Rounding Guarantees for Message-Passing MAP Inference with Logical Dependencies

Stephen H. Bach
Computer Science Dept.
University of Maryland
College Park, MD 20742
bach@cs.umd.edu

Bert Huang
Computer Science Dept.
University of Maryland
College Park, MD 20742
bert@cs.umd.edu

Lise Getoor
Computer Science Dept.
UC Santa Cruz
Santa Cruz, CA 95064
getoor@soe.ucsc.edu

Abstract

We present the equivalence of the first-order local consistency relaxation and the MAX SAT relaxation of Goemans and Williamson [1] for a class of MRFs we refer to as logical MRFs. This allows us to combine the advantages of both approaches into a single technique: solving the local consistency relaxation with any of a number of message-passing algorithms, and then improving the solution quality via a guaranteed rounding procedure when the relaxation is not tight. Logical MRFs are a general class of models that can incorporate many common dependencies, such as mixtures of submodular and supermodular potentials, and logical implications. They can be used for many tasks, including natural language processing, computer vision, and computational social science.

1 Introduction

One of the canonical problems for probabilistic modeling is finding the most probable assignment to the unobserved random variables, i.e., maximum a posteriori (MAP) inference. For Markov random fields (MRFs), MAP inference is NP-hard in general [2], so approximations are required in practice. In this paper, we provide a new analysis of approximate MAP inference for a particularly flexible and broad class of MRFs we refer to as *logical MRFs*. In these models, potentials are defined by truth tables of disjunctive logical clauses with non-negative weights. This class includes many common types of models, such as mixtures of submodular and supermodular potentials, and many of the models that can be defined using the language Markov logic [3].[1] Such models are useful for the many domains that require expressive dependencies, such as natural language processing, computer vision, and computational social science. MAP inference for logical MRFs is still NP-hard [4], so we consider two main approaches for approximate inference, each with distinct advantages.

The first approach uses *local consistency relaxations* [5]. Instead of solving a combinatorial optimization over discrete variables, MAP inference is first viewed equivalently as the optimization of marginal distributions over variable and potential states. The marginals are then relaxed to *pseudo-marginals*, which are only consistent among local variables and potentials. The primary advantage of local consistency relaxations is that they lead to highly scalable message-passing algorithms, such as dual decomposition [6]. However—except for a few special cases—local consistency relaxations produce fractional solutions, which require some rounding or decoding procedure to find discrete solutions. For most MRFs, including logical MRFs, there are no previously known guarantees on the quality of these solutions.

The second approach to tractable MAP inference for logical MRFs is *weighted maximum satisfiability (MAX SAT) relaxation*, in which one views MAP inference as the classical MAX SAT problem

[1] Markov logic can also include potentials defined by clauses with conjunctions and with negative weights.

and relaxes it to a convex program from that perspective. Given a set of disjunctive logical clauses with associated nonnegative weights, MAX SAT is the problem of finding a Boolean assignment that maximizes the sum of the weights of the satisfied clauses. Convex programming relaxations for MAX SAT also produce fractional solutions, but unlike local consistency relaxations, they offer theoretically guaranteed rounding procedures [1]. However, though these relaxations are tractable in principle, general-purpose convex program solvers do not scale well to large graphical models [7].

In this paper, we unite these two approaches. Our contribution is the following theoretical result: for logical MRFs, the first-order local consistency relaxation and the MAX SAT relaxation of Goemans and Williamson [1] are equivalent. We sketch a proof of this equivalence that analyzes the local consistency relaxation as a hierarchical optimization and reasons about KKT conditions of the optimizations at lower levels of the hierarchy. This new, compact, hierarchical form is easily seen to be equivalent to the MAX SAT relaxation.

This proof of equivalence is important because it reveals that one can combine the advantages of both approaches into a single algorithm that offers scalable and accurate inference by using the message-passing algorithms developed in the graphical models community and the guaranteed rounding procedures of the MAX SAT relaxation.

2 Preliminaries

2.1 Markov random fields

MRFs are probabilistic graphical models that factor according to the structure of an undirected graph. For the purposes of this paper, we consider MRFs with discrete domains.

Definition 1. Let $\boldsymbol{x} = (x_1, \dots, x_n)$ be a vector of n random variables, where each variable x_i has discrete domain $\mathcal{X}_i = \{0, 1, \dots, K_i - 1\}$. Then, let $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)$ be a vector of m potentials, where each potential $\phi_j(\boldsymbol{x})$ maps states of a subset of the variables $\boldsymbol{x}_j$ to real numbers. Finally, let $\boldsymbol{w} = (w_1, \dots, w_m)$ be a vector of m real-valued parameters. Then, a **Markov random field** over $\boldsymbol{x}$ is a probability distribution of the form $P(\boldsymbol{x}) \propto \exp\left(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x})\right)$.

2.2 Local consistency relaxations

A popular approach for tractable inference in MRFs is local consistency relaxation [5]. This approach starts by viewing MAP inference as an equivalent optimization over marginal probabilities. For each $\phi_j \in \boldsymbol{\phi}$, let $\boldsymbol{\theta}_j$ be a marginal distribution over joint assignments to the variable subset $\boldsymbol{x}_j$. For example, $\theta_j(\boldsymbol{x}_j)$ is the probability that $\boldsymbol{x}_j$ is in a particular joint state. Also, let $x_j(i)$ denote the setting of the variable with index i in the state $\boldsymbol{x}_j$.

With this variational formulation, inference can be relaxed to an optimization over the *first-order local polytope* $\mathbb{L}$. Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ be a vector of probability distributions, where $\mu_i(k)$ is the marginal probability that x_i is in state k. The first-order local polytope is

$$\mathbb{L} \triangleq \left\{ \boldsymbol{\theta}, \boldsymbol{\mu} \geq \mathbf{0} \;\middle|\; \begin{array}{ll} \sum_{\boldsymbol{x}_j | x_j(i) = k} \theta_j(\boldsymbol{x}_j) = \mu_i(k) & \forall i, j, k \\ \sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j) = 1 & \forall j \\ \sum_{k=0}^{K_i - 1} \mu_i(k) = 1 & \forall i \end{array} \right\},$$

which constrains each marginal distribution $\boldsymbol{\theta}_j$ over joint states $\boldsymbol{x}_j$ to be consistent only with the marginal distributions $\boldsymbol{\mu}$ over individual variables that participate in the potential ϕ_j.

MAP inference can then be approximated with the *first-order local consistency relaxation*:

$$\arg\max_{(\boldsymbol{\theta}, \boldsymbol{\mu}) \in \mathbb{L}} \sum_{j=1}^{m} w_j \sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j)\, \phi_j(\boldsymbol{x}_j), \tag{1}$$

which is an upper bound on the true MAP objective. The first-order local consistency relaxation is a much more tractable linear program than exact inference, and it can be applied to any MRF. Much work has focused on solving the first-order local consistency relaxation for large-scale MRFs, which we discuss further in Section 4. However, in general, the solutions are fractional, and there are no guarantees on the approximation quality of a discretization of these fractional solutions.

2.3 Logical Markov random fields

We now turn to the focus of this paper: logical MRFs, which are MRFs whose potentials $\boldsymbol{\phi}$ are defined by disjunctive Boolean clauses with associated nonnegative weights, formally defined as follows.

Definition 2. Let $\boldsymbol{C} = (C_1, \ldots, C_m)$ be a vector of logical clauses, where each clause $C_j \in \boldsymbol{C}$ is a disjunction of literals and each literal is a variable x or its negation $\neg x$ such that each variable $x_i \in \boldsymbol{x}$ appears at most once in C_j. Let I_j^+ (resp. I_j^-) $\subseteq \{1, \ldots, n\}$ be the set of indices of the variables that are not negated (resp. negated) in C_j. Then C_j can be written as $\left(\bigvee_{i \in I_j^+} x_i\right) \vee \left(\bigvee_{i \in I_j^-} \neg x_i\right)$. A **logical Markov random field** is an MRF in which each variable x_i has Boolean domain $\{0, 1\}$, i.e., $K_i = 2$, each potential $\phi_j(\boldsymbol{x}) = 1$ if $\boldsymbol{x}$ satisfies C_j and 0 otherwise, and each parameter $w_j \geq 0$.

Logical MRFs are very expressive. A clause C_j can be viewed equivalently as an implication from conditions to consequences: $\bigwedge_{i \in I_j^-} x_i \implies \bigvee_{i \in I_j^+} x_i$. If multiple sets of conditions should imply the same set of possible consequences, or if one set of conditions should imply multiple sets of possible consequences, then additional clauses can be added to the set $\boldsymbol{C}$, covering the cases that cannot be expressed in a single clause. Moreover, the generality of logical MRFs can be stated more broadly: MAP inference for any discrete distribution of bounded factor size can be converted to a MAX SAT problem—and therefore MAP inference for a logical MRF—of size polynomial in the variables and clauses [8].

2.4 MAX SAT relaxations

The MAP problem for a logical MRF can also be viewed as an instance of MAX SAT and approximately solved from this perspective. The MAX SAT problem is to find a Boolean assignment to the variables $\boldsymbol{x}$ that maximizes the sum of the weights of the satisfied clauses from a set of clauses $\boldsymbol{C}$. A solution to MAX SAT is also the MAP state of the logical MRF defined via $\boldsymbol{C}$. Since MAX SAT is NP-hard, convex programming relaxations are a tractable approach.

Goemans and Williamson [1] introduced a linear programming relaxation that provides rounding guarantees for the solution. We review their technique and the results of their analysis here. For each variable x_i, associate with it a continuous variable $y_i \in [0, 1]$. Then, let $\boldsymbol{y}^\star$ be the solution to the linear program

$$\arg\max_{\boldsymbol{y} \in [0,1]^n} \sum_{C_j \in \boldsymbol{C}} w_j \min \left\{ \sum_{i \in I_j^+} y_i + \sum_{i \in I_j^-} (1 - y_i), 1 \right\} . \tag{2}$$

After solving the linear program, each variable x_i is independently set to 1 according to a rounding probability function f, i.e., $p_i = f(y_i^\star)$. Many functions can be chosen for f, but a simple one [1] analyze is the linear function $f(y_i^\star) = \frac{1}{2} y_i^\star + \frac{1}{4}$. Let $\hat{W}$ be the expected total weight of the satisfied clauses from using this randomized rounding procedure. Let $W^\star$ be the maximum total weight of the satisfied clauses over all assignments to $\boldsymbol{x}$, i.e., the weight of the MAX SAT solution. Goemans and Williamson [1] showed that $\hat{W} \geq \frac{3}{4} W^\star$. The method of conditional probabilities [9] can deterministically find an assignment to $\boldsymbol{x}$ that achieves a total weight of at least $\hat{W}$. Each variable x_i is greedily set to the value that maximizes the expected weight over the unassigned variables, conditioned on either possible value of x_i and the previously assigned variables. This greedy maximization can be applied quickly because, in many models, variables only participate in a small fraction of the clauses, making the change in expectation quick to compute for each variable.

3 Equivalence analysis

In this section, we prove the equivalence of the first-order consistency relaxation and the MAX SAT relaxation of Goemans and Williamson [1] for logical MRFs (Theorem 1). Our proof analyzes the local consistency relaxation to derive an equivalent, more compact optimization over only the variable pseudomarginals $\boldsymbol{\mu}$. We show that this compact form is identical to the MAX SAT relaxation.

Since the variables are Boolean, we refer to each pseudomarginal $\mu_i(1)$ as simply μ_i. Let $\boldsymbol{x}_j^F$ denote the unique setting such that $\phi_j(\boldsymbol{x}_j^F) = 0$.

We begin by reformulating the local consistency relaxation as a hierarchical optimization, first over the variable pseudomarginals $\boldsymbol{\mu}$ and then over the factor pseudomarginals $\boldsymbol{\theta}$. Due to the structure of local polytope $\mathbb{L}$, the pseudomarginals $\boldsymbol{\mu}$ parameterize inner linear programs that decompose over the structure of the MRF, such that—given fixed $\boldsymbol{\mu}$—there is an independent linear program $\hat{\phi}_j(\boldsymbol{\mu})$ over $\boldsymbol{\theta}_j$ for each clause C_j. We rewrite objective (1) as

$$\arg\max_{\boldsymbol{\mu}\in[0,1]^n} \sum_{C_j \in \boldsymbol{C}} \hat{\phi}_j(\boldsymbol{\mu}), \tag{3}$$

where

$$\hat{\phi}_j(\boldsymbol{\mu}) = \max_{\boldsymbol{\theta}_j \geq \boldsymbol{0}} \; w_j \sum_{\boldsymbol{x}_j | \boldsymbol{x}_j \neq \boldsymbol{x}_j^F} \theta_j(\boldsymbol{x}_j) \tag{4}$$

$$\text{s.t.} \sum_{\boldsymbol{x}_j | x_j(i)=1} \theta_j(\boldsymbol{x}_j) = \mu_i \qquad \forall i \in I_j^+ \tag{5}$$

$$\sum_{\boldsymbol{x}_j | x_j(i)=0} \theta_j(\boldsymbol{x}_j) = 1 - \mu_i \qquad \forall i \in I_j^- \tag{6}$$

$$\sum_{\boldsymbol{x}_j} \theta_j(\boldsymbol{x}_j) = 1 \tag{7}$$

It is straightforward to verify that objectives (1) and (3) are equivalent for logical MRFs. All constraints defining $\mathbb{L}$ can be derived from the constraint $\boldsymbol{\mu} \in [0,1]^n$ and the constraints in the definition of $\hat{\phi}_j(\boldsymbol{\mu})$. We have omitted redundant constraints to simplify analysis.

To make this optimization more compact, we replace each inner linear program $\hat{\phi}_j(\boldsymbol{\mu})$ with an expression that gives its optimal value for any setting of $\boldsymbol{\mu}$. Deriving this expression requires reasoning about any maximizer $\boldsymbol{\theta}_j^\star$ of $\hat{\phi}_j(\boldsymbol{\mu})$, which is guaranteed to exist because program (4) is bounded and feasible for any parameters $\boldsymbol{\mu} \in [0,1]^n$ and w_j.

Theorem 1. *For a logical MRF, the first-order local consistency relaxation of MAP inference is equivalent to the MAX SAT relaxation of Goemans and Williamson [1]. Specifically, any partial optimum $\boldsymbol{\mu}^\star$ of objective (1) is an optimum $\boldsymbol{y}^\star$ of objective (2), and vice versa.*

Proof. (Sketch) To derive a simplified expression for $\hat{\phi}_j(\boldsymbol{\mu})$ that makes objectives (3) and (2) equivalent, we first identify a condition on $\boldsymbol{\mu}$ that is sufficient for program (4) to not be fully satisifiable, i.e., it cannot achieve the maximum value of w_j. Throughout, we assume $w_j > 0$, since the derivation is trivial if $w_j = 0$. If $\sum_{i\in I_j^+} \mu_i + \sum_{i \in I_j^-}(1-\mu_i) < 1$, then $\theta_j^\star(\boldsymbol{x}_j^F) > 0$, by constraints (5), (6), and (7). In this case, program (4) equals $\sum_{i\in I_j^+} \mu_i + \sum_{i \in I_j^-}(1-\mu_i)$, which can be shown via the Karush-Kuhn-Tucker (KKT) conditions [10, 11] for program (4). If instead $\sum_{i\in I_j^+} \mu_i + \sum_{i \in I_j^-}(1-\mu_i) \geq 1$, then program (4) equals w_j. Together, these results give the solution to program (4) as a piecewise-linear expression. Substituting this expression into the local consistency relaxation objective (3) gives a projected optimization over only $\boldsymbol{\mu}$ which is identical to the MAX SAT relaxation objective (2). □

4 Related work

A large body of work has focused on solving local consistency relaxations of MAP inference quickly. The theory developed in this paper is applicable to any algorithm that can find the optimal variable pseudomarginals $\boldsymbol{\mu}^\star$. We consider two families of such approaches.

The first approach is dual decomposition [6, DD], which solves a dual problem of (1). Only some DD algorithms can be used to find the optimum $\boldsymbol{\mu}^\star$ in order to compute rounding probabilities. Subgradient methods for DD (e.g., Jojic et al. [12], Komodakis et al. [13], and Schwing et al. [14]) can find $\boldsymbol{\mu}^\star$ in many ways, including those described by Anstreicher and Wolsey [15], Nedić and

Ozdaglar [16], and Shor [17]. Other DD algorithms, such as TRW-S [18], MSD [19], MPLP [20], and ADLP [21], use coordinate descent to solve the dual objective. In general, there is no known way to find the primal solution $\boldsymbol{\mu}^\star$ with coordinate descent DD.

The second approach uses message-passing algorithms to solve objective (1) directly in its primal form and therefore always finds $\boldsymbol{\mu}^\star$. One well-known algorithm is that of Ravikumar et al. [22], which uses proximal optimization, a general approach that iteratively improves the solution by searching for nearby improvements. The authors also provide rounding guarantees for when the relaxed solution is integral, i.e., the relaxation is tight, allowing the algorithm to converge faster. Such guarantees are complementary to ours, since we consider the case when the relaxation is not tight. Another message-passing algorithm that solves the primal objective is AD^3 [23], which uses the alternating direction method of multipliers [24, ADMM]. AD^3 optimizes objective (1) for binary, pairwise MRFs and supports the addition of certain deterministic constraints on the variables. A third example of a primal message-passing algorithm is APLP [21], which is the primal analog of ADLP. Like AD^3, it uses ADMM to optimize the objective.

In addition to the many approaches for solving the first-order local consistency relaxation, other approaches to approximating MAP inference include tighter linear programming relaxations [25, 26]. These tighter relaxations enforce local consistency on variable subsets that are larger than individual variables, which makes them *higher-order local consistency relaxations*. Mezuman et al. [27] developed techniques for special cases of higher-order relaxations, such as when the MRF contains cardinality potentials, in which the probability of a configuration depends on the number of variables in a particular state. Some papers have also explored nonlinear convex programming relaxations, e.g., Ravikumar and Lafferty [28] and Kumar et al. [29].

Previous analyses have identified particular subclasses whose local consistency relaxations are tight, i.e., the maximum of the relaxed program is exactly the maximum of the original problem. These special classes include graphical models with tree-structured dependencies, models with submodular potential functions, models encoding bipartite matching problems, and those with *nand* potentials and perfect graph structures [5, 30, 31, 32]. These tightness guarantees are powerful, but they require more restrictive conditions on the distributions than our analysis. Our results complement these types of analyses by identifying a larger class of problems with an approximation-quality guarantee.

Researchers have studied performance guarantees of other subclasses of the first-order local consistency relaxation. Kleinberg and Tardos [33] and Chekuri et al. [34] considered the metric labeling problem. Feldman et al. [35] used the local consistency relaxation to decode binary linear codes.

Finally, we note the work of Huynh and Mooney [36], which introduced a linear programming relaxation for Markov logic networks [3] inspired by MAX SAT relaxations. Markov logic networks subsume logical MRFs, but the relaxation of general Markov logic provides no guarantees on the quality of solutions.

5 Conclusion

We presented the equivalence of the first-order local consistency relaxation and the MAX SAT relaxation of Goemans and Williamson [1] for logical MRFs. This result is important because the local consistency relaxation can first be solved with any of a number of scalable message-passing algorithms, and the quality of the results can be improved with a guaranteed rounding procedure.

Acknowledgments This work was supported by NSF grant IIS1218488, and IARPA via DoI/NBC contract number D12PC00337. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

References

[1] M. X. Goemans and D. P. Williamson. New 3/4-approximation algorithms for the maximum satisfiability problem. *SIAM J. Discrete Math.*, 7(4):656–666, 1994.

[2] S. E. Shimony. Finding MAPs for belief networks is NP-hard. *Artificial Intel.*, 68(2):399–410, 1994.

[3] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.

[4] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, 1976.

[5] M. J. Wainwright and M. I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers, 2008.

[6] D. Sontag, A. Globerson, and T. Jaakkola. Introduction to dual decomposition for inference. In S. Sra, S. Nowozin, and S. J. Wright, editors, *Opt. for Machine Learning*, pages 219–254. MIT Press, 2011.

[7] C. Yanover, T. Meltzer, and Y. Weiss. Linear programming relaxations and belief propagation – An empirical study. *J. Mach. Learn. Res.*, 7:1887–1907, 2006.

[8] J. D. Park. Using weighted MAX-SAT engines to solve MPE. In *National Conference on Artificial Intelligence (AAAI)*, 2002.

[9] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley-Interscience, third edition, 2008.

[10] W. Karush. Minima of Functions of Several Variables with Inequalities as Side Constraints. Master's thesis, University of Chicago, 1939.

[11] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Berkeley Symp. Math. Stat. and Prob.*, 1951.

[12] V. Jojic, S. Gould, and D. Koller. Accelerated dual decomposition for MAP inference. In *International Conference on Machine Learning (ICML)*, 2010.

[13] N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 33(3):531–552, 2011.

[14] A. G. Schwing, T. Hazan, M. Pollefeys, and R. Urtasun. Globally convergent dual MAP LP relaxation solvers using Fenchel-Young margins. In *Advances in Neural Information Processing Systems*, 2012.

[15] K. M. Anstreicher and L. A. Wolsey. Two "well-known" properties of subgradient optimization. *Mathematical Programming*, 120(1):213–220, 2009.

[16] A. Nedić and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optimization*, 19(4):1757–1780, 2009.

[17] N. Z. Shor. *Minimization Methods for Non-Differentiable Functions*. Springer-Verlag, 1985.

[18] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 28(10):1568–1583, 2006.

[19] T. Werner. A linear programming approach to max-sum problem: A review. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 29(7):1165–1179, 2007.

[20] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *Advances in Neural Information Processing Systems*, 2007.

[21] O. Meshi and A. Globerson. An alternating direction method for dual MAP LP relaxation. In *European Conference on Machine learning (ECML)*, 2011.

[22] P. Ravikumar, A. Agarwal, and M. J. Wainwright. Message-passing for graph-structured linear programs: Proximal methods and rounding schemes. *J. Mach. Learn. Res.*, 11:1043–1080, 2010.

[23] A. Martins, M. Figueiredo, P. Aguiar, N. Smith, and E. Xing. An augmented Lagrangian approach to constrained MAP inference. In *International Conference on Machine Learning (ICML)*, 2011.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*. Now Publishers, 2011.

[25] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss. Tightening LP relaxations for MAP using message passing. In *Uncertainty in Aritificial Intelligence (UAI)*, 2008.

[26] D. Sontag, D. K. Choe, and Y. Li. Efficiently searching for frustrated cycles in MAP inference. In *Uncertainty in Aritificial Intelligence (UAI)*, 2012.

[27] E. Mezuman, D. Tarlow, A. Globerson, and Y. Weiss. Tighter linear program relaxations for high order graphical models. In *Uncertainty in Aritificial Intelligence (UAI)*, 2013.

[28] P. Ravikumar and J. Lafferty. Quadratic programming relaxations for metric labeling and Markov random field MAP estimation. In *International Conference on Machine Learning (ICML)*, 2006.

[29] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Solving Markov random fields using second order cone programming relaxations. In *Computer Vision and Pattern Recognition (CVPR)*, 2006.

[30] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer-Verlag, 2003.

[31] T. Jebara. MAP estimation, message passing, and perfect graphs. In *Uncertainty in Artificial Intelligence (UAI)*, 2009.

[32] J. Foulds, N. Navaroli, P. Smyth, and A. Ihler. Revisiting MAP estimation, message passing and perfect graphs. In *AI & Statistics*, 2011.

[33] J. Kleinberg and É. Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *J. ACM*, 49(5):616–639, 2002.

[34] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM J. Discrete Math.*, 18(3):608–625, 2005.

[35] J. Feldman, M. J. Wainwright, and D. R. Karger. Using linear programming to decode binary linear codes. *Information Theory, IEEE Trans. on*, 51(3):954–972, 2005.

[36] T. Huynh and R. Mooney. Max-margin weight learning for Markov logic networks. In *European Conference on Machine Learning (ECML)*, 2009.