Equitable Partitions of Concave Free Energies

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

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

Abstract

Significant progress has recently been made towards formalizing symmetry-aware variational inference approaches into a coherent framework. With the exception of TRW for marginal inference, however, this framework resulted in approximate MAP algorithms only, based on equitable and orbit partitions of the graphical model. Here, we deepen our understanding of it for marginal inference. We show that a large class of concave free energies admits equitable partitions, of which orbit partitions are a special case, that can be exploited for lifting. Although already interesting on its own, we go one step further. We demonstrate that concave free energies of pairwise models can be reparametrized so that existing convergent algorithms for lifted marginal inference can be used without modification.

1 INTRODUCTION

Computing likelihoods and marginals using graphical models [24] is an important task for many applications in biology, information retrieval, and computer vision, among other fields. If the graphical models are defined over trees, marginals can be efficiently computed using belief propagation. For models with cycles, however, exact inference is generally intractable. This motivates approximate inference algorithms, favoring algorithms which are as accurate as possible while being guaranteed to converge. One prominent example are variational inference approaches [24, 7, 14], where one aims to approximate a given distribution by a simpler one, i.e., one whose marginals are easier to read off. If a good fit is found, the marginals of the approximating distribution can be used as approximations to the marginals of the original one. Such approaches are typically obtained in two steps: (1) one selects an approximation criterion (a free energy), which is a function of the approximating marginals, and (2) designs optimization algorithms to minimize that free energy efficiently.

A recent development in probabilistic inference has been the use of symmetry [18, 1, 19, 23] as a basis for efficient algorithms. Detecting and utilizing symmetry is establishing itself as an important component of inference. On one hand, there are classes of models where symmetry provides the only means for tractable inference [4]. On the other, in approximate inference algorithms (which tend to be tractable by design) symmetry usually translates to significant improvements in running time as a result of reducing the number of variables of the problem. Symmetry-aware inference approaches are often referred to as lifted inference approaches [20, 9], and one of the first lifted variational inference approaches was lifted loopy belief propagation [8, 22, 10]. These works, however, are largely of algorithmic nature and specific to loopy belief propagation. More recently, Bui et al. [1] proposed a general, algebraic framework for lifted variational inference. It formalizes the notion of symmetry in graphical models via lifting partitions and shows how to exploit them within corresponding variational optimization problems. With the exception of lifted TRW via Frank-Wolfe optimization [2], however, the framework resulted in approximate MAP inference algorithms only, based on equitable and orbit partitions of the graphical model [1, 17, 16].

Our goal in this paper is to deepen our understanding of the lifted variational framework for marginal inference. We do so by extending the notion of equitable partitions — a formalization of symmetry — of models to equitable partitions of energies. We show that within a well-known class of concave energies, **a)** given a concave energy that admits an equitable partition, the number of variables in the resulting optimization problem can be reduced in a way that an exact solution can still be found. Moreover, **b)** given a model that admits an equitable partition, a concave energy that admits the same equitable partition can always be constructed. In combination, these two results allow us to perform concave inference without breaking the symmetry of the model. Although already interesting on its

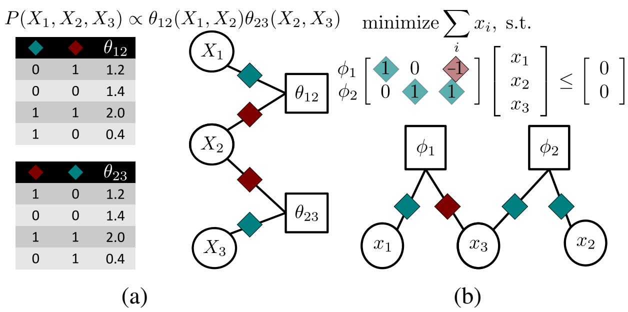


Figure 1: Representations. (a) An MRF and its factor graph. (b) An LP and its factor graph. (Best viewed in color)

own, we go one step further. We demonstrate that for the case of pairwise models, "lifted" concave free energies can be reparametrized to "ground" energies of smaller models. That is, for any pair $(G, \boldsymbol{c})$ of pairwise model and energy parameters as well as an appropriate partition, we can find a pair $(G', \boldsymbol{c}')$ of smaller size such that the resulting energies are equivalent, regardless of the solver. This enables us to use existing highly efficient and distributed convergent message passing algorithms for lifted marginal inference such (as [21]) without modification. Furthermore, we provide a novel angle on the open question raised by Bui et al. regarding the applicability equitable partitions within lifted variational inference. While they point out that partitions coarser than an orbit partition of the model generally cannot lift the TRW energy [2] faithfully, we show that a large class of other energies in the same class do admit any equitable partition of the model.

To achieve the above goal(s) we start off by reviewing the required tools from (lifted) variational inference. Section 3 then introduces the notion of an equitable partition of a convex energy and shows that this partition is a lifting partition. In Section 4, we show to reparametrize energies and models modulo equitable partitions in order to eliminate the necessity of a lifted solver. Before concluding, we illustrate our theoretical results empirically.

2 BACKGROUND

We will start with reviewing variational inference in Markov random fields (MRF). Then we will touch upon the basics of the lifting framework for variational problems.

Variational 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_α represent the possible realizations of a subset α of these random variables. Markov random fields (MRFs) compactly represent a joint distribution over $\mathbf{X}$ as a factorization $P(\mathbf{X} = \mathbf{x}) = Z^{-1} \exp\left[\sum_\alpha \theta_\alpha(x_\alpha) + \sum_i \theta_i(x_i)\right]$, see [24] for more details.

It is often convenient to represent MRFs by their factor graphs. In this paper, however, we will slightly modify the standard definition of a factor graph. For our purposes, a **factor graph** G is a colored tri-partite graph, whose nodes represent the variables, factors and the positions of variables in factors within an MRF. In contrast to standard factor graphs and as illustrated in Fig. 1a, we connect a variable X_i to factor θ_α via a dummy position node $\lozenge_{i\alpha}$, which we color according to the symmetry of θ_α. More precisely, if the positions of X_i and X_j are compatible, that is, $\theta_\alpha(\ldots, x_i, \ldots, x_j, \ldots) = \theta_\alpha(\ldots, x_j, \ldots, x_i, \ldots)$ for all realizations x_i, x_j, we color $\lozenge_{i\alpha}$ and $\lozenge_{j\alpha}$ with the same color. If the positions are not compatible, they receive different colors. Moreover, we assume that factors, variables and positions use different color spaces, e.g. a position and factor node cannot share the same color. While this is not the most compact representation, it will allow us to use a common graphical representation across various kinds of partitions and optimization problems.

Inference in MRFs is generally intractable, hence, inference tasks are often addressed via approximations. One prominent class of approximate inference algorithms arises from the following optimization problem:

$$\boldsymbol{\mu}^* = \operatorname*{argmax}_{\boldsymbol{\mu} \in \mathcal{L}(G)} \underbrace{\left[\boldsymbol{\theta}^T \boldsymbol{\mu} + T \cdot \widehat{H}(\boldsymbol{\mu})\right]}_{=:F(\boldsymbol{\mu})}, \qquad (1)$$

where F is the free energy and the set $\mathcal{L}(G)$, defined as

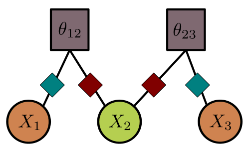
$$\mathcal{L}(G) = \left\{ \boldsymbol{\mu} \geqslant 0 \;\middle|\; \begin{array}{l} \sum_{x_i} \mu_i(x_i) = 1 \\ \sum_{x_\alpha \backslash x_i} \mu_\alpha(x_\alpha) = \mu_i(x_i) \end{array} \right\}, \qquad (2)$$

is known as the local polytope [24]. The problem in Eq. 1 is at the heart of many message-passing inference algorithms. For instance, if we set $T = 0$ (or sufficiently small in the sense of [13]), $\boldsymbol{\mu}^*$ in Eq. 1 yields a linear programming approximation of the Maximum a-Posteriori (MAP) problem and prominent MAP algorithms such as MPLP and MSD are typically derived as specialized solvers for the latter. If, on the other hand, we choose T to be 1 and $\widehat{H}$ to be an approximation of the entropy function, $\boldsymbol{\mu}^*$ approximates the vector of single-node and factor marginals of the distribution P. For example, we can choose $\widehat{H} = \widehat{H}_{\boldsymbol{c}}$ as

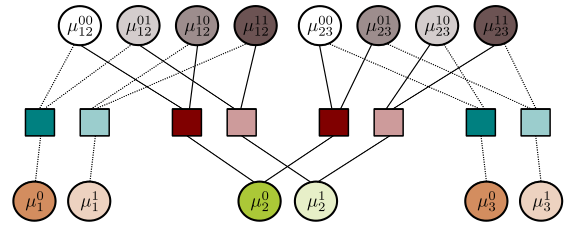
$$\widehat{H}_{\boldsymbol{c}}(\boldsymbol{\mu}) = \sum\nolimits_i c_i H_i(\mu_i) + \sum\nolimits_\alpha c_\alpha H_\alpha(\mu_\alpha), \qquad (3)$$

where H_i and H_α are local entropies. For $c_i = 1 - |\operatorname{nb}(i)|$ and $c_\alpha = 1$, F becomes the Bethe energy, F_{Bethe}. In this case, solving the set of saddle-point equations of Eq. 1 by means of fixed-point iteration yields the popular Loopy Belief Propagation algorithm. The Bethe Energy often gives surprisingly good approximations to the true marginals, however, it is rather difficult to optimize over. Thus, one may prefer to consider instances of $\widehat{H}_{\boldsymbol{c}}$, where maximization is efficient.

Naturally, a class of such energies results from $\widehat{H}$ being concave. In particular, we are interested in values of $\boldsymbol{c}$ that



$$
\begin{aligned}
&\text{minimize } \langle \boldsymbol{\theta}, \boldsymbol{\mu} \rangle \quad \text{subject to:}\\
&\mu_{12}^{00} + \mu_{12}^{01} = \mu_1^0,\ \mu_{12}^{11} + \mu_{12}^{10} = \mu_1^1\\
&\mu_{12}^{00} + \mu_{12}^{10} = \mu_2^0,\ \mu_{12}^{11} + \mu_{12}^{01} = \mu_2^1\\
&\mu_{23}^{00} + \mu_{23}^{01} = \mu_2^0,\ \mu_{23}^{11} + \mu_{23}^{10} = \mu_2^1\\
&\mu_{23}^{00} + \mu_{23}^{10} = \mu_3^0,\ \mu_{23}^{11} + \mu_{23}^{01} = \mu_3^1
\end{aligned}
$$



(a) (b) (c)

Figure 2: Model symmetry of G propagates to its MAP-LP. (a) G – colored by the classes of the CEP. (b) The MAP-LP of G (nonnegativity and normalization constraints as well as position nodes have been omitted for clarity). (c) The factor graph of the MAP-LP of G colored by its CEP. The colors indicate that the classes can be deduced from the CEP of G.. Note that the darker and lighter version of each color are not grouped together. (Best viewed in color)

make $\widehat{H}_{\boldsymbol{c}}$ in Eq. 3 concave, as the structure of these energies gives rise to message-passing algorithms with desirable theoretical properties, cf. [14, 6]. A sufficient condition for $\widehat{H}_{\boldsymbol{c}}$ to be concave is that nonnegative auxiliary numbers $c_{\alpha\alpha}, c_{ii}$, and $c_{i\alpha}$ exist, called **counting numbers**, obeying to

$$
\mathcal{C}(G) = \left\{ \ c_\alpha, c_i \ \middle| \ \begin{array}{l} \exists c_{\alpha\alpha}, c_{ii}, c_{i\alpha} \geqslant 0 \ , \\ c_\alpha = c_{\alpha\alpha} + \sum_{i \in \alpha} c_{i\alpha}, \\ c_i = c_{ii} - \sum_{\alpha : i \in \alpha} c_{i\alpha} \end{array} \right\} . \tag{4}
$$

Finally, note that for $T = 0$, Eq. 1 becomes a linear program (LP), called the MAP-LP of G. An LP is an optimization problem of the form $\text{maximize}\, \boldsymbol{c}^T \boldsymbol{x}$ subject to $\boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}$. While LPs are not the focus of this paper, they will be an important tool in the analysis of variational inference problems. Note that linear programs, like MRFs, can be represented by factor graphs. We represent a constraint α (row α of $\boldsymbol{A}$) as a factor node ϕ_α, LP variable i as a variable node x_i and the coefficient $A_{\alpha i}$ as position node $\Diamond_{\alpha i}$. Our notion of compatibility of positions here is that $\Diamond_{\alpha i}$ is colored with the same color as $\Diamond_{\beta j}$ if $A_{\alpha i} = A_{\beta j}$. Additionally, x_i and x_j are colored with the same color if $c_i = c_j$, while ϕ_α and ϕ_β are colored the same if $b_\alpha = b_\beta$. An example is given in Fig. 1b. Having set-up the variational inference problem, we now give a short review of lifted variational inference.

Lifted Variational Inference via Lifting Partitions. Lifted inference approaches essentially amount to reducing the size a model by grouping together indistinguishable variables and factors. In other words, they exploit symmetries. To formalize the notion of symmetry more concisely, we follow [1].

Consider the linearly constrained concave program

$$
\boldsymbol{x}^* = \text{argmax}_{\boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}} \, J(\boldsymbol{x}) \ . \tag{$\clubsuit$}
$$

We are interested in partitioning the variables of the program by a partition $\mathcal{P} = \{P_1, \ldots, P_p\}$, $P_i \cap P_j = \varnothing$, $\bigcup_i P_i = [x_1, \ldots, x_n]$, such that there exists at least one solution that **respects** the partition. More formally, $\mathcal{P}$ is a **lifting partition** of ($\clubsuit$) if ($\clubsuit$) admits a solution with $x_i = x_j$ whenever x_i and x_j are in the same class in $\mathcal{P}$. We call the linear subspace defined by the latter condition $\mathbb{R}_{\mathcal{P}}$.

Having obtained a lifting partition of the ground variational problem, we can now restrict the solution space to $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\} \cap \mathbb{R}_{\mathcal{P}}$. That is, we constrain equivalent variables to be equal, knowing that at least one solution will be preserved in this space of lower dimension. Since ground variables of the same class are now equal, they can be replaced with a single aggregated (lifted) variable. The resulting lifted problem has one variable per equivalence class, thus, if the lifting partition is coarse enough, significant compression and in turn run-time savings can be achieved. To recover a ground solution from the lifted solution, one assigns the value of the lifted variable to every ground variable in the class.

For linear programs, Mladenov et al. [15, 17, 16, 5] have shown that equitable partitions [3] act as lifting partitions. An **equitable partition of a graph** is a partition $\mathcal{P}$ of the vertex set such that for any pair of vertices u and v in the same class P_n and any other class P_m, $|\,\text{nb}(u) \cap P_m| = |\,\text{nb}(v) \cap P_m|$[1]. For colored graphs, we additionally require that the color vector respects the partition. We call the quantity $|\,\text{nb}(v) \cap P_m|$ the degree of P_n to P_m, $\deg(P_n, P_m)$. For notational convenience, we will introduce equitable partitions of factor graphs as $\mathcal{P} = \{P_1, \ldots, P_p, Q_1, \ldots, Q_q, D_1, \ldots, D_d\}$, where the P-classes refer to the variable classes, the Q-classes to factor classes and the D-classes to position classes.

For the purposes of our discussion, an **equitable partition of a linear program** is an equitable partition of its factor graph. The existence of an equitable partition of a linear program implies the existence of certain doubly-stochastic

[1] The orbit partitions discussed in [1, 2] are a special kind of equitable partitions.

matrices $(\boldsymbol{\Sigma}, \boldsymbol{\Pi})$ such that $\boldsymbol{c}^T\boldsymbol{\Pi} = \boldsymbol{c}^T$, $\boldsymbol{\Sigma b} = \boldsymbol{b}$ and $\boldsymbol{\Sigma A} = \boldsymbol{A\Pi}$ [5].

Recall that a concave energy inference problem as defined here is essentially a linear program (the MAP-LP) plus a linear combination of local entropies in the objective. The symmetries of the MAP-LP have already received attention [1, 17, 16], and it is understood that the MAP-LP preserves the symmetries present in the model. We will use this understanding as a starting point for our discussion of concave energies. We briefly formalize the claim.

Lemma 1. *Any equitable partition $\mathcal{P}$ of an MRF G induces an equitable partition $\mathcal{P}'$ of the resulting MAP-LP.*

Let us briefly sketch how this works. Suppose we are given an equitable partition $\mathcal{P}$ of G. To obtain an equitable partition $\mathcal{P}'$ on $\mathcal{L}(G)$, we group together $\mu_i(0)$ with $\mu_j(0)$, resp. $\mu_i(1)$ with $\mu_j(1)$ if X_i is grouped with X_j in $\mathcal{P}$. To partition the joint state pseudomarginals, we use the following rule. Let θ_α and θ_β be two factors grouped together in $\mathcal{P}$ ($\alpha = \beta$ is also allowed). Then, for all permutations $\pi : \alpha \to \beta$ such that $\pi(i) = j$ only if $\Diamond_{i\alpha}$ is grouped together with $\Diamond_{j\beta}$ in $\mathcal{P}$, we group together $\mu_\alpha(\boldsymbol{x})$ with $\mu_\beta(\pi(\boldsymbol{x}))$ for every joint configuration $\boldsymbol{x}$. This grouping of the LP variables also induces a grouping of the constraints and positions that completes the partition.

Due to lack of space, we will not prove this here. Instead, we give an example of how model symmetry propagates to MAP-LP symmetry. Fig. 2 shows an MRF G colored by its CEP (a) and the resulting MAP-LP (b) (some constraints and the position nodes have been omitted for clarity). In Fig. 2c, we see the correspondence between the CEP of G and the CEP of the MAP-LP as indicated by the colors.

3 EQUITABLE PARTITIONS OF CONCAVE FREE ENERGIES

With the basics of lifted variational inference at hand, we can now begin our main discussion. We proceed as follows. We start off by defining an equitable partition of a concave energy. Then, as the first main result of this section, we show that any concave energy that admits an equitable partition has a solution that **respects** that partition. This establishes that equitable partitions of concave energies are lifting partitions. Next, we show that given an equitable partition of an MRF, concave energies that admit this partition are guaranteed to exist. That is, if we want to do convergent inference on a model with symmetries, we can always find a suitable energy that does not break the symmetries. Finally, we will look at some heuristics used for selecting concave energies and examine their relationship to equitable partitions.

Definition 2. *An* **equitable partition of a concave energy** *$\widehat{H}_{\boldsymbol{c}}$ (as in Eq. 3) with $\boldsymbol{c} \in \mathcal{C}(G)$ is an equitable partition of G such that X_i and X_j are grouped together only if $c_i = c_j$ and $\theta_\alpha, \theta_\beta$ are grouped together only if $c_\alpha = c_\beta$.*

We will shortly show that equitable partitions of concave energies preserve optimal solutions of the variational problem. Before we do so, however, we will formalize what we consider to be the symmetries of $\widehat{H}_{\boldsymbol{c}}$.

If we set aside the constraints and ignore semantics of $\boldsymbol{\mu}$, $\widehat{H}_{\boldsymbol{c}}$ is just a linear combination of $x \log x$ terms, i.e. $\widehat{H}_{\boldsymbol{c}}(\boldsymbol{x}) = \sum_k c_k x_k \log x_k$. Observe that if we permute any two variables whose c's are the same, we do not change $\widehat{H}_{\boldsymbol{c}}$. In other words, any permutation $\boldsymbol{\Pi}$ with $\boldsymbol{\Pi c} = \boldsymbol{c}$ is an automorphism of $\widehat{H}_{\boldsymbol{c}}$, i.e., $\widehat{H}_{\boldsymbol{c}}(\boldsymbol{\Pi x}) = \widehat{H}_{\boldsymbol{c}}(\boldsymbol{x})$. Moreover, switching any pair of variables can be done independently of other pairs. We now restate the above in formal terms. If we introduce $\mathcal{R} = \{R_1, \ldots, R_r\}$ that partitions the variables into classes having equal c, then the following holds:

Observation 3. *The group $\Gamma = \bigotimes_{R \in \mathcal{R}} \mathbb{S}_{|R|}$ is isomorphic to a subgroup of $\mathbb{AUT}(\widehat{H})$.*

Here $\bigotimes$ denotes the group product and $\mathbb{S}_n$ is the symmetric group over n elements. Now, let us "switch on" the semantics of the argument and interpret the automorphism group of $\widehat{H}_{\boldsymbol{c}}$ in terms of pseudomarginals. We can see that the following operations are automorphisms of $\widehat{H}_{\boldsymbol{c}}$: we can exchange the pseudomarginals of any two variables with the same c's, e.g., $(\mu_i(0), \mu_i(1)) \mapsto (\mu_j(0), \mu_j(1))$; similarly, we can exchange any two sets of factor beliefs; we could also exchange states within a set of beliefs: $\mu_i(0) \mapsto \mu_i(1)$, or even exchange states across variables, $\mu_i(0) \mapsto \mu_j(1)$ (given that the c's are compatible). All of these operations can be done independently of each other. Of course, many symmetries of $\widehat{H}_{\boldsymbol{c}}$ are not symmetries of Eq. 1, as they are not symmetries of the constraints. For example, Eq. 1 would generally not admit, say the reordering of states of a variable without reordering the states of its neighbors as a symmetry, since marginalization constraints tie adjacent nodes in the factor graph.

In summary, $\widehat{H}_{\boldsymbol{c}}$ is a highly symmetric object given that we have equal c's. Hence, what we really have to be careful about are the symmetries of the constraints. However, as we will discuss now, equitability takes care of the constraints and we end up with lifting partitions for Eq. 1.

Theorem 4 (EPs of Concave Energies are Lifting Partitions)**.** *Let G be an MRF and $\boldsymbol{c} \in \mathcal{C}(G)$ a vector of counting numbers. If $\mathcal{P}$ is an equitable partition of $\widehat{H}_{\boldsymbol{c}}$, then $\mathcal{P}'$ obtained from $\mathcal{P}$ via Lemma 1 is a lifting partition of Eq. 1.*

Proof. To prove that $\mathcal{P}'$ is a lifting partition of Eq. 1, we need to prove that Eq. 1 admits a solution, where equivalent variables take on equal values. We will establish this in the following way. Given any feasible vector $\boldsymbol{\mu}$ of Eq. 1, we will produce a vector $\boldsymbol{\mu}'$ by replacing each variable by the average of its class. E.g., if the variable $\mu_i(x)$ is in some class P, then $\mu_i'(x) = 1/|P| \sum_{j \in P} \mu_j(x)$. Thus averaged, the vector $\boldsymbol{\mu}'$ respects $\mathcal{P}'$. Then, we need that **a)** $\boldsymbol{\mu}'$ is feasible as well and that **b)** $F(\boldsymbol{\mu}') \geqslant F(\boldsymbol{\mu})$. Having established **a)** an **b)**, the rest is simple: we take any optimal solution of Eq. 1 and average over the classes. By **a)** we know the average is feasible. By **b)** we know that it will not decrease the objective value. Since we started with somethin that was already optimal, it must be that averaged vector is of equal objective value, as improvement over the optimum is not possible by definition. Thus we have found a new optimum that respects the partition. It now only remains to verify that **a)** and **b)** indeed hold.

Proof of a). The averaging operation over the partition classes can be represented in a linear algebraic way, as multiplying $\boldsymbol{\mu}$ with the doubly stochastic matrix $\boldsymbol{X}$ defined as:

$$X_{ij} = \begin{cases} 1/|C| & \text{if } (\boldsymbol{\mu})_i, (\boldsymbol{\mu})_j \text{ are both in some } C \in \mathcal{P}', \\ 0 & \text{otherwise.} \end{cases}$$

The brackets in the equation indicate that the indices above are used in a generic sense (not bound to factors, variables or particular states). The theory of equitable partitions of LPs tells us that if $\mathcal{P}'$ is equitable, then $\boldsymbol{\mu}$ being feasible implies $\boldsymbol{\mu}' = \boldsymbol{X}\boldsymbol{\mu}$ is feasible as well, as $\boldsymbol{X}$ is a fractional automorphism of the LP [5].

Proof of b). The Birkhoff-von-Neumann Theorem [12] allows one to decompose the doubly stochastic $\boldsymbol{X}$ as a convex combination of permutation matrices $\sum_i \lambda_i \boldsymbol{\Pi}_i$. Note that any of these permutation matrices will exchange only variables that had been grouped together in $\mathcal{P}'$. This follows from the fact that the λ's form a convex combination. If $\boldsymbol{\Pi}_k$ has a nonzero element, then $\sum_k \lambda_k (\Pi_k)_{ij} = X_{ij}$ has to be strictly greater than 0 as well, as the convex combination has at least one nonzero element. By definition $\boldsymbol{c}$ respects $\mathcal{P}'$, hence all $\boldsymbol{\Pi}$'s are automorphisms of $\widehat{H}_{\boldsymbol{c}}$. Moreover, $\boldsymbol{\theta}$ also respects $\mathcal{P}'$ due to our definition of equitable partitions, hence the $\boldsymbol{\Pi}$'s are automorphisms of $\boldsymbol{\theta}^T$ as well. Taken together, we establish that all $\boldsymbol{\Pi}_i$'s are automorphisms of F, $F(\boldsymbol{\mu}) = F(\boldsymbol{\Pi}_i\boldsymbol{\mu})$. With this in mind, the concavity of F gives us the result: $F(\boldsymbol{\mu}') = F(\sum_i \lambda_i \boldsymbol{\Pi}_i \boldsymbol{\mu}) \geqslant \sum_i \lambda_i F(\boldsymbol{\Pi}_i\boldsymbol{\mu}) = F(\boldsymbol{\mu})$. □

Thus, we have established that equitable partitions of concave energies are lifting partitions for the variational problem of interest. However, an important question that remains is: do they actually exist? That is, if we want to do inference, can we find counting numbers that permit lifting at all? As we will show now, not only do such numbers exist, but also most heuristics presented in literature for finding counting numbers will yield $\boldsymbol{c}$ that respect equitable partitions of G. Let us first show the existence of liftable counting numbers.

Lemma 5 (Existence of liftable counting numbers)**.** *Let $\mathcal{P}$ be an equitable partition of the MRF G. If $\mathcal{C}(G)$ is not empty, then there exists a $\boldsymbol{c}$-vector that respects $\mathcal{P}$. In other words, $\mathcal{P}$ is an equitable partition of $\widehat{H}_{\boldsymbol{c}}$ for at least one $\boldsymbol{c}$.*

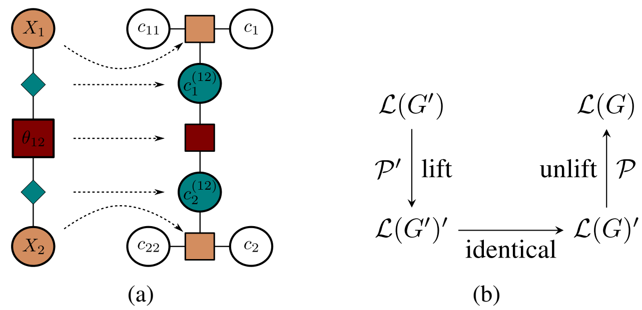


Figure 3: (a) Illustration of Lemma 5. (b) Commutative diagram underlying the "lifted inference by reparametrization" paradigm. (Best viewed in color)

Proof (sketch). Observe that the set of counting numbers is defined by a linear program, as Eq. 4 consistes of linear constraints. Thus, we can again rely on the fact that equitable partitions of linear programs act as lifting partitions, given that we manage to translate $\mathcal{P}$ into an equitable partition $\mathcal{P}'$ of $\mathcal{C}(G)$. Let us show the translation in question. We group together c_i with c_j, c_{ii} with c_{jj}, and ϕ_i (the constraint generated by X_i) with ϕ_j in $\mathcal{P}'$ if X_i and X_j are grouped together in $\mathcal{P}$. Similarly, if $\mathcal{P}$ groups θ_α and θ_β, we group c_α with c_β, $c_{\alpha\alpha}$ with $c_{\beta\beta}$, and ϕ_α (the constraint generated by θ_α) with ϕ_β . Finally, $c_{i\alpha}$ and $c_{j\beta}$ are grouped together if $\Diamond_{i\alpha}$ and $\Diamond_{j\beta}$ are grouped together in $\mathcal{P}$. Now let us argue that $\mathcal{P}'$ is indeed equitable on $\mathcal{C}(G)$. The main idea is as follows: we will show that the factor graph of G is isomorphic to a "skeleton" subgraph of the factor graph of $\mathcal{C}(G)$. As such, any equitable partition of G is also equitable on the skeleton of $\mathcal{C}(G)$. Then, we will complete the partition on the remaining elements of $\mathcal{C}(G)$ in a way that preserves equitability.

To obtain the skeleton, we temporarily ignore the variables c_i, c_{ii}, c_α, $c_{\alpha\alpha}$. Then, the map $M : G \to \mathcal{C}(G)$, which maps the position node $\Diamond_{\alpha i}$ to the LP variable $c_{i\alpha}$, the factor θ_α to the constraint of the factor and the variable X_i to the constraint of the variable, is an isomorphism. This follows directly from Eq. 4. An LP variable $c_{i\alpha}$ appears in the factor constraint of θ_α if and only if θ_α is connected to position node $\Diamond_{\alpha i}$ (in other words variable i participates in factor θ_α). Moreover a variable $c_{i\alpha}$ appears in the variable constraint of X_i if and only if X_i is connected to position node $\Diamond_{\alpha i}$. Thus, an equitable partition of G yields an equitable partition of the constraints and the variables $c_{i\alpha}$. If we reintroduce now c_i, c_{ii}, c_α, $c_{\alpha\alpha}$, we see that each appears in exactly one constraint, so they can be partitioned in a way to preserve equitability. Fig. 3a provides an illustration. Note that we have so far ignored the position nodes in the FG of $\mathcal{C}(G)$. It can be verified that they can be partitioned without breaking the equitability of the partition obtained thus far. □

We have established that if G admits an equitable partition, then there will be at least one energy that admits the same equitable partition. The question now is, how do we compute the appropriate counting numbers? Naturally, as in Thm 4, we can take any vector of counting numbers, any equitable partition of G and simply average $\boldsymbol{c}$ over the classes. However, as $\mathcal{C}(G)$ is a polyhedron, there are infinitely many vectors of counting numbers and the usefulness of the resulting energies in terms of the approximate inference problem will vary. In the following, we show that several heuristics for picking counting numbers naturally yield counting numbers that allow nice equitable partitions of the energy. The heuristics that we will discuss first are the following two: **(a)** [6] following the principle of insufficient reason one tries to make $\boldsymbol{c}$'s as uniform as possible by minimizing either $\sum_\alpha (c_\alpha - 1)^2$ or $\sum_\alpha c_\alpha \log c_\alpha$ over $\mathcal{C}(G)$; **(b)** [14] finding the least-squares projection of the $\boldsymbol{c}$'s of the Bethe energy onto $\mathcal{C}(G)$, i.e. $\boldsymbol{c}^* = \operatorname{argmin}_{\boldsymbol{c}\in\mathcal{C}(G)} \sum_\alpha (c_\alpha - 1)^2 + \sum_i (c_i - |\operatorname{nb}(X_i)| + 1)^2$. Both of them lead to liftable energies.

Proposition 6 (Finding lifted counting numbers)**.** *Let $\mathcal{P}$ be an EP of G and $\mathcal{P}'$ is an EP of $\mathcal{C}(G)$ obtained as in Thm. 5. Then $\mathcal{P}'$ is a lifting partition of both (a) and (b).*

Proof. The proof is identical to the proof of Thm. 4. As, Lemma 5 already established, given any $\boldsymbol{c} \in \mathcal{C}(G)$, we obtain a new $\boldsymbol{c}' \in \mathcal{C}(G)$ by assigning to any c'_α or c'_i the average of $\boldsymbol{c}$ over the respective class. Now we need to show that this averaging does not increase the respective objective values. We argue in the same way as in Thm. 4. For both cases in **(a)**, the automorphism group consists of the product of two symmetric groups - we can exchange any two c_i, c_j and any two c_α, c_β independently. Thus averaging over $\mathcal{P}'$ is equivalent to averaging over automorphisms of the objectives, which by convexity does not increase the value. For the case of **(b)**, the automorphism group is smaller. We are allowed to exchange any two c_α, c_β independently, but we can exchange c_i, c_j only if $|\operatorname{nb}(X_i)| = |\operatorname{nb}(X_j)|$. Recall however, that in any equitable partition, for any class Q, we group X_i and X_j if $\deg(X_i, Q) = \deg(X_j, Q)$. This implies $|\operatorname{nb}(X_i)| = |\operatorname{nb}(X_j)|$. Thus, the averaging matrix X as in Thm. 4 will have a non-zero value only if $|\operatorname{nb}(X_i)| = |\operatorname{nb}(X_j)|$, and the resulting Birkhoff-von-Neumann decomposition will consist of automorphisms of the objective. □

In essence, this proposition tells us that optimal numbers (w.r.t. **(a)** and **(b)**) can be found that turn any equitable partition of G into an equitable partition of the energy. Could this be the case for all heuristics? As it turns out, some heuristics can impose restrictions on what partitions can be EPs of their respective energies.

To see this, consider for example as a further option **(c)** Tree-Reweighted BP (in pairwise models). We obtain $\boldsymbol{c} \in \mathcal{C}(G)$ by setting c_α to be the number of spanning trees passing through θ_α divided by the number of all spanning trees in G. If we translate the result of Bui et al. [2] in the language of the present paper, it states that the equitable partition of G coarser than its orbit partition may generally not be turned into an equitable partition of the TRW energy. However, for the orbit partition, they give an efficient algorithm to produce the appropriate $\boldsymbol{c}$.

Finally, one could also follow Meshi et al. [14] as option **(d)**: instead of approximating the numbers of F_{Bethe}, we project F_{Bethe} itself onto the set of concave energies $F_{\boldsymbol{c}\in\mathcal{C}(G)}$. More precisely, we take $\boldsymbol{c}^* = \operatorname{argmin}_{\boldsymbol{c}\in\mathcal{C}(G)} \int_{\boldsymbol{\mu}\in\mathcal{L}(G)} \left(F_{\text{Bethe}}(\boldsymbol{\mu}) - F_{\boldsymbol{c}}(\boldsymbol{\mu})\right)^2 d\boldsymbol{\mu}$. As a matter of fact, it is even an open question whether F_{Bethe} admits stationary points that respect[2] any $\mathcal{P}$. It should be noted that if we want a coarser partition than what **(c)** or **(d)** permit, we could still take the counting numbers and average them over a coarser equitable partition of G. However, the question of whether this operation preserves the quality of approximation remains open.

4 LIFTING AS REPARAMETRIZATION

One issue that arises with the above approach is that in certain cases, compression changes the structure of the optimization problem. For example, given an equitable partition $\mathcal{P} = \{Q_1, \dots, Q_q, P_1, \dots, P_p, D_1, \dots, D_d\}$ of G, the compressed inference problem will take on the following form: $\boldsymbol{\mu}^* =$

$$\operatorname*{argmax}_{\boldsymbol{\mu}\in\mathcal{L}(G)'} E'(\boldsymbol{\mu}) + \sum_{P\in\mathcal{P}} |P| c_P H_P(\mu_P) + \sum_{Q\in\mathcal{P}} |Q| c_Q H_Q(\mu_Q) ,$$

where

$$E'(\boldsymbol{\mu}) = \sum_{P\in\mathcal{P}} |P|\theta_P \mu_P + \sum_{Q\in\mathcal{P}} |Q|\theta_Q \mu_Q \quad \text{and}$$

$$\mathcal{L}(G)' = \left\{ \boldsymbol{\mu} \geqslant 0 \;\middle|\; \begin{array}{l} \mu_P^0 + \mu_P^1 = 1 \\ \mu_Q^{00} + \mu_D^{01} = \mu_P^0 \\ \mu_Q^{11} + \mu_{D'}^{01} = \mu_P^1 \\ \mu_Q^{00} + \mu_{D'}^{01} = \mu_{P'}^0 \\ \mu_Q^{11} + \mu_D^{01} = \mu_{P'}^1 \end{array} \right\} . \tag{5}$$

Here, Q is the representative of a factor class, P and P' are the representatives of its neighboring variable classes and D and D' are the representatives of the position classes that connect variables in P and P' to factors in Q. Note that our notation treats the beliefs of factor Q being in state 00 or 11 differently from the beliefs of factor Q being in state 01 or 01. This becomes important in the following situation: in an equitable partition, it could happen that $P = P'$. That is, for any ground factor in the class Q, both participating variables are in the same class of the lifting partition. For the variational problem, this means that we need to unify the

[2] Works on lifted loopy belief propagation [8, 22, 10] do indeed show that BP admits such solutions. However, the question of whether this is due to F_{Bethe} or an artifact of the way BP optimizes it is unclear.

variables $\mu^{01}_{D'}$ and μ^{01}_{D}, ending up with a peculiar-looking set of constraints such as:

$$\mu^{00}_Q + \mu^{01}_D = \mu^0_P \text{ and } \mu^{11}_Q + \mu^{01}_D = \mu^1_P \quad .$$

This example illustrates why we cannot simply view the lifted inference problem as a standard inference problem on a smaller "proper" factor graph. We have a binary factor that has the same variable in both positions (which is, unfortunately, not equivalent to a unary factor). Now, in terms of message-passing, to send a message to P, Q would have to first eliminate P – an operation which is not supported by standard message-passing frameworks.

To circumvent the problem, at least for the case of $T = 0$ in Eq. 1, we can follow the "lifted inference as reparametrization" paradigm recently advocated by Mladenov et al. [16]. The main idea is based on the fact that equitable partitions allow one to not only recover a solution of the ground variational problem from the lifted one, but also a solution of the ground problem can be projected onto $\mathbb{R}_{\mathcal{P}}$ by averaging the variables within each equivalence class. Thus, what we do is the following: given an G and an equitable partition $\mathcal{P}$, we find a smaller G' and a partition $\mathcal{P}'$ such that the variational problem of G lifted with $\mathcal{P}$ is identical to the variational problem of G' lifted with $\mathcal{P}'$. So, we solve the smaller G' with a ground solver, average over $\mathcal{P}'$ to obtain a solution to the lifted problem of G', transfer that to the lifted variational problem of G, since they are identical, and then finally unlift according to $\mathcal{P}$. The idea is illustrated in Fig. 3. The work of [16] shows how to find G' and $\mathcal{P}$ and then reparametrize the factors of G' such that its lifted MAP-LPs are same the one of G. Our goal here is to show that if in addition we reparametrize the $\boldsymbol{c}$'s, their lifted concave energies will also be the same. Thus, in essence, we can make use of any ground concave energy solver that allows manually setting $\boldsymbol{c}$'s for lifted inference.

As we would like to avoid introducing again the technical arguments of [16], we will introduce a notion of weak partition equivalence, which subsumes the equivalence in [16], and build upon that. We will show that reparametrization of $\widehat{H}_{\boldsymbol{c}}$ is possible among partition-equivalent pairs. Since the pairs produced by the algorithm of [16] are partition-equivalent, the result applies.

Definition 7. *Let G and G' be MRFs. We call G and G'* **weakly partition-equivalent** *if they admit equitable partitions $\mathcal{P} = \{P_1, \ldots, P_p, Q_1, \ldots, Q_q, D_1, \ldots, D_d\}$ resp. $\mathcal{P}' = \{P'_1, \ldots, P'_p, Q'_1, \ldots, Q'_q, D'_1, \ldots, D'_d\}$ having the same number of variable, factor and position classes. Moreover, for any two classes $X, Y \in \mathcal{P}$, we have* $\deg(X, Y) \neq 0$ if and only if $\deg(X', Y') \neq 0$ *(♠).*

Note that the actual number of nodes in the class may very well be different, we do not require $|Q| = |Q'|$.

Lemma 8. *If G and G' are weakly partition equivalent w.r.t. $\mathcal{P}$ and $\mathcal{P}'$, then $\deg(Q, D) \neq 0$ implies $\frac{|Q|}{|Q'|}\deg(Q, D) = \frac{|D|}{|D'|}\deg(Q', D')$, resp. $\deg(P, D) \neq 0$ implies $\frac{|P|}{|P'|}\deg(P, D) = \frac{|D|}{|D'|}\deg(P', D')$.*

Proof. For any equitable partition it will hold that $|Q|\deg(Q, D) = |D|\deg(D, Q)$, resp. $|Q'|\deg(Q', D') = |D'|\deg(D, Q)$. Since D consists of only position nodes, and every position node can be connected to exactly one factor (and one variable), $\deg(D, Q) = \deg(D', Q') = 1$. With this in mind, we take the quotient of both equations and obtain $\frac{|Q|\deg(Q,D)}{|Q'|\deg(Q',D')} = \frac{|D|}{|D'|}$. Multiplying both sides by $\deg(Q', D')$ yields the result. The reasoning for the variable classes is identical. □

Now, suppose counting numbers for G which respect $\mathcal{P}$ have been found, that is, for every θ_α in the class Q, $c_\alpha = c_Q$ and so on, as in Thm. 5. Then, we can construct a vector of counting numbers for G' by the following procedure: for every vertex in G' (regardless of whether it is a variable, factor, or position), we take the size of its class, $|X'|$, the size of the corresponding class in $\mathcal{P}$, $|X|$ and then normalize the counting number c_X (from G) by their ratio. That is, $c'_k = (|X|/|X'|)c_X$.

Theorem 9. *Suppose G and G' are weakly partition equivalent with respect to $\mathcal{P}$ and $\mathcal{P}'$ and $\boldsymbol{c} \in \mathcal{C}(G)$ respects $\mathcal{P}$. Then, the vector $\boldsymbol{c}'$ having $c'_k = c_{X'} = \frac{|X|}{|X'|}c_X$ consists of counting numbers for G'. I.e. $\boldsymbol{c}' \in \mathcal{C}(G')$.*

Proof. We have assumed $\boldsymbol{c}$ respects $\mathcal{P}$. This allows us to rewrite the conditions of Eq. 4 as

$$\begin{aligned} C_Q &= C_{QQ} + \textstyle\sum_D \deg(Q, D)c_D \quad \text{and,} \\ C_P &= C_{PP} + \textstyle\sum_D \deg(P, D)c_D \ . \end{aligned}$$

The above describes the lifted LP of $\mathcal{C}(G)$ after unification of all equivalent variables. Now, observe that

$$\begin{aligned} C_{Q'} &= \frac{|Q|}{|Q'|}C_Q = \frac{|Q|}{|Q'|}\left[C_{QQ} + \sum_D \deg(Q, D)c_D\right] \\ &= \frac{|Q|}{|Q'|}C_{QQ} + \textstyle\sum_{D:\deg(Q,D)\neq 0} \dfrac{|Q|}{|Q'|}\deg(Q, D)c_D \\ &= C_{Q'Q'} + \textstyle\sum_{D:\deg(Q,D)\neq 0} \deg(Q', D')\dfrac{|D|}{|D'|}c_D \\ &= C_{Q'Q'} + \textstyle\sum_{D':\deg(Q',D')\neq 0} \deg(Q', D')c_{D'} \ . \end{aligned}$$

Note, we were allowed to switch the index in the last line due to (♠). Similarly,

$$\begin{aligned} C_{P'} &= \frac{|P|}{|P'|}C_P = \frac{|P|}{|P'|}\left[C_{PP} - \sum_D \deg(P, D)c_D\right] \\ &= \frac{|P|}{|P'|}C_{PP} - \textstyle\sum_{D:\deg(P,D)\neq 0} \dfrac{|P|}{|P'|}\deg(P, D)c_D \\ &= C_{P'P'} - \textstyle\sum_{D:\deg(P,D)\neq 0} \deg(P', D')\dfrac{|D|}{|D'|}c_D \\ &= C_{P'P'} - \textstyle\sum_{D':\deg(P',D')\neq 0} \deg(P', D')c_{D'} \ . \end{aligned}$$

Complete Graph		Friends-Smokers		Clique-Cycle	
W	$V(x)$	W	$x \neq y \wedge \neg Fr(x,y)$	W	$x \neq y \wedge (Q1(x) \Leftrightarrow \neg Q2(y))$
-0.1	$x \neq y \wedge (V(x) \Leftrightarrow V(y))$	-0.8	$Ca(x)$	W	$x \neq y \wedge (Q2(x) \Leftrightarrow \neg Q3(y))$
		-12.4	$Aux(x,y)$	W	$x \neq y \wedge (Q3(x) \Leftrightarrow \neg Q1(y))$
		1.5	$(Sm(x) \Rightarrow Ca(x))$	$-W$	$x \neq y \wedge (Q1(x) \Leftrightarrow Q2(y))$
		6.2	$(x \neq y \wedge Aux(x,y) \wedge Smokes(x))$	$-W$	$x \neq y \wedge (Q2(x) \Leftrightarrow Q3(y))$
		6.2	$(x \neq y \wedge Aux(x,y) \wedge Smokes(y))$	$-W$	$x \neq y \wedge (Q3(x) \Leftrightarrow Q1(y))$
		6.2	$(x \neq y \wedge Aux(x,y) \wedge Friends(x,y))$		
		-3.1	$(x \neq y \wedge (Smokes(x) \wedge Friends(x,y))$		
		-3.1	$(x \neq y \wedge (Smokes(y) \wedge Friends(x,y))$		

Table 1: The Markov Logic Network models used to illustrate our theoretical results. Details are given in the main text.

What this tells us is that $\boldsymbol{c}'$ is a solution of $\mathcal{C}(G')$ lifted according to $\mathcal{P}'$. This implies that we can recover a vector of ground counting numbers by assigning to c'_k the corresponding c'_X. □

So now, given $\boldsymbol{c}$ obtained as above, let us examine the lifted entropy of G' with respect to $\mathcal{P}'$,

$$\begin{aligned}\widehat{H}' &= \sum\nolimits_{P' \in \mathcal{P}'} |P'| c_{P'} H_{P'} + \sum\nolimits_{Q' \in \mathcal{P}'} |Q'| c_{Q'} H_{Q'} \\ &= \sum\nolimits_{P' \in \mathcal{P}'} |P'| \frac{|P|}{|P'|} c_P H_{P'} + \sum\nolimits_{Q' \in \mathcal{P}'} |Q'| \frac{|Q|}{|Q'|} c_Q H_{Q'} \\ &= \sum\nolimits_{P \in \mathcal{P}'} |P| c_P H_P + \sum\nolimits_{Q' \in \mathcal{P}'} |Q| c_Q H_Q = \widehat{H} \ .\end{aligned}$$

As we see, this is exactly the lifted entropy of G with respect to $\mathcal{P}$. To conclude the argument, we note that the algorithm of [16] takes care that the linear $E'(\cdot)$ part and the constraints of the lifted variational problems are the same. Thus, with the addition of the reparametrized $\boldsymbol{c}$'s, we achieve full equivalence of both lifted variational problems.

5 EMPIRICAL ILLUSTRATION

We will now illustrate our theoretical results by demonstrating that the "lifting by reparametrization" paradigm allows one to lift Schwing et al.'s distributed message-passing algorithm[3] for marginal inference [21] without any overhead. As far as we know, this presents the first lifted convergent variational marginal inference that can handle problems efficiently by distributing and parallelizing the inference computation and the memory requirements. The convergence and optimality guarantees are preserved by consistency messages, sent between the distributed cores, that our lifted variant directly inherits from Schwing et al.'s original version. Together with parallelizable lifting approaches [11], this shows for the first time that each step of the lifted inference pipeline is readily parallelizable.

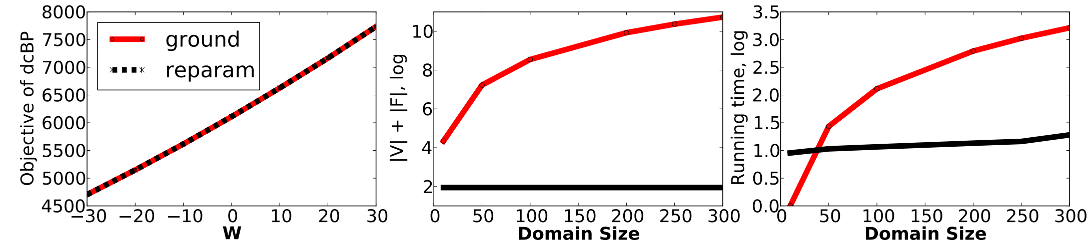
The experimental protocol is inspired by [2] and uses three of their test models in Markov Logic Network syntax, cf. Tab. 1: complete graph, friends-smokers, and clique-cycle.

[3] http://alexander-schwing.de/projects.php

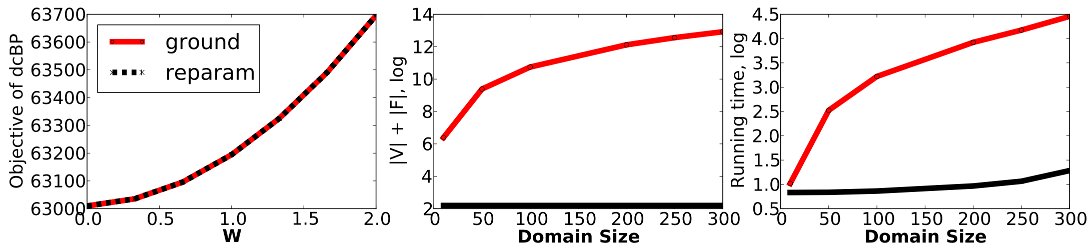
We focus on the repulsive case, i.e. the weight of interaction clauses is set to a negative values. The parameter W denotes the weight that will be varying across the experiments. As Bui et al. [2] point out, in all models except clique-cycle, W acts like the "local field" potential in an Ising model; a positive value of W means that variables tend to be in the 1 state, whereas a negative value favors the 0 state. The complete graph model is an Ising model over the complete graph over n nodes (the domain size); all parameters are the same. The weight of the interaction clause is repulsive with -0.1. The friends-smokers is a pairwise version of the one negated one used by Bui et al., where we also used a repulsive interaction between smokers: Here the domain size is the number of people. Finally, the clique-cycle model encodes a model with 3 cliques, each consisting of n nodes (the domain size), and 3 bipartite graphs between them. For more details we refer to [2].

Since we reparametrize an existing variational approach, we will not report on the accuracy of the marginals and the quality of the objective as they have been investigated already in the corresponding literature. Rather, we perform a "within model" comparison of the objectives achieved, the size of the models, and the running times for inference since they are what our theoretical results are about. Specifically, we evaluated the lifted and the ground version on several instances of the three test models, varying the parameter W and the domain size. We assume no evidence has been observed, which results in a large amount of symmetries. As Bui et al. [2] argue this is a sensible setting since performing marginal inference in relational probabilistic models can be very useful for maximum-likelihood parameter estimation.

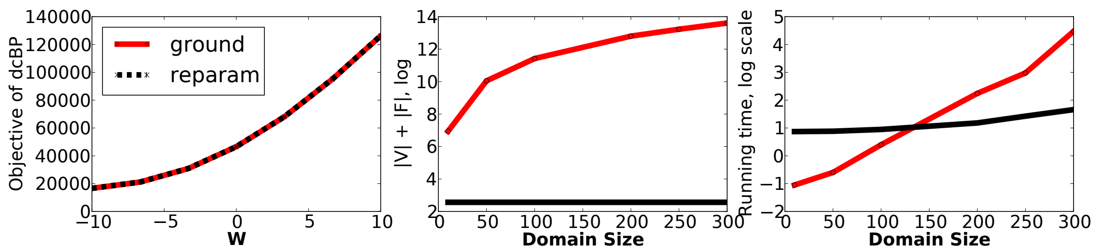
The experimental results are summarized in Fig. 4. As one can see the lifted inference can be orders of magnitude faster than its ground version. We also ran a single-core lifted loopy belief propagation. However, due to its lack of convergence, it is difficult to have a meaningful comparison of running times. Nevertheless, we observed cases where it did not converge but actually started to oscillate.



(a) Complete Graph MLN.



(b) Clique-Cycle MLN.



(c) Friends-smokers MLN.

Figure 4: Experimental results on the test models from Tab. 1. Each row shows from left to right the objective for different weights, the size (number of nodes and factors) in log-space and the running time in seconds in log-space for ground (red) versus lifted (black). As one case see, lifted variational marginal inference can be orders of magnitude faster than it ground version without sacrificing the objective. (best viewed in color)

6 CONCLUSIONS

We have established a "lifted inference by reparametrization" paradigm for variational marginal inference. More precisely, we have introduced the notion of equitable partitions of concave free energies and shown how to use them to reparameterize the corresponding variational optimization problems. In turn, a large class of existing variational marginal inference algorithms can directly be made aware of symmetries without modifications. We illustrated this by lifting Schwing et al.'s distributed message-passing algorithm for marginal inference, resulting in the first lifted, distributed, convergent message passing algorithms for marginal inference. Moreover, the paradigm of reparametrization allows us to address the observation of Bui et al. [2] about their Frank-Wolfe TRW solver running slower than BP. At least in the case where no extra tightening is required, one can just compute the TRW counting numbers with lifted Kruskal, reparametrize and apply a generic convergent message-passing algorithm.

Our work provides several avenues for future work. For instance, one should explore what other constraints we can posed on counting numbers to enforce exactness while we can still optimize over the set in a lifted fashion. Since the dimensionality reduction changes the geometry of the variational optimization problem, one should also investigate its interaction with the solvers. It is interesting to explore features of relational languages to speed up lifted variational marginal inference even more. One of the most interesting open question raised by our work is whether non-trivial reparametrizations of F_{Bethe} and of energies in general exists and are exploitable for speeding up optimization, at least in an approximate sense. An affirmative answer would have deep implications not only for probabilistic inference but for many tasks in computer vision, machine learning, and AI in general.

Acknowledgements: The authors would like to thank the reviewers for their feedback. This research was partly supported by the German-Israeli Foundation (GIF) for Scientific Research and Development, 1180-218.6/2011.

References

[1] 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.

[2] H.H. Bui, T.N. Huynh, and D. Sontag. Lifted tree-reweighted variational inference. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

[3] C. Godsil and G. Royle. *Algebraic Graph Theory*. Springer, 2001.

[4] Eric Gribkoff, Guy Van den Broeck, and Dan Suciu. Understanding the complexity of lifted inference and asymmetric weighted model counting. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, July 2014.

[5] M. Grohe, K. Kersting, M. Mladenov, and E. Selman. Dimension reduction via colour refinement. In *Proceedings of the 22th Annual European Symposium on Algorithms (ESA)*, pages 505–516, 2014.

[6] T. Hazan and A. Shashua. Convergent message-passing algorithms for inference over general graphs with convex free energies. In *Proceedings of the Twenty-Forth Conference in Uncertainty in Artificial Intelligence, (UAI)*, pages 264–273, 2008.

[7] T. Heskes. Convexity arguments for efficient minimization of the bethe and kikuchi free energies. *J. Artif. Intell. Res. (JAIR)*, 26:153–190, 2006.

[8] A. Jaimovich, O. Meshi, and N. Friedman. Template-based Inference in Symmetric Relational Markov Random Fields. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence (UAI-07)*, pages 191–199, 2007.

[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] K. Kersting, M. Mladenov, R. Garnett, and M. Grohe. Power iterated color refinement. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, pages 1904–1910, 2014.

[12] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities : theory of majorization and its applications*. Springer series in statistics. Springer, New York, 2011.

[13] T. Meltzer, A. Globerson, and Y. Weiss. Convergent message passing algorithms - a unifying view. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 393–401, 2009.

[14] O. Meshi, A. Jaimovich, A. Globerson, and N. Friedman. Convexifying the bethe free energy. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 402–410, 2009.

[15] 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.

[16] M. Mladenov, A. Globerson, and K. Kersting. Lifted message passing as reparametrization of graphical models. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

[17] M. Mladenov, K. Kersting, and A. Globerson. Efficient lifting of MAP LP relaxations using k-locality. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics (AISTATS) 22-25, 2014*, pages 623–632, 2014.

[18] M. Niepert. Markov chains on orbits of permutation groups. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2012.

[19] M. Niepert and G. Van den Broeck. Tractability through exchangeability: A new perspective on efficient probabilistic inference. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, pages 2467–2475, 2014.

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

[21] A.G. Schwing, T. Hazan, M. Pollefeys, and R. Urtasun. Distributed message passing for large scale graphical models. In *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition,(CVPR)*, pages 1833–1840, 2011.

[22] 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.

[23] G. Van den Broeck and M. Niepert. Lifted probabilistic inference for asymmetric graphical models. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[24] 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.