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: $f_{X,Y} \Rightarrow (sm(X) \leftrightarrow sm(Y))$ 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 formula 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 aris-
ing 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 \(X = (X_1, X_2, \ldots, X_n)\) 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 \(X\) 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 \(I = (G, \theta)\) on a graph \(G = (V, E)\), all variables are binary, i.e., \(X_i \in \{0, 1\}\). 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:

\[
p(x) \propto \exp \sum_{ij \in E} \theta_{ij} x_i x_j + \sum_i \theta_i x_i
\]

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 \(X_i\) and a factor vertex \(\phi_{ij}\) for each \(\theta_{ij}\). Moreover, \(\phi_{ij}\) is connected to \(v_i\) and \(\phi_{ij}\) 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 \(I = (G, \theta)\), 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

\[
\mu^* = \arg \max_{\mu \in \mathcal{M}(G)} \sum_{ij \in E} \mu_{ij} \theta_{ij} + \sum_i \mu_i \theta_i = \theta \cdot \mu
\]
and the resulting LP may have optima which are not valid
marginals. The vectors with 
\(\mathcal{M}(G)\) as an outer bound on 
\(\mathcal{M}\), which is an outer bound on 
\(\mathcal{M}(G)\) to describe 
\([25]\), and is NP-complete to maximize over. However, we present them in a slightly different manner, 
equivalent to the standard local consistency bounds typical-
ly requires an exponential number of inequalities 
for every pair 
\(v,v'\) connecting them. The neighbors of a node 
\(v\) in Fig. 1(a), or the row and column indices of a matrix. 
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 
\(B \in \{0,1\}^{\mathcal{V} \times \mathcal{F}}\). 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 \(\hat{B} \in \mathbb{Q}^{\mathcal{P} \times |\mathcal{V}|}\) and define as 
\[\hat{B}_{pu} = 1/|P_p| \text{ if } u \in P_p \text{ and } 0 \text{ if } u \notin P_p.\]
Algebraically, \(B\) and \(\hat{B}\) are related as 
\(\hat{B} = (B^TB)^{-1}B^T\), i.e., \(\hat{B}\) is the left 
pseudoinverse of \(B; \hat{B}B = I_{|\mathcal{P}|}\).
The partitions we consider will never group elements of \(V\) 
with elements in \(F\). Thus, the matrix \(B\) will always be of 
the form \(B = \begin{pmatrix} B_P & 0 \\ 0 & B_Q \end{pmatrix}\), 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 \(\hat{B} = (\hat{B}_P, \hat{B}_Q)\).
Let \(u \in \mathbb{R}^{\mathcal{V}}\) be a real vector composed as 
\(u = [c, b]^T, c \in \mathbb{R}^{\mathcal{V}}, b \in \mathbb{R}^{\mathcal{F}}\). The values of \(u\) 
can be thought of as labels for the elements of \(U\). We say that a partition \(\mathcal{P}\) 
respects \(u\) 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 \(u\), then 
\((c^TB_P)p_i = |P_i|c_x\) where \(x\) is any member of \(P_i\) (and similarly 
for \(B_Q, b\)). Moreover, \((\hat{B}_P)c_i = c_x\) where \(x\) is any 
member of \(P_i\) (and similarly for \(\hat{B}_Q, b\)).
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 \(nb(v)\).

**Definition 1 (Equitable partition of a bipartite graph).**
An equitable partition of a bipartite graph \(G = (V \cup F, E)\) 
given a vector \(u \in \mathbb{R}^{\mathcal{V} \times |\mathcal{F}|}\) 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 \(Q_n\), \(|nb(v) \cap Q_n| = \|nb(v') \cap Q_n\|\); (b) for 
every pair \(f, f' \in F\) in some \(Q_m\), and for every class \(P_n\), 
\(|nb(f) \cap P_n| = \|nb(f') \cap P_n\|\). Furthermore, \(\mathcal{P}\) must 
respect the vector \(u\).

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

\textit{Definition 2 (Equitable partition of a matrix).} An equitable partition of a matrix \( A \in \mathbb{R}^{m \times n} \) given a vector \( u \) is a partition \( \mathcal{P} = \{P_1, \ldots, P_p, Q_1, \ldots, Q_q\} \) of the sets \( V = \{\text{row}[1], \ldots, \text{row}[n]\} \) and \( F = \{\text{col}[1], \ldots, \text{col}[n]\} \) 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 \( P_n \), \( \sum_{v \in P_n} A_{vf} = \sum_{v \in P_n} A_{vff} \). In addition, \( \mathcal{P} \) must respect \( u \).

Note that Def. 1 is an instance of Def. 2 when we take as \( u \) the (bipartite) 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, \( \text{DM}(G, \mathcal{P}) \), has \( |\mathcal{P}| \times |\mathcal{P}| \) entries, where each entry represents how members of different classes interact. More precisely, \( \text{DM}(G, \mathcal{P})_{ij} = |\text{nb}(u) \cap P_j| \), where \( u \) is any element of \( P_i \) due to the bipartiteness of \( G \), this matrix will have the block form \( \text{DM}(G, \mathcal{P}) = \left( \begin{array}{cc} 0 & \text{DF} \ni \mathcal{P} \\ \text{DV} \end{array} \right) \), where \( \text{DV} \) represents the relationship of the \( P \)-classes to the \( Q \)-classes and \( \text{DF} \) vice-versa. As a shorthand, we use the notation \( \text{DM} = (\text{DF}, \text{DV}) \). 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 \( |\text{nb}(u) \cap P_j|, u \in P_i \) many edges going from \( u \) to \( v \). \( \text{DM}(G, \mathcal{P}) \) 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 \( G/\mathcal{P} \) of \( G \), which corresponds only to the DF-block of \( \text{DM}(G, \mathcal{P}) \) 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 \( G/\mathcal{P} \) 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]).

\textbf{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 \( L = (A, b, c) \) be a linear program, corresponding to the optimization problem \( x^* = \arg\max_{x \leq b} c^T x \).

\textbf{Theorem 3 (Lifted Linear Programs [7])}. 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 \( u = (c, b)^T \). Then, \( L' = (\hat{B}_Q A B_P, \hat{B}_Q b, B_P^T c) \) is an LP with fewer variables and constraints. The following relates \( L \) and \( L' \):

\begin{enumerate}[(a)]
  \item If \( x' \) is a feasible point in \( L' \), then \( B_P x' \) is feasible in \( L \).
  \item If in addition \( x' \) is optimal in \( L' \), \( B_P x' \) is optimal in \( L \) with the same objective value.
  \item If \( x \) is a feasible point in \( L \), then \( \hat{B}_P x \) is feasible in \( L' \).
  \item If in addition \( x \) is optimal in \( L \), \( \hat{B}_P x \) is optimal in \( L' \) with the same objective value.
\end{enumerate}

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).

\footnote{We restrict the theorem to what is relevant for the discussion.}
Given an MRF with graph $G$ and parameter $\theta$, which we denote by $I = (G, \theta)$, denote its standard MAP-LP relaxation by the LP defined via $(A, b, c)$. To obtain a potentially smaller LP, we calculate the equitable partition of the variables and factors of the original factor graph. We denote those by $\{1, \ldots, |G|\}$, whose corresponding $\alpha$-classes generalize edge orbits while the $\beta$-classes generalize superfactor. Notably, the argument is that the reparametrization approach to lifted inference.

Proposition 4. Let $P = \{p_1, \ldots, p_2\} \cup \{q_1, \ldots, q_2\}$ be an equitable partition of the variables and factors of the MRF specified by $I = (G, \theta)$. Let $G \mid P$ be the factor quotient of $I$. Then, LMAP-LP(I) can be written as:

$$\mu^* = \arg\max_{\mu \in L'(G)} \sum_{P \in V(G \mid P)} \theta_P |P| \mu_P + \sum_{Q \in F(G \mid P)} \theta_Q |Q| \mu_Q.$$  

Where $\theta_P, \theta_Q$ are the parameters of $I$ for the corresponding partition elements. The constraints $L'(G)$ are defined as the set of $\mu \geq 0$ such that:

$$\left\{ \begin{array}{l} \forall P, P' \in V(G \mid P), Q \in F(G \mid P) \\ \text{s.t. } P, P' \in \text{nb}(Q) \\ \alpha'(Q) \equiv \mu_Q \leq \mu_P ; \quad \beta'(Q) \equiv \mu_Q \leq \mu_P' \\ \gamma'(Q) \equiv \mu_P + \mu_P' - \mu_Q \leq 1 \end{array} \right\} \quad (3)$$

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 $\mu_P$ for every variable class $P \in V(G \mid P)$ (i.e., every supervariable) (2) we instantiate an LP variable $\mu_Q$ for every factor class $Q \in F(G \mid P)$ (i.e. superfactor); (3) for every pair

![Figure 3: Commutative diagram established by Thm. 6 underlying our reparametrization approach to lifted inference.](image-url)
Algorithm 1: Solving MRF $I$ using an equivalent MRF $J$

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

1. Solve MRF $J$, i.e., compute $\tau = \text{argmax}_\mu \text{MAP-LP}(J)$;
2. Lift the solution $\tau$ to LMAP-LP($J$). That is, compute $\tau' = B^T_P\tau$ (Thm. 3(b));
3. Recover solution of $I$, i.e., compute $\mu = B^T_P\tau'$ (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 $G \sslash P$ 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 $L(G \sslash P)$, 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 $\phi_{ij}$ in some $Q$, whose adjacent variables $x_i, x_j$ fall into the same class, $P = P'$. In terms of constraints, the corresponding triple $P, P', Q$, with $P = P'$ yields the constraint $2\mu_P - \mu_Q \leq 1$. Graphically, this situation occurs whenever $G \sslash P$ 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 $\phi_{ij}, P(i) \neq P(j)$, in other words $G \sslash P$ 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 $\theta' \in \mathbb{R}^3$ as $\theta'_Q = |Q|\theta_Q$, $\theta'_P = |P|\theta_P$ (cf. Eq. 3). In this case, the MRF $I' = (G \sslash P, \theta')$ 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 $G \sslash P$ 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, TRWB, 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 $I = (G, \theta^I)$ and $J = (H, \theta^J)$ having simple graphs are LP-equivalent if we can find an equitable partition $P$ of $G$ with incidence matrix $B = (B_P, B_Q)$ and an equitable partition $P'$ of $H$ with incidence matrix $B' = (B'_P, B'_Q)$ such that LMAP-LP($I$) := $(B^T_Q AB_P, B^T_Q b_Q, B'_P c) = ((B^T_Q A B'_P, B^T_Q b_P, (B'_P)^T c)) =: \text{LMAP-LP}(J)$.

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 $B^T_P \mu$ 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 $\tau$ is feasible in MAP-LP($J$), $\mu = B_P B_P \tau$ is feasible in MAP-LP($I$). Moreover, if $\tau$ is optimal, $\mu$ is optimal as well. (B) if $\mu$ is feasible in MAP-LP($I$), $\tau = B^T_P B_P \mu$ is feasible in MAP-LP($J$). Moreover, if $\mu$ is optimal, $\tau$ is optimal as well.

Proof. We prove only (A) due to the symmetry of the statement. Let $\tau$ be feasible in MAP-LP($J$). By Thm. 3(b), $\tau' = B^T_P B_P \tau$ is feasible in LMAP-LP($J$). Due to LP-equivalence, LMAP-LP($J$) = LMAP-LP($I$), $\tau'$ is also a solution to LMAP-LP($I$). Now, we unlift $\tau'$ with respect to LMAP-LP($I$). Due to Thm. 3(b), $\mu = B_P B_P \tau$ is feasible in MAP-LP($I$). Moreover, if $\tau$ is optimal in MAP-LP($J$), Thm. 3 tells us that optimality 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 $2|G \sslash P|$.

Let $I = (G, \theta)$ be an MRF and $P$ be an EP of its variables and factors. We will introduce the algorithm in two
Let us now see how to find weights $\theta'$ as illustrated in Fig. 4(c) consists of adding for every (2)-superfactor in $G \parallel P$ 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 $G \parallel P$. 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 $G \parallel P$. Finally, Step (C) as shown in Fig. 4(e) introduces one representative for every other supernode and connects it to other representatives based on $G \parallel P$. If it happens that the represented supernode is connected to a (2)-supervariable or (1,2)-superfactor in $G \parallel P$, 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 $P'$ is the EP of $G'$ induced by Alg. 2 (the partition which groups nodes in $G'$ together if they represent the same supernode of $G \parallel P$). Let $Q$ be any factor class in $P$ and $Q'$ be the corresponding class in $P'$. We then compute the weight $\theta_{Q'}$ of the factors $\phi' \in Q'$ of $I'$ as

$$\theta_{Q'} = (|Q|/|Q'|) \theta_Q,$$

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

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

**Proof.** Following Def. 5 we must show that given $I$ and its EP $P$, there is a partition $P'$ of $I'$ such that the lifted
LPs are equal. We take the partition \( \mathcal{P}' \) to be the one induced by Alg. 2. \( \mathcal{P}' \) is equal on \( G' \) by construction: we can go through Alg. 2 to verify that every two nodes in \( G' \) representing the same supernode of \( G \) \( \triangledown_{P} \) are connected to the same number of representatives of every other supernode of \( G \) \( \triangledown_{P} \) we omit this due to space restrictions). Now, to show that LMAP-LP(I) has the same constraints as LMAP-LP(I'), we need \( G \triangledown_{P} = G' \triangledown_{P'} \). To see that this holds, observe that Alg. 2 connects \( p \) to \( q \) in \( G' \) if only if \( P \) is connected to \( Q \) in \( G \triangledown_{P} \). If \( P \) is a \( (2) \)-supervariable, \( P \triangledown_{P} Q \) is a \( (2) \)-supervariable \( - q \) will be connected to \( p \) in Step (A). If \( P \triangledown_{P} Q \) is a \( (1,2) \)-supervariable \( \triangledown_{P} \) \( p \) and \( q \) will be connected in Step (B). If \( Q \) is \( (1,2) \)-of \( P \) and \( Q \) is \( (1) \)-supervariable \( \triangledown_{P} \) \( p \) and \( q \) will be connected in Step (C). There are no other possible combinations. Hence, as \( \mathcal{P}' \) consists of all representatives of \( P \) and \( \mathcal{P}' \) consists of all representatives of \( Q \), \( \mathcal{P}' \) and \( \mathcal{P}' \) are connected in \( G' \triangledown_{P'} \) if \( P \) is connected to \( Q \). Moreover, representatives of \( (2) \)-supervariables 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 \( G'/\triangledown_{P'} \) if and only if \( Q \) is connected to \( P \) via a double edge in \( G \triangledown_{P} \).

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 \( \sum_{Q' \in \mathcal{P}'} |Q'| \theta_{Q' \mu_{Q'}} = \sum_{Q' \in \mathcal{P}'} |Q'| (\sum_{Q \in \mathcal{P}} |Q| |Q'|) \theta_{Q' \mu_{Q'}} = \sum_{Q' \in \mathcal{P}'} |Q| \theta_{Q' \mu_{Q'}} = \sum_{Q \in \mathcal{P}} |Q| \theta_{Q \mu_{Q}}. \) 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').

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 \( I = (G, \theta) \) be an Ising model and an \( I' = (G', \theta') \) be computed as above. Then there is no other LP-Equivalent MRF with less factors or less vertices than \( G' \). Moreover, \( |V(G')| \leq 2|V(G \triangledown_{P})| \) and \( |E(G')| \leq 2|E(G \triangledown_{P})| \), 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 \( (2) \)-supervariable in \( G \triangledown_{P} \) adjacent to some \( (2) \)-supervariable \( P \). Due to equivalence, \( Q' \) is a \( (2) \)-supervariable in \( H \triangledown_{P'} \) as well and \( P' \) is a \( (2) \)-supervariable. Hence, the class \( P' \) \( \triangledown_{P} \) must have at least two ground variables from \( H \). Next, let \( Q \) be a \( (1,2) \)-factor in \( G \triangledown_{P} \) adjacent to a \( (2) \)-supervariable. Analogously, \( Q' \) is a \( (1,2) \)-factor in \( H \triangledown_{P'} \) 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 single 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 \( G \triangledown_{P} \) (step 1) and exactly two representatives of any \( (1,2) \)-supervariable (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 \( G \triangledown_{P} \) by at most 2 nodes in \( G' \), \( G' \) can have at most twice as many factors and variables as \( G \).}

\[ \sum_{Q \in \mathcal{P}} |Q| \theta_{Q \mu_{Q}}. \]
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 respectively. The third set considers a simple FR(X, Y) ⇒ (SM(X) ⇔ SM(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


