An Analysis of Convex Relaxations for MAP Estimation

M. Pawan Kumar
Dept. of Computing
Oxford Brookes University
pkmudigonda@brookes.ac.uk

V. Kolmogorov
Computer Science
University College London
vnk@adastral.ucl.ac.uk

P.H.S. Torr
Dept. of Computing
Oxford Brookes University
philiptorr@brookes.ac.uk

Abstract

The problem of obtaining the maximum a posteriori estimate of a general discrete random field (i.e. a random field defined using a finite and discrete set of labels) is known to be NP-hard. However, due to its central importance in many applications, several approximate algorithms have been proposed in the literature. In this paper, we present an analysis of three such algorithms based on convex relaxations: (i) LP-S: the linear programming (LP) relaxation proposed by Schlesinger [25] for a special case and independently in [4, 17, 31] for the general case; (ii) QP-RL: the quadratic programming (QP) relaxation by Ravikumar and Lafferty [22]; and (iii) SOCP-MS: the second order cone programming (SOCP) relaxation first proposed by Muramatsu and Suzuki [20] for two label problems and later extended in [18] for a general label set.

We show that the SOCP-MS and the QP-RL relaxations are equivalent. Furthermore, we prove that despite the flexibility in the form of the constraints/objective function offered by QP and SOCP, the LP-S relaxation *strictly dominates* (i.e. provides a better approximation than) QP-RL and SOCP-MS. We generalize these results by defining a large class of SOCP (and equivalent QP) relaxations which is dominated by the LP-S relaxation. Based on these results we propose some novel SOCP relaxations which strictly dominate the previous approaches.

1 Introduction

Discrete random fields are a powerful tool to obtain a probabilistic formulation for various applications in Computer Vision and related areas [3, 5]. Hence, developing accurate and efficient algorithms for performing inference on a given discrete random field is of fundamental importance. In this work, we will focus on the problem of maximum a posteriori (MAP) estimation. MAP estimation is a key step in obtaining the solutions to many applications such as stereo, image stitching and segmentation [29]. Furthermore, it is closely related to many important Combinatorial Optimization problems such as MAXCUT [8], multi-way cut [6], metric labelling [3, 14] and 0-extension [3, 12].

Given data **D**, a discrete random field models the distribution (i.e. either the joint or the conditional probability) of a labelling for a set of random variables. Each of these variables $\mathbf{v} = \{v_0, v_1, \cdots, v_{n-1}\}$ can take a label from a discrete set $\mathbf{l} = \{l_0, l_1, \cdots, l_{h-1}\}$. A particular labelling of variables $\mathbf{v}$ is specified by a function f whose domain corresponds to the indices of the random variables and whose range is the index of the label set, i.e.

$$f : \{0, 1, \cdots, n-1\} \rightarrow \{0, 1, \cdots, h-1\}. \tag{1}$$

In other words, random variable v_a takes label $l_{f(a)}$. For convenience, we assume the model to be a conditional random field (CRF) while noting that all the results of this paper also apply to Markov random fields (MRF).

A CRF specifies a neighbourhood relationship $\mathcal{E}$ between the random variables, i.e. $(a, b) \in \mathcal{E}$ if, and only if, v_a and v_b are neighbouring random variables. Within this framework, the conditional probability of a labelling f given data $\mathbf{D}$ is specified as

$$\Pr(f|\mathbf{D},\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\exp(-Q(f;\mathbf{D},\boldsymbol{\theta}). \tag{2}$$

Here $\boldsymbol{\theta}$ represents the parameters of the CRF and $Z(\boldsymbol{\theta})$ is a normalization constant which ensures that the probability sums to one (also known as the partition function). The energy $Q(f;\mathbf{D},\boldsymbol{\theta})$ is given by

$$Q(f;\mathbf{D},\boldsymbol{\theta}) = \sum_{v_a\in\mathbf{v}} \theta^1_{a;f(a)} + \sum_{(a,b)\in\mathcal{E}} \theta^2_{ab;f(a)f(b)}. \tag{3}$$

The term $\theta^1_{a;f(a)}$ is called a unary potential since its value depends on the labelling of one random variable at a time. Similarly, $\theta^2_{ab;f(a)f(b)}$ is called a pairwise potential as it depends on a pair of random variables. For simplicity, we assume that $\theta^2_{ab;f(a)f(b)} = w(a,b)d(f(a),f(b))$ where $w(a,b)$ is the weight that indicates the strength of the pairwise relationship between variables v_a and v_b, with $w(a,b)=0$ if $(a,b)\notin\mathcal{E}$, and $d(\cdot,\cdot)$ is a distance function on the labels[1]. As will be seen later, this formulation of the pairwise potentials would allow us to concisely describe our results.

We note that a subclass of this problem where $w(a,b)\geq 0$ and the distance function $d(\cdot,\cdot)$ is a semi-metric or a metric has been well-studied in the literature [3, 4, 14]. However, we will focus on the general MAP estimation problem. In other words, unless explicitly stated, we do not place any restriction on the form of the unary and pairwise potentials.

The problem of MAP estimation is well known to be NP-hard in general. Since it plays a central role in several applications, many approximate algorithms have been proposed in the literature. In this work, we analyze three such algorithms which are based on convex relaxations. Specifically, we consider: (i) LP-S, the linear programming (LP) relaxation of [4, 17, 25, 31]; (ii) QP-RL, the quadratic programming (QP) relaxation of [22]; and (iii) SOCP-MS, the second order cone programming (SOCP) relaxation of [18, 20]. In order to provide an outline of these relaxations, we formulate the problem of MAP estimation as an Integer Program (IP).

1.1 Integer Programming Formulation

We define a binary variable vector $\mathbf{x}$ of length nh. We denote the element of $\mathbf{x}$ at index $a\cdot h+i$ as $x_{a;i}$ where $v_a\in\mathbf{v}$ and $l_i\in\mathbf{l}$. These elements $x_{a;i}$ specify a labelling f such that

$$x_{a;i} = \begin{cases} 1 & \text{if} \quad f(a)=i, \\ -1 & \text{otherwise.} \end{cases} \tag{4}$$

We say that the variable $x_{a;i}$ *belongs to* variable v_a since it defines which label v_a does (or does not) take. Let $\mathbf{X}=\mathbf{x}\mathbf{x}^\top$. We refer to the $(a\cdot h+i, b\cdot h+j)^{th}$ element of the matrix $\mathbf{X}$ as $X_{ab;ij}$ where $v_a,v_b\in\mathbf{v}$ and $l_i,l_j\in\mathbf{l}$. Clearly the sum of the unary potentials for a labelling specified by $(\mathbf{x},\mathbf{X})$ is given by

$$\sum_{v_a,l_i} \theta^1_{a;i}\frac{(1+x_{a;i})}{2}. \tag{5}$$

Similarly the sum of the pairwise potentials for a labelling $(\mathbf{x},\mathbf{X})$ is given by

$$\sum_{(a,b)\in\mathcal{E},l_i,l_j} \theta^2_{ab;ij}\frac{(1+x_{a;i})}{2}\frac{(1+x_{b;j})}{2} = \sum_{(a,b)\in\mathcal{E},l_i,l_j} \theta^2_{ab;ij}\frac{(1+x_{a;i}+x_{b;j}+X_{ab;ij})}{4}. \tag{6}$$

---

[1] The pairwise potentials for any CRF can be represented in the form $\theta^2_{ab;ij} = w(a,b)d(i,j)$. This can be achieved by using a larger set of labels $\hat{\mathbf{l}} = \{l_{0;0},\cdots,l_{0;h_1},\cdots,l_{n-1;h_1}\}$ such that the unary potential of v_a taking label $l_{b;i}$ is $\theta^1_{a;i}$ if $a=b$ and ∞ otherwise. In other words, a variable v_a can only take labels from the set $\{l_{a;0},\cdots,l_{a;h-1}\}$ since all other labels will result in an energy value of ∞. The pairwise potential for variables v_a and v_b taking labels $l_{a;i}$ and $l_{b;j}$ respectively can then be represented in the form $w(a,b)d(a;i,b;j)$ where $w(a,b)=1$ and $d(a;i,b;j)=\theta^2_{ab;ij}$. Note that using a larger set of labels $\hat{\mathbf{l}}$ will increase the time complexity of MAP estimation algorithms, but does not effect the analysis presented in this paper.

Hence, the following IP finds the labelling with the minimum energy, i.e. it is equivalent to the MAP estimation problem:

$$\text{IP:} \quad \mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{v_a, l_i} \theta^1_{a;i} \frac{(1+x_{a;i})}{2} + \sum_{(a,b)\in\mathcal{E}, l_i, l_j} \theta^2_{ab;ij} \frac{(1+x_{a;i}+x_{b;j}+X_{ab;ij})}{4}$$

$$\text{s.t.} \quad \mathbf{x} \in \{-1, 1\}^{nh}, \tag{7}$$

$$\sum_{l_i \in \mathbf{l}} x_{a;i} = 2 - h, \tag{8}$$

$$\mathbf{X} = \mathbf{x}\mathbf{x}^\top. \tag{9}$$

Constraints (7) and (9) specify that the variables $\mathbf{x}$ and $\mathbf{X}$ are binary such that $X_{ab;ij} = x_{a;i}x_{b;j}$. We will refer to them as the *integer constraints*. Constraint (8), which specifies that each variable should be assigned only one label, is known as the *uniqueness constraint*. Note that one uniqueness constraint is specified for each variable v_a. Solving the above IP is in general NP-hard. It is therefore common practice to obtain an approximate solution using convex relaxations. We describe four such convex relaxations below.

1.2 Linear Programming Relaxation

The LP relaxation, proposed by Schlesinger [25] for a special case (where the pairwise potentials specify a hard constraint, i.e. they are either 0 or ∞) and independently in [4, 17, 31] for the general case, is given as follows:

$$\text{LP-S:} \quad \mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{v_a, l_i} \theta^1_{a;i} \frac{(1+x_{a;i})}{2} + \sum_{(a,b)\in\mathcal{E}, l_i, l_j} \theta^2_{ab;ij} \frac{(1+x_{a;i}+x_{b;j}+X_{ab;ij})}{4}$$

$$\text{s.t.} \quad \mathbf{x} \in [-1, 1]^{nh}, \mathbf{X} \in [-1, 1]^{nh\times nh}, \tag{10}$$

$$\sum_{l_i \in \mathbf{l}} x_{a;i} = 2 - h, \tag{11}$$

$$\sum_{l_j \in \mathbf{l}} X_{ab;ij} = (2-h)x_{a;i}, \tag{12}$$

$$X_{ab;ij} = X_{ba;ji}, \tag{13}$$

$$1 + x_{a;i} + x_{b;j} + X_{ab;ij} \geq 0. \tag{14}$$

In the above relaxation, which we call LP-S, only those elements $X_{ab;ij}$ of $\mathbf{X}$ are used for which $(a, b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$. Unlike the IP, the feasibility region of the above problem is relaxed such that the variables $x_{a;i}$ and $X_{ab;ij}$ lie in the interval $[-1, 1]$. Further, the constraint (9) is replaced by equation (12) which is called the *marginalization constraint* [31]. One marginalization constraint is specified for each $(a, b) \in \mathcal{E}$ and $l_i \in \mathbf{l}$. Constraint (13) specifies that $\mathbf{X}$ is symmetric. Constraint (14) ensures that $\theta^2_{ab;ij}$ is multiplied by a number between 0 and 1 in the objective function. These constraints (13) and (14) are defined for all $(a, b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$. The formulation of the LP-S relaxation presented here uses a slightly different notation to the ones described in [15, 31]. However, it can easily be shown that the two formulations are equivalent by using the variables $\mathbf{y}$ and $\mathbf{Y}$ instead of $\mathbf{x}$ and $\mathbf{X}$ such that $y_{a;i} = \frac{1+x_{a;i}}{2}, Y_{ab;ij} = \frac{1+x_{a;i}+x_{b;j}+X_{ab;ij}}{4}$. Note that the above constraints are not exhaustive, i.e. it is possible to specify other constraints for the problem of MAP estimation (as will be seen in the different relaxations described in the subsequent sections).

Properties of the LP-S Relaxation:

- Since the LP-S relaxation specifies a linear program it can be solved in polynomial time. A labelling f can then be obtained by rounding the (possibly fractional) solution of the LP-S.
- Using the rounding scheme of [14], the LP-S provides a multiplicative bound[2] of 2 when the pairwise potentials form a Potts model [4].
- Using the rounding scheme of [4], LP-S obtains a multiplicative bound of $2 + \sqrt{2}$ for truncated linear pairwise potentials.

[2]Consider a set of optimization problems $\mathcal{A}$ and a relaxation scheme defined over this set $\mathcal{A}$. In other words, for every optimization problem $\text{A} \in \mathcal{A}$, the relaxation scheme provides a relaxation $\text{B} \in \mathcal{B}$ of A. Let e^A denote the optimal value of the optimization problem A. Further, let $\hat{e}^A$ denote the value of the objective function of A at the point obtained by rounding the optimal solution of its relaxation B. The relaxation scheme is said to provide a multiplicative bound of ρ for the set $\mathcal{A}$ if, and only if, the following condition is satisfied: $E(\hat{e}^A) \leq \rho e^A, \forall \text{A} \in \mathcal{A}$, where $E(\cdot)$ denotes the expectation of its argument under the rounding scheme employed.

- LP-S provides a multiplicative bound of 1 when the energy function $Q(\cdot; \mathbf{D}, \boldsymbol{\theta})$ of the CRF is submodular [26] (also see [11, 24] for the st-MINCUT graph construction for minimizing submodular energy functions).
- The LP-S relaxation provides the same optimal solution for all reparameterizations $\overline{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ (i.e. for all $\overline{\boldsymbol{\theta}} \equiv \boldsymbol{\theta}$) [15, 34].

Although the LP-S relaxation can be solved in polynomial time, the state of the art Interior Point algorithms can only handle up to a few thousand variables and constraints. In order to overcome this deficiency several efficient algorithms have been proposed in the literature for approximately solving the Lagrangian dual of LP-S [15, 16, 27, 28, 31, 34].

1.3 Quadratic Programming Relaxation

We now describe the QP relaxation for the MAP estimation IP which was proposed by Ravikumar and Lafferty [22]. To this end, it would be convenient to reformulate the objective function of the IP using a vector of unary potentials of length nh (denoted by $\hat{\boldsymbol{\theta}}^1$) and a matrix of pairwise potentials of size $nh \times nh$ (denoted by $\hat{\boldsymbol{\theta}}^2$). The element of the unary potential vector at index $(a \cdot h + i)$ is defined as

$$\hat{\theta}^1_{a;i} = \theta^1_{a;i} - \sum_{v_c \in \mathbf{v}} \sum_{l_k \in \mathbf{l}} |\theta^2_{ac;ik}|, \tag{15}$$

where $v_a \in \mathbf{v}$ and $l_i \in \mathbf{l}$. The $(a \cdot h + i, b \cdot h + j)^{th}$ element of the pairwise potential matrix $\hat{\boldsymbol{\theta}}^2$ is defined such that

$$\hat{\theta}^2_{ab;ij} = \begin{cases} \sum_{v_c \in \mathbf{v}} \sum_{l_k \in \mathbf{l}} |\theta^2_{ac;ik}|, & \text{if} \quad a = b, i = j, \\ \theta^2_{ab;ij} & \text{otherwise,} \end{cases} \tag{16}$$

where $v_a, v_b \in \mathbf{v}$ and $l_i, l_j \in \mathbf{l}$. In other words, the potentials are modified by defining a pairwise potential $\hat{\theta}^2_{aa;ii}$ and subtracting the value of that potential from the corresponding unary potential $\theta^1_{a;i}$. The advantage of this reformulation is that the matrix $\hat{\boldsymbol{\theta}}^2$ is guaranteed to be positive semidefinite, i.e. $\hat{\boldsymbol{\theta}}^2 \succeq 0$. Using the fact that for $x_{a;i} \in \{-1, 1\}$,

$$\left(\frac{1 + x_{a;i}}{2} \right)^2 = \frac{1 + x_{a;i}}{2}, \tag{17}$$

it can be shown that the following is equivalent to the MAP estimation problem [22]:

$$\text{QP-RL:} \qquad \mathbf{x}^* = \arg\min_{\mathbf{x}} \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right)^\top \hat{\boldsymbol{\theta}}^1 + \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right)^\top \hat{\boldsymbol{\theta}}^2 \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right), \tag{18}$$

$$\text{s.t.} \qquad \sum_{l_i \in \mathbf{l}} x_{a;i} = 2 - h, \forall v_a \in \mathbf{v}, \tag{19}$$

$$\mathbf{x} \in \{-1, 1\}^{nh}, \tag{20}$$

where $\mathbf{1}$ is a vector of appropriate dimensions whose elements are all equal to 1. By relaxing the feasibility region of the above problem to $\mathbf{x} \in [-1, 1]^{nh}$, the resulting QP can be solved in polynomial time since $\hat{\boldsymbol{\theta}}^2 \succeq 0$ (i.e. the relaxation of the QP (18)-(20) is convex). We call the above relaxation QP-RL. Note that in [22], the QP-RL relaxation was described using the variable $\mathbf{y} = \frac{\mathbf{1}+\mathbf{x}}{2}$. However, the above formulation can easily be shown to be equivalent to the one presented in [22].

Ravikumar and Lafferty [22] proposed a rounding scheme for QP-RL (different from the ones used in [4, 14]) that provides an additive bound[3] of $\frac{S}{4}$ for the MAP estimation problem, where $S = \sum_{(a,b) \in \mathcal{E}} \sum_{l_i, l_j \in \mathbf{l}} |\theta^2_{ab;ij}|$ (i.e. S is the sum of the absolute values of all pairwise potentials) [22]. Under their rounding scheme, this bound can be shown to be tight[4] using a random

[3] A relaxation scheme defined over the set of optimization problems $\mathcal{A}$ is said to provide an additive bound of σ for $\mathcal{A}$ if, and only if, the following holds true: $E(\hat{e}^A) \leq e^A + \sigma, \forall \text{A} \in \mathcal{A}$. Here e^A is the optimal value of A and $\hat{e}^A$ is the value obtained by rounding the solution of B.

[4] The multiplicative bound specified by a relaxation scheme defined over the set of optimization problems $\mathcal{A}$ is said to be *tight* if, and only if, there exists an $\text{A} \in \mathcal{A}$ such that $E(\hat{e}^A) = \rho e^A$. Similarly, the additive bound specified by a relaxation scheme defined over $\mathcal{A}$ is said to be tight if, and only if, there exists an $\text{A} \in \mathcal{A}$ such that $E(\hat{e}^A) = e^A + \sigma$.

field defined over two random variables which specifies uniform unary potentials and Ising model pairwise potentials. Further, they also proposed an efficient iterative procedure for solving the QP-RL relaxation approximately. However, unlike LP-S, no multiplicative bounds have been established for the QP-RL formulation for special cases of the MAP estimation problem.

1.4 Semidefinite Programming Relaxation

The SDP relaxation of the MAP estimation problem replaces the non-convex constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ by the convex semidefinite constraint $\mathbf{X} - \mathbf{x}\mathbf{x}^\top \succeq 0$ [7, 8, 19] which can be expressed as

$$\begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succeq 0, \tag{21}$$

using Schur's complement [2]. Further, like LP-S, it relaxes the integer constraints by allowing the variables $x_{a;i}$ and $X_{ab;ij}$ to lie in the interval $[-1, 1]$ with $X_{aa;ii} = 1$ for all $v_a \in \mathbf{v}, l_i \in \mathbf{l}$. Note that the value of $X_{aa;ii}$ is derived using the fact that $X_{aa;ii} = x_{a;i}^2$. Since $x_{a;i}$ can only take the values -1 or 1 in the MAP estimation IP, it follows that $X_{aa;ii} = 1$. The SDP relaxation is a well-studied approach which provides accurate solutions for the MAP estimation problem (e.g. see [33]). However, due to its computational inefficiency, it is not practically useful for large scale problems with $nh > 1000$. See however [21, 23, 30].

1.5 Second Order Cone Programming Relaxation

We now describe the SOCP relaxation that was proposed by Muramatsu and Suzuki [20] for the MAXCUT problem (i.e. MAP estimation with $h = 2$) and later extended for a general label set [18]. This relaxation, which we call SOCP-MS, is based on the technique of Kim and Kojima [13]. For completeness we first describe the general technique of [13] and later show how SOCP-MS can be derived using it.

SOCP **Relaxations:** Kim and Kojima [13] observed that the SDP constraint $\mathbf{X} - \mathbf{x}\mathbf{x}^\top \succeq 0$ can be further relaxed to second order cone (SOC) constraints. Their technique uses the fact that the Frobenius inner product of two semidefinite matrices is non-negative. For example, consider the inner product of a fixed matrix $\mathbf{C} = \mathbf{U}\mathbf{U}^\top \succeq 0$ with $\mathbf{X} - \mathbf{x}\mathbf{x}^\top$ (which, by the SDP constraint, is also positive semidefinite). This inner product can be expressed as an SOC constraint as follows:

$$\mathbf{C} \bullet (\mathbf{X} - \mathbf{x}\mathbf{x}^\top) \geq 0, \tag{22}$$

$$\Rightarrow \|(\mathbf{U})^\top \mathbf{x}\|^2 \leq \mathbf{C} \bullet \mathbf{X}. \tag{23}$$

Hence, by using a set of matrices $\mathcal{S} = \{\mathbf{C}^k | \mathbf{C}^k = \mathbf{U}^k(\mathbf{U}^k)^\top \succeq 0, k = 1, 2, \ldots, n_C\}$, the SDP constraint can be further relaxed to n_C SOC constraints, i.e.

$$\Rightarrow \|(\mathbf{U}^k)^\top \mathbf{x}\|^2 \leq \mathbf{C}^k \bullet \mathbf{X}, k = 1, \cdots, n_C. \tag{24}$$

It can be shown that, for the above set of SOC constraints to be equivalent to the SDP constraint, $n_C = \infty$. However, in practice, we can only specify a finite set of SOC constraints. Each of these constraints may involve some or all variables $x_{a;i}$ and $X_{ab;ij}$. For example, if $C^k_{ab;ij} = 0$, then the k^{th} SOC constraint will not involve $X_{ab;ij}$ (since its coefficient will be 0).

The SOCP-MS **Relaxation:** Consider a pair of neighbouring variables v_a and v_b, i.e. $(a, b) \in \mathcal{E}$, and a pair of labels l_i and l_j. These two pairs define the following variables: $x_{a;i}$, $x_{b;j}$, $X_{aa;ii} = X_{bb;jj} = 1$ and $X_{ab;ij} = X_{ba;ji}$ (since $\mathbf{X}$ is symmetric). For each such pair of variables and labels, the SOCP-MS relaxation specifies two SOC constraints which involve only the above variables [18, 20]. In order to specify the exact form of these SOC constraints, we need the following definitions.

Using the variables v_a and v_b (where $(a, b) \in \mathcal{E}$) and labels l_i and l_j, we define the submatrices $\mathbf{x}^{(a,b,i,j)}$ and $\mathbf{X}^{(a,b,i,j)}$ of $\mathbf{x}$ and $\mathbf{X}$ respectively as:

$$\mathbf{x}^{(a,b,i,j)} = \begin{pmatrix} x_{a;i} \\ x_{b;j} \end{pmatrix}, \mathbf{X}^{(a,b,i,j)} = \begin{pmatrix} X_{aa;ii} & X_{ab;ij} \\ X_{ba;ji} & X_{bb;jj} \end{pmatrix}. \tag{25}$$

The SOCP-MS relaxation specifies SOC constraints of the form:

$$\|(\mathbf{U}^k_{MS})^\top \mathbf{x}^{(a,b,i,j)}\|^2 \leq \mathbf{C}^k_{MS} \bullet \mathbf{X}^{(a,b,i,j)}, \tag{26}$$

for all pairs of neighbouring variables $(a,b) \in \mathcal{E}$ and labels $l_i, l_j \in \mathbf{l}$. To this end, it uses the following two matrices:

$$\mathbf{C}^1_{MS} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{C}^2_{MS} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{27}$$

In other words SOCP-MS specifies a total of $2|\mathcal{E}|h^2$ SOC constraints. Note that both the matrices $\mathbf{C}^1_{MS}$ and $\mathbf{C}^2_{MS}$ defined above are positive semidefinite, and hence can be written as $\mathbf{C}^1_{MS} = \mathbf{U}^1_{MS}(\mathbf{U}^1_{MS})^\top$ and $\mathbf{C}^1_{MS} = \mathbf{U}^1_{MS}(\mathbf{U}^1_{MS})^\top$ where

$$\mathbf{U}^1_{MS} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \mathbf{U}^2_{MS} = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \tag{28}$$

Substituting these matrices in inequality (26) we see that the constraints defined by the SOCP-MS relaxation are given by

$$\begin{aligned}
& \|(\mathbf{U}^1_{MS})^\top \mathbf{x}^{(a,b,i,j)}\|^2 \leq \mathbf{C}^1_{MS} \bullet \mathbf{X}^{(a,b,i,j)}, \\
& \|(\mathbf{U}^2_{MS})^\top \mathbf{x}^{(a,b,i,j)}\|^2 \leq \mathbf{C}^2_{MS} \bullet \mathbf{X}^{(a,b,i,j)}, && (29)\\
\Rightarrow \quad & \left\| \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_{a;i} \\ x_{b;j} \end{pmatrix} \right\|^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \bullet \begin{pmatrix} X_{aa;ii} & X_{ab;ij} \\ X_{ba;ji} & X_{bb;jj} \end{pmatrix}, \\
& \left\| \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_{a;i} \\ x_{b;j} \end{pmatrix} \right\|^2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \bullet \begin{pmatrix} X_{aa;ii} & X_{ab;ij} \\ X_{ba;ji} & X_{bb;jj} \end{pmatrix}, && (30)\\
\Rightarrow \quad & (x_{a;i} + x_{b;j})^2 \leq X_{aa;ii} + X_{bb;jj} + X_{ab;ij} + X_{ba;ji}, \\
& (x_{a;i} - x_{b;j})^2 \leq X_{aa;ii} + X_{bb;jj} - X_{ab;ij} - X_{ba;ji}, && (31)\\
\Rightarrow \quad & (x_{a;i} + x_{b;j})^2 \leq 2 + 2X_{ab;ij}, \\
& (x_{a;i} - x_{b;j})^2 \leq 2 - 2X_{ab;ij}. && (32)
\end{aligned}$$

The last expression is obtained using the fact that $\mathbf{X}$ is symmetric and $X_{aa;ii} = 1$, for all $v_a \in \mathbf{v}$ and $l_i \in \mathbf{l}$. Hence, in the SOCP-MS formulation, the MAP estimation IP is relaxed to

$$\begin{aligned}
\text{SOCP-MS:} \quad & \mathbf{x}^* = \arg\min_{\mathbf{x}} \textstyle\sum_{v_a, l_i} \theta^1_{a;i} \frac{(1+x_{a;i})}{2} + \sum_{(a,b)\in\mathcal{E}, l_i, l_j} \theta^2_{ab;ij} \frac{(1+x_{a;i}+x_{b;j}+X_{ab;ij})}{4} \\
\text{s.t.} \quad & \mathbf{x} \in [-1,1]^{nh}, \mathbf{X} \in [-1,1]^{nh \times nh}, && (33)\\
& \textstyle\sum_{l_i \in \mathbf{l}} x_{a;i} = 2 - h, && (34)\\
& (x_{a;i} - x_{b;j})^2 \leq 2 - 2X_{ab;ij}, && (35)\\
& (x_{a;i} + x_{b;j})^2 \leq 2 + 2X_{ab;ij}, && (36)\\
& X_{ab;ij} = X_{ba;ji}. && (37)
\end{aligned}$$

We refer the reader to [18, 20] for details. The SOCP-MS relaxation yields the supremum and infimum for the elements of the matrix $\mathbf{X}$ using constraints (35) and (36) respectively, i.e.

$$\frac{(x_{a;i} + x_{b;j})^2}{2} - 1 \leq X_{ab;ij} \leq 1 - \frac{(x_{a;i} - x_{b;j})^2}{2}. \tag{38}$$

These constraints are specified for all $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$. When the objective function of SOCP-MS is minimized, one of the two inequalities would be satisfied as an equality. This can be proved by assuming that the value for the vector $\mathbf{x}$ has been fixed. Hence, the elements of the matrix $\mathbf{X}$ should take values such that it minimizes the objective function subject to the constraints (35) and (36). Clearly, the objective function would be minimized when $X_{ab;ij}$ equals either its supremum or infimum value, depending on the sign of the corresponding pairwise potential $\theta^2_{ab;ij}$, i.e.

$$X_{ab;ij} = \begin{cases} \frac{(x_{a;i}+x_{b;j})^2}{2} - 1 & \text{if} \quad \theta^2_{ab;ij} > 0, \\ 1 - \frac{(x_{a;i}-x_{b;j})^2}{2} & \text{otherwise.} \end{cases} \tag{39}$$

Similar to the LP-S and QP-RL relaxations defined above, the SOCP-MS relaxation can also be solved in polynomial time. To the best our knowledge, no bounds have been established for the SOCP-MS relaxation in earlier work. Furthermore, no previous specialized algorithms exist for solving SOCP-MS (or indeed any other SOCP relaxation) efficiently.

2 Comparing Relaxations

In order to compare the relaxations described above, we require the following definitions. We say that a relaxation A *dominates* the relaxation B (alternatively, B is dominated by A) if and only if

$$\min_{(\mathbf{x},\mathbf{X})\in\mathcal{F}(\mathrm{A})} e(\mathbf{x},\mathbf{X};\boldsymbol{\theta}) \geq \min_{(\mathbf{x},\mathbf{X})\in\mathcal{F}(\mathrm{B})} e(\mathbf{x},\mathbf{X};\boldsymbol{\theta}), \forall\boldsymbol{\theta}, \tag{40}$$

where $\mathcal{F}(\mathrm{A})$ and $\mathcal{F}(\mathrm{B})$ are the feasibility regions of the relaxations A and B respectively. The term $e(\mathbf{x},\mathbf{X};\boldsymbol{\theta})$ denotes the value of the objective function at $(\mathbf{x},\mathbf{X})$ (i.e. the energy of the possibly fractional labelling $(\mathbf{x},\mathbf{X})$) for the MAP estimation problem defined over the CRF with parameter $\boldsymbol{\theta}$. Thus the optimal value of the dominating relaxation A is always greater than or equal to the optimal value of relaxation B. We note here that the concept of domination has been used previously in [4] (to compare LP-S with the linear programming relaxation in [14]).

Relaxations A and B are said to be *equivalent* if A dominates B and B dominates A, i.e. their optimal values are equal to each other for all CRFs. A relaxation A is said to *strictly dominate* relaxation B if A dominates B but B does not dominate A. In other words there exists at least one CRF with parameter $\boldsymbol{\theta}$ such that

$$\min_{(\mathbf{x},\mathbf{X})\in\mathcal{F}(\mathrm{A})} e(\mathbf{x},\mathbf{X};\boldsymbol{\theta}) > \min_{(\mathbf{x},\mathbf{X})\in\mathcal{F}(\mathrm{B})} e(\mathbf{x},\mathbf{X};\boldsymbol{\theta}). \tag{41}$$

Note that, by definition, the optimal value of any relaxation would always be less than or equal to the energy of the optimal (i.e. the MAP) labelling. Hence, the optimal value of a strictly dominating relaxation A is closer to the optimal value of the MAP estimation IP compared to that of relaxation B. In other words, A provides a tighter lower bound for MAP estimation than B.

We now describe two special cases of domination which are used extensively in the remainder of this paper.

Case I: Consider two relaxations A and B which share a common objective function. For example, the objective functions of the LP-S and the SOCP-MS relaxations described in the previous section have the same form. Further, let A and B differ in the constraints that they specify such that $\mathcal{F}(\mathrm{A}) \subseteq \mathcal{F}(\mathrm{B})$, i.e. the feasibility region of A is a subset of the feasibility region of B.

Given two such relaxations, we claim that A dominates B. This can be proved by contradiction. To this end, we assume that A does not dominate B. Therefore, by definition of domination, there exists at least one parameter $\boldsymbol{\theta}$ for which B provides a greater value of the objective function than A. Let an optimal solution of A be $(\mathbf{x}_A,\mathbf{X}_A)$. Similarly, let $(\mathbf{x}_B,\mathbf{X}_B)$ be an optimal solution of B. By our assumption, the following holds true:

$$e(\mathbf{x}_A,\mathbf{X}_A;\boldsymbol{\theta}) < e(\mathbf{x}_B,\mathbf{X}_B;\boldsymbol{\theta}). \tag{42}$$

However, since $\mathcal{F}(\mathrm{A}) \subseteq \mathcal{F}(\mathrm{B})$ it follows that $(\mathbf{x}_A,\mathbf{X}_A) \in \mathcal{F}(\mathrm{B})$. Hence, from equation (42), we see that $(\mathbf{x}_B,\mathbf{X}_B)$ cannot be an optimal solution of B. This proves our claim.

We can also consider a case where $\mathcal{F}(\mathrm{A}) \subset \mathcal{F}(\mathrm{B})$, i.e. the feasibility region of A is a strict subset of the feasibility region of B. Using the above argument we see that A dominates B. Further, assume that there exists a parameter $\boldsymbol{\theta}$ such that the intersection of the set of all optimal solutions of A and the set of all optimal solutions of B is null. In other words if $(\mathbf{x}_B,\mathbf{X}_B)$ is an optimal solution of B then $(\mathbf{x}_B,\mathbf{X}_B) \notin \mathcal{F}(\mathrm{A})$. Clearly, if such a parameter $\boldsymbol{\theta}$ exists then A strictly dominates B.

Case II: Consider two relaxations A and B such that they share a common objective function. Further, let the constraints of B be a subset of the constraints of A. We claim that A dominates B. This follows from the fact that $\mathcal{F}(\mathrm{A}) \subseteq \mathcal{F}(\mathrm{B})$ and the argument used in Case I above.

2.1 Our Results

We prove that LP-S strictly dominates SOCP-MS (see section 3). Further, in section 4, we show that QP-RL is equivalent to SOCP-MS. This implies that LP-S strictly dominates the QP-RL relaxation. In section 5 we generalize the above results by proving that a large class of SOCP (and equivalent QP) relaxations is dominated by LP-S. Based on these results, we propose a novel set of constraints which result in SOCP relaxations that dominate LP-S, QP-RL and SOCP-MS. These relaxations introduce SOC constraints on cycles and cliques formed by the neighbourhood relationship of the CRF.

3 LP-S vs. SOCP-MS

We now show that for the MAP estimation problem the linear constraints of LP-S, i.e.

$$\mathbf{x} \in [-1,1]^{nh}, \mathbf{X} \in [-1,1]^{nh \times nh}, \tag{43}$$
$$\textstyle\sum_{l_i \in \mathbf{l}} x_{a;i} = 2-h, \tag{44}$$
$$\textstyle\sum_{l_j \in \mathbf{l}} X_{ab;ij} = (2-h)x_{a;i}, \tag{45}$$
$$X_{ab;ij} = X_{ba;ji}, \tag{46}$$
$$1 + x_{a;i} + x_{b;j} + X_{ab;ij} \geq 0. \tag{47}$$

are stronger than the SOCP-MS constraints, i.e.

$$\mathbf{x} \in [-1,1]^{nh}, \mathbf{X} \in [-1,1]^{nh \times nh}, \tag{48}$$
$$\textstyle\sum_{l_i \in \mathbf{l}} x_{a;i} = 2-h, \tag{49}$$
$$(x_{a;i} - x_{b;j})^2 \leq 2 - 2X_{ab;ij}, \tag{50}$$
$$(x_{a;i} + x_{b;j})^2 \leq 2 + 2X_{ab;ij}, \tag{51}$$
$$X_{ab;ij} = X_{ba;ji}. \tag{52}$$

In other words the feasibility region of LP-S is a strict subset of the feasibility region of SOCP-MS (i.e. $\mathcal{F}(\text{LP-S}) \subset \mathcal{F}(\text{SOCP-MS})$). This in turn would allow us to prove the following theorem.

Theorem 1: The LP-S relaxation strictly dominates the SOCP-MS relaxation.

Proof: The LP-S and the SOCP-MS relaxations differ only in the way they relax the non-convex constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$. While LP-S relaxes $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ using the marginalization constraint (45), SOCP-MS relaxes it to constraints (50) and (51). The SOCP-MS constraints provide the supremum and infimum of $X_{ab;ij}$ as

$$\frac{(x_{a;i} + x_{b;j})^2}{2} - 1 \leq X_{ab;ij} \leq 1 - \frac{(x_{a;i} - x_{b;j})^2}{2}. \tag{53}$$

Consider a pair of neighbouring variables v_a and v_b and a pair of labels l_i and l_j. Recall that SOCP-MS specifies the constraints (50) and (51) for all such pairs of random variables and labels, i.e. for all $x_{a;i}, x_{b;j}, X_{ab;ij}$ such that $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$. In order to prove this theorem we use the following two lemmas.

Lemma 3.1: If $x_{a;i}$, $x_{b;j}$ and $X_{ab;ij}$ satisfy the LP-S constraints, i.e. constraints (43)-(47), then

$$|x_{a;i} - x_{b;j}| \leq 1 - X_{ab;ij}. \tag{54}$$

The above result holds true for all $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$.

Proof: From the LP-S constraints, we get

$$\begin{aligned} \frac{1 + x_{a;i}}{2} &= \sum_{l_k \in \mathbf{l}} \frac{1 + x_{a;i} + x_{b;k} + X_{ab;ik}}{4}, \\ \frac{1 + x_{b;j}}{2} &= \sum_{l_k \in \mathbf{l}} \frac{1 + x_{a;k} + x_{b;j} + X_{ab;kj}}{4}. \end{aligned} \tag{55}$$

Therefore,

$$\begin{aligned} |x_{a;i} - x_{b;j}| &= 2\left|\tfrac{1+x_{a;i}}{2} - \tfrac{1+x_{b;j}}{2}\right|, \\ &= 2\left|\left(\tfrac{1+x_{a;i}}{2} - \tfrac{1+x_{a;i}+x_{b;j}+X_{ab;ij}}{4}\right) - \left(\tfrac{1+x_{b;j}}{2} - \tfrac{1+x_{a;i}+x_{b;j}+X_{ab;ij}}{4}\right)\right|, \\ &\leq 2\left(\tfrac{1+x_{a;i}}{2} - \tfrac{1+x_{a;i}+x_{b;j}+X_{ab;ij}}{4}\right) + 2\left(\tfrac{1+x_{b;j}}{2} - \tfrac{1+x_{a;i}+x_{b;j}+X_{ab;ij}}{4}\right), \\ &= 1 - X_{ab;ij}. \end{aligned} \tag{56}$$

Note that the inequality holds since both the expressions in the parantheses, i.e.

$$\left(\frac{1 + x_{a;i}}{2} - \frac{1 + x_{a;i} + x_{b;j} + X_{ab;ij}}{4}\right), \left(\frac{1 + x_{b;j}}{2} - \frac{1 + x_{a;i} + x_{b;j} + X_{ab;ij}}{4}\right), \tag{57}$$

are non-negative, as follows from equations (47) and (55). $\blacksquare$

Using the above lemma, we get

$$
\begin{aligned}
& (x_{a;i} - x_{b;j})^2 \leq (1 - X_{ab;ij})(1 - X_{ab;ij}), && (58)\\
\Rightarrow \quad & (x_{a;i} - x_{b;j})^2 \leq 2(1 - X_{ab;ij}), && (59)\\
\Rightarrow \quad & (x_{a;i} - x_{b;j})^2 \leq 2 - 2X_{ab;ij}. && (60)
\end{aligned}
$$

Inequality (59) is obtained using the fact that $-1 \leq X_{ab;ij} \leq 1$ and hence, $1 - X_{ab;ij} \leq 2$. Using inequality (58), we see that the necessary condition for the equality to hold true is $(1 - X_{ab;ij})(1 - X_{ab;ij}) = 2 - 2X_{ab;ij}$, i.e. $X_{ab;ij} = -1$. Note that inequality (60) is equivalent to the SOCP-MS constraint (50). Thus LP-S provides a smaller supremum of $X_{ab;ij}$ when $X_{ab;ij} > -1$.

Lemma 3.2: If $x_{a;i}$, $x_{b;j}$ and $X_{ab;ij}$ satisfy the LP-S constraints then

$$|x_{a;i} + x_{b;j}| \leq 1 + X_{ab;ij}. \tag{61}$$

This holds true for all $(a, b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$.

Proof: According to constraint (47),

$$-(x_{a;i} + x_{b;j}) \leq 1 + X_{ab;ij}. \tag{62}$$

Using Lemma 3.1, we get the following set of inequalities:

$$|x_{a;i} - x_{b;k}| \leq 1 - X_{ab;ik}, l_k \in \mathbf{l}, k \neq j \tag{63}$$

Adding the above set of inequalities, we get

$$
\begin{aligned}
& \textstyle\sum_{l_k \in \mathbf{l}, k \neq j} |x_{a;i} - x_{b;j}| \leq \sum_{l_k \in \mathbf{l}, k \neq j} (1 - X_{ab;ik}), && (64)\\
\Rightarrow \quad & \textstyle\sum_{l_k \in \mathbf{l}, k \neq j} (x_{a;i} - x_{b;k}) \leq \sum_{l_k \in \mathbf{l}, k \neq j} (1 - X_{ab;ik}), && (65)\\
\Rightarrow \quad & \textstyle (h-1)x_{a;i} - \sum_{l_k \in \mathbf{l}, k \neq j} x_{b;k} \leq (h-1) - \sum_{l_k \in \mathbf{l}, k \neq j} X_{ab;ik}, && (66)\\
\Rightarrow \quad & (h-1)x_{a;i} + (h-2) + x_{b;j} \leq (h-1) + (h-2)x_{a;i} + X_{ab;ij}. && (67)
\end{aligned}
$$

The last step is obtained using constraints (44) and (45), i.e.

$$\sum_{l_k \in \mathbf{l}} x_{b;k} = (2 - h), \sum_{l_k \in \mathbf{l}} X_{ab;ik} = (2 - h)x_{a;i}. \tag{68}$$

Rearranging the terms, we get

$$(x_{a;i} + x_{b;j}) \leq 1 + X_{ab;ij}. \tag{69}$$

Thus, according to inequalities (62) and (69)

$$|x_{a;i} + x_{b;j}| \leq 1 + X_{ab;ij}. \quad \blacksquare \tag{70}$$

Using the above lemma, we obtain

$$
\begin{aligned}
& (x_{a;i} + x_{b;j})^2 \leq (1 + X_{ab;ij})(1 + X_{ab;ij}), && (71)\\
\Rightarrow \quad & (x_{a;i} + x_{b;j})^2 \leq 2 + 2X_{ab;ij}. && (72)
\end{aligned}
$$

where the necessary condition for the equality to hold true is $1 + X_{ab;ij} = 2$ (i.e. $X_{ab;ij} = 1$). Note that the above constraint is equivalent to the SOCP-MS constraint (51). Together with inequality (60), this proves that the LP-S relaxation provides smaller supremum and larger infimum of the elements of the matrix **X** than the SOCP-MS relaxation. Thus, $\mathcal{F}(\text{LP-S}) \subset \mathcal{F}(\text{SOCP-MS})$.

One can also construct a parameter $\boldsymbol{\theta}$ for which the set of all optimal solutions of SOCP-MS do not lie in the feasibility region of LP-S. In other words the optimal solutions of SOCP-MS belong to the non-empty set $\mathcal{F}(\text{SOCP-MS}) - \mathcal{F}(\text{LP-S})$. Using the argument of Case I in section 2, this implies that LP-S strictly dominates SOCP-MS. $\blacksquare$

Note that the above theorem does not apply to the variation of SOCP-MS described in [18, 20] which include *triangular inequalities* [1]. However, since triangular inequalities are linear constraints, LP-S can be extended to include them. The resulting LP relaxation would strictly dominate the SOCP-MS relaxation with triangular inequalities.

4 QP-RL vs. SOCP-MS

We now prove that QP-RL and SOCP-MS are equivalent (i.e. their optimal values are equal for MAP estimation problems defined over all CRFs). Specifically, we consider a vector $\mathbf{x}$ which lies in the feasibility regions of the QP-RL and SOCP-MS relaxations, i.e. $\mathbf{x} \in [-1,1]^{nh}$. For this vector, we show that the values of the objective functions of the QP-RL and SOCP-MS relaxations are equal. This would imply that if $\mathbf{x}^*$ is an optimal solution of QP-RL for some CRF with parameter $\boldsymbol{\theta}$ then there exists an optimal solution $(\mathbf{x}^*, \mathbf{X}^*)$ of the SOCP-MS relaxation. Further, if e^Q and e^S are the optimal values of the objective functions obtained using the QP-RL and SOCP-MS relaxation, then $e^Q = e^S$.

Theorem 2: The QP-RL relaxation and the SOCP-MS relaxation are equivalent.

Proof: Recall that the QP-RL relaxation is defined as follows:

$$\text{QP-RL:} \qquad \mathbf{x}^* = \arg\min_{\mathbf{x}} \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right)^\top \hat{\boldsymbol{\theta}}^1 + \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right)^\top \hat{\boldsymbol{\theta}}^2 \left(\tfrac{\mathbf{1}+\mathbf{x}}{2}\right), \tag{73}$$

$$\text{s.t.} \qquad \textstyle\sum_{l_i \in \mathbf{l}} x_{a;i} = 2 - h, \forall v_a \in \mathbf{v}, \tag{74}$$

$$\mathbf{x} \in \{-1,1\}^{nh}, \tag{75}$$

where the unary potential vector $\hat{\boldsymbol{\theta}}^1$ and the pairwise potential matrix $\hat{\boldsymbol{\theta}}^2 \succeq 0$ are defined as

$$\hat{\theta}^1_{a;i} = \theta^1_{a;i} - \sum_{v_c \in \mathbf{v}} \sum_{l_k \in \mathbf{l}} |\theta^2_{ac;ik}|, \tag{76}$$

$$\hat{\theta}^2_{ab;ij} = \begin{cases} \sum_{v_c \in \mathbf{v}} \sum_{l_k \in \mathbf{l}} |\theta^2_{ac;ik}|, & \text{if} \quad a = b, i = j, \\ \theta^2_{ab;ij} & \text{otherwise.} \end{cases} \tag{77}$$

Here, the terms $\theta^1_{a;i}$ and $\theta^2_{ac;ik}$ are the (original) unary potentials and pairwise potentials for the given CRF. Consider a feasible solution $\mathbf{x}$ of the QP-RL and the SOCP-MS relaxations. Further, let $\mathbf{X}$ be the solution obtained when minimizing the objective function of the SOCP-MS relaxation whilst keeping $\mathbf{x}$ fixed. We prove that the value of the objective functions for both relaxations at the above feasible solution is the same by equating the coefficient of $\theta^1_{a;i}$ and $\theta^2_{ab;ij}$ for all $v_a \in \mathbf{v}$, $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$ in both formulations. Using equation (76), we see that $\theta^1_{a;i}$ is multiplied by $\frac{1+x_{a;i}}{2}$ in the objective function of the QP-RL. Similarly, $\theta^1_{a;i}$ is multiplied by $\frac{1+x_{a;i}}{2}$ in the SOCP-MS relaxation. Therefore the coefficients of $\theta^1_{a;i}$ in both relaxations are equal for all $v_a \in \mathbf{v}$ and $l_i \in \mathbf{l}$.

We now consider the pairwise potentials, i.e. $\theta^2_{ab;ij}$ and show that their coefficients are the same when obtaining the minimum of the objective function. We consider the following two cases.

Case I: Let $\theta^2_{ab;ij} = \theta^2_{ba;ji} \geq 0$. Using equation (77) we see that, in the QP-RL relaxation, $\theta^2_{ab;ij} + \theta^2_{ba;ji}$ is multiplied by the following term:

$$\left(\frac{1+x_{a;i}}{2}\right)^2 + \left(\frac{1+x_{b;j}}{2}\right)^2 + 2\left(\frac{1+x_{a;i}}{2}\right)\left(\frac{1+x_{b;j}}{2}\right) - \frac{1+x_{a;i}}{2} - \frac{1+x_{b;j}}{2}. \tag{78}$$

In the case of SOCP-MS relaxation, since $\theta^2_{ab;ij} \geq 0$, the minimum of the objective function is obtained by using the minimum value that $X_{ab;ij}$ would take given the SOC constraints. Since $\mathbf{X}$ is symmetric, we see that $\theta^2_{ab;ij} + \theta^2_{ba;ji}$ is multiplied by the following term:

$$\tfrac{1+x_{a;i}+x_{b;j}+\inf\{X_{ab;ij}\}}{2} \tag{79}$$

$$= \tfrac{1+x_{a;i}+x_{b;j}+(x_{a;i}+x_{b;j})^2/2-1}{2}, \tag{80}$$

where the infimum of $X_{ab;ij}$ is defined by constraint (51) in the SOCP-MS relaxation. It can easily be verified that the terms (78) and (80) are equal.

Case II: Now consider the case where $\theta^2_{ab;ij} = \theta^2_{ba;ji} < 0$. In the QP-RL relaxation, the term $\theta^2_{ab;ij} + \theta^2_{ba;ji}$ is multiplied by

$$\frac{1+x_{a;i}}{2} + \frac{1+x_{b;j}}{2} + 2\left(\frac{1+x_{a;i}}{2}\right)\left(\frac{1+x_{b;j}}{2}\right) - \left(\frac{1+x_{a;i}}{2}\right)^2 - \left(\frac{1+x_{b;j}}{2}\right)^2. \tag{81}$$

In order to obtain the minimum of the objective function, the SOCP-MS relaxation uses the maximum value that $X_{ab;ij}$ would take given the SOC constraints. Thus, $\theta^2_{ab;ij} + \theta^2_{ba;ji}$ is multiplied by

$$\frac{1+x_{a;i}+x_{b;j}+\sup\{X_{ab;ij}\}}{2} \tag{82}$$

$$= \frac{1+x_{a;i}+x_{b;j}+1-(x_{a;i}-x_{b;j})^2/2}{2}, \tag{83}$$

where the supremum of $X_{ab;ij}$ is defined by constraint (50). Again, the terms (81) and (83) can be shown to be equivalent. $\blacksquare$

Theorems 1 and 2 prove that the LP-S relaxation strictly dominates the QP-RL and SOCP-MS relaxations. A natural question that now arises is whether the additive bound of QP-RL (proved in [22]) is applicable to the LP-S and SOCP-MS relaxations. Our next theorem answers this question in an affirmative. To this end, we use the rounding scheme proposed in [22] to obtain the labelling f for all the random variables of the given CRF. This rounding scheme is summarized below:

- Pick a variable v_a which has not been assigned a label.
- Assign the label l_i to v_a (i.e. f(a) = i) with probability $\frac{1+x_{a;i}}{2}$.
- Continue till all variables have been assigned a label.

Recall that $\sum_{i=0}^{h-1} \frac{1+x_{a;i}}{2} = 1$ for all $v_a \in \mathbf{v}$. Hence, once v_a is picked it is guaranteed to be assigned a label. In other words the above rounding scheme terminates after $n = |\mathbf{v}|$ steps. To the best of our knowledge, this additive bound was previously known only for the QP-RL relaxation [22].

Theorem 3: For the above rounding scheme, LP-S and SOCP-MS provide the same additive bound as the QP-RL relaxation of [22], i.e. $\frac{S}{4}$ where $S = \sum_{(a,b)\in\mathcal{E}} \sum_{l_i,l_j\in\mathbf{l}} |\theta^2_{ab;ij}|$ (i.e. the sum of the absolute values of all pairwise potentials). Furthermore, this bound is tight.

Proof: The QP-RL and SOCP-MS relaxations are equivalent. Thus the above theorem holds true for SOCP-MS. We now consider the LP-S relaxation of [4, 17, 25, 31]. We denote the energy of the optimal labelling as e^*. Recall that e^L denotes the optimal value of the LP-S which is obtained using possibly fractional variables $(\mathbf{x}^*, \mathbf{X}^*)$. Clearly, $e^L \leq e^*$. The energy of the labelling $\hat{\mathbf{x}}$, obtained after rounding the solution of the LP-S relaxation, is represented by the term $\hat{e}^L$,

Using the above notation, we now show that the LP-S relaxation provides an additive bound of $\frac{S}{4}$ for the above rounding scheme. We first consider the unary potentials and observe that

$$E\left(\theta^1_{a;i}\left(\frac{1+\hat{x}_{a;i}}{2}\right)\right) = \theta^1_{a;i}\left(\frac{1+x^*_{a;i}}{2}\right), \tag{84}$$

where $E(\cdot)$ denotes the expectation of its argument under the above rounding scheme. Similarly, for the pairwise potentials,

$$E\left(\theta^2_{ab;ij}\left(\frac{1+\hat{x}_{a;i}}{2}\right)\left(\frac{1+\hat{x}_{b;j}}{2}\right)\right) = \theta^2_{ab;ij}\left(\frac{1+x^*_{a;i}+x^*_{b;ij}+x^*_{a;i}x^*_{b;j}}{4}\right). \tag{85}$$

We analyze the following two cases:

(i) $\theta^2_{ab;ij} \geq 0$: Using the fact that $X^*_{ab;ij} \geq |x^*_{a;i} + x^*_{b;j}| - 1$ (see Lemma 3.2), we get

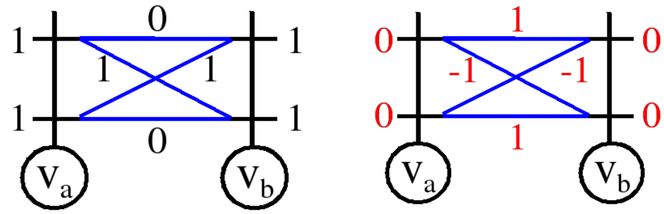
$$\begin{aligned} & 1+x^*_{a;i}+x^*_{b;j}+x^*_{a;i}x^*_{b;j} - (1+x^*_{a;i}+x^*_{b;j}+X^*_{ab;ij}) \\ = \ & x^*_{a;i}x^*_{b;j} - X^*_{ab;ij} \\ \leq \ & x^*_{a;i}x^*_{b;j} + 1 - |x^*_{a;i}+x^*_{b;j}| \\ \leq \ & 1, \end{aligned} \tag{86}$$

where the equality holds when $x^*_{a;i} = x^*_{b;j} = 0$. Therefore,

$$E\left(\theta^2_{ab;ij}\left(\frac{1+\hat{x}_{a;i}}{2}\right)\left(\frac{1+\hat{x}_{b;j}}{2}\right)\right) \leq \theta^2_{ab;ij}\frac{(1+x^*_{a;i}+x^*_{b;ij}+X^*_{ab;ij})}{4} + \frac{|\theta^2_{ab;ij}|}{4}. \tag{87}$$

(ii) $\theta^2_{ab;ij} < 0$: Using the fact that $X^*_{ab;ij} \leq 1 - |x^*_{a;i} - x^*_{b;j}|$ (see Lemma 3.1), we get

$$\begin{aligned} & 1+x^*_{a;i}+x^*_{b;j}+x^*_{a;i}x^*_{b;j} - (1+x^*_{a;i}+x^*_{b;j}+X^*_{ab;ij}) \\ \geq \ & x^*_{a;i}x^*_{b;j} - 1 + |x^*_{a;i}-x^*_{b;j}| \\ \geq \ & -1, \end{aligned} \tag{88}$$



(a) (b)

Figure 1: *An example* CRF *for proving the tightness of the* LP-S *additive bound of* $\frac{S}{4}$. **(a)** *The two random variables* v_a *and* v_b *are shown as unfilled circles. Their two putative labels are shown as branches (i.e. the horizontal lines) of the trellises (i.e. the vertical lines). The value of the unary potential* $\theta^1_{a;i}$ *is shown next to the* i^{th} *branch of the trellis on top of* v_a. *The pairwise potential* $\theta^2_{ab;ij}$ *is shown next to the connection between the* i^{th} *and* j^{th} *branches of the trellises on top of* v_a *and* v_b *respectively. Note that the unary potentials are uniform while the pairwise potentials form an Ising model.* **(b)** *An optimal solution of the* LP-S *relaxation for the* CRF *shown in (a). This solution is shown in red to differentiate it from the potentials shown in (a). The values of the variables* $x_{a;i}$ *are shown next to the* i^{th} *branch of the trellis of* v_a. *Note that all variables* $x_{a;i}$ *have been assigned to* 0. *The values of the variables* $X_{ab;ij}$ *are shown next to the connection between the* i^{th} *and* j^{th} *branch of the trellises of* v_a *and* v_b. *Note that* $X_{ab;ij} = -1$ *if* $\theta^2_{ab;ij} > 0$ *and* $X_{ab;ij} = 1$ *otherwise.*

where the equality holds when $x^*_{a;i} = x^*_{b;j} = 0$. Therefore,

$$E\left(\theta^2_{ab;ij}\left(\frac{1+\hat{x}_{a;i}}{2}\right)\left(\frac{1+\hat{x}_{b;j}}{2}\right)\right) \leq \theta^2_{ab;ij}\frac{(1+x^*_{a;i}+x^*_{b;ij}+X^*_{ab;ij})}{4} + \frac{|\theta^2_{ab;ij}|}{4}. \quad (89)$$

Summing the expectation of the unary and pairwise potentials for all $v_a \in \mathbf{v}$, $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$, we get

$$e^* \leq E(\hat{e}^L) \leq e^L + \frac{S}{4} \leq e^* + \frac{S}{4}, \quad (90)$$

which proves the additive bound for LP-S.

This additive bound can indeed be shown to be tight by using the following simple example. Consider an instance of the MAP estimation problem for a CRF defined on two variables $\mathbf{v} = \{v_a, v_b\}$ each of which can take one of two labels from the set $\mathbf{l} = \{l_0, l_1\}$. Let the unary and pairwise potentials be as defined in Fig. 1(a), i.e. the unary potentials are uniform and the pairwise potentials follow the Ising model.

An optimal solution of the LP-S relaxation is given in Fig. 1(b). Clearly, $e^* = 2$ (e.g. for the labelling $f = \{0,0\}$ or $f = \{1,1\}$) while $E(\hat{e}^L) = 2 + \frac{2}{4} = e^* + \frac{S}{4}$. Thus the additive bounds obtained for the LP-S, QP-RL and SOCP-MS algorithms are the same. In fact, one can construct arbitrarily large CRFs (i.e. CRF defined over a large number of variables) with uniform unary potentials and Ising model pairwise potentials for which the bound can be shown to be tight. ∎

The above bound was proved for the case of binary variables (i.e. h = 2) in [10] using a slightly different rounding scheme. Our result can be viewed as a generalization of this for any arbitrary number of labels. We note here that better bounds can be obtained for some special cases of the MAP estimation problem using the LP-S relaxation together with more clever rounding schemes (such as those described in [4, 14]).

5 QP and SOCP Relaxations over Trees and Cycles

We now generalize the results of Theorem 1 by defining a large class of SOCP relaxations which is dominated by LP-S. Specifically, we consider the SOCP relaxations which relax the non-convex constraint $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ using a set of second order cone (SOC) constraints of the form

$$||(\mathbf{U}^k)^\top \mathbf{x}|| \leq \mathbf{C}^k \bullet \mathbf{X}, k = 1, \cdots, n_C \quad (91)$$

where $\mathbf{C}^k = \mathbf{U}^k(\mathbf{U}^k)^\top \succeq 0$, for all $k = 1, \cdots, n_C$. In order to make the proofs of the subsequent theorems easier, we make two assumptions. However, the theorems would hold true even without these assumptions as discussed below.

Assumption 1: We assume that the integer constraints

$$\mathbf{x} \in \{-1, +1\}^{nh}, \mathbf{X} \in \{-1, +1\}^{nh \times nh}, \tag{92}$$

are relaxed to

$$\mathbf{x} \in [-1, +1]^{nh}, \mathbf{X} \in [-1, +1]^{nh \times nh}, \tag{93}$$

with $X_{aa;ii} = 1$, for all $v_a \in \mathbf{v}, l_i \in \mathbf{l}$. The constraints (93) provide the convex hull for all the points defined by the integer constraints (92). Recall that the convex hull of a set of points is the smallest convex set which contains all the points. We now discuss how the above assumption is not restrictive with respect to the results that follow. Let A be a relaxation which contains constraints (93). Further, let B be a relaxation which differs from A only in the way it relaxes the integer constraints. Then by definition of convex hull $\mathcal{F}(\text{A}) \subset \mathcal{F}(\text{B})$. In other words A dominates B (see Case I in section 2). Hence, if A is dominated by the LP-S relaxation, then LP-S would also dominate B.

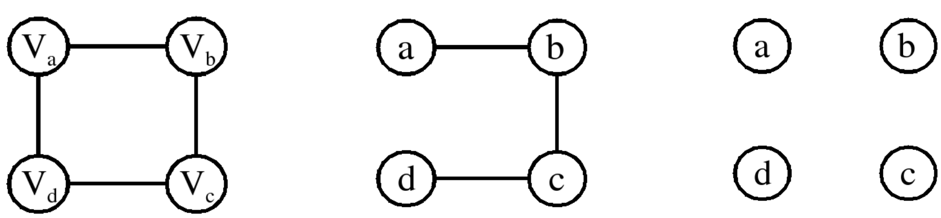
Assumption 2: We assume that the set of constraints (91) contains all the constraints specified in the SOCP-MS relaxation. Recall that for a given pair of neighbouring random variables, i.e. $(a, b) \in \mathcal{E}$, and a pair of labels $l_i, l_j \in \mathbf{l}$, SOCP-MS specifies SOC constraints using two matrices (say $\mathbf{C}^1$ and $\mathbf{C}^2$) which are 0 everywhere except for the following 2×2 submatrices:

$$\begin{pmatrix} C^1_{aa;ii} & C^1_{ab;ij} \\ C^1_{ba;ji} & C^1_{bb;jj} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} C^2_{aa;ii} & C^2_{ab;ij} \\ C^2_{ba;ji} & C^2_{bb;jj} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{94}$$

In the case where a given relaxation A does not contain the SOCP-MS constraints, we can define a new relaxation B. This new relaxation B is obtained by adding all the SOCP-MS constraints to A. By definition, B dominates A (although not necessarily strictly, see Case II in section 2). Hence, if B is dominated by the LP-S relaxation then it follows that LP-S would also dominate A. Hence, our assumption about including the SOCP-MS constraints is not restrictive for the results presented in this section.

Note that each SOCP relaxation belonging to this class would define an equivalent QP relaxation (similar to the equivalent QP-RL relaxation defined by the SOCP-MS relaxation). Hence, all these QP relaxations will also be dominated by the LP-S relaxation. Before we begin to describe our results in detail, we need to set up some notation as follows.

5.1 Notation



(a) (b) (c)

Figure 2: **(a)** *An example* CRF *defined over four variables which form a cycle. Note that the observed nodes are not shown for the sake of clarity of the image.* **(b)** *The set E^k specified by the matrix $\mathbf{C}^k$ shown in equation (96), i.e. $E^k = \{(a, b), (b, c), (c, d)\}$.* **(c)** *The set $V^k = \{a, b, c, d\}$. See text for definitions of these sets.*

We consider an SOC constraint which is of the form described in equation (91), i.e.

$$||(\mathbf{U}^k)^\top \mathbf{x}|| \leq \mathbf{C}^k \bullet \mathbf{X}, \tag{95}$$

where $k \in \{1, \cdots, n_C\}$. In order to help the reader understand the notation better, we use an example CRF shown in Fig. 2(a). This CRF is defined over four variables $\mathbf{v} = \{v_a, v_b, v_c, v_d\}$ (connected to form a cycle of length 4), each of which take a label from the set $\mathbf{l} = \{l_0, l_1\}$. For this CRF we specify a constraint using a matrix $\mathbf{C}^k \succeq 0$ which is 0 everywhere, except for the following 4×4 submatrix:

$$\begin{pmatrix} C^k_{aa;00} & C^k_{ab;00} & C^k_{ac;00} & C^k_{ad;00} \\ C^k_{ba;00} & C^k_{bb;00} & C^k_{bc;00} & C^k_{bd;00} \\ C^k_{ca;00} & C^k_{cb;00} & C^k_{cc;00} & C^k_{cd;00} \\ C^k_{da;00} & C^k_{db;00} & C^k_{dc;00} & C^k_{dd;00} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} \tag{96}$$

Using the SOC constraint shown in equation (95) we define the following sets:

- The set E^k is defined such that $(a, b) \in E^k$ if, and only if, it satisfies the following conditions:

$$(a, b) \in \mathcal{E}, \tag{97}$$

$$\exists l_i, l_j \in \mathbf{l} \text{ such that } C^k_{ab;ij} \neq 0. \tag{98}$$

  Recall that $\mathcal{E}$ specifies the neighbourhood relationship for the given CRF. In other words E^k is the subset of the edges in the graphical model of the CRF such that $\mathbf{C}^k$ specifies constraints for the random variables corresponding to those edges. For the example CRF (shown in Fig. 2(a)) and $\mathbf{C}^k$ matrix (in equation (96)), the set E^k obtained is shown in Fig. 2(b).

- The set V^k is defined as $a \in V^k$ if, and only if, there exists a $v_b \in \mathbf{v}$ such that $(a, b) \in E^k$. In other words V^k is the subset of hidden nodes in the graphical model of the CRF such that $\mathbf{C}^k$ specifies constraints for the random variables corresponding to those hidden nodes. Fig. 2(c) shows the set V^k for our example SOC constraint.

- The set $\mathcal{T}^k$ consists of elements $a; i \in \mathcal{T}^k$ which satisfy

$$a \in V^k, l_i \in \mathbf{l}, \tag{99}$$

$$\exists b \in V^k, l_j \in \mathbf{l}, \text{ such that } C^k_{ab;ij} \neq 0. \tag{100}$$

  In other words the set $\mathcal{T}^k$ consists of the set of indices for the vector $\mathbf{x}$ which are constrained by inequality (95), i.e. the coefficient of $x_{a;i}$ where $a; i \in \mathcal{T}^k$ are non-zero in the LHS of inequality (95). Note that in equation (96) the constraint is specified using only the label l_0 for all the random variables $\mathbf{v}$. Thus the set $\mathcal{T}^k$ is given by

$$\mathcal{T}^k = \{(a; 0), (b; 0), (c; 0), (d; 0)\}. \tag{101}$$

  For each set $\mathcal{T}^k$ we define three disjoint subsets of $\mathcal{T}^k \times \mathcal{T}^k$ as follows.

- The set $\mathcal{T}^k_0$ is defined as

$$\mathcal{T}^k_0 = \{(a; i, b; j) | (a; i, b; j) \in \mathcal{T}^k \times \mathcal{T}^k, (a, b) \in \mathcal{E}, (a, b) \notin E^k\}. \tag{102}$$

  Note that by definition $C^k_{ab;ij} = 0$ if $(a; i, b; j) \in \mathcal{T}^k_0$. Thus $\mathcal{T}^k_0$ indexes the elements of matrix $\mathbf{X}$ which are not constrained by inequality (95) but are present in the set $\mathcal{T}^k \times \mathcal{T}^k$. For the matrix $\mathbf{C}^k$ in equation (96), the set $\mathcal{T}^k_0$ is given by

$$\mathcal{T}^k_0 = \{(a; 0, d; 0)\} \tag{103}$$

- The set $\mathcal{T}^k_1$ is defined as

$$\mathcal{T}^k_1 = \{(a; i, b; j) | (a; i, b; j) \in \mathcal{T}^k \times \mathcal{T}^k, (a, b) \notin \mathcal{E}\}. \tag{104}$$

  In other words the set $\mathcal{T}^k_1$ indexes the elements of matrix $\mathbf{X}$ which are constrained by inequality (95) but do not belong to any pair of neighbouring random variables. Note that the variables $X_{ab;ij}$ such that $(a; i, b; j) \in \mathcal{T}^k_1$ were not present in the LP-S relaxation. For the matrix $\mathbf{C}^k$ in equation (96), the set $\mathcal{T}^k_1$ is given by

$$\mathcal{T}^k_1 = \{(a; 0, c; 0), (b; 0, d; 0)\} \tag{105}$$

- The set $\mathcal{T}_2^k$ is defined as

$$\mathcal{T}_2^k = \{(a;i,b;j)|(a;i,b;j) \in \mathcal{T}^k \times \mathcal{T}^k, (a,b) \in E^k\}. \tag{106}$$

In other words the set $\mathcal{T}_2^k$ indexes the elements of matrix $\mathbf{X}$ which are constrained by inequality (95) and belong to a pair of neighbouring random variables. For the matrix $\mathbf{C}^k$ in equation (96), the set $\mathcal{T}_1^k$ is given by

$$\mathcal{T}_2^k = \{(a;0,b;0),(b;0,c;0)(c;0,d;0)\} \tag{107}$$

Note that $\mathcal{T}_0^k \bigcup \mathcal{T}_1^k \bigcup \mathcal{T}_2^k = \mathcal{T}^k \times \mathcal{T}^k$. For a given set of pairwise potentials $\theta^2_{ab;ij}$ we define two disjoint sets of $\mathcal{T}_2^k$ as follows.

- The set $\mathcal{T}_{2+}^k$ corresponds to non-negative pairwise potentials, i.e.

$$\mathcal{T}_{2+}^k = \{(a;i,b;j)|(a;i,b;j) \in \mathcal{T}_2^k, \theta^2_{ab;ij} \geq 0\}, \tag{108}$$

Thus the set $\mathcal{T}_{2+}^k$ indexes the elements of matrix $\mathbf{X}$ which belong to $\mathcal{T}_2^k$ and are multiplied by a non-negative pairwise potential in the objective function of the relaxation.

- The set $\mathcal{T}_{2-}^k$ corresponds to negative pairwise potentials, i.e.

$$\mathcal{T}_{2-}^k = \{(a;i,b;j)|(a;i,b;j) \in \mathcal{T}_2^k, \theta^2_{ab;ij} < 0\}, \tag{109}$$

Thus the set $\mathcal{T}_{2-}^k$ indexes the elements of matrix $\mathbf{X}$ which belong to $\mathcal{T}_2^k$ and are multiplied by a negative pairwise potential in the objective function of the relaxation. Note that $\mathcal{T}_2^k = \mathcal{T}_{2+}^k \bigcup \mathcal{T}_{2-}^k$. For the purpose of illustration let us assume that, for the example CRF in Fig. 2(a), $\theta^2_{ab;00} \geq 0$ while $\theta^2_{bc;00} < 0$ and $\theta^2_{cd;00} < 0$. Then,

$$\mathcal{T}_{2+}^k = \{(a;0,b;0)\}, \tag{110}$$
$$\mathcal{T}_{2-}^k = \{(b;0,c;0),(c;0,d;0)\}, \tag{111}$$

We also define a weighted graph $G^k = (V^k, E^k)$ whose vertices are specified by the set V^k and whose edges are specified by the set E^k. The weight of an edge $(a,b) \in E^k$ is given by $w(a,b)$. Recall that $w(a,b)$ specifies the strength of the pairwise relationship between two neighbouring variables v_a and v_b. Thus, for our example SOC constraint, the vertices of this graph are given in Fig. 2(c) while the edges are shown in Fig. 2(b). This graph can be viewed as a subgraph of the graphical model representation for the given CRF.

Further, we define the submatrices $\mathbf{x}_T^k$ and $\mathbf{X}_T^k$ of $\mathbf{x}$ and $\mathbf{X}$ respectively such that

$$\mathbf{x}_T^k = \{x_{a;i}|a;i \in \mathcal{T}^k\}, \tag{112}$$
$$\mathbf{X}_T^k = \{X_{ab;ij}|(a;i,b;j) \in \mathcal{T}^k \times \mathcal{T}^k\}. \tag{113}$$

For our example, these submatrices will be given by

$$\mathbf{x}_T^k = \begin{pmatrix} x_{a;0} \\ x_{b;0} \\ x_{c;0} \\ x_{d;0} \end{pmatrix}, \mathbf{X}_T^k = \begin{pmatrix} X_{aa;00} & X_{ab;00} & X_{ac;00} & X_{ad;00} \\ X_{ba;00} & X_{bb;00} & X_{bc;00} & X_{bd;00} \\ X_{ca;00} & X_{cb;00} & X_{cc;00} & X_{cd;00} \\ X_{da;00} & X_{db;00} & X_{dc;00} & X_{dd;00} \end{pmatrix}. \tag{114}$$

Using the above notation, we are now ready to describe our results in detail.

5.2 QP and SOCP Relaxations over Trees

We begin by considering those relaxations where the SOC constraints are defined such that the graphs $G^k = (V^k, E^k)$ form trees. For example, the graph G^k defined by the SOC constraint in equation (96) forms a tree as shown in Fig. 2(b). We denote such a relaxation, which specifies SOC constraints only over trees, by SOCP-T. Note that SOCP-MS (and hence, QP-RL) can be considered a special case of this class of relaxations where the number of vertices in each tree is equal to two (since the constraints are defined for all $(a,b) \in \mathcal{E}$).

We will remove this restriction by allowing the number of vertices in each tree to be arbitrarily large (i.e. between 1 and n). We consider one such tree $G = (V, E)$. Note that for a given relaxation

SOCP-T, there may be several SOC constraints defined using this tree G (or its subtree). Without loss of generality, we assume that the constraints

$$||(\mathbf{U}^k)^\top \mathbf{x}|| \leq \mathbf{C}^k \bullet \mathbf{X}, k = 1, \cdots, n'_C \tag{115}$$

are defined on the tree G. In other words,

$$G^k \subseteq G, k = 1, \cdots, n'_C, \tag{116}$$

where $G^k \subseteq G$ implies that G^k is a subtree of G. In order to make the notation less cluttered, **we will drop the superscript k from the sets defined in the previous section** (since we will consider only one tree G in our analysis).

We will now show that SOCP-T is dominated by the LP-S relaxation. This result is independent of the choice of trees G and matrices $\mathbf{C}^k$. To this end, we define the term $e_1(\mathbf{x}_T)$ for a given value of $\mathbf{x}_T$ as

$$e_1(\mathbf{x}_T) = \sum_{(a;i) \in \mathcal{T}} \left(\theta^1_{a;i} + \sum_{(b;j) \in \mathcal{T}} \frac{\theta^2_{ab;ij}}{2} \right) x_{a;i}. \tag{117}$$

Further, for a fixed $\mathbf{x}_T$ we also define the following two terms:

$$e_2^S(\mathbf{x}_T) = \min_{(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{SOCP-T})} \sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} X_{ab;ij}, \tag{118}$$

$$e_2^L(\mathbf{x}_T) = \min_{(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{LP-S})} \sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} X_{ab;ij}, \tag{119}$$

where $\mathcal{F}(\text{SOCP-T})$ and $\mathcal{F}(\text{LP-S})$ are the feasibility regions of SOCP-T and LP-S respectively. We use the notation $(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{SOCP-T})$ loosely to mean that we can obtain a feasible solution $(\mathbf{x}, \mathbf{X})$ of SOCP-T such that the values of the variables $x_{a;i}$ where $a;i \in \mathcal{T}$ and $X_{ab;ij}$ where $(a;i,b;j) \in \mathcal{T} \times \mathcal{T}$ are equal to the values specified by $\mathbf{x}_T$ and $\mathbf{X}_T$. The notation $(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{LP-S})$ is used similarly. Note that for a given $\mathbf{x}_T$ the possible values of $\mathbf{X}_T$ are constrained such that $(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{SOCP-T})$ and $(\mathbf{x}_T, \mathbf{X}_T) \in \mathcal{F}(\text{LP-S})$ (in the case of SOCP-T and LP-S respectively). Hence different values of $\mathbf{x}_T$ will provide different values of $e_2^S(\mathbf{x}_T)$ and $e_2^L(\mathbf{x}_T)$.

The contribution of the tree G to the objective function of SOCP-T and LP-S is given by

$$e^S = \min_{\mathbf{x}_T} \frac{e_1(\mathbf{x}_T)}{2} + \frac{e_2^S(\mathbf{x}_T)}{4}, \tag{120}$$

$$e^L = \min_{\mathbf{x}_T} \frac{e_1(\mathbf{x}_T)}{2} + \frac{e_2^L(\mathbf{x}_T)}{4} \tag{121}$$

respectively. Assuming that the trees G do not overlap, the total value of the objective function would simply be the sum of e^S (for SOCP-T) or e^L (for LP-S) over all trees G. However, since we use an arbitrary parameter $\boldsymbol{\theta}$ in our analysis, it follows that the results do not depend on this assumption of non-overlapping trees. In other words if two trees G^1 and G^2 share an edge $(a,b) \in \mathcal{E}$ then we can simply consider two MAP estimation problems defined using arbitrary parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 = \boldsymbol{\theta}$. We can then add the contribution of G_1 for the MAP estimation problem with parameter $\boldsymbol{\theta}_1$ to the contribution of G_2 for the MAP estimation problem with parameter $\boldsymbol{\theta}_2$. This would then provide us with the total contribution of G_1 and G_2 for the original MAP estimation defined using parameter $\boldsymbol{\theta}$.

Using the above argument it follows that if, for all G and for all $\boldsymbol{\theta}$, the following holds true:

$$\frac{e_1(\mathbf{x}_T)}{2} + \frac{e_2^S(\mathbf{x}_T)}{4} \leq \frac{e_1(\mathbf{x}_T)}{2} + \frac{e_2^L(\mathbf{x}_T)}{4}, \forall \mathbf{x}_T \in [-1,1]^{|\mathcal{T}|} \tag{122}$$

$$\Rightarrow \quad e_2^S(\mathbf{x}_T) \leq e_2^L(\mathbf{x}_T), \forall \mathbf{x}_T \in [-1,1]^{|\mathcal{T}|}, \tag{123}$$

then LP-S dominates SOCP-T (since this would imply that $e^S \leq e^L$, for all G and for all $\boldsymbol{\theta}$). This is the condition that we will use to prove that LP-S dominates all SOCP relaxations whose constraints are defined over trees. To this end, we define a vector $\boldsymbol{\omega} = \{\omega_k, k = 1, \cdots, n'_C\}$ of non-negative real numbers such that

$$\sum_k \omega_k C^k_{ab;ij} = \theta^2_{ab;ij}, \forall (a;i,b;j) \in \mathcal{T}_2. \tag{124}$$

Due to the presence of the matrices $\mathbf{C}^k$ defined in equation (94) (which result in the SOCP-MS constraints for all $(a,b) \in \mathcal{E}$ and $l_i, l_j \in \mathbf{l}$), such a vector $\boldsymbol{\omega}$ would always exist for any CRF parameter $\boldsymbol{\theta}$. We denote the matrix $\sum_k \omega_k \mathbf{C}^k$ by $\mathbf{C}$. Clearly, $\mathbf{C} \succeq 0$, and hence can be written as $\mathbf{C} = \mathbf{U}\mathbf{U}^\top$.

Using the constraints $||(\mathbf{U}^k)^\top \mathbf{x}||^2 \leq \mathbf{C}^k \bullet \mathbf{X}_T$ together with the the fact that $\omega_k \geq 0$, we get the following inequality[5]:

$$
\begin{aligned}
& \textstyle\sum_k \omega_k ||(\mathbf{U}^k)^\top \mathbf{x}||^2 \leq \sum_k \omega_k \mathbf{C}^k \bullet \mathbf{X}, \\
\Rightarrow \quad & ||\mathbf{U}^\top \mathbf{x}||^2 \leq \mathbf{C} \bullet \mathbf{X}, \\
\Rightarrow \quad & ||\mathbf{U}^\top \mathbf{x}||^2 \leq \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii} X_{aa;ii} + \sum_{(a;i,b;j) \in \mathcal{T}_1} C_{ab;ij} X_{ab;ij} + \sum_{(a;i,b;j) \in \mathcal{T}_2} C_{ab;ij} X_{ab;ij} \\
\Rightarrow \quad & \|\mathbf{U}^\top \mathbf{x}\|^2 - \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j) \in \mathcal{T}_1} C_{ab;ij} X_{ab;ij} \leq \sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} X_{ab;ij},
\end{aligned}
\tag{125}
$$

where the last expression is obtained using the fact that $C_{ab;ij} = \theta^2_{ab;ij}$ for all $(a;i,b;j) \in \mathcal{T}_2$ and $X_{aa;ii} = 1$ for all $v_a \in \mathbf{v}$ and $l_i \in \mathbf{l}$. Note that, in the absence of any other constraint (which is our assumption), the value of $e_2^S(\mathbf{x}_T)$ after the minimization would be exactly equal to the LHS of the inequality given above (since the objective function containing $e_2^S(\mathbf{x}_T)$ is being minimized). In other words,

$$
\begin{aligned}
e_2^S(\mathbf{x}_T) &= \min \sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} X_{ab;ij}, \\
&= \min \|\mathbf{U}^\top \mathbf{x}\|^2 - \sum_{a;i \in \mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j) \in \mathcal{T}_1} C_{ab;ij} X_{ab;ij}.
\end{aligned}
\tag{126}
$$

For the LP-S relaxation, from Lemmas 3.1 and 3.2, we obtain the following value of $e_2^L(\mathbf{x}_T)$:

$$
|x_{a;i} + x_{b;j}| - 1 \leq X_{ab;ij} \leq 1 - |x_{a;i} - x_{b;j}|, \tag{127}
$$

$$
\begin{aligned}
\Rightarrow \quad & e_2^L(\mathbf{x}_T) = \min \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} X_{ab;ij}, \\
& = \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_{2+}} \theta^2_{ab;ij} (|x_{a;i} + x_{b;j}|) - \sum_{(a;i,b;j) \in \mathcal{T}_{2-}} \theta^2_{ab;ij} (|x_{a;i} - x_{b;j}|) - \\
& \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} |\theta^2_{ab;ij}|
\end{aligned}
\tag{128}
$$

We are now ready to prove the following theorem.

Theorem 4: SOCP relaxations (and the equivalent QP relaxations) which define constraints only using graphs $G = (V, E)$ which form (arbitrarily large) trees are dominated by the LP-S relaxation.

Proof: We begin by assuming that $d(i,j) \geq 0$ for all $l_i, l_j \in \mathbf{l}$ and later drop this constraint on the distance function[6]. We will show that for any arbitrary tree G and any matrix $\mathbf{C}$, the value of $e_2^L(\mathbf{x}_T)$ is greater than the value of $e_2^S(\mathbf{x}_T)$ for all $\mathbf{x}_T$. This would prove inequality (123) which in turn would show that the LP-S relaxation dominates SOCP-T (and the equivalent QP relaxation which we call QP-T) whose constraints are defined over trees.

It is assumed that we do not specify any additional constraints for all the variables $X_{ab;ij}$ where $(a;i,b;j) \in \mathcal{T}_1$ (i.e. for $X_{ab;ij}$ not belonging to any of our trees). In other words these variables $X_{ab;ij}$ are bounded only by the relaxation of the integer constraint, i.e. $-1 \leq X_{ab;ij} \leq +1$. Thus in equation (126) the minimum value of the RHS (which is equal to the value of $e_2^S(\mathbf{x}_T)$) is obtained by using the following value of $X_{ab;ij}$ where $(a;i,b;j) \in \mathcal{T}_1$:

$$
X_{ab;ij} = \begin{cases} 1 & \text{if} \quad C_{ab;ij} \geq 0, \\ -1 & \text{otherwise.} \end{cases} \tag{129}
$$

Substituting these values in equation (126) we get

$$
\begin{aligned}
& e_2^S(\mathbf{x}_T) = ||\mathbf{U}^\top \mathbf{x}||^2 - \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j) \in \mathcal{T}_1} |C_{ab;ij}|, \\
\Rightarrow \quad & e_2^S(\mathbf{x}_T) = \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii} x^2_{a;i} + \sum_{(a;i,b;j) \in \mathcal{T}_1} C_{ab;ij} x_{a;i} x_{b;j} + \sum_{(a;i,b;j) \in \mathcal{T}_2} \theta^2_{ab;ij} x_{a;i} x_{b;j} \\
& \quad - \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j) \in \mathcal{T}_1} |C_{ab;ij}|,
\end{aligned}
\tag{130}
$$

[5] Note that there are no terms corresponding to $(a;i,b;j) \in \mathcal{T}_0$ in inequality (125) since $C_{ab;ij} = 0$ if $(a;i,b;j) \in \mathcal{T}_0$. In other words, $X_{ab;ij}$ vanishes from the above inequality if $(a;i,b;j) \in \mathcal{T}_0$.

[6] Recall that $d(\cdot,\cdot)$ is a distance function on the labels. Together with the weights $w(\cdot,\cdot)$ defined over neighbouring random variables, it specifies the pairwise potentials as $\theta^2_{ab;ij} = w(a,b)d(i,j)$.

where the last expression is obtained using the fact that $\mathbf{C} = \mathbf{U}^\top \mathbf{U}$. Consider the term $\sum_{(a;i,b;j)\in\mathcal{T}_1} C_{ab;ij} x_{a;i} x_{b;j}$ which appears in the RHS of the above equation. For this term, clearly the following holds true

$$\sum_{(a;i,b;j)\in\mathcal{T}_1} C_{ab;ij} x_{a;i} x_{b;j} \leq \sum_{(a;i,b;j)\in\mathcal{T}_1} \frac{|C_{ab;ij}|}{2}(x_{a;i}^2 + x_{b;j}^2), \tag{131}$$

since for all $(a;i,b;j) \in \mathcal{T}_1$

$$C_{ab;ij} \leq |C_{ab;ij}|, \tag{132}$$

$$x_{a;i} x_{b;j} \leq \frac{(x_{a;i}^2 + x_{b;j}^2)}{2}. \tag{133}$$

Inequality (131) provides us with an upper bound on the value of $e_2^S(\mathbf{x}_T)$ as follows:

$$\begin{aligned} e_2^S(\mathbf{x}_T) \leq & \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii} x_{a;i}^2 + \sum_{(a;i,b;j)\in\mathcal{T}_1} \frac{|C_{ab;ij}|}{2}(x_{a;i}^2 + x_{b;j}^2) + \sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i} x_{b;j} \\ & - \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j)\in\mathcal{T}_1} |C_{ab;ij}|. \end{aligned} \tag{134}$$

Note that in order to prove inequality (123), i.e.

$$e_2^S(\mathbf{x}_T) \leq e_2^L(\mathbf{x}_T), \forall \mathbf{x}_T \in [-1,1]^{|\mathcal{T}|}, \tag{135}$$

it would be sufficient to show that $e_2^L(\mathbf{x}_T)$ specified in equation (128) is greater than the RHS of inequality (134) (since the RHS of inequality (134) is greater than $e_2^S(\mathbf{x}_T)$). We now simplify the two infimums $e_2^L(\mathbf{x}_T)$ and $e_2^S(\mathbf{x}_T)$ as follows.

LP-S Infimum: Let $z_{a;i} = \sqrt{|x_{a;i}|(1-|x_{a;i}|)}$. From equation (128), we see that the infimum provided by the LP-S relaxation is given by

$$\begin{aligned} & \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_{2+}} \theta^2_{ab;ij}(|x_{a;i}+x_{b;j}|) - \sum_{(a;i,b;j)\in\mathcal{T}_{2-}} \theta^2_{ab;ij}(|x_{a;i}-x_{b;j}|) - \sum_{(a;i,b;j)\in\mathcal{T}_2} |\theta^2_{ab;ij}| \\ & = - \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_{2+}} |\theta^2_{ab;ij}|(1 - |x_{a;i}+x_{b;j}| + x_{a;i}x_{b;j}) \\ & \quad - \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_{2-}} |\theta^2_{ab;ij}|(1 - |x_{a;i}-x_{b;j}| - x_{a;i}x_{b;j}) \\ & \quad + \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i}x_{b;j} \\ & \geq - \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_2} |\theta^2_{ab;ij}|(1-|x_{a;i}|)(1-|x_{b;j}|) - 2\sum_{(a;i,b;j)\in\mathcal{T}_2} |\theta^2_{ab;ij}| z_{a;i} z_{b;j} + \\ & \quad + \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i}x_{b;j}. \end{aligned} \tag{136}$$

The last expression is obtained using the fact that

$$(1 - |x_{a;i}+x_{b;j}| + x_{a;i}x_{b;j}) \leq (1-|x_{a;i}|)(1-|x_{b;j}|) + 2z_{a;i}z_{b;j}, \tag{137}$$

$$(1 - |x_{a;i}-x_{b;j}| - x_{a;i}x_{b;j}) \leq (1-|x_{a;i}|)(1-|x_{b;j}|) + 2z_{a;i}z_{b;j}. \tag{138}$$

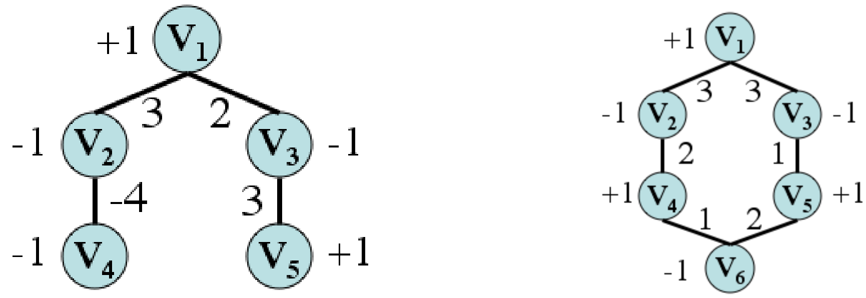
SOCP Infimum: From inequality (134), we see that the infimum provided by the SOCP-T relaxation is given by

$$\begin{aligned} & \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii} x_{a;i}^2 + \sum_{(a;i,b;j)\in\mathcal{T}_1} \frac{|C_{ab;ij}|}{2}(x_{a;i}^2 + x_{b;j}^2) + \sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i} x_{b;j} \\ & \quad - \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii} - \sum_{(a;i,b;j)\in\mathcal{T}_1} |C_{ab;ij}| \\ = & \quad - \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii}(1-x_{a;i}^2) - \sum_{(a;i,b;j)\in\mathcal{T}_1} |C_{ab;ij}|(1 - \frac{x_{a;i}^2}{2} - \frac{x_{b;j}^2}{2}) \\ & \quad + \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i} x_{b;j} \\ \leq & \quad - \textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii}(1-|x_{a;i}|)^2 - \sum_{(a;i,b;j)\in\mathcal{T}_1} |C_{ab;ij}|(1-|x_{a;i}|)(1-|x_{b;j}|) \\ & \quad -2\textstyle\sum_{a;i\in\mathcal{T}} C_{aa;ii} z_{a;i}^2 - 2\sum_{(a;i,b;j)\in\mathcal{T}_1} |C_{ab;ij}| z_{a;i} z_{b;j} \\ & \quad + \textstyle\sum_{(a;i,b;j)\in\mathcal{T}_2} \theta^2_{ab;ij} x_{a;i} x_{b;j}. \end{aligned} \tag{139}$$

The last expression is obtained using

$$1 - x_{a;i}^2 \geq (1-|x_{a;i}|)^2 + 2z_{a;i}^2, \tag{140}$$

$$1 - \frac{x_{a;i}^2}{2} - \frac{x_{b;j}^2}{2} \geq (1-|x_{a;i}|)(1-|x_{b;j}|) + 2z_{a;i}z_{b;j}. \tag{141}$$



(a) (b)

Figure 3: **(a)** *An example subgraph G which forms a tree. The weights of the edges and corresponding elements of the vector* **m** *are also shown.* **(b)** *An example subgraph G which forms an even cycle where all weights are positive. The elements of* $\mathbf{s}$ *are defined using the* $\{+1, -1\}$ *assignments of the vertices.*

In order to prove the theorem, we use the following two lemmas.

Lemma 5.1: The following inequality holds true for any matrix $\mathbf{C} \succeq 0$:

$$\begin{aligned} \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii}(1 - |x_{a;i}|)^2 + \sum_{(a;i,b;j) \in \mathcal{T}_1} |C_{ab;ij}|(1 - |x_{a;i}|)(1 - |x_{b;j}|) \\ \geq \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} |\theta^2_{ab;ij}|(1 - |x_{a;i}|)(1 - |x_{b;j}|). \end{aligned} \tag{142}$$

In other words, the first term in the RHS of inequality (136) exceeds the sum of the first two terms in the RHS of inequality (139).

Proof: The proof relies on the fact that $\mathbf{C}$ is positive semidefinite. We construct a vector $\mathbf{m} = \{m_a, a = 1, \cdots, n\}$ where n is the number of variables. Let $p(a)$ denote the parent of a non-root vertex a of tree G (where the root vertex can be chosen arbitrarily). The vector **m** is defined such that

$$m_a = \begin{cases} 0 & \text{if} \quad a \text{ does not belong to tree } G, \\ 1 & \text{if} \quad a \text{ is the root vertex of } G, \\ -m_{p(a)} & \text{if} \quad w(a, p(a)) > 0, \\ m_{p(a)} & \text{if} \quad w(a, p(a)) < 0. \end{cases} \tag{143}$$

Here $w(\cdot, \cdot)$ are the weights provided for a given CRF. Fig. 3(a) shows an example of a graph which forms a tree together with the corresponding elements of $\mathbf{m}$. Using the vector **m**, we define a vector $\mathbf{s}$ of length nh (where $h = |\mathbf{l}|$) such that $s_{a;i} = 0$ if $a; i \notin \mathcal{T}$ and $s_{a;i} = m_a(1 - |x_{a;i}|)$ otherwise. Since $\mathbf{C}$ is positive semidefinite, we get

$$\mathbf{s}^\top \mathbf{C} \mathbf{s} \geq 0 \tag{144}$$

$$\begin{aligned} \Rightarrow \quad & \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii}(1 - |x_{a;i}|)^2 + \sum_{(a;i,b;j) \in \mathcal{T}_1} m_a m_b C_{ab;ij}(1 - |x_{a;i}|)(1 - |x_{b;j}) \\ & \quad + \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} m_a m_b \theta^2_{ab;ij}(1 - |x_{a;i}|)(1 - |x_{b;j}) \geq 0, \end{aligned} \tag{145}$$

$$\begin{aligned} \Rightarrow \quad & \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii}(1 - |x_{a;i}|)^2 + \sum_{(a;i,b;j) \in \mathcal{T}_1} m_a m_b C_{ab;ij}(1 - |x_{a;i}|)(1 - |x_{b;j}) \\ & \quad \geq \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} |\theta^2_{ab;ij}|(1 - |x_{a;i}|)(1 - |x_{b;j}), \end{aligned} \tag{146}$$

$$\begin{aligned} \Rightarrow \quad & \textstyle\sum_{a;i \in \mathcal{T}} C_{aa;ii}(1 - |x_{a;i}|)^2 + \sum_{(a;i,b;j) \in \mathcal{T}_1} |C_{ab;ij}|(1 - |x_{a;i}|)(1 - |x_{b;j}|) \\ & \quad \geq \textstyle\sum_{(a;i,b;j) \in \mathcal{T}_2} |\theta^2_{ab;ij}|(1 - |x_{a;i}|)(1 - |x_{b;j}|). \quad \blacksquare \end{aligned} \tag{147}$$

Lemma 5.2: The following inequality holds true for any matrix $\mathbf{C} \succeq 0$:

$$\sum_{a;i \in \mathcal{T}} C_{aa;ii} z^2_{a;i} + \sum_{(a;i,b;j) \in \mathcal{T}_1} |C_{ab;ij}| z_{a;i} z_{b;j} \geq \sum_{(a;i,b;j) \in \mathcal{T}_2} |\theta^2_{ab;ij}| z_{a;i} z_{b;j}. \tag{148}$$

In other words the second term in the RHS of inequality (136) exceeds the sum of the third and fourth terms in inequality (139).

Proof: Similar to Lemma 5.1, we construct a vector $\mathbf{s}$ of length nh such that $s_{a;i} = 0$ if $a; i \notin \mathcal{T}$ and $s_{a;i} = m_a z_{a;i}$ otherwise. The proof follows by observing that $\mathbf{s}^\top \mathbf{C}\mathbf{s} \geq 0$. ∎

Using the above two lemmas, we see that the sum of the first two terms of inequality (136) exceed the sum of the first four terms of inequality (139). Further, the third and the fifth terms of inequalities (136) and (139) are the same. Since inequality (136) provides the lower limit of $e_2^L(\mathbf{x}_T)$ and inequality (139) provides the upper limit of $e_2^S(\mathbf{x}_T)$, it follows that $e_2^L(\mathbf{x}_T) \geq e_2^S(\mathbf{x}_T)$ for all $\mathbf{x}_T \in [-1, 1]^{|\mathcal{T}|}$. Using condition (123), this proves the theorem. ∎

The proofs of Lemmas 5.1 and 5.2 make use of the fact that for any neighbouring random variables v_a and v_b (i.e. $(a, b) \in \mathcal{E}$), the pairwise potentials $\theta^2_{ab;ij}$ have the same sign for all $l_i, l_j \in \mathbf{l}$. This follows from the non-negativity property of the distance function. However, Theorem 4 can be extended to the case where the distance function does not obey the non-negativity property. To this end, we define a parameter $\overline{\boldsymbol{\theta}}$ which satisfies the following condition:

$$Q(f; \mathbf{D}, \overline{\boldsymbol{\theta}}) = Q(f; \mathbf{D}, \boldsymbol{\theta}), \forall f. \tag{149}$$

Such a parameter $\overline{\boldsymbol{\theta}}$ is called the reparameterization of $\boldsymbol{\theta}$ (i.e. $\overline{\boldsymbol{\theta}} \equiv \boldsymbol{\theta}$). Note that there exist several reparameterizations of any parameter $\boldsymbol{\theta}$. We are interested in a parameter $\overline{\boldsymbol{\theta}}$ which satisfies

$$\sum_{l_i, l_j \in \mathbf{l}} |\overline{\theta}^2_{ab;ij}| = |\sum_{l_i, l_j \in \mathbf{l}} \theta^2_{ab;ij}|, \forall (a, b) \in \mathcal{E}. \tag{150}$$

It can easily be shown that such a reparameterization always exists. Specifically, consider the general form of reparameterization described in [15], i.e.

$$\overline{\theta}^1_{a;i} = \theta^1_{a;i} + M_{ba;i}, \tag{151}$$

$$\overline{\theta}^2_{ab;ij} = \theta^2_{ab;ij} - M_{ba;i} - M_{ab;j}. \tag{152}$$

Clearly one can set the values of the terms $M_{ba;i}$ and $M_{ab;j}$ such that equation (150) is satisfied. Further, the optimal value of LP-S for the parameter $\overline{\boldsymbol{\theta}}$ is equal to its optimal value obtained using $\boldsymbol{\theta}$. For details, we refer the reader to [15]. Using this parameter $\overline{\boldsymbol{\theta}}$, we obtain an LP-S infimum which is similar in form to the inequality (136) for any distance function (i.e. without the positivity constraint $d(i, j) \geq 0$ for all $l_i, l_j \in \mathbf{l}$). This LP-S infimum can then be easily compared to the SOCP-T infimum of inequality (139) (using slight extensions of Lemmas 5.1 and 5.2), thereby proving the results of Theorem 4 for a general distance function. We omit details.

As an upshot of the above theorem, we see that the feasibility region of LP-S is always a subset of the feasibility region of SOCP-T (for any general set of trees and SOC constraints), i.e. $\mathcal{F}(\text{LP-S}) \subset \mathcal{F}(\text{SOCP-T})$. This implies that $\mathcal{F}(\text{LP-S}) \subset \mathcal{F}(\text{QP-T})$, where QP-T is the equivalent QP relaxation defined by SOCP-T.

We note that the above theorem can also be proved using the results of [32] on *moment constraints* (which imply that LP-S provides the exact solution for the MAP estimation problems defined over tree-structured random fields). However, the proof presented here allows us to generalize the results of Theorem 4 for certain cycles as follows.

5.3 QP and SOCP Relaxations over Cycles

We now prove that the above result also holds true when the graph G forms an *even cycle*, i.e. cycles with even number of vertices, whose weights are all non-negative or all non-positive provided $d(i, j) \geq 0$, for all $l_i, l_j \in \mathbf{l}$.

Theorem 5: When $d(i, j) \geq 0$ for all $l_i, l_j \in \mathbf{l}$, the SOCP relaxations which define constraints only using non-overlapping graphs G which form (arbitrarily large) even cycles with all positive or all negative weights are dominated by the LP-S relaxation.

Proof: It is sufficient to show that Lemmas 5.1 and 5.2 hold for a graph $G = (V, E)$ which forms an even cycle. We first consider the case where $\theta^2_{ab;ij} > 0$. Without loss of generality, we assume that $V = \{1, 2, \ldots, t\}$ (where t is even) such that $(i, i + 1) \in E$ for all $i = 1, \cdots, t - 1$. Further, $(t, 1) \in E$ thereby forming an even cycle. We construct a vector $\mathbf{m}$ of size n such that $m_a = -1^a$ if $a \in V$ and $m_a = 0$ otherwise. When $\theta^2_{ab;ij} < 0$, we define a vector $\mathbf{m}$ such that $m_a = 1$ if $a \in V$

and $m_a = 0$ otherwise. Fig. 3(b) shows an example of a graph G which forms an even cycle together with the corresponding elements of $\mathbf{m}$. Using $\mathbf{m}$, we construct a vector $\mathbf{s}$ of length nh (similar to the proofs of Lemmas 5.1 and 5.2). Lemmas 5.1 and 5.2 follow from the fact that $\mathbf{s}^\top \mathbf{C}\mathbf{s} \geq 0$. We omit details. ∎

The above theorem can be proved for cycles of any length whose weights are all negative by a similar construction. Further, it also holds true for *odd cycles* (i.e. cycles of odd number of variables) which have only one positive or only one negative weight. However, as will be seen in the next section, unlike trees it is not possible to extend these results for any general cycle.

6 Some Useful SOC Constraints

We now describe two SOCP relaxations which include all the marginalization constraints specified in LP-S. Note that the marginalization constraints can be incorporated within the SOCP framework but not in the QP framework.

6.1 The SOCP-C Relaxation

The SOCP-C relaxation (where C denotes cycles) defines second order cone (SOC) constraints using positive semidefinite matrices $\mathbf{C}$ such that the graph G (defined in § 5.1) form cycles. Let the variables corresponding to vertices of one such cycle G of length c be denoted as $\mathbf{v}_C = \{v_b | b \in \{a_1, a_2, \cdots, a_c\}\}$. Further, let $\mathbf{l}_C = \{l_j | j \in \{i_1, i_2, \cdots, i_c\}\} \in \mathbf{l}^c$ be a set of labels for the variables $\mathbf{v}_C$. The SOCP-C relaxation specifies the following constraints:

- The marginalization constraints, i.e.

$$\sum_{l_j \in \mathbf{l}} X_{ab;ij} = (2-h)x_{a;i}, \forall (a,b) \in \mathcal{E}, l_i \in \mathbf{l}. \tag{153}$$

- A set of SOC constraints

$$||\mathbf{U}^\top \mathbf{x}|| \leq \mathbf{C} \bullet \mathbf{X}, \tag{154}$$

  such that the graph G defined by the above constraint forms a cycle. The matrix $\mathbf{C}$ is 0 everywhere except the following elements:

$$C_{a_k,a_l,i_k,i_l} = \begin{cases} \lambda_c & \text{if} \quad k = l, \\ D_c(k,l) & \text{otherwise.} \end{cases} \tag{155}$$

  Here $\mathbf{D}_c$ is a $c \times c$ matrix which is defined as follows:

$$D_c(k,l) = \begin{cases} 1 & \text{if} \quad |k-l| = 1 \\ (-1)^{c-1} & \text{if} \quad |k-l| = c-1 \\ 0 & \text{otherwise,} \end{cases} \tag{156}$$

  and λ_c is the absolute value of the smallest eigenvalue of $\mathbf{D}_c$.

In other words the submatrix of $\mathbf{C}$ defined by $\mathbf{v}_C$ and $\mathbf{l}_C$ has diagonal elements equal to λ_c and off-diagonal elements equal to the elements of $\mathbf{D}_c$. As an example we consider two cases when $c = 3$ and $c = 4$. In these cases the matrix $\mathbf{D}_c$ is given by

$$\mathbf{D}_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \text{ and } \mathbf{D}_4 = \begin{pmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \tag{157}$$

respectively, while $\lambda_3 = 1$ and $\lambda_4 = \sqrt{2}$. Clearly, $\mathbf{C} = \mathbf{U}^\top \mathbf{U} \succeq 0$ since its only non-zero submatrix $\lambda_c I + \mathbf{D}_c$ (where I is a $c \times c$ identity matrix) is positive semidefinite. This allows us to define a valid SOC constraint as shown in inequality (154). We choose to define the SOC constraint (154) for only those set of labels $\mathbf{l}_C$ which satisfy the following:

$$\sum_{(a_k,a_l) \in \mathcal{E}} D_c(k,l)\theta^2_{a_k a_l; i_k i_l} \geq \sum_{(a_k,a_l) \in \mathcal{E}} D_c(k,l)\theta^2_{a_k a_l; j_k j_l}, \forall \{j_1, j_2, \cdots, j_c\}. \tag{158}$$

Note that this choice is motivated by the fact that the variables $X_{a_k a_l;i_k i_l}$ corresponding to these sets $\mathbf{v}_C$ and $\mathbf{l}_C$ are assigned trivial values by the LP-S relaxation in the presence of non-submodular terms (see example below), i.e.

$$X_{a_k a_l;i_k i_l} = \begin{cases} -1 & \text{if} \quad \theta^2_{a_k a_l;i_k i_l} \geq 0, \\ 1 & \text{otherwise.} \end{cases} \tag{159}$$

In order to avoid this trivial solution, we impose the SOC constraint (154) on them.

Since marginalization constraints are included in the SOCP-C relaxation, the value of the objective function obtained by solving this relaxation would at least be equal to the value obtained by the LP-S relaxation (i.e. SOCP-C dominates LP-S, see Case II in section 2). We can further show that in the case where $|\mathbf{l}| = 2$ and the constraint (154) is defined over a frustrated cycle[7] SOCP-C strictly dominates LP-S. One such example is given below. Note that if the given CRF contains no frustrated cycle, then it can be solved exactly using the method described in [9].

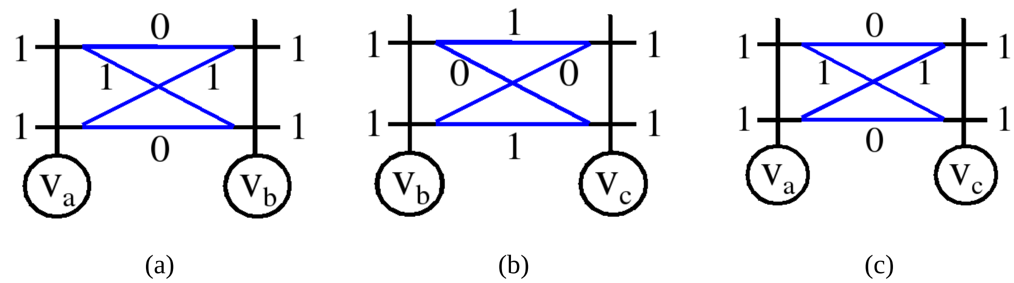


Figure 4: *An example CRF defined over three random variables $\mathbf{v} = \{v_a, v_b, v_c\}$ shown as unfilled circles. Each of these variables can take one of two labels from the set $\mathbf{l} = \{l_0, l_1\}$ which are shown as branches (i.e. the horizontal lines) of trellises (i.e. the vertical lines) on top of the random variables. The unary potentials are shown next to the corresponding branches. The pairwise potentials are shown next to the edges connecting the branches of two neighbouring variables. Note that the pairwise potentials defined for (a, b) and (a, c) form a submodular Ising model (in* **(a)** *and* **(b)** *respectively). The pairwise potentials defined for (b, c) are non-submodular (in* **(c)***).*

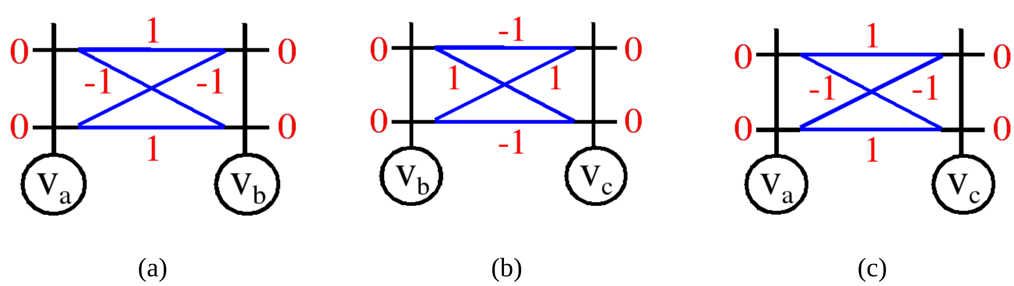


Figure 5: *An optimal solution provided by the LP-S relaxation for the CRF shown in Fig. 4. This solution is shown in red to avoid confusing it with the potentials shown in Fig. 4. The value of variable $x_{a;i}$ is shown next to the i^{th} branch of the trellis on top of v_a. In this optimal solution, all such variables $x_{a;i}$ are equal to 0. The value of the variable $X_{ab;ij}$ is shown next to the connection joining the i^{th} and the j^{th} branch of the trellises on top of v_a and v_b respectively. Note that $X_{ab;ij} = -1$ when $\theta^2_{ab;ij} > 0$ and $X_{ab;ij} = 1$ otherwise. This provides us with the minimum value of the objective function of LP-S, i.e. 3.*

Example: We consider a frustrated cycle and show that SOCP-C strictly dominates LP-S. Specifically, we consider a CRF with $\mathbf{v} = \{v_a, v_b, v_c\}$ and $\mathbf{l} = \{l_0, l_1\}$. The neighbourhood of this CRF is defined such that the variables form a cycle of length 3, i.e. $\mathcal{E} = \{(a, b), (b, c), (c, a)\}$. We define a

[7] A cycle is called frustrated if it contains an odd number of non-submodular terms.

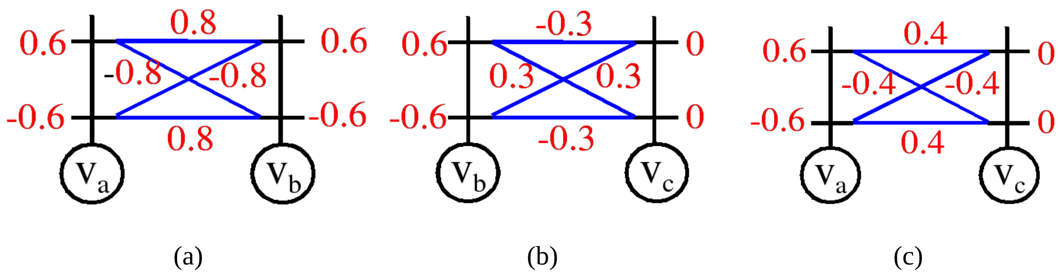


Figure 6: *An optimal solution provided by the* SOCP-C *relaxation for the* CRF *shown in Fig. 4. This optimal solution provides us with the optimal value of* 3.75 *which greater than the* LP-S *optimal value for the solution shown in Fig. 5. Note that the optimal solution of* LP-S *does not belong to the feasibility region of* SOCP-C *as it violates constraint (160). This example proves that* SOCP-C *strictly dominates* LP-S.

frustrated cycle which consists of all 3 variables of this CRF using the unary and pairwise potentials shown in Fig. 4, i.e. the unary potentials are uniform and the pairwise potentials define only one non-submodular term (between the vertices b and c). Clearly, the energy of the optimal labelling for the above problem is 4. The value of the objective function obtained by solving the LP-S relaxation is 3 at an optimal solution shown in Fig. 5.

The LP-S optimal solution is no longer feasible when the SOCP-C relaxation is used. Specifically, the constraint

$$(x_{a;0} + x_{b;1} + x_{c;1})^2 \leq 3 + 2(X_{ab;01} + X_{ac;01} + X_{bc;11}) \tag{160}$$

is violated. In fact, the value of the objective function obtained using the SOCP-C relaxation is 3.75. Fig. 6 shows an optimal solution of the SOCP-C relaxation for the CRF in Fig. 4. The above example can be generalized to a frustrated cycle of any length. This proves that SOCP-C strictly dominates the LP-S relaxation (and hence, the QP-RL and SOCP-MS relaxations).

The constraint defined in equation (154) is similar to the (linear) cycle inequality constraints [1] which are given by

$$\sum_{k,l} D_c(k,l) X_{a_k a_l; i_k i_l} \geq 2 - c. \tag{161}$$

We believe that the feasibility region defined by cycle inequalities is a strict subset of the feasibility region defined by equation (154). In other words a relaxation defined by adding cycle inequalities to LP-S would strictly dominate SOCP-C. We are not aware of a formal proof for this. We now describe the SOCP-Q relaxation.

6.2 The SOCP-Q Relaxation

In this previous section we saw that LP-S dominates SOCP relaxations whose constraints are defined on trees. However, the SOCP-C relaxation, which defines its constraints using cycles, strictly dominates LP-S. This raises the question whether matrices **C**, which result in more complicated graphs G, would provide an even better relaxation for the MAP estimation problem. In this section, we answer this question in an affirmative. To this end, we define an SOCP relaxation which specifies constraints such that the resulting graph G from a clique. We denote this relaxation by SOCP-Q (where Q indicates cliques).

The SOCP-Q relaxation contains the marginalization constraint and the cycle inequalities (defined above). In addition, it also defines SOC constraints on graphs G which form a clique. We denote the variables corresponding to the vertices of clique G as $\mathbf{v}_Q = \{v_b | b \in \{a_1, a_2, \cdots, a_q\}\}$. Let $\mathbf{l}_Q = \{l_j | j \in \{i_1, i_2, \cdots, i_q\}\}$ be a set of labels for these variables $\mathbf{v}_Q$. Given this set of variables $\mathbf{v}_Q$ and labels $\mathbf{l}_Q$, we define an SOC constraint using a matrix $\mathbf{C}$ of size $nh \times nh$ which is zero everywhere except for the elements $C_{a_k a_l; i_k i_l} = 1$. Clearly, $\mathbf{C}$ is a rank 1 matrix with eigenvalue 1 and eigenvector $\mathbf{u}$ which is zero everywhere except $u_{a_k;i_k} = 1$ where $v_{a_k} \in \mathbf{v}_Q$ and $l_{i_k} \in \mathbf{l}_Q$. This

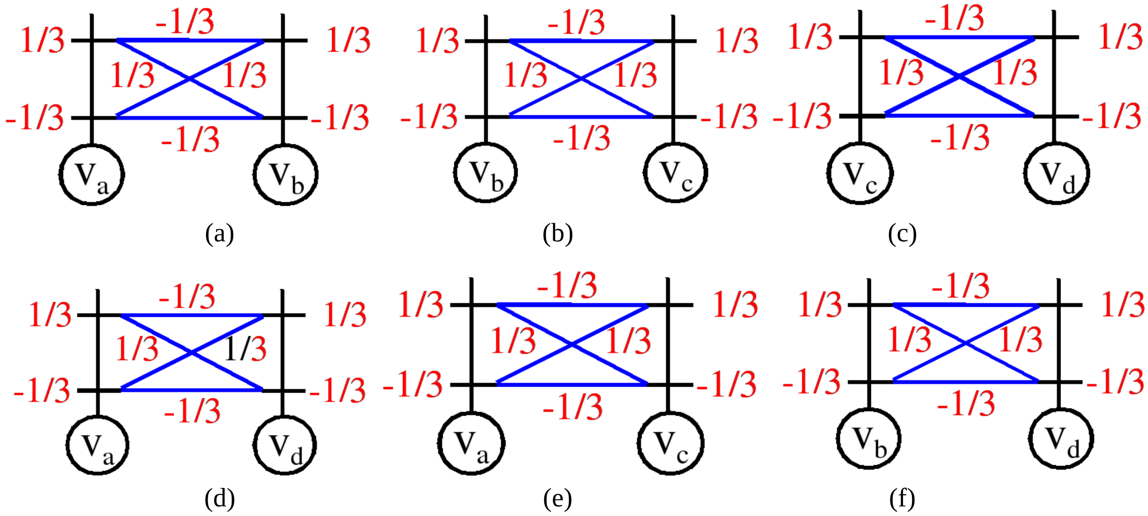


Figure 7: *An infeasible solution for* SOCP-Q. *The value of the variable $x_{a;i}$ is shown next to the i^{th} branch of the trellis on top of v_a. The value of $X_{ab;ij}$ is shown next to the connection between the i^{th} and the j^{th} branches of the trellises on top of v_b and v_b respectively. It can easily be verified that these variables satisfy all cycle inequalities. However, they do not belong to the feasibility region of* SOCP-Q *since they violate constraint (166).*

implies that $\mathbf{C} \succeq 0$, which enables us to obtain the following SOC constraint:

$$\left(\sum_k x_{a_k;i_k}\right)^2 \leq q + \sum_{k,l} X_{a_k a_l;i_k i_l}. \tag{162}$$

We choose to specify the above constraint only for the set of labels $\mathbf{l}_Q$ which satisfy the following condition:

$$\sum_{(a_k,a_l)\in\mathcal{E}} \theta^2_{a_k a_l;i_k i_l} \geq \sum_{(a_k,a_l)\in\mathcal{E}} \theta^2_{a_k a_l;j_k j_l}, \forall\{j_1, j_2, \cdots, j_q\}. \tag{163}$$

Again, this choice is motivated by the fact that the variables $X_{a_k a_l;i_k i_l}$ corresponding to these sets $\mathbf{v}_Q$ and $\mathbf{l}_Q$ are assigned trivial values by the LP-S relaxation in the presence of non-submodular pairwise potentials.

When the clique contains a frustrated cycle, it can be shown that SOCP-Q dominates the LP-S relaxation (similar to SOCP-C). Further, using a counter-example, it can proved that the feasibility region given by cycle inequalities is not a subset of the feasibility region defined by constraint (162). One such example is given below.

Example: We present an example to prove that the feasibility region given by cycle inequalities is not a subset of the feasibility region defined by the SOC constraint

$$\left(\sum_k x_{a_k;i_k}\right)^2 \leq q + \sum_{k,l} X_{a_k a_l;i_k i_l}, \tag{164}$$

which is used in SOCP-Q. Note that it would be sufficient to provide a set of variables $(\mathbf{x}, \mathbf{X})$ which satisfy the cycle inequalities but not constraint (164).

To this end, we consider a CRF defined over the random variables $\mathbf{v} = \{v_a, v_b, v_c, v_d\}$ which form a clique of size 4 with respect to the neighbourhood relationship $\mathcal{E}$, i.e.

$$\mathcal{E} = \{(a,b), (b,c), (c,d), (a,d), (a,c), (b,d)\}. \tag{165}$$

Each of these variables takes a label from the set $\mathbf{l} = \{l_0, l_1\}$. Consider the set of variables $(\mathbf{x}, \mathbf{X})$ shown in Fig. 7 which do not belong to the feasibility region of SOCP-Q. It can be easily shown that these variables satisfy all the cycle inequalities (together with all the constraints of the LP-S relaxation). However, $(\mathbf{x}, \mathbf{X})$ defined in Fig. 7 does not belong to the feasibility region of the SOCP-Q relaxation since it does not satisfy the following SOC constraint:

$$\left(\sum_{v_a \in \mathbf{v}} x_{a;0}\right)^2 \leq 4 + 2\left(\sum_{(a,b)\in\mathcal{E}} X_{ab;00}\right). \tag{166}$$

7 Discussion

We presented an analysis of approximate algorithms for MAP estimation which are based on convex relaxations. The surprising result of our work is that despite the flexibility in the form of the objective function/constraints offered by QP and SOCP, the LP-S relaxation dominates a large class of QP and SOCP relaxations. It appears that the authors who have previously used SOCP relaxations in the Combinatorial Optimization literature [20] and those who have reported QP relaxation in the Machine Learning literature [22] were unaware of this result. We also proposed two new SOCP relaxations (SOCP-C and SOCP-Q) and presented some examples to prove that they provide a better approximation than LP-S. An interesting direction for future research would be to determine the best SOC constraints for a given MAP estimation problem (e.g. with truncated linear/quadratic pairwise potentials).

Acknowledgments: We thank Pradeep Ravikumar and John Lafferty for careful reading of the manuscript and for pointing out an error in our description of the SOCP-MS relaxation.

References

[1] F. Barahona and A. Mahjoub. On the cut polytope. *Mathematical Programming*, 36:157–173, 1986.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.

[4] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labelling problem via a new linear programming formulation. In *SODA*, 2001.

[5] F. Cohen. *Modeling and Applications of Stochastic Processes*. Kluwer, 1986.

[6] E. Dalhaus, D. Johnson, C. Papadimitriou, P. Seymour, and M. Yannakakis. The complexity of multi-terminal cuts. *SICOMP*, 23(4):864–894, 1994.

[7] E. de Klerk, D. Pasechnik, and J. Warners. Approximate graph colouring and max-k-cut algorithms based on the theta function. *Journal of Combinatorial Optimization*, 8(3):267–294, 2004.

[8] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of ACM*, 42:1115–1145, 1995.

[9] P. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, 28:121–155, 1984.

[10] P. Hammer and B. Kalantari. A bound on the roof duality gap. Technical Report RRR 46, Rutgers Center for Operations Research, Rutgers University, 1987.

[11] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *PAMI*, 25(10):1333–1336, October 2003.

[12] A. Karzanov. Minimum 0-extension of graph metrics. *European Journal of Combinatorics*, 19:71–101, 1998.

[13] S. Kim and M. Kojima. Second-order cone programming relaxation of nonconvex quadratic optimization problems. Technical report, Tokyo Institute of Technology, 2000.

[14] J. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. In *STOC*, pages 14–23, 1999.

[15] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.

[16] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007.

[17] A. Koster, C. van Hoesel, and A. Kolen. The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, 23(3-5):89–97, 1998.

[18] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Solving Markov random fields using second order cone programming relaxations. In *CVPR*, volume I, pages 1045–1052, 2006.

[19] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal of Optimization*, 11:796–817, 2001.

[20] M. Muramatsu and T. Suzuki. A new second-order cone programming relaxation for max-cut problems. *Journal of Operations Research of Japan*, 43:164–177, 2003.

[21] C. Olsson, A.P. Eriksson, and F. Kahl. Solving large scale binary quadratic problems: Spectral methods vs. semidefinite programming. In *CVPR*, pages 1–8, 2007.

[22] P. Ravikumar and J. Lafferty. Quadratic programming relaxations for metric labelling and Markov random field MAP estimation. In *ICML*, 2006.

[23] C. Schellewald and C. Schnorr. Subgraph matching with semidefinite programming. In *IWCIA*, 2003.

[24] D. Schlesinger and B. Flach. Transforming an arbitrary minsum problem into a binary one. Technical Report TUD-F106-01, Dresden University of Technology, 2006.

[25] M. Schlesinger. Sintaksicheskiy analiz dvumernykh zritelnikh singnalov v usloviyakh pomekh (syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika*, 4:113–130, 1976.

[26] M. Schlesinger and B. Flach. Some solvable subclass of structural recognition problems. In *Czech Pattern Recognition Workshop*, 2000.

[27] M. Schlesinger and V. Giginyak. Solution to structural recognition (MAX,+)-problems by their equivalent transformations. Part 1. *Control Systems and Computers*, 1:3–15, 2007.

[28] M. Schlesinger and V. Giginyak. Solution to structural recognition (MAX,+)-problems by their equivalent transformations. Part 2. *Control Systems and Computers*, 2:3–18, 2007.

[29] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *ECCV*, pages II: 16–29, 2006.

[30] P. H. S. Torr. Solving Markov random fields using semidefinite programming. In *AISTATS*, 2003.

[31] M. Wainwright, T. Jaakola, and A. Willsky. MAP estimation via agreement on trees: Message passing and linear programming. *IEEE Trans. on Information Theory*, 51(11):3697–3717, 2005.

[32] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, University of California, Berkeley, 2003.

[33] M. Wainwright and M. Jordan. Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations. Technical Report 671, University of California, Berkeley, 2004.

[34] T. Werner. A linear programming approach to max-sum problem: A review. *PAMI*, 2007.