Conditions Beyond Treewidth for Tightness of Higher-order LP Relaxations

Mark Rowland
University of Cambridge

Aldo Pacchiano
UC Berkeley

Adrian Weller
University of Cambridge

Abstract

Linear programming (LP) relaxations are a popular method to attempt to find a most likely configuration of a discrete graphical model. If a solution to the relaxed problem is obtained at an integral vertex then the solution is guaranteed to be exact and we say that the relaxation is tight. We consider binary pairwise models and introduce new methods which allow us to demonstrate refined conditions for tightness of LP relaxations in the Sherali-Adams hierarchy. Our results include showing that for higher order LP relaxations, treewidth is not precisely the right way to characterize tightness. This work is primarily theoretical, with insights that can improve efficiency in practice.

1 INTRODUCTION

Discrete undirected graphical models are widely used in machine learning, providing a powerful and compact way to model relationships between variables. A key challenge is to identify a most likely configuration of variables, termed *maximum a posteriori* (MAP) or *most probable explanation* (MPE) inference. There is an extensive literature on this problem from various communities, where it may be described as energy minimization (Kappes et al., 2013) or solving a valued constraint satisfaction problem (VCSP, Schiex et al., 1995).

Throughout this paper, we focus on the important class of binary pairwise models (Ising models), allowing arbitrary singleton and edge potentials. For this class, the MAP problem is sometimes described as quadratic pseudo-Boolean optimization (QPBO, e.g. Hammer et al., 1984). In these models, each edge potential may be characterized as either *attractive* (tending to pull its end variables toward the same value; equivalent to a submodular cost function) or *repulsive*. Eaton and Ghahramani (2013) showed that any discrete model may be arbitrarily well approximated by a binary pairwise model, though this may require a large increase in the number of variables.

MAP inference is NP-hard for a general binary pairwise model, hence much work has attempted to identify settings where polynomial-time methods are feasible. We call such settings *tractable* and the methods *efficient*.

In this work, we consider a popular approach which first expresses the MAP problem as an integer linear program (ILP) then relaxes this to a linear program (LP), see §2 for details. An LP attains an optimum at a vertex of the feasible region: if the vertex is integral then it provides an exact solution to the original problem and we say that the LP is *tight*. If the LP is performed over the *marginal polytope*, which enforces global consistency (Wainwright and Jordan, 2008), then the LP will always be tight but exponentially many constraints are required, hence the method is not efficient. The marginal polytope $\mathbb{M}$ is typically relaxed to the *local polytope* $\mathbb{L}_2$, which enforces consistency only over each pair of variables (thus yielding *pseudomarginals* which are pairwise consistent but may not correspond to any true global distribution), requiring only a number of constraints which is linear in the number of edges.

LP relaxations are widely used in practice. However, the most common form LP+$\mathbb{L}_2$ often yields a fractional solution, thus motivating more accurate approaches which enforce higher order cluster consistency (Batra et al., 2011). A well-studied example is foreground-background image segmentation. If edge potentials are learned from data, because objects in the real world are contiguous, most edges will be attractive (typically neighboring pixels will be pulled toward the same identification of foreground or background unless there is strong local data from color or intensity). On the horses dataset considered by Domke (2013), LP+$\mathbb{L}_2$ is loose but if triplet constraints are added, then the LP relaxation is often tight. Our work helps to explain and understand such phenomena. This has clear theoretical value and can improve efficiency in practice.

Sherali and Adams (1990) introduced a series of successively tighter relaxations of the marginal polytope: for any integer r, $\mathbb{L}_r$ enforces consistency over all clusters of vari-

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

ables of size $\leq r$. For any fixed r, LP+$\mathbb{L}_r$ is solvable in polynomial time: higher r leads to improved accuracy but higher runtime. Most earlier work considered the case $r = 2$, though recently there has been progress in understanding conditions for tightness for LP+$\mathbb{L}_3$ (Weller et al., 2016; Weller, 2016b,a).

Here we significantly improve on the result for $\mathbb{L}_3$ of Weller et al. (2016), and provide important new results for when LP+$\mathbb{L}_4$ is guaranteed to be tight, employing an interesting geometric perspective. Our main contributions are summarized in §1.3. We first develop the background and context, see (Deza and Laurent, 1997) for a more extensive survey.

Most previous work considers separately two different types of restricted settings that guarantee tightness, either: (i) constraining the potentials to particular families; or (ii) placing structural restrictions on the topology of connections between variables. As an example of the first type of restriction, it is known that if all edge potentials are attractive (equivalently, if all cost functions are submodular), then the basic relaxation LP+$\mathbb{L}_2$ is tight. In fact, Thapper and Živný (2016) showed that for discrete models with variables with any finite number of labels, and potentials of any order: if no restriction is placed on topology, then for a given family of potentials, either the LP relaxation on the natural local polytope is always tight, and hence solves all such problems efficiently, or the problem set is NP-hard.

1.1 Treewidth, Minors and Conditions for Tightness

Exploring the class of structural restrictions, Chandrasekaran et al. (2008) showed that subject to mild assumptions, if no restriction is placed on types of potentials, then the structural constraint of *bounded treewidth* is needed for tractable *marginal* inference.[1] Indeed, Wainwright and Jordan (2004) proved that if a model topology has treewidth $\leq r-1$ then this is sufficient to guarantee tightness for LP+$\mathbb{L}_r$. As a well-known simple example, if a connected model has treewidth 1 (equivalent to being a tree), then the standard relaxation LP+$\mathbb{L}_2$ is always tight.

The graph property of treewidth $\leq r-1$ is *closed under taking minors* (definitions in §5.1, examples in Figures 1 and 2), hence by the celebrated *graph minor theorem* (Robertson and Seymour, 2004), the property may be characterized by forbidding a unique finite set of minimal *forbidden minors*. Said differently, of all the graphs with treewidth $\geq r$, there is a unique finite set T_r of graphs which are minimal with respect to minor operations. Hence, the sufficient condition of Wainwright and Jordan (2004) for tightness of LP+$\mathbb{L}_r$ may be reframed as: if the graph of a model does not contain any graph in the set T_r as a minor, then LP+$\mathbb{L}_r$ is guaranteed to be tight for any potentials.

The relevant sets of forbidden minors for $\mathbb{L}_2$ and $\mathbb{L}_3$ are particularly simple with just one member each: $T_2 = \{K_3\}$ and $T_3 = \{K_4\}$, where K_n is the complete graph on n vertices. For higher values of r, T_r always contains K_r but there are also other forbidden minors, and their number grows rapidly: T_4 has 4 members (Arnborg et al., 1990) while T_5 has over 70 (Sanders, 1993).

Weller (2016a) showed that, for any r, the graph property of LP+$\mathbb{L}_r$ being *tight for all valid potentials* on the graph is also closed under taking minors. Hence, by Robertson-Seymour, the property for LP+$\mathbb{L}_r$ may be characterized by forbidding a unique set of minimal forbidden minors U_r. It was shown that, in fact, $U_2 = T_2 = \{K_3\}$ and $U_3 = T_3 = \{K_4\}$. However, until this work, all that was known about U_4 is that it contains the complete graph K_5: it has been an open question whether or not $U_4 = T_4$.

One of our main contributions here is to show that $U_4 \neq T_4$. Indeed, in §5 we show that $U_4 \cap T_4 = \{K_5\}$ and that U_4 must contain at least one other forbidden minor, which we cannot yet identify. This progress on understanding U_4 is a significant theoretical development, demonstrating that in general, treewidth is not precisely the right way to characterize tightness of LP relaxations.

Whereas Weller (2016a) made extensive use of powerful earlier results in combinatorics in order to identify U_3, including two results which won the prestigious Fulkerson prize (Lehman, 1990; Guenin, 2001), our analysis takes a different, geometric approach (developed in §4 and §5), which may be of independent interest.

1.2 Stronger Hybrid Conditions

Throughout §1.1, we considered only the graph topology of a model's edge potentials. If we also have access to the *signs* of each edge (attractive or repulsive), then stronger results may be derived. By combining restrictions on both classes of potentials *and* structure, these are termed *hybrid* conditions (Cooper and Živný, 2011).

In this direction, Weller (2016a) showed that for a signed graph, the property that LP+$\mathbb{L}_r$ is tight for all valid potentials (now respecting the graph structure *and* edge signs), is again closed under taking minors, hence again may be characterized by a finite set of minimal forbidden minors U'_r. Further, Weller showed that for $r = 2$ and $r = 3$, the forbidden minors are precisely only the *odd* versions of the forbidden minors for a standard unsigned graph, where an odd version of a graph G means the signed version of G where every edge is repulsive (a repulsive edge is sometimes called odd). That is, $U'_2 = \{\text{odd-}K_3\}$ and $U'_3 = \{\text{odd-}K_4\}$. To see the increased power of these results, observe for example that this means that LP+$\mathbb{L}_3$ is

[1]The treewidth of a graph is one less than the minimum size of a maximum clique in a triangulation of the graph, as used in the junction tree algorithm. Marginal inference seeks the marginal probability distribution for a subset of variables, which is typically harder than MAP inference.

tight for any model of any treewidth, even if it contains K_4 minors, provided only that it does not contain the particular signing of K_4 where all edges are odd (or an equivalent resigning thereof, see §5.1).

In §5, we show somewhat similarly that of all possible signings of K_5, it is only an odd-K_5 which leads to non-tightness of LP+$\mathbb{L}_4$.

1.3 Main Contributions

Given the background in §1.1 and §1.2, here we highlight key contributions.

In §3, we significantly strengthen the result of Weller et al. (2016) for an *almost balanced* model (which contains a distinguished variable s.t. removing it renders the model balanced, see §3.1 for precise definitions). For such a model, it was shown that LP+$\mathbb{L}_3$ is always tight. Here we show that we may relax the polytope from $\mathbb{L}_3$ to a variant we call $\mathbb{L}_3^s$, which enforces triplet constraints only for those triplets including the one distinguished variable s, while still guaranteeing tightness. This has important practical implications since we are guaranteed tightness with dramatically fewer linear constraints, and thus much faster runtime.

We also show in §3 that for any model (no restriction on potentials), enforcing all triplet constraints of $\mathbb{L}_3$ is equivalent to enforcing only those involving the edges of any triangulation (chordal envelope) of its graph, which may significantly improve runtime.

In §4, we introduce a geometric perspective on the tightness of LP relaxations, which may be of independent interest.

In §5, we use these geometric methods to provide powerful new conditions on the tightness of LP+$\mathbb{L}_4$. These show that the relationship which holds between forbidden minors characterizing treewidth and LP+$\mathbb{L}_r$ tightness for $r = 2$ and $r = 3$ breaks down for $r = 4$, hence demonstrating that treewidth is not precisely the right condition for analyzing tightness of higher-order LP relaxations.

1.4 Related Work

We discuss related work throughout the text. To our knowledge, aside from (Weller, 2016a), there is little prior work which considers conditions on signed minors for inference in graphical models. Watanabe (2011) derives a similar characterization to identify when belief propagation has a unique fixed point.

2 BACKGROUND AND PRELIMINARIES

2.1 The MAP Inference Problem

A binary pairwise graphical model is a collection of random variables $(X_i)_{i \in V}$, each taking values in $\{0, 1\}$, such that the joint probability distribution may be written in a minimal representation (Wainwright and Jordan, 2008) as

$$\mathbb{P}(X_i = x_i \; \forall i \in V) \propto \exp\left(\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_i x_j \right), \quad (1)$$

for some potentials $\theta_i \in \mathbb{R}$ for all $i \in V$, and $W_{ij} \in \mathbb{R}$ for all $ij \in E \subseteq V^{(2)}$. We identify the topology of the model as the graph $G = (V, E)$. When $W_{ij} > 0$, there is a preference for X_i and X_j to take the same setting and the edge ij is *attractive*; when $W_{ij} < 0$, the edge is *repulsive* (see Weller, 2014, §2 for details).

A fundamental problem for graphical models is *maximum a posteriori* (MAP) inference, which asks for the identification of a most likely joint state of all the random variables $(X_i)_{i \in V}$ under the probability distribution specified in (1).

The MAP inference problem is clearly equivalent to maximizing the argument of the exponential in (1), yielding the following integer quadratic program:

$$\max_{x \in \{0,1\}^V} \left[\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_i x_j \right]. \quad (2)$$

2.2 LP Relaxations for MAP Inference

A widely used approach to solving Problem (2) is first to replace the integer programming problem with an equivalent linear program (LP), and then to optimize the objective over a relaxed polytope with a polynomial number of linear constraints. This leads to an LP which is efficient to solve but may return a fractional solution (in which case, branching or cutting approaches are often used in practice).

In detail, from Problem 2, an auxiliary variable $x_{ij} = x_i x_j$ is introduced for each edge, as in Problem (3):

$$\max_{\substack{x \in \{0,1\}^{V \cup E} \\ x_{ij} = x_i x_j \forall ij \in E}} \left[\sum_{i \in V} \theta_i x_i + \sum_{ij \in E} W_{ij} x_{ij} \right]. \quad (3)$$

With the objective now linear, an LP may be formed by optimizing over the convex hull of the optimization domain of Problem (3). This convex hull is called the *marginal polytope* (Wainwright and Jordan, 2008) and denoted $\mathbb{M}(G)$. Thus we obtain the equivalent problem:

$$\max_{q \in \mathbb{M}(G)} \left[\sum_{i \in V} \theta_i q_i + \sum_{ij \in E} W_{ij} q_{ij} \right]. \quad (4)$$

This LP (4) is in general no more tractable than Problem (2), due to the number of constraints needed to describe $\mathbb{M}(G)$. It is standard to obtain a tractable problem by *relaxing* the domain of optimization $\mathbb{M}(G)$ to some larger polytope $\mathbb{L}$ which is easier to describe. The resulting relaxed optimization yields an upper bound on the optimal value of Problem (2), though if the $\arg\max$ over $\mathbb{L}$ is an extremal point of $\mathbb{M}(G)$, then the approximation must be exact, and we say that the relaxation is *tight* for this problem instance.

We focus on the family of relaxations introduced by Sherali and Adams (1990), defined below. We first consider a probabilistic interpretation of the marginal polytope.

Notation. For any finite set Z, let $\mathscr{P}(Z)$ be the set of probability measures on Z.

The marginal polytope can then be written

$$\begin{aligned}\mathbb{M}(G) = \{((q_i)_{i\in V}, (q_{ij})_{ij\in E}) \mid \exists \mu \in \mathscr{P}(\{0,1\}^V) \\ \text{s.t. } q_i = \mathbb{P}_\mu(X_i = 1)\ \forall i \in V\,, \\ q_{ij} = \mathbb{P}_\mu(X_i = 1, X_j = 1)\ \forall ij \in E\}.\end{aligned} \tag{5}$$

Intuitively the constraints of the marginal polytope enforce a global consistency condition for the set of parameters $((q_i)_{i\in V}, (q_{ij})_{ij\in E})$, so that together they describe marginal distributions of some global distribution over the entire set of variables. A natural approach is to relax this condition of global consistency to a less stringent notion of consistency only for smaller clusters of variables.

Definition 1. The Sherali-Adams polytope $\mathbb{L}_r(G)$ of order r for a binary pairwise graphical model on $G = (V, E)$ is

$$\begin{aligned}\mathbb{L}_r(G) = \{((q_i)_{i\in V}, (q_{ij})_{ij\in E}) \mid \forall \alpha \subseteq V \text{ with } |\alpha| \leq r, \\ \exists \tau_\alpha \in \mathscr{P}(\{0,1\}^\alpha) \text{ s.t. } \tau_\alpha\big|_\beta = \tau_\beta\ \forall \beta \subset \alpha, \text{ and} \\ \tau_{\{i,j\}}(1,1) = q_{ij}\ \forall ij \in E,\ \tau_{\{i\}}(1) = q_i\ \forall i \in V\},\end{aligned} \tag{6}$$

where for $\beta \subset \alpha$, $\tau_\alpha\big|_\beta$ denotes the marginal distribution of τ_α on $\{0,1\}^\beta$.

Considering successive fixed values of $r \in \mathbb{N}$, the Sherali-Adams polytopes therefore yield the following sequence of tractable approximations to Problem (2):

$$\max_{q\in\mathbb{L}_r(G)} \left[\sum_{i\in V} \theta_i q_i + \sum_{ij\in E} W_{ij} q_{ij}\right]. \tag{7}$$

As r is increased, so too does the number of linear constraints required to define $\mathbb{L}_r$, leading to a tighter polytope and a more accurate solution, but at the cost of greater computationally complexity. For $r = |V|$, $\mathbb{L}_r(G)$ is exactly equal to the marginal polytope $\mathbb{M}(G)$.

A fundamental question concerning LP relaxations for MAP inference is when it is possible to use a computationally cheap relaxation $\mathbb{L}_r(G)$ and still obtain an exact answer to the original inference problem. This is of great practical importance, as it leads to tractable algorithms for particular classes of problems that in full generality are NP-hard. In this paper we investigate this question for a variety of problem classes for the Sherali-Adams relaxations: $\mathbb{L}_2(G)$, $\mathbb{L}_3(G)$, and $\mathbb{L}_4(G)$; also referred to as the local, triplet, and quad polytopes respectively.

3 REFINING TIGHTNESS RESULTS FOR $\mathbb{L}_3(G)$

Here we first derive new results for the triplet polytope $\mathbb{L}_3$ that significantly strengthen earlier work (Weller et al., 2016) for *almost balanced* models. Then in §3.3, we demonstrate that for any model, triplet constraints need be applied only over edges in some triangulation of its graph (that is, over any chordal envelope).

3.1 Graph-theoretic Preliminaries

A *signed graph* is a graph $G = (V, E)$ together with a function $\Sigma : E \to \{\text{even, odd}\}$, where attractive edges are even and repulsive edges are odd. A signed graph is *balanced* (Harary, 1953) if its vertices V may be partitioned into two exhaustive sets V_1, V_2 such that all edges with both endpoints in V_1 or both endpoints in V_2 are even, whilst all edges with one endpoint in each of V_1 and V_2 is odd.

A signed graph $G = (V, E, \Sigma)$ is *almost balanced* (Weller, 2015b) if there exists some distinguished vertex $s \in V$ such that removing s leaves the remainder balanced (thus any balanced graph is almost balanced). To detect if a graph is almost balanced, and if so then to find a distinguished vertex, may be performed efficiently (simply hold out one variable at a time and test the remainder to see if it is balanced, Harary and Kabell, 1980). A graphical model is almost balanced if the signed graph corresponding to its edge potentials is almost balanced.

3.2 Tightness of LP Relaxations for Almost Balanced Models

The following result was shown by Weller et al. (2016).

Theorem 2. *Given a graph G, the triplet polytope $\mathbb{L}_3(G)$ is tight on the class of almost balanced models on G.*

We present a significant strengthening of Theorem 2 in a new direction, which identifies exactly which constraints of $\mathbb{L}_3(G)$ are required to ensure tightness for the class of almost balanced models.

Given an almost balanced model with distinguished variable s (that is, deleting s renders the model balanced), define the polytope $\mathbb{L}_3^s(G)$ by taking all the pairwise constraints of $\mathbb{L}_2(G)$, and adding triplet constraints *only* for triplets of variables that include s (see §10.4.1 in the Supplement for more details).

$\mathbb{L}_3^s(G)$ is a significant relaxation of $\mathbb{L}_3(G)$, requiring only $\mathcal{O}(|V|^2)$ linear constraints, rather than the $\mathcal{O}(|V|^3)$ constraints needed for $\mathbb{L}_3(G)$. Thus it is substantially faster to optimize over $\mathbb{L}_3^s(G)$. Nevertheless, our next result shows that $\mathbb{L}_3^s(G)$ is still tight for an almost balanced model, and indeed is the 'loosest' possible polytope with this property.

Theorem 3. *The polytope $\mathbb{L}_3^s(G)$ is tight on the class of almost balanced models on graph G with distinguished variable s. Further, no linear constraint of $\mathbb{L}_3^s(G)$ may be removed to yield a polytope which is still tight on all models in this class of potentials. Proof in the Supplement §10.*

Considering cutting plane methods, Theorem 3 demonstrates exactly which constraints from $\mathbb{L}_3(G)$ may be necessary to add to the polytope $\mathbb{L}_2(G)$ in order to achieve tightness for an almost balanced model.

3.3 Chordality and Extending Partial Marginals

By its definition, the polytope $\mathbb{L}_2$ enforces pairwise constraints on every pair of variables in a model, whether or not they are connected by an edge, yet typically one enforces constraints only for edges E in the model. This is sufficient because it is not hard to see that if one has edge marginals for the graph $G(V, E)$, it is always possible to extend these to edge marginals for the complete graph on V while remaining within $\mathbb{L}_2$ (and the values of these additional marginals are irrelevant to the score by assumption).

Here we provide an analogous result for $\mathbb{L}_3$, which shows that for any model (no restriction on potentials), one need only enforce triplet constraints over any triangulation (chordal envelope) of its graph.

Theorem 4. *For a chordal graph G, the polytope $\mathbb{L}_2(G)$ together with the triplet constraints only for those triplets of variables that form 3-cliques in G, is equal to $\mathbb{L}_3(G)$, i.e. the polytope given by enforcing constraints on all triplets of G. Proof in the Supplement §9.*

4 THE GEOMETRY OF SHERALI-ADAMS RELAXATIONS

Here we introduce several geometric notions for the Sherali-Adams polytopes which we shall apply in §5.

4.1 The Geometry of the Sherali-Adams Polytopes

The study of tightness of LP relaxations is naturally expressed in the language of polyhedral combinatorics. We introduce key notions from the literature, then provide new proofs of characterizations of tightness for the local polytope $\mathbb{L}_2(G)$ with these geometric ideas.

Given a polytope $P \subset \mathbb{R}^m$, and an extremal point (vertex) $v \in \text{Ext}(P)$, the *normal cone* to P at v, denoted $N_P(v)$, is the polyhedral cone defined by

$$N_P(v) = \left\{ c \in \mathbb{R}^m | v \in \arg\max_{x \in P} \langle c, x \rangle \right\}. \qquad (8)$$

We define the *conical hull* of a finite set X as: $\text{Cone}(X) = \{\sum_{x\in X} \lambda_x x \mid \lambda_x \geq 0 \ \forall x \in X\}$. The following characterization of normal cones will be particularly useful.

Lemma 5 (Theorem 2.4.9, Schneider, 1993)**.** *Let $P = \{x \in \mathbb{R}^m | Ax \leq b\}$ be a polytope for some $A = [a_1, \dots, a_k]^\top \in \mathbb{R}^{k \times m}$, $b \in \mathbb{R}^k$ (for some $k \in \mathbb{N}$). Then for $v \in \text{Ext}(P)$, we have*

$$N_P(v) = \text{Cone}(\{a_i | \langle a_i, v\rangle = b\}). \qquad (9)$$

Further, if the representation $\{x \in \mathbb{R}^m | Ax \leq b\}$ has no redundant constraints, then $\{a_i | \langle a_i, v\rangle = b\}$ is a complete set of extremal rays of $N_P(v)$ (up to scalar multiplication).

With these geometric notions in hand, we have a succinct characterization of the set of potentials for which a given Sherali-Adams relaxation is tight.

Lemma 6. *The set of potentials which are tight with respect to $\mathbb{L}_r$ is exactly given by the following union of cones*

$$\bigcup_{v \in \text{Ext}(\mathbb{M}(G))} N_{\mathbb{L}_r(G)}(v). \qquad (10)$$

This concise characterization, together with the explicit parametrization of normal cones given by Lemma 5 and the form of the linear constraints defining the local polytope $\mathbb{L}_2(G)$, yields an efficient algorithm for generating arbitrary potentials which are tight with respect to $\mathbb{L}_2(G)$.

We would like also to identify classes of potentials for which $\mathbb{L}_2(G)$ is guaranteed to be tight. We demonstrate the power of our geometric approach by providing new proofs in §7 of the Supplement of the following earlier results.

Lemma 7. *If G is a tree, then $\mathbb{L}_2(G)$ is tight for all potentials, that is $\mathbb{L}_2(G) = \mathbb{M}(G)$.*

Lemma 8. *For an arbitrary graph G, $\mathbb{L}_2(G)$ is tight for the set of balanced models.*

The proofs proceed by explicitly demonstrating that a given potential lies in a cone $N_{\mathbb{L}_2(G)}(v)$ for some vertex v of the marginal polytope, by expressing the potential as a conical combination of the extremal rays of the cone.

4.2 The Symmetry of the Sherali-Adams Polytopes

The Sherali-Adams polytopes have rich symmetries which can be exploited when classifying tightness of LP relaxations using the tools discussed in §4.1. Intuitively, these symmetries arise either by considering relabellings of the vertices of the graph G (permutations), or relabellings of the state space of individual variables (flippings). The key result is that a Sherali-Adams polytope is tight a for potential c iff it is tight for any permutation or flipping of c.

4.2.1 The Permutation Group

Let $\sigma \in S_V$ be a relabeling of the vertices of the graph G. This permutation then induces a bijective map Y_σ :

$\mathbb{L}_r(G) \to \mathbb{L}_r(G)$ (which naturally lifts to a linear map on $\mathbb{R}^{V \cup E}$), given by applying the corresponding relabeling to the components of the pseudomarginal vectors:

$$Y_\sigma((q_i)_{i \in V}, (q_{ij})_{ij \in E}) = ((q_{\sigma(i)})_{i \in V}, (q_{\sigma(i)\sigma(j)})_{ij \in E}) \quad \forall ((q_i)_{i \in V}, (q_{ij})_{ij \in E}) \in \mathbb{L}_r(G)\,.$$

The element $\sigma \in S_V$ also naturally induces a linear map on the space of potentials, which is formally the dual space $(\mathbb{R}^{V \cup E})^\star$, although we will frequently identify it with $\mathbb{R}^{V \cup E}$. We denote the map on the space of potentials by $Y_\sigma^\dagger : (\mathbb{R}^{V \cup E})^\star \to (\mathbb{R}^{V \cup E})^\star$, given by

$$Y_\sigma^\dagger((c_i)_{i \in V}, (c_{ij})_{ij \in E}) = ((c_{\sigma(i)})_{i \in V}, (c_{\sigma(i)\sigma(j)})_{ij \in E}) \quad \forall ((c_i)_{i \in V}, (c_{ij})_{ij \in E}) \in \mathbb{R}^{V \cup E}\,.$$

The sets $\{Y_\sigma | \sigma \in S_V\}$ and $\{Y_\sigma^\dagger | \sigma \in S_V\}$ obey the group axioms (under the operation of composition), and hence form groups of symmetries on $\mathbb{L}_r(G)$ and $(\mathbb{R}^{V \cup E})^\star$ respectively; they are both naturally isomorphic to S_V.

These symmetry groups form a useful formalism for thinking about tightness of Sherali-Adams relaxations. We provide one such result in this language, proof in §8 of the Supplement.

Lemma 9. *$\mathbb{L}_r(G)$ is tight for a given potential $c \in \mathbb{R}^{V \cup E}$ iff it is tight for all potentials $Y_\sigma^\dagger(c)$, $\sigma \in S_V$.*

4.2.2 The Flipping Group

Whilst the permutation group described above corresponds to permuting the labels of vertices in the graph, it is also useful to consider the effect of permuting the labels of the states of individual variables. In the case of binary models, the label set is $\{0, 1\}$, so permuting labels corresponds to switching $0 \leftrightarrow 1$, which we refer to as 'flipping'. Given a variable $v \in V$, define the affine map $F_{(v)} : \mathbb{R}^{V \cup E} \to \mathbb{R}^{V \cup E}$ which acts on any pseudomarginal $q \in \mathbb{R}^{V \cup E}$ to flip v as follows (see Weller et al., 2016 for details):

$$\left[F_{(v)}(q)\right]_v = 1 - q_v,$$
$$\left[F_{(v)}(q)\right]_{vw} = q_w - q_{vw},\ \forall vw \in E,$$

while $F_{(v)}$ leaves unchanged all other coordinates of q. Note that $F_{(v)}$ restricts to a bijection on $\mathbb{L}_r(G)$. The flipping maps commute and have order 2, hence the group generated by them, $\langle F_{(v)} | v \in V \rangle$, is isomorphic to $\mathbb{Z}_2^{|V|}$. A general element of this group can be thought of as simultaneously flipping a subset $I \subseteq V$ of variables, written as $F_{(I)} : \mathbb{L}_r(G) \to \mathbb{L}_r(G)$.

Flipping a subset of variables $I \subseteq V$ also naturally induces a map $F_{(I)}^\dagger : (\mathbb{R}^{V \cup E})^\star \to (\mathbb{R}^{V \cup E})^\star$ on the space of potentials. We give a full description of this map in §8 of the Supplement. An analogous result to Lemma 9 also holds for the group of flipping symmetries.

Lemma 10. *$\mathbb{L}_r(G)$ is tight for a given potential $c \in \mathbb{R}^{V \cup E}$ iff it is tight for all potentials $F_{(I)}^\dagger(c)$, $I \subseteq V$.*

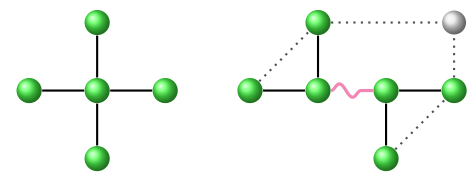


Figure 1: The left graph is a *minor* (unsigned) of the right graph, obtained by deleting the grey dotted edges and resulting isolated grey vertex, and contracting the purple wavy edge. See §5.1.

4.2.3 The Joint Symmetry Group of the Sherali-Adams Polytopes

Tying together the remarks of §4.2.1 and §4.2.2, note that in general the symmetries of the flipping and permutation groups on $\mathbb{L}_r(G)$ do not commute. In fact, observe that

$$Y_\sigma^{-1} \circ F_{(I)} \circ Y_\sigma = F_{(\sigma^{-1}(I))}\,. \tag{11}$$

Thus the group of symmetries of $\mathbb{L}_r(G)$ generated by permutations and flippings is isomorphic to the semidirect product $S_V \rtimes \mathbb{Z}_2^{|V|}$.

5 FORBIDDEN MINOR CONDITIONS FOR TIGHTNESS OF $\mathbb{L}_4(G)$

We first introduce graph minors and their application to the characterizations of both treewidth and tightness of LP relaxations over $\mathbb{L}_r$. While these characterizations are the same for $r \leq 3$, a key contribution in this section is to use the geometric perspective of §4 to show that the characterizations are not the same for $r = 4$.

5.1 Graph Minor Theory

For further background, see (Diestel, 2010, Chapter 12). Given a graph $G = (V, E)$, a graph H is a *minor* of G, written $H \leq G$, if it can be obtained from G via a series of edge deletions, vertex deletions, and edge *contractions*. The result of contracting an edge $uv \in E$ is the graph $G'(V', E')$ where u and v are 'merged' to form a new vertex w which is adjacent to any vertex that was previously adjacent to either u or v. That is $V' = V \setminus \{u, v\} \cup \{w\}$, and $E' = \{e \in E \mid u, v \notin e\} \cup \{wx \mid ux \in E \text{ or } vx \in E\}$. This is illustrated in Figure 1.

A property of a graph is *closed under taking minors* or *minor-closed* if whenever G has the property and $H \leq G$, then H also has the property.

The celebrated *graph minor theorem* of Robertson and Seymour (2004) proves that any minor-closed graph property may be characterized by a unique finite set $\{H_1, \ldots, H_m\}$ of minimal (wrt minor operations) forbidden minors; that is, a graph G has the property iff it does not contain any H_i as a minor. Checking to see if a graph contains some

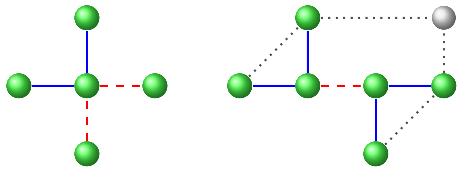


Figure 2: The left graph is a *signed minor* of the right signed graph, obtained similarly to Figure 1 except that before contracting the repulsive edge, first flip the vertex at its right end. Solid blue (dashed red) edges are attractive (repulsive). Grey dotted edges on the right are deleted and may be of any sign. See §5.1.

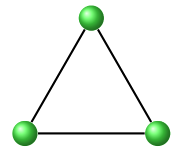


Figure 3: K_3 (unsigned), the only element of T_2, the set of minimal forbidden minors for treewidth ≤ 1. $T_2 = U_2$, the set of forbidden minors for LP+$\mathbb{L}_2$ to be tight. See §5.2.

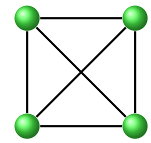
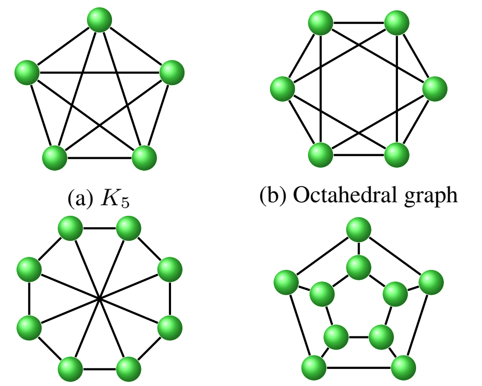


Figure 4: K_4 (unsigned), the only element of T_3, the set of minimal forbidden minors for treewidth ≤ 2. $T_3 = U_3$, the set of forbidden minors for LP+$\mathbb{L}_3$ to be tight. See §5.2.



(a) K_5 (b) Octahedral graph (c) Wagner graph (d) Pentagonal prism graph

Figure 5: The four members of T_4, the set of minimal forbidden minors (unsigned) for treewidth ≤ 3. We show the new result that $T_4 \neq U_4$, in fact $T_4 \cap U_4 = \{K_5\}$, where U_4 is the set of forbidden minors for LP+$\mathbb{L}_4$ to be tight. See §5.2.

H as a minor may be performed efficiently (Robertson and Seymour, 1995).

We shall also consider signed graphs (see §3.1) and their respective *signed minors*. A signed minor of a signed graph is obtained as before by edge deletion, vertex deletion, and edge contraction but now: contraction may be performed only on even (attractive) edges; and any *resigning* operation is also allowed, in which a subset of vertices $S \subseteq V$ is selected then all edges with exactly one end in S are flipped even $\leftrightarrow$ odd.[2] An example is shown in Figure 2. This notion of flipping is closely related to the notion of flipping of potentials, introduced in §4.2.2.

The graph minor theorem of Robertson and Seymour generalizes to signed graphs (Huynh, 2009; Geelen et al., 2014): any property of a signed graph which is closed under taking signed minors, may be characterized by a unique finite set of minimal forbidden signed minors.

5.2 Treewidth Characterizations of Tightness

A fundamental result in the study of LPs over Sherali-Adams relaxations is the following sufficient condition.

Theorem 11 (Wainwright and Jordan, 2004)**.** *If G has treewidth $\leq r-1$, then $\mathbb{L}_r(G) = \mathbb{M}(G)$; equivalently, LP+$\mathbb{L}_r$ is tight for all valid potentials on G.*

The goal of this section is to study to what extent a partial converse to this result holds; that is, to what extent tightness of a Sherali-Adams polytope can hold for graphs of high treewidth. We focus on the case of $\mathbb{L}_4(G)$, and proceed based on the graph minor properties of G.

See §1.1 for a quick review of results that the properties, for any r, of treewidth$\leq r-1$, and of LP+$\mathbb{L}_r$ being *tight for any valid potentials*, are both minor-closed (Weller, 2016a). Thus, by the graph minor theorem, both are characterized by a finite set of minimal forbidden minors.

We shall often be interested in the set of potentials $\widetilde{\Sigma} := \{c \in \mathbb{R}^{V \cup E} \,|\, \text{sign}(c_e) = \Sigma(e)\ \forall e \in E\}$ that are consistent with a given signing Σ of the graph G. We say that the polytope $\mathbb{L}_r(G)$ is *tight* for a signing Σ if LP+$\mathbb{L}_r$ is tight for all potentials in the set $\widetilde{\Sigma} \subset \mathbb{R}^{V \cup E}$.

The property of a signed graph (G, Σ) that $\mathbb{L}_r(G)$ is tight for the signing Σ is also minor-closed (Weller, 2016a), hence can also be characterized by forbidden minimal signed minors.

Recalling the notation introduced in §1.1: for unsigned graphs, we call the respective sets of minimal forbidden minors T_r (for treewidth $\leq r-1$) and U_r (for LP tightness over $\mathbb{L}_r$); for signed graphs, U'_r is the set of minimal forbidden signed minors for tightness of LP+$\mathbb{L}_r$. It is known (Weller, 2016a) that in fact, $T_r = U_r$ for $r = 2, 3$, and that U'_r is exactly the set of odd signings (i.e. signings where all edges of the graph are odd/repulsive) of the graphs of T_r for $r = 2, 3$. We shall show in §5.3 that both of these relationships break down for $r = 4$.

The sets T_2, T_3 and T_4 are shown in Figures 3 to 5, whilst the sets U'_2 and U'_3 are shown in Figures 6 and 7.

[2]Hence, to contract an odd edge, one may first flip either end of the edge to make it even, then contract. In our context of binary pairwise models, flipping a subset S is equivalent to switching from a model with variables $\{X_i\}$ to a new model with variables $\{Y_i : Y_i = 1 - X_i\ \forall i \in S, Y_i = X_i\ \forall i \in V \setminus S\}$ and setting potentials to preserve the distribution, which flips the sign of W_{ij} for any edge ij with one end in S; details in (Weller, 2015a, §2.4).

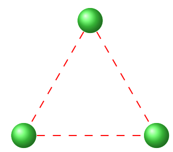


Figure 6: Odd-K_3, the unique element of U_2', the set of minimal forbidden signed minors for tightness of $\mathbb{L}_2(G)$. Red dashed edges represent repulsive edges. See §5.2.

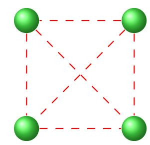


Figure 7: Odd-K_4, the unique element of U_3', the set of minimal forbidden signed minors for tightness of $\mathbb{L}_3(G)$. Red dashed edges represent repulsive edges. See §5.2.

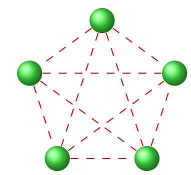


Figure 8: Odd-K_5, the unique signing of an element of T_4, the set of minimal forbidden unsigned minors for treewidth ≤ 3, that appears in U_4', the set of minimal forbidden signed minors for $\mathbb{L}_4(G)$, as shown by our new Theorem 12. It was previously known that for $\mathbb{L}_r, r \leq 3$, the minimal forbidden signed minors for LP-tightness are exactly the odd versions of the minimal forbidden unsigned minors for treewidth $\leq r-1$. We believe there must be at least one other forbidden minor for tightness of $\mathbb{L}_4$, see §5.4. Red dashed edges represent repulsive edges.

5.3 Identifying Forbidden Signed Minors

$\mathbb{L}_4(G)$ is tight for a potential $c \in \mathbb{R}^{V \cup E}$ if and only if

$$\max_{q \in \mathbb{L}_4(G)} \langle c, q \rangle = \max_{x \in \mathbb{M}(G)} \langle c, x \rangle \,, \tag{12}$$

or equivalently if

$$\max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] = 0 \,. \tag{13}$$

Since $\mathbb{M}(G) \subseteq \mathbb{L}_4(G)$, we have $\max_{q \in \mathbb{L}_4(G)} \langle c, q \rangle \geq \max_{x \in \mathbb{M}(G)} \langle c, x \rangle \;\; \forall c \in \mathbb{R}^{V \cup E}$. Hence, it follows that $\mathbb{L}_4(G)$ is not tight for some potential $c \in \widetilde{\Sigma}$ (the set of potentials respecting a signing Σ of G, see §5.2) iff the following optimization problem has a non-zero optimum:

$$\max_{c \in \widetilde{\Sigma}} \max_{q \in \mathbb{L}_4(G)} \min_{x \in \mathbb{M}(G)} \left[\langle c, q \rangle - \langle c, x \rangle \right] \,. \tag{14}$$

For the graphs in T_4, this is a high-dimensional indefinite quadratic program which is intractable to solve. However, using the geometric ideas of §4, we may decompose this problem into a sequence of tractable linear programs. This process involves computing vertex representations (V-representations) for a variety of polytopes using the ideas of §4, and computing the orbits of the set of signings of G under the natural action of the group described in §4.2.3, see the Supplement §11 for full details. Solving these linear programs then allows the exact set of non-tight signings of the graphs in Figure 5 to be identified; see Figure 8.

Theorem 12. *The only non-tight signing for $\mathbb{L}_4$ of any minimal forbidden minor for treewidth ≤ 3 is the odd-K_5.*

5.4 Discussion: Other Forbidden Minors

Previous work showed that tightness for all valid potentials of LP+$\mathbb{L}_2$ may be characterized exactly by forbidding just an odd-K_3 as signed minor, and that a similar result for LP+$\mathbb{L}_3$ holds by forbidding just an odd-K_4 (Weller, 2016a). These are precisely the odd versions of the forbidden minors for the respective treewidth conditions.

A natural conjecture for $\mathbb{L}_4$ was that one must forbid just some signings of the four graphs in T_4, see Figure 5. Now given Theorem 12, it would seem sensible to wonder if LP+$\mathbb{L}_4$ is tight for all valid potentials iff a model's graph does not contain an odd-K_5?

However, this must be false (unless P=NP), since if it were true: We would have LP+$\mathbb{L}_4$ is tight for any model not containing K_5 (as an unsigned minor). It is well-known that planar graphs are those without K_5 or $K_{3,3}$ as a minor ($K_{3,3}$ is the complete bipartite graph where each partition has 3 vertices), i.e. a subclass of graphs which are K_5-free. Hence, we would have a polytime method to solve MAP inference for any planar binary pairwise model. Yet it is not hard to see that we may encode minimal vertex cover in such a model, and it is known that planar minimum vertex cover is NP-hard (Lichtenstein, 1982).

We have not yet been able to identify any other minimal forbidden minor for tightness of LP+$\mathbb{L}_4$, but note that one natural candidate is some signing of a $k \times k$ grid of sufficient size, since this is planar with treewidth k.

6 CONCLUSION

LP relaxations are widely used for the fundamental task of MAP inference for graphical models. Considering binary pairwise models, we have provided important theoretical results on when various relaxations are guaranteed to be tight, which guarantees that in practice, an exact solution can be found efficiently.

A key result focuses on the connection between tightness of LP relaxations of a model and the treewidth of its graph. For the first two levels of the Sherali-Adams hierarchy, that is for the pairwise and triplet relaxations, it was known that the characterizations are essentially identical. However, we have shown that this pattern does not hold for the next level in the hierarchy, that is for the quadruplet polytope $\mathbb{L}_4$.

We refined this result by considering the signed graph of a model and its signed minors. To derive these results we introduced geometric methods which may be of independent interest.

Acknowledgements

We thank Francis Bach, Maria Lomeli and the anonymous reviewers for helpful comments. MR acknowledges support by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L016516/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. AW acknowledges support by the Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

S. Arnborg, A. Proskurowski, and D. Corneil. Forbidden minors characterization of partial 3-trees. *Discrete Mathematics*, 80 (1), February 1990.

D. Batra, S. Nowozin, and P. Kohli. Tighter relaxations for MAP-MRF inference: A local primal-dual gap based separation algorithm. In *AISTATS*, 2011.

V. Chandrasekaran, N. Srebro, and P. Harsha. Complexity of inference in graphical models. In *UAI*, pages 70–78, 2008.

M. Cooper and S. Živný. Hybrid tractability of valued constraint problems. *Artificial Intelligence*, 175(9):1555–1569, 2011.

M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Springer Publishing Company, Incorporated, 1st edition, 1997. ISBN 978-3-642-04294-2.

R. Diestel. *Graph Theory*. Springer, fourth edition, 2010.

J. Domke. Learning graphical model parameters with approximate marginal inference. *TPAMI*, 35(10):2454–2467, 2013.

F. Eaton and Z. Ghahramani. Model reductions for inference: Generality of pairwise, binary, and planar factor graphs. *Neural Computation*, 25(5):1213–1260, 2013.

J. Geelen, B. Gerards, and G. Whittle. Solving Rota's conjecture. *Notices of the AMS*, 61(7):736–743, 2014.

B. Guenin. A characterization of weakly bipartite graphs. *Journal of Combinatorial Theory, Series B*, 83(1):112–168, 2001.

P. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0–1 optimization. *Mathematical programming*, 28(2):121–155, 1984.

F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2:143–146, 1953.

F. Harary and J. Kabell. A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences*, 1(1):131–136, 1980.

T. Huynh. *The linkage problem for group-labelled graphs*. PhD thesis, University of Waterloo, 2009.

J. Kappes, B. Andres, F. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. Kausler, J. Lellmann, N. Komodakis, and C. Rother. A comparative study of modern inference techniques for discrete energy minimization problems. In *CVPR*, 2013.

A. Lehman. On the width-length inequality and degenerate projective planes. In *Polyhedral Combinatorics*, pages 101–105. American Mathematical Society, 1990.

D. Lichtenstein. Planar formulae and their uses. *SIAM Journal on Computing*, 11(2):329–343, 1982.

N. Robertson and P. Seymour. Graph minors. XIII. The disjoint paths problem. *Journal of Combinatorial Theory, Series B*, 63 (1):65–110, 1995.

N. Robertson and P. Seymour. Graph minors. XX. Wagner's conjecture. *Journal of Combinatorial Theory, Series B*, 92(2):325–357, 2004.

D. Sanders. *Linear algorithms for graphs of tree-width at most four*. PhD thesis, Georgia Tech, 1993.

T. Schiex, H. Fargier, and G. Verfaillie. Valued constraint satisfaction problems: Hard and easy problems. *IJCAI (1)*, 95: 631–639, 1995.

R. Schneider. *Convex bodies : the Brunn-Minkowski theory*. Cambridge University Press, 1993.

H. Sherali and W. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

J. Thapper and S. Živný. The complexity of finite-valued CSPs. *Journal of the ACM*, 63(4), 2016.

M. Wainwright and M. Jordan. Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations. *Univ. California, Berkeley, Technical Report*, 671:4, 2004.

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

Y. Watanabe. Uniqueness of belief propagation on signed graphs. In *Neural Information Processing Systems*, 2011.

A. Weller. *Methods for Inference in Graphical Models*. PhD thesis, Columbia University, 2014.

A. Weller. Bethe and related pairwise entropy approximations. In *Uncertainty in Artificial Intelligence (UAI)*, 2015a.

A. Weller. Revisiting the limits of MAP inference by MWSS on perfect graphs. In *Artificial Intelligence and Statistics (AISTATS)*, 2015b.

A. Weller. Characterizing tightness of LP relaxations by forbidding signed minors. In *Uncertainty in Artifical Intelligence (UAI)*, 2016a.

A. Weller. Uprooting and rerooting graphical models. In *International Conference on Machine Learning (ICML)*, 2016b.

A. Weller, M. Rowland, and D. Sontag. Tightness of LP relaxations for almost balanced models. In *Artificial Intelligence and Statistics (AISTATS)*, 2016.