor graph and not on the LP matrix — which is larger than the factor graph — for discovering symmetries. Moreover, in all cases the same objective was achieved, that is, reparametrization does not sacrifice quality. In turn, question **(Q1)** can clearly be answered affirmatively. Fig. 5(b) summarizes the performance of MPLP on the reparametrized models. As one can see, MPLP can be significantly faster than LLP for solving MAP-LPs without sacrificing the objective; it was always identical to the LP solutions. To illustrate than one may also run other LP-based message-passing solvers, Figs. 5(c) summarizes the performance of TRW on CORA. As one can see, lifting TRW by reparametrization is possible and differences in time are likely due to initialization, stopping criterion, etc. In any case, question **(Q2)** can clearly be answered affirmatively. All results so far show that lifted LP-based MP solvers can be significantly faster than generic LP solvers. Figs. 5(d,e) summarize the results for low-rank evidence approximation. As one can see in (d), significant reduction in model size can be achieved even at rank 100, which in turn can lead to faster MPLP running times (e). For each low-rank model, the ground and the reparametrized MPLP achieved the same objective. Plot (e), however, omits the time for performing BMF. It can be too costly to first run BMF canceling the benefits of lifted LP-based inference (in contrast to exact inference as in [23]). Nevertheless, w.r.t. **(Q3)** these results illustrate that evidence approximation can result in major speed-ups.

5 CONCLUSIONS

In this paper, we proved that lifted MAP-LP inference in MRFs with symmetries can be reduced to MAP-LP inference in standard models of reduced size. In turn, we can use any off-the-shelf MAP-LP inference algorithm — in particular approaches based on message-passing — for lifted inference. This incurs no major overhead: for given evidence, the reduced MRF is at most twice as large than the corresponding fully lifted MRF. By plugging in different existing MAP-LP inference algorithms, our approach yields a family of lifted MAP-LP inference algorithms. We illustrated this empirically for MPLP and tree-reweighted BP. In fact, running MPLP yields the first provably convergent lifted MP approach for MAP-LP relaxations. More importantly, our result suggests a novel view on lifted inference: *lifted inference can be viewed as standard inference in a reparametrized model*. Exploring this view for marginal inference as well as for branch-and-bound MAP inference approaches are the most attractive avenue for future work.

Acknowledgments: The authors would like to thank the anonymous reviewers and Udi Apsel for their feedback and Guy Van den Broeck for providing the evidence approximation code. This research was partly supported by the German-Israeli Foundation (GIF) for Scientific Research and Development, 1180-218.6/2011, and by the German Science Foundation (DFG), KE 1686/2-1.

References

[1] B. Ahmadi, K. Kersting, M. Mladenov, and S. Natarajan. Exploiting symmetries for scaling loopy belief propagation and relational training. *Machine Learning*, 92(1):91–132, 2013.

[2] C. Berkholz, P. Bonsma, and M. Grohe. Tight lower and upper bounds for the complexity of canonical colour refinement. In *Proceedings of the 21st Annual European Symposium on Algorithms*, 2013.

[3] H.H. Bui, T.N. Huynh, and S. Riedel. Automorphism groups of graphical models and lifted variational inference. In *Proc. of the 29th Conference on Uncertainty in Artificial Intelligence (UAI-2013)*, 2013.

[4] D. Fierens, K. Kersting, J. Davis, J. Chen, and M. Mladenov. Pairwise markov logic. In *Postproc. of the 22nd International Conference on Inductive Logic Programming (ILP-2012).*, 2012.

[5] L. Getoor and B. Taskar, editors. *An Introduction to Statistical Relational Learning*. MIT Press, 2007.

[6] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for map LP-relaxations. In *Proc. of the 21st Annual Conf. on Neural Inf. Processing Systems (NIPS)*, 2007.

[7] M. Grohe, K. Kersting, M. Mladenov, and E. Selman. Dimension reduction via colour refinement. *CoRR*, abs/1307.5697, 2013.

[8] F. Hadiji and K. Kersting. Reduce and re-lift: Bootstrapped lifted likelihood maximization for map. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI)*, 2013.

[9] K. Kersting. Lifted probabilistic inference. In *Proceedings of ECAI-2012*. IOS Press, 2012.

[10] K. Kersting, B. Ahmadi, and S. Natarajan. Counting Belief Propagation. In *Proc. of the 25th Conf. on Uncertainty in Artificial Intelligence (UAI–09)*, 2009.

[11] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[12] O. Meshi, T. Jaakkola, and A. Globerson. Convergence rate analysis of MAP coordinate minimization algorithms. In *Advances in Neural information Processing Systems 25*. MIT Press, 2012.

[13] M. Mladenov, B. Ahmadi, and K. Kersting. Lifted linear programming. In *15th Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2012)*, pages 788–797, 2012. Volume 22 of JMLR: W&CP 22.

[14] M. Mladenov, A. Globerson, and K. Kersting. Efficient lifting of map lp relaxations using k-locality. In *17th Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2014)*, 2014. JMLR: W&CP volume 33.

[15] J. Noessner, M. Niepert, and H. Stuckenschmidt. Rockit: Exploiting parallelism and symmetry for map inference in statistical relational models. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI)*, 2013.

[16] D. Poole. First-Order Probabilistic Inference. In *Proc. of the 18th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 985–991, 2003.

[17] M. Richardson and P. Domingos. Markov Logic Networks. *Machine Learning*, 62:107–136, 2006.

[18] S. Sarkhel, D. Venugopal, P. Single, and V. Gogate. Lifted map inference for markov logic networks. In *17th Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2014)*, 2014. JMLR: W&CP volume 33.

[19] P. Singla and P. Domingos. Lifted First-Order Belief Propagation. In *Proc. of the 23rd AAAI Conf. on Artificial Intelligence (AAAI-08)*, pages 1094–1099, Chicago, IL, USA, July 13-17 2008.

[20] D. Sontag, D.K. Choe, and Y. Li. Efficiently searching for frustrated cycles in map inference. In *Proc. of the 28th Conference on Uncertainty in Artificial Intelligence (UAI-2012)*, pages 795–804, 2012.

[21] D. Sontag and T. Jaakkola. New outer bounds on the marginal polytope. In *Proc. of the 21st Annual Conference on Neural Information Processing Systems (NIPS-2007)*, 2007.

[22] G. Van den Broeck, A. Choi, and A. Darwiche. Lifted relax, compensate and then recover: From approximate to exact lifted probabilistic inference. In *Proc. of the 28th Conf. on Uncertainty in Artificial Intelligence (UAI)*, 2012.

[23] G. Van den Broeck and A. Darwiche. On the complexity and approximation of binary evidence in lifted inference. In *Proc. of the 27th Annual Conf. on Neural Information Processing Systems (NIPS)*, 2013.

[24] M.J. Wainwright, T. Jaakkola, and A.S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717, 2005.

[25] M.J. Wainwright and M.I. Jordan. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1-2):1–305, January 2008.