Analyzing Convex Relaxations for MAP Estimation

M. Pawan Kumar, Stanford University, pawan@cs.stanford.edu
Vladimir Kolmogorov, University College London, vnk@cs.ucl.ac.uk
Philip Torr, Oxford Brookes University, philiptorr@brookes.ac.uk

When faced with a difficult-to-solve problem, it is standard practice in the optimization community to resort to convex relaxations, that is, approximate the original problem using an easier (convex) problem. The widespread use of convex relaxations can be attributed to two factors: (i) recent advances in convex optimization imply that the relaxation can be solved efficiently; and (ii) the relaxed problem readily lends itself to theoretical analysis that often relieves interesting properties. In this chapter, we will focus on theoretical analysis for MAP estimation of discrete MRFs, with the aim of highlighting the importance of designing tight relaxations. The next chapter will illustrate the efficiency afforded by this approach.

We review three standard convex relaxation approaches that have been proposed in the literature:

- The linear programming (LP) relaxation first proposed by Schlesinger for the satisfiability problem [15], and later developed independently for the general case in [2, 8, 19]. We denote this relaxation by LP-S where ‘S’ stands for Schlesinger.
- The quadratic programming (QP) relaxation proposed by Ravikumar and Lafferty [14], which we denote by QP-RL (where ‘RL’ stands for Ravikumar and Lafferty).
- The second-order cone programming (SOCP) relaxation first proposed by Muramatsu and Suzuki for the maximum cut problem [13], and later extended for general MAP estimation of discrete MRFs in [12]. We denote this relaxation by SOCP-MS where ‘MS’ stands for Muramatsu and Suzuki.
Note that it is well-known that any LP can be expressed as a QP and any QP can be expressed as an SOCP. Yet, despite the expressive power of QP and SOCP, we show that the relaxations of [13, 14] are dominated by (that is, provide a weaker approximation than) the LP relaxation. Furthermore, we show that this result can be generalized to a large class of QP and SOCP relaxations. Before proceeding with the analysis, we describe some basic concepts in mathematical optimization that will be used throughout the chapter.

1.1 Preliminaries

**Optimization Problem:** An optimization problem has the following form:

\[
y^* = \arg \min_y g(y), \\
\text{s.t. } h_i(y) \leq 0, i = 1, \ldots, C.
\]

(1.1)

Here \( y \) is called the variable, \( g(\cdot) \) is the objective function, and \( h_i(\cdot), i = 1, \ldots, C \), which restrict the values that \( y \) can take, are called the constraints. The set of all \( y \) that satisfy all the constraints of the optimization problem is called the feasible region. The value \( y^* \) is called an optimal solution and \( g(y^*) \) is called the optimal value. Note that, in general, there may be more than one optimal solution. However, as indicated in the above problem, our task is to find one optimal solution.

**Integer Program:** Integer Programs (IP) refers to the class of optimization problems where the elements of the variable \( y \) are constrained to take integer values. In other words, the feasible region of an IP consists of points \( y \in \mathbb{Z}^n \) (where \( \mathbb{Z} \) is the set of integers and \( n \) is the dimensionality of \( y \)). The class of problems defined by IP is generally NP-hard to solve. In contrast, the following four types of problems have convex objective functions and convex feasible regions. This allows us to solve them efficiently [1].

**Linear Program:** A linear program (LP) is an optimization problem with a linear objective function and linear constraints. Formally, an LP is specified as follows:

\[
y^* = \arg \min_y c^\top y \\
\text{s.t. } Ay \leq b.
\]

(1.2)
1.1 Preliminaries

The vector $c$ defines the objective function, while the matrix $A$ and vector $b$ specify the constraints.

**Convex Quadratic Program:** A convex quadratic program (QP) is an optimization problem with a convex quadratic objective function and linear constraints, i.e.

$$
y^* = \arg\min_y \|Ay + b\|^2 + c^\top y
$$

s.t. $A'y \leq b'$.  \hspace{1cm} (1.3)

**Second-Order Cone Program:** A second-order cone program (SOCP) is an optimization problem with a linear objective function and convex quadratic constraints, i.e.

$$
y^* = \arg\min_y c^\top y,
$$

s.t. $\|A_iy + b_i\|^2 \leq c_i^\top y + d_i, i = 1, \cdots, C$. \hspace{1cm} (1.4)

Convex quadratic constraints are also known as second-order cone (SOC) constraints, Lorentz cone constraints or (because of their shape) ice-cream cone constraints [1].

**Semidefinite Program:** A semidefinite program (SDP) is an optimization problem with a linear objective function and linear constraints defined over a variable matrix $Y$ that is restricted to be positive semidefinite\(^1\) (denoted by $Y \succeq 0$). Formally, an SDP is written as:

$$
Y^* = \arg\min_Y C \bullet Y,
$$

s.t. $Y \succeq 0$,

$$
A_i \bullet Y \leq b_i, i = 1, \cdots, C. \hspace{1cm} (1.5)
$$

Here the operator ($\bullet$) denotes the Frobenius inner product, that is, $A \bullet B = \sum_i \sum_j A_{ij}B_{ij}$.

**Relaxation:** A relaxation of an optimization problem $A$ is another optimization problem $B$ such that: (i) the feasible region of $B$ is a superset of the feasible region of $A$; and (ii) if $y$ belongs to the feasible region

\(^1\) An $n \times n$ matrix $Y$ is said to be positive semidefinite if all its $n$ eigenvalues are non-negative. Equivalently, $Y$ is positive semidefinite if $c^\top Yc \geq 0$, for all $c \in \mathbb{R}^n$. It is worth noting that any positive semidefinite matrix $Y$ can be written as $Y = UU^\top$ for an appropriate matrix $U$. 

STENNING: “CVXRELAX” — 2009/12/15 — 14:09 — PAGE 3 — #3
of A and B then the value of the objective function of B is less than or
equal to the value of the objective function of A at \( y \).

In the next section, we formulate MAP estimation as an IP and
describe its various convex relaxations. All the relaxations discussed
here are special cases of relaxations (as defined above), in that their
value is exactly equal to the value of the integer program for feasible
integer solutions. The subsequent sections identify the best relaxation
amongst them.

1.2 MAP Estimation and its Convex Relaxations

Integer Programming Formulation

We define a binary variable vector \( y \) of length \( NK \) where \( N \) is the
number of random variables and \( K \) is the number of possible labels.
We denote the element of \( y \) at index \( i \cdot K + a \) as \( y_{i:a} \) where \( i \in \mathcal{V} \) and
\( l_a \in \mathcal{L} \). The vector \( y \) specifies a labeling \( x \) such that
\[
y_{i:a} = \begin{cases} 
1 & \text{if } \ x_i = l_a, \\
0 & \text{otherwise}. 
\end{cases}
\]

Let \( Y = y y^\top \). We refer to the \((i \cdot K + a, j \cdot K + b)^{th}\) element of the matrix
\( Y \) as \( y_{ij:ab} \) where \( i, j \in \mathcal{V} \) and \( l_a, l_b \in \mathcal{L} \). Using the above definitions,
we see that the MAP estimation problem is equivalent to the following
IP:
\[
\begin{align*}
\mathbf{y}^* &= \arg\min_{\mathbf{y}} \sum_{i,a} \Phi_i(l_a)y_{i:a} + \sum_{(i,j) \in E, l_a, l_b} \Psi_{ij}(l_a, l_b)y_{ij:ab} \\
\text{s.t.} & \quad \mathbf{y} \in \{0, 1\}^{NK}, \mathbf{Y} = \mathbf{y}\mathbf{y}^\top, \quad (1.6) \\
& \quad \sum_{l_a \in \mathcal{L}} y_{i:a} = 1. \quad (1.7)
\end{align*}
\]

Constraint (1.6) specifies that the variables \( y \) and \( Y \) are binary such
that \( y_{ij:ab} = y_{i:a}y_{j:b} \). We will refer to them as the integer constraints.
Constraint (1.7), which specifies that each variable should be assigned
only one label, is known as the uniqueness constraint. Since the above
IP is equivalent to MAP estimation, it follows that in general it is
NP-hard to solve.
1.2 MAP Estimation and its Convex Relaxations

Linear Programming Relaxation

The LP relaxation, proposed by [15] for a special case (where the pairwise potentials specify a hard constraint, i.e. they are either 0 or $\infty$) and independently in [2, 8, 19] for the general case, is given as follows:

$$
egin{aligned}
\mathbf{y}^* &= \arg \min \mathbf{y} \sum_{i,l_a} \Phi_i(l_a) y_{i; a} + \sum_{(i,j) \in \mathcal{E}, l_a,l_b} \Psi_{ij}(l_a,l_b) y_{ij; ab} \\
\text{s.t.} \quad &y_{i; a} \geq 0, y_{ij; ab} = y_{ji; ba} \geq 0, \sum_{l_a \in \mathcal{L}} y_{i; a} = 1, (1.8) \\
&\sum_{l_a \in \mathcal{L}} y_{ij; ab} = y_{i; a}. (1.9)
\end{aligned}
$$

In the above relaxation, which we call LP-S, only those elements $y_{ij; ab}$ of $\mathbf{Y}$ are used for which $(i,j) \in \mathcal{E}$. Unlike the IP, the feasible region of the above problem is relaxed such that the variables $y_{i; a}$ and $y_{ij; ab}$ lie in the interval $[0,1]$. Further, the constraint $\mathbf{Y} = \mathbf{y}\mathbf{y}^\top$ is replaced by equation (1.9), which is called the marginalization constraint [19]. One marginalization constraint is specified for each $(i,j) \in \mathcal{E}$ and $l_a \in \mathcal{L}$. Note that the above convex constraints are not exhaustive. In other words, it is possible to specify other convex constraints for the problem of MAP estimation such as those used by the relaxations below.

Quadratic Programming Relaxation

We now describe the QP relaxation proposed by [14], which we call QP-RL. To this end, it would be convenient to reformulate the objective function of the IP using a vector of unary potentials of length $NK$ (denoted by $\hat{\Phi}$) and a matrix of pairwise potentials of size $NK \times NK$ (denoted by $\hat{\Psi}$). The elements of the unary potential vector and the pairwise potential matrix are defined as:

$$
\begin{aligned}
\hat{\Phi}_{i; a} &= \Phi_i(l_a) - \sum_{k \in \mathcal{V}, l_c \in \mathcal{L}} |\Psi_{ik}(l_a,l_c)|, \\
\hat{\Psi}_{ij; ab} &= \begin{cases} 
\sum_{k \in \mathcal{V}, l_c \in \mathcal{L}} |\Psi_{ik}(l_a,l_c)|, & \text{if } i = j, a = b, \\
\Psi_{ij}(l_a,l_b) & \text{otherwise,}
\end{cases}
\end{aligned}
$$

where $i, j \in \mathcal{V}$ and $l_a, l_b \in \mathcal{L}$. In other words, the potentials are modified by defining a pairwise potential $\hat{\Psi}_{i; a a}$ and subtracting the value of that potential from the corresponding unary potential $\Phi_{i; a}$. The advantage
of this reformulation is that the matrix $\hat{\Psi}$ is guaranteed to be positive semidefinite. This can be seen by observing that for any vector $y \in \mathbb{R}^{NK}$ the following holds true:

$$y^T \hat{\Psi} y = \sum_{(i,j) \in E} \sum_{l_a,l_b \in L} \left( |\Psi_{ij}(l_a,l_b)| y_{i;a}^2 + |\Psi_{ij}(l_a,l_b)| y_{j;b}^2 + 2\Psi_{ij}(l_a,l_b) y_{i;a} y_{j;b} \right)$$

$$= \sum_{(i,j) \in E} \sum_{l_a,l_b \in L} \left( |\Psi_{ij}(l_a,l_b)| \left( y_{i;a} + \frac{\Psi_{ij}(l_a,l_b)}{|\Psi_{ij}(l_a,l_b)|} y_{j;b} \right)^2 \right)$$

$$\geq 0.$$  

Using the fact that for $y_{i;a} \in \{0,1\}$, $y_{i;a}^2 = y_{i;a}$, it can be shown that the following is equivalent to the MAP estimation problem [14]:

$$y^* = \arg \min_y y^T \hat{\Phi} + y^T \hat{\Psi} y,$$

s.t.

$$\sum_{l_a \in L} y_{i;a} = 1, \forall i \in V,$$

$$y \in \{0,1\}^{NK},$$

(1.12)

By relaxing the feasible region of the above problem to $y \in [0,1]^{NK}$, the resulting QP can be solved in polynomial time since $\hat{\Psi} \succeq 0$ (thereby ensuring that the above relaxation is convex).

Semidefinite Programming Relaxation

In order to describe the SDP relaxation of MAP estimation, it would be useful to reformulate the IP using binary variables $y_{i;a}^\prime = 2y_{i;a} - 1$ and $y_{ij;ab}^\prime = 4y_{ij;ab} - y_{i;a}^\prime - y_{j;b}^\prime - 1$ that take values $\{-1,1\}$. Clearly, the above IP is equivalent to the previous one:

$$\min_{\hat{y}'} \left( \sum_{i,a} \Phi_i(l_a) \left( \frac{1 + y_{i;a}^\prime}{2} \right) + \right.$$

$$\sum_{(i,j) \in E, l_a,l_b} \Psi_{ij}(l_a,l_b) \left( \frac{1 + y_{i;a}^\prime + y_{j;b}^\prime + y_{ij;ab}^\prime}{4} \right) \right)$$

$$\text{s.t.}$$

$$\sum_{l_a \in L} y_{i;a}^\prime = 1, \forall i \in V,$$

$$y^\prime \in \{-1,1\}^{NK},$$

(1.13)
1.2 MAP Estimation and its Convex Relaxations

\[ y' \in \{-1, 1\}^{NK}, Y' = y' y'^T, \sum_{l_a \in \mathcal{L}} y'_{i;a} = 2 - K. \quad (1.14) \]

The SDP relaxation replaces the non-convex constraint \( Y' = y' y'^T \) by the convex semidefinite constraint \( Y' - y' y'^T \succeq 0 \) \[5\]. Further, it relaxes the integer constraints by allowing the variables to lie in the interval \([-1, 1]\). Finally, using the fact that in the IP \( y'_{ij;ab} = y'_{i;a} y'_{j;b} \), it ensures that \( 1 + y'_{i;a} + y'_{j;b} + y'_{ij;ab} \geq 0 \), \( y'_{ij;ab} = y'_{ji;ba} \) and \( y'_{ii;aa} = 1 \). The SDP relaxation is a well-studied approach that provides accurate solutions for the MAP estimation problem (e.g. see \[18, 20\]). However, due to its computational inefficiency, it is not practically useful for large scale problems with \( NK > 1000 \).

Second-Order Cone Programming Relaxation

We now describe the SOCP relaxation that was proposed by \[13\] for the special case where \( K = 2 \) and later extended for a general label set \[12\]. This relaxation, which we call SOCP-MS, is based on the technique of \[6\]. For completeness we first describe the general technique of \[6\] and later show how SOCP-MS can be derived using it.

**SOCP Relaxations:** In \[6\], the authors observed that the SDP constraint \( Y' - y' y'^T \succeq 0 \) can be further relaxed to 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 \( S = UU^\top \succeq 0 \) with \( Y' - y' y'^T \) (which, by the SDP constraint, is also positive semidefinite). The non-negativity of this inner product can be expressed as an SOC constraint as follows:

\[ S \bullet (Y' - y' y'^T) \geq 0 \Rightarrow \| (U)^\top y' \|^2 \leq S \bullet Y'. \quad (1.15) \]

Hence, by using a set of matrices \( S = \{S^c|S^c = U^c (U^c)^\top \succeq 0, c = 1, 2, \ldots, C\} \), the SDP constraint can be further relaxed to \( C \) SOC constraints, i.e. \( \| (U^c)^\top y' \|^2 \leq S^c \bullet Y', c = 1, \ldots, C \). It can be shown that, for the above set of SOC constraints to be equivalent to the SDP constraint, \( C = \infty \). However, in practice, we can only specify a finite set of SOC constraints. Each of these constraints may involve some or all variables \( y'_{i;a} \) and \( y'_{ij;ab} \). For example, if \( S^c_{ij;ab} = 0 \), then the \( c^{th} \) SOC constraint will not involve \( y'_{ij;ab} \) (since its coefficient will be 0).

**The SOCP-MS Relaxation:** Consider a pair of neighbouring variables \( (i, j) \in \mathcal{E} \), and a pair of labels \( l_a \) and \( l_b \). These two pairs define the following variables: \( y'_{i;a}, y'_{j;b}, y'_{i;i;aa} = y'_{j;j;bb} = 1 \) and \( y'_{ij;ab} = y'_{ji;ba} \) (since
Analyzing Convex Relaxations for MAP Estimation

\( Y' \) is symmetric). For each such pair of variables and labels, the SOCP-MS relaxation specifies the following two SOC constraints [12, 13]:

\[
(y'_{ia} + y'_{jb})^2 \leq 2 + 2y'_{ij;ab}, \quad (y'_{ia} - y'_{jb})^2 \leq 2 - 2y'_{ij;ab}.
\]

(1.16)

It can be verified that the above constraints correspond to the following semidefinite matrices \( S^1 \) and \( S^2 \) respectively:

\[
S^1_{i'j'a'b'} = \begin{cases} 
1 & \text{if } i' = i, j' = i, a' = a, b' = a, \\
1 & \text{if } i' = j, j' = j, a' = b, b' = b, \\
1 & \text{if } i' = i, j' = j, a' = a, b' = b, \\
0 & \text{otherwise,}
\end{cases}
\]

\[
S^2_{i'j'a'b'} = \begin{cases} 
1 & \text{if } i' = i, j' = i, a' = a, b' = a, \\
1 & \text{if } i' = j, j' = j, a' = b, b' = b, \\
-1 & \text{if } i' = i, j' = j, a' = a, b' = b, \\
-1 & \text{if } i' = j, j' = i, a' = b, b' = a, \\
0 & \text{otherwise,}
\end{cases}
\]

(1.17)

Hence, the SOCP-MS formulation is given by

\[
y^{**} = \arg \min_{y'} \left( \sum_{i,a} \Phi_i(l_a) \left( \frac{1 + y'_{ia}}{2} \right) + \sum_{(i,j) \in E, l_a, l_b} \Psi_{ij}(l_a, l_b) \left( \frac{1 + y'_{ia} + y'_{jb} + y'_{ij;ab}}{4} \right) \right)
\]

\[
\text{s.t. } y'_{ia} \in [-1, 1], y'_{ij;ab} \in [-1, 1],
\]

\[
y'_{ij;ab} = y'_{ji;ba} + \sum_{l_a \in \mathcal{E}} y'_{ia} = 2 - K,
\]

\[
(y'_{ia} - y'_{jb})^2 \leq 2 - 2y'_{ij;ab},
\]

\[
(y'_{ia} + y'_{jb})^2 \leq 2 + 2y'_{ij;ab},
\]

(1.18) (1.19) (1.20) (1.21)

We refer the reader to [12, 13] for details. The SOCP-MS relaxation yields the supremum and infimum for the elements of the matrix \( Y' \) using constraints (1.20) and (1.21) respectively, i.e.

\[
\frac{(y'_{ia} + y'_{jb})^2}{2} - 1 \leq y'_{ij;ab} \leq 1 - \frac{(y'_{ia} - y'_{jb})^2}{2}.
\]

(1.22)
These constraints are specified for all \((i, j) \in \mathcal{E}\) and \(l_a, l_b \in \mathcal{L}\). 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 \(y'\) has been fixed. Hence, the elements of the matrix \(Y'\) should take values such that it minimizes the objective function subject to the constraints (1.20) and (1.21). Clearly, the objective function would be minimized when \(y'_{ij;ab}\) equals either its supremum or infimum value, depending on the sign of the corresponding pairwise potential \(\Psi_{ij}(l_a, l_b)\), i.e.

\[
y'_{ij;ab} = \begin{cases} 
\frac{(y'_{i;a}+y'_{j;b})^2 - 1}{(y'_{i;a}+y'_{j;b})^2} & \text{if } \Psi_{ij}(l_a, l_b) > 0, \\
1 - \frac{(y'_{i;a}-y'_{j;b})^2}{(y'_{i;a}-y'_{j;b})^2} & \text{otherwise.}
\end{cases}
\] (1.23)

### 1.3 Comparing the Relaxations

**A Criterion for Comparison**

In order to compare the relaxations described above, we require the following definitions. We say that a relaxation \(R_1\) dominates \([2]\) a relaxation \(R_2\) (alternatively, \(R_2\) is dominated by \(R_1\)) if and only if the optimal value of \(R_1\) is greater than or equal to the optimal value of \(R_2\) for all MAP estimation problems. We note here that the concept of domination has been used previously by \([2]\) (to compare LP-S with the linear programming relaxation of \([7]\)).

Relaxations \(R_1\) and \(R_2\) are said to be equivalent if \(R_1\) dominates \(R_2\) and \(R_2\) dominates \(R_1\), that is, their optimal values are equal to each other for all MAP estimation problems. A relaxation \(R_1\) is said to strictly dominate relaxation \(R_2\) if \(R_1\) dominates \(R_2\) but \(R_2\) does not dominate \(R_1\). In other words, \(R_1\) dominates \(R_2\) and there exists at least one MAP estimation problem for which the optimal value of \(R_1\) is strictly greater than the optimal value of \(R_2\). Note that, by definition, the optimal value of any relaxation would always be less than or equal to the energy of the MAP labeling. Hence, the optimal value of a strictly dominating relaxation \(R_1\) is closer to the optimal value of the MAP estimation IP compared to that of relaxation \(R_2\). In other words, \(R_1\) provides a tighter lower bound than \(R_2\) and, in that sense, is a better relaxation than \(R_2\).

It is worth noting that the concept of domination (or strict domination) applies to the optimal (possibly fractional) solutions of the relaxations. In practice, the optimal solution is rounded to a feasible integer solution in order to obtain a labeling of the MRF. For example,
the simplest rounding scheme would be to treat the fractional value $y_{i,a}$ as the probability of assigning the random variable $i \in V$ the label $l_a \in L$, and sample labelings from this distribution. In such a scenario, it is possible that the dominating relaxation $R_1$ may provide a labeling with higher energy than $R_2$. The natural question that arises is “Why not compare the final solutions directly?”. Unfortunately, although this would be the ideal comparison criterion, there exist too many rounding schemes in the literature to make such an approach practical feasible. Having said that, the concept of domination as described above provides a nice balance between computability and usefulness. In fact, one could argue that if a dominating relaxation provides worse final integer solutions, then the deficiency of the method may be rooted in the rounding technique used.

Using domination as the criterion for comparison, we now analyze the convex relaxations described in the previous section. Note that we only provide a sketch of the proofs here. We refer the reader to [10] for the details.

**LP-S vs. SOCP-MS**

We now provide a comparison of the LP-S and SOCP-MS relaxations. To this end, it would be helpful to formulate the constraints of both the relaxations using the same variables. Hence, we rewrite the LP-S constraints as

\[
\begin{align*}
y'_{i,a} & \in [-1,1], \quad y'_{ij;ab} \in [-1,1], \quad (1.24) \\
\sum_{l_a \in L} y'_{i,a} & = 2 - K, \quad (1.25) \\
\sum_{l_b \in L} y'_{ij;ab} & = (2 - K)y'_{i,a}, \quad (1.26) \\
y'_{ij;ab} & = y'_{ji;ab}, \quad (1.27) \\
1 + y'_{i,a} + y'_{j;b} + y'_{ij;ab} & \geq 0. \quad (1.28)
\end{align*}
\]
1.3 Comparing the Relaxations

Recall that the SOCP-MS constraints are given by

\[ y'_{i;a} \in [-1, 1], y'_{ij;ab} \in [-1, 1], \]
\[ \sum_{l_a \in L} y'_{i;a} = 2 - K, \]  
\[ (y'_{i;a} - y'_{j;ib})^2 \leq 2 - 2y'_{ij;ab}, \]  
\[ (y'_{i;a} + y'_{j;ib})^2 \leq 2 + 2y'_{ij;ab}, \]  
\[ y'_{ij;ab} = y'_{ji;ba}. \]  

(1.30)  
(1.31)  
(1.32)  
(1.33)

We show that the feasible region of LP-S is a strict subset of the feasible region of SOCP-MS. This would allow us to prove the following.

**Theorem 1:** LP-S strictly dominates SOCP-MS.

**Sketch of the Proof:** The LP-S and the SOCP-MS relaxations differ only in the way they relax the non-convex constraint \( Y' = y' y'^\top \). While LP-S relaxes \( Y' = y' y'^\top \) using the marginalization constraint (1.26), SOCP-MS relaxes it to constraints (1.31) and (1.32). The SOCP-MS constraints provide the supremum and infimum of \( y'_{ij;ab} \) as

\[ \frac{(y'_{i;a} + y'_{j;ib})^2}{2} - 1 \leq y'_{ij;ab} \leq 1 - \frac{(y'_{i;a} - y'_{j;ib})^2}{2}. \]  

(1.34)

In order to prove this theorem, we use the following Lemmas.

**Lemma 1:** If \( y'_{i;a}, y'_{j;ib} \) and \( y'_{ij;ab} \) satisfy the LP-S constraints, that is, constraints (1.24)-(1.28), then \( |y'_{i;a} - y'_{j;ib}| \leq 1 - y'_{ij;ab} \). The above result holds true for all \((i, j) \in \mathcal{E} \) and \( l_a, l_b \in \mathcal{L} \).

**Lemma 2:** If \( y'_{i;a}, y'_{j;ib} \) and \( y'_{ij;ab} \) satisfy the LP-S constraints then \( |y'_{i;a} + y'_{j;ib}| \leq 1 + y'_{ij;ab} \). This holds true for all \((i, j) \in \mathcal{E} \) and \( l_a, l_b \in \mathcal{L} \).

Squaring both the sides of the above inequalities, we can show that LP-S provides smaller supremum and larger infimum of the elements of the matrix \( Y' \) than the SOCP-MS relaxation. Thus, the feasible region of LP-S and is a strict subset of the feasible region of SOCP-MS.

One can also construct potentials for which the set of all optimal solutions of SOCP-MS do not lie in the feasible region of LP-S. For example, see Fig. 1.1. Together with the above two Lemmas, this implies that LP-S strictly dominates SOCP-MS.

Note that the above Theorem does not apply to the variation of SOCP-MS described in [12, 13] that includes triangular inequalities [3]. 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.
Figure 1.1  (a) An example MRF defined using two neighbouring random variables. Note that the observed nodes are not shown for the sake of clarity of the image. Each random variable can take one of two labels, represented by the branches (i.e. the horizontal lines) of the trellis (i.e. the vertical lines) on top of the variables. The value of the unary potential $\Phi_i(l_a)$ is shown next to the $a^{th}$ branch of the trellis on top of $i$. For example, $\Phi_i(l_1) = 10$ (shown next to the lower branch of the trellis on top of $i$) and $\Phi_j(l_2) = 3$ (shown next to the upper branch of the trellis on top of $j$). The pairwise potential $\Psi_{ij}(l_a,l_b)$ is shown next to the connection between the $a^{th}$ and $b^{th}$ branches of the trellises on top of $i$ and $j$ respectively. For example, $\Psi_{ij}(l_1,l_1) = -10$ (shown next to the bottom-most connection between the two trellises) and $\Psi_{ij}(l_1,l_2) = -5$ (shown next to the diagonal connection between the two trellises).  (b) The optimal solution obtained using the LP-S relaxation. The value of $y_{ia}$ is shown next to the $a^{th}$ branch of the trellis on top of $i$. Similarly, the value of $y_{ij,ab}$ is shown next to the connection between the $a^{th}$ and $b^{th}$ branches of the trellises on top of $i$ and $j$ respectively. Note that the value of the objective function for the optimal solution is 6. (c) A feasible solution of the SOCP-MS relaxation that does not belong to the feasible region of LP-S and has an objective function value of 2. It follows that the optimal solution of SOCP-MS would lie outside the feasible region of LP-S and have a value of at most 2. Together with Lemmas 1 and 2, this proves that LP-S strictly dominates SOCP-MS.

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 MRFs). To aid the comparison, we rewrite QP-RL using the variables
1.3 Comparing the Relaxations

\[ \min_{y'} \left( \frac{1 + y'}{2} \right) \Phi + \left( \frac{1 + y'}{2} \right) \Psi \left( \frac{1 + y'}{2} \right), \]

\[ \sum_{l_i \in L} y_i' = 2 - K, \forall i \in V, \quad (1.35) \]

\[ y \in [-1, 1]^{NK}, \quad (1.36) \]

where \( 1 \) is a vector of dimension \( NK \times 1 \) whose elements are all equal to 1. We consider a vector \( y' \) that lies in the feasible regions of the QP-RL and SOCP-MS relaxations, i.e. \( y' \in [-1, 1]^{NK} \). 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 \( y^* \) is an optimal solution of QP-RL for some MRF then there exists an optimal solution \((y^*, Y^*)\) 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:** QP-RL and SOCP-MS are equivalent.

**Sketch of the Proof:** Recall that in the QP-RL relaxation

\[ \hat{\Phi}_{i,a} = \Phi_i(l_a) - \sum_{k \in V} \sum_{l_c \in L} |\Psi_{ik}(a,c)|, \quad (1.37) \]

\[ \hat{\Psi}_{ij,ab} = \left\{ \begin{array}{ll} \sum_{k \in V} \sum_{l_c \in L} |\Psi_{ik}(a,c)|, & \text{if } i = j, a = b, \\ \Psi_{ij}(l_a, l_b), & \text{otherwise}, \end{array} \right. \quad (1.38) \]

Here, the terms \( \Phi_i(l_a) \) and \( \Psi_{ij}(a, m) \) are the (original) unary potentials and pairwise potentials for the given MRF. Consider a feasible solution \( y' \) of the QP-RL and the SOCP-MS relaxations. Further, let \( y' \) be the solution obtained when minimizing the objective function of the SOCP-MS relaxation whilst keeping \( y' \) fixed. In order to prove the theorem, we compare the coefficient of \( \Phi_i(l_a) \) and \( \Psi_{ij}(l_a, l_b) \) for all \( i \in V, (i,j) \in E \). It can be verified that the coefficients are the same for both QP-RL and SOCP-MS [10]. This proves the theorem.

Theorems 1 and 2 prove that the LP-S relaxation strictly dominates the QP-RL and SOCP-MS relaxations.

**SOCP Relaxations over Trees and Cycles**

We now generalize the above results to a large class of SOCP (and equivalent QP) relaxations. Recall that SOCP relaxations further relax
the SDP constraint to the following SOC constraints:

\[
\| (U^c)^\top y' \|^2 \leq S^c \bullet Y', c = 1, \cdots, C. \tag{1.39}
\]

Consider one such SOC constraint defined by the semidefinite matrix \( S^c = U^c (U^c)^\top \). Using this constraint, we define a graph \( G^c = (\mathcal{V}^c, \mathcal{E}^c) \) as follows:

- The set \( \mathcal{E}^c \) is defined such that \((i, j) \in \mathcal{E}^c\) if and only if it satisfies the following conditions:

\[
(i, j) \in \mathcal{E}, \exists l_a, l_b \in L \text{ such that } S_{ij}^{c, ab} \neq 0. \tag{1.40}
\]

In other words \( \mathcal{E}^c \) is the subset of the edges in the graphical model of the MRF such that \( S^c \) specifies constraints for the random variables corresponding to those edges.

- The set \( \mathcal{V}^c \) is defined as \( i \in \mathcal{V}^c \) if and only if there exists a \( j \in \mathcal{V} \) such that \((i, j) \in \mathcal{E}^c\). In other words \( \mathcal{V}^c \) is the subset of hidden nodes in the graphical model of the MRF such that \( S^c \) specifies constraints for the random variables corresponding to those hidden nodes.

We say that the constraint specified by \( S^c \) is defined over the graph \( G^c \). Note that, according to the above definitions, the constraints used in SOCP-MS are defined over graphs that consist of a single edge, that is, \( \mathcal{V}^c = \{i, j\} \) and \( \mathcal{E}^c = \{(i, j)\} \). Hence, the following theorem provides a generalization of Theorem 1.

**Theorem 3:** Any SOCP relaxation whose constraints are defined over arbitrarily large trees is dominated by LP-S.

We can go a step further and characterize those SOCP relaxations that define constraints over arbitrarily large cycles and are dominated by LP-S. For example, consider the case when the pairwise potentials are of the form \( \Psi_{ij}(l_a, l_b) = w_{ij} d(l_a, l_b) \) where \( w_{ij} \) is the weight of the edge \((i, j)\) and \( d(\cdot, \cdot) \) is a non-negative distance function. Then the following theorem holds true.

**Theorem 4:** An SOCP relaxation whose constraints are defined over non-overlapping graphs \( G^c \) that form arbitrarily large even cycles with all positive or all negative weights is dominated by LP-S. The result also holds if \( G^c \) form arbitrarily large odd cycles with only one positive or only one negative weight.

The proofs of Theorems 3 and 4 are given in [10].
1.4 Discussion

We reviewed three standard relaxations for MAP estimation of discrete MRFs: LP-S, QP-RL and SOCP-MS. We showed that, despite the flexibility in the form of the objective function and constraints afforded by QP and SOCP, both QP-RL and SOCP-MS are dominated by LP-S. In fact, the domination of LP-S can be extended to a large class of QP and SOCP relaxations.

It is worth noting that, although LP-S is provably tighter than several relaxations, it is by no means the tightest possible relaxation. In fact, recent work has focused on obtaining tighter relaxations by adding more constraints to LP-S, such as cycle inequalities, clique-web inequalities and positive semidefinite inequalities [4]. Amongst them, the most popular and well-studied class of constraints is cycle inequalities. These constraints are defined over subgraphs of $G$ that forms a cycle. Although there exist exponentially many cycle inequalities (in terms of the number of random variables) for a given MAP estimation problem, it has been shown that they yield to a polynomial time separation algorithm [16]. In other words, given a fractional labeling specified by $(y, Y)$, we can find a cycle inequality that it violates in polynomial time. This allows us to design efficient cutting plane algorithms where we initially enforce only the LP-S constraints and keep adding a violated cycle inequality after each iteration until we obtain a good approximation to MAP labeling. We refer the interested reader to [16] for more details.

Another way to strengthen the LP-S relaxation is to add constraints defined over a clique containing a small number of nodes $(3, 4, \ldots)$. This is sometimes known as a cluster-based LP relaxation. Such an approach has been successfully used, for example, in some bio-informatics problems [17].

We hope that this chapter has helped illustrate the amenability of convex relaxations to theoretical analysis. Such an analysis often makes it possible to choose the right approach for the problem at hand. While several other properties of convex relaxations for MAP estimation have been reported in the literature (for example, its integrality gap [2] for special cases), there still remain some open questions. One such question concerns the worst case multiplicative bounds obtained by LP-S (or any convex relaxation in general). It has recently been shown that all the known multiplicative bounds of LP-S for $N \geq K$ can be obtained via some recently proposed move-making algorithms that only employ the efficient minimum cut procedure [11, 9]. However, it is not clear whether this result can be generalized to all MAP estimation
problems (for example, when $N < K$ or when the pairwise potentials do not form a semi-metric distance).
Bibliography


