Improved Moves for Truncated Convex Models

M. Pawan Kumar
Dept. of Engineering Science
University of Oxford
pawan@robots.ox.ac.uk

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

Abstract

We consider the problem of obtaining the approximate maximum *a posteriori* estimate of a discrete random field characterized by pairwise potentials that form a truncated convex model. For this problem, we propose an improved st-MINCUT based *move making* algorithm. Unlike previous move making approaches, which either provide a loose bound or no bound on the quality of the solution (in terms of the corresponding Gibbs energy), our algorithm achieves the same guarantees as the standard linear programming (LP) relaxation. In other words, for the case of truncated linear metric we obtain a multiplicative bound of $2+\sqrt{2}$, while for truncated quadratic semi-metric we obtain a multiplicative bound of $O(\sqrt{M})$ (where M is the truncation factor). Compared to previous approaches based on the LP relaxation, e.g. interior-point algorithms or tree-reweighted message passing (TRW), our method is faster as it uses only the efficient st-MINCUT algorithm in its design. Furthermore, it directly provides us with a primal solution (unlike TRW and other related methods which solve the dual of the LP). We demonstrate the effectiveness of the proposed approach on both synthetic and standard real data problems.

Our analysis also opens up an interesting question regarding the relationship between move making algorithms (such as α-expansion and the algorithms presented in this paper) and the randomized rounding schemes used with convex relaxations. We believe that further explorations in this direction would help design efficient algorithms for more complex relaxations.

1 Introduction

Discrete random fields are a powerful tool for formulating several problems in Computer Vision such as stereo reconstruction, segmentation, image stitching and image denoising [28]. Given data $\mathbf{D}$ (e.g. an image or a video), random fields model the probability of a set of random variables $\mathbf{v}$, i.e. either the joint distribution of $\mathbf{v}$ and $\mathbf{D}$ as in the case of Markov random fields (MRF) [2] or the conditional distribution of $\mathbf{v}$ given $\mathbf{D}$ as in the case of conditional random fields (CRF) [21]. The word 'discrete' refers to the fact that each of the random variables $v_a \in \mathbf{v} = \{v_0, \cdots, v_{n-1}\}$ can take one label from a discrete set $\mathbf{l} = \{l_0, \cdots, l_{h-1}\}$. Throughout this paper, we will assume a MRF framework while noting that our results are equally applicable for an CRF.

An MRF defines a neighbourhood relationship (denoted by $\mathcal{E}$) over the random variables such that $(a, b) \in \mathcal{E}$ if, and only if, v_a and v_b are neighbouring random variables. Given an MRF, a *labelling* refers to a function f such that

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

In other words, the function f assigns to each random variable $v_a \in \mathbf{v}$, a label $l_{f(a)} \in \mathbf{l}$. The probability of the labelling is given by the following Gibbs distribution:

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

where $\boldsymbol{\theta}$ is the parameter of the MRF and $Z(\boldsymbol{\theta})$ is the normalization constant (i.e. the partition function). Assuming a pairwise MRF, the Gibbs energy is given by:

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

where $\theta^1_{a;f(a)}$ and $\theta^2_{ab;f(a)f(b)}$ are the unary and pairwise potentials respectively. The superscripts '1' and '2' indicate that the unary potential depends on the labelling of one random variable at a time, while the pairwise potential depends on the labelling of two neighbouring random variables.

Using equation (2) it follows that the labelling f which maximizes the posterior $\Pr(f, \mathbf{D}|\boldsymbol{\theta})$ can be obtained by minimizing the Gibbs energy. The problem of obtaining such a labelling f is known as maximum *a posteriori* (MAP) estimation. In this paper, we consider the problem of MAP estimation of random fields where the pairwise potentials are defined by *truncated convex models* [5]. Formally speaking, the pairwise potentials are of the form

$$\theta^2_{ab;f(a)f(b)} = w_{ab} \min\{d(f(a) - f(b)), M\} \tag{4}$$

where $w_{ab} \geq 0$ for all $(a, b) \in \mathcal{E}$, $d(\cdot)$ is a convex function and $M > 0$ is the truncation factor. Recall that, by the definition of Ishikawa [10], a function $d(\cdot)$ defined at discrete points (specifically over integers) is convex if, and only if,

$$d(x + 1) - 2d(x) + d(x - 1) \geq 0, \forall x \in \mathbb{Z}. \tag{5}$$

It is assumed that $d(x) = d(-x)$. Otherwise, it can be replaced by $(d(x) + d(-x))/2$ without changing the energy of any of the possible labellings of the random field [29]. Examples of pairwise potentials of this form include the truncated linear metric and the truncated quadratic semi-metric, i.e.

$$\begin{aligned} \theta^2_{ab;f(a)f(b)} &= w_{ab} \min\{|f(a) - f(b)|, M\}, \\ \theta^2_{ab;f(a)f(b)} &= w_{ab} \min\{(f(a) - f(b))^2, M\}. \end{aligned} \tag{6}$$

Before proceeding further, we would like to note here that the method presented in this paper can be trivially extended to *truncated submodular models* (a generalization of truncated convex models). However, we will restrict our discussion to truncated convex models for two reasons: (i) it makes the analysis of our approach easier; and (ii) truncated convex pairwise potentials are commonly used in several problems such as stereo reconstruction, image denoising and inpainting [28]. Note that in the absence of a truncation factor (i.e. when we only have convex pairwise potentials) the exact MAP estimation can be obtained efficiently using the methods of Ishikawa [10] or Veksler [29]. However, minimizing the Gibbs energy in the presence of a truncation factor is well-known to be NP-hard. Given their widespread use, it is not surprising that several approximate MAP estimation algorithms have been proposed in the literature for the truncated convex model. Below, we review such algorithms.

1.1 Related Work

Given a random field with truncated convex pairwise potentials, Felzenszwalb and Huttenlocher [7] improved the efficiency of the popular max-product belief propagation (BP) algorithm [22] to obtain the MAP estimate. BP provides the exact MAP estimate when the neighbourhood structure $\mathcal{E}$ of the MRF defines a tree (i.e. it contains no loops). However, for a general MRF, BP provides no bounds on the quality of the approximate MAP labelling obtained. In fact, it is not even guaranteed to converge.

The results of [7] can be used directly to speed-up the tree-reweighted message passing algorithm (TRW) [30] and its sequential variant TRW-S [13]. Both TRW and TRW-S attempt to optimize the Lagrangian dual of the standard linear programming (LP) relaxation of the MAP estimation problem [6, 19, 25, 30]. Unlike BP and TRW, TRW-S is guaranteed to converge. However, it is well-known that TRW-S and other related algorithms [8, 17, 26, 27, 31] suffer from the following problems: (i) they are slower than algorithms based on efficient graph-cuts [28]; and (ii) they only provide a dual solution [13]. The primal solution (i.e. the labelling f) is often obtained from the dual solution in an unprincipled manner[1]. Furthermore, it was also observed that, unlike graph-cuts based approaches,

[1]We note here that the recently proposed algorithm in [23] directly provides the primal solution. However, it is much slower than the methods which solve the dual.

TRW-S does not work well when the random field models long range interactions (i.e. when the neighbourhood relationship $\mathcal{E}$ is highly connected) [15]. However, due to the lack of experimental results, it is not clear whether this observation applies to the methods described in [8, 17, 26, 27, 31].

Another way of solving the LP relaxation is to resort to interior point algorithms [4]. Although interior point algorithms are much slower in practice than TRW-S, they have the advantage of providing the primal (possibly fractional) solution of the LP relaxation. Chekuri *et al.* [6] showed that when using certain randomized rounding schemes on the primal solution (to get the final labelling f), the following guarantees hold true: (i) for Potts model (i.e. $d(f(a)-f(b)) = |f(a)-f(b)|$ and $M = 1$), we obtain a multiplicative bound[2] of 2 by using the rounding scheme of [12]; (ii) for the truncated linear metric (i.e. $d(f(a) - f(b)) = |f(a) - f(b)|$ and a general $M > 0$), we obtain a multiplicative bound of $2 + \sqrt{2}$ using the rounding scheme of [6]; and (iii) for the truncated quadratic semi-metric (i.e. $d(f(a) - f(b)) = (f(a) - f(b))^2$ and a general $M > 0$), we obtain a multiplicative bound of $O(\sqrt{M})$ using the rounding scheme of [6].

The algorithms most related to our approach are the so-called move making methods which rely on solving a series of graph-cut (specifically st-MINCUT) problems. Move making algorithms start with an initial labelling f_0 and iteratively minimize the Gibbs energy by moving to a better labelling. At each iteration, (a subset of) random variables have the option of either retaining their old label or taking a new label from a subset of the labels **l**. For example, in the $\alpha\beta$-swap algorithm [5] the variables currently labelled l_α or l_β can either retain their labels or swap them (i.e. some variables labelled l_α can be relabelled as l_β and vice versa). The recently proposed range move algorithm [29] modifies this approach such that any variable currently labelled l_i where $i \in [\alpha, \beta]$ can be assigned any label l_j where $j \in [\alpha, \beta]$. Note that the new label l_j can be different from the old label l_i, i.e. $i \neq j$. Both these algorithms (i.e. $\alpha\beta$-swap and range move) do not provide any guarantees on the quality of the solution.

In contrast, the α-expansion algorithm [5] (where each variable can either retain its label or get assigned the label l_α at an iteration) provides a multiplicative bound of 2 for the Potts model and $2M$ for the truncated linear metric[3]. Gupta and Tardos [9] generalized the α-expansion algorithm for the truncated linear metric and obtained a multiplicative bound of 4. Both α-expansion [5] and its generalization [9] do not provide any bounds for the truncated quadratic semi-metric. Komodakis and Tziritas [18] designed a primal-dual algorithm which provides a bound of $2M$ for the truncated quadratic semi-metric. Note that these bounds are inferior to the bounds obtained by the LP relaxation. However, all the above move making algorithms use only a single st-MINCUT at each iteration and are hence, much faster than interior point algorithms, TRW, TRW-S and BP.

1.2 Our Results

We further extend the approach of Gupta and Tardos [9] in two ways (section 2). The first extension allows us to handle any truncated convex model (and not just truncated linear). The second extension allows us to consider a potentially larger subset of labels at each iteration compared to [9]. As will be seen in the subsequent analysis (section 3), these two extensions allow us to solve the MAP estimation problem efficiently using st-MINCUT whilst obtaining the same guarantees as the LP relaxation [6]. Furthermore, similar to other move making algorithms, our approach does not suffer from the problems of TRW-S mentioned above. In order to demonstrate its practical use, we provide

[2]Let f be the labelling obtained by an algorithm A (e.g. in this case the LP relaxation followed by the rounding scheme) for a class of MAP estimation problems (e.g. in this case when the pairwise potentials form a Potts model). Let f^* be the optimal labelling. The algorithm A is said to achieve a multiplicative bound of σ, if for every instance in the class of MAP estimation problems the following holds true:

$$E\left(\frac{Q(f, \mathbf{D}; \boldsymbol{\theta})}{Q(f^*, \mathbf{D}; \boldsymbol{\theta})}\right) \leq \sigma,$$

where $E(\cdot)$ denotes the expectation of its argument under the rounding scheme.

[3]Note that since α-expansion does not involve any randomized rounding, it is said to provide a multiplicative bound of σ for a class of MAP estimation problems if, and only if, the following holds true for all instances of that class:

$$\frac{Q(f, \mathbf{D}; \boldsymbol{\theta})}{Q(f^*, \mathbf{D}; \boldsymbol{\theta})} \leq \sigma.$$

Initialization
- Initialize the labelling to some function f_1. For example, $f_1(a) = 0$ for all $v_a \in \mathbf{v}$.
Iteration
- Choose an interval $I_m = [i_m + 1, j_m]$ where $(j_m - i_m) = L$ such that $d(L) \geq M$.
- Move from current labelling f_m to a new labelling f_{m+1} such that $f_{m+1}(a) = f_m(a)$ or $f_{m+1}(a) \in I_m, \forall v_a \in \mathbf{v}$.
The new labelling is obtained by solving the st-MINCUT problem on a graph described in § 2.1.
Termination
- Stop when there is no further decrease in the Gibbs energy for any interval I_m.

Table 1: *Our Algorithm. As is typical with move making methods, our approach iteratively goes from one labelling to the next by solving an st-*MINCUT *problem. It converges when there remain no moves which reduce the Gibbs energy further.*

a favourable comparison of our method with several state of the art MAP estimation algorithms (section 4).

2 Description of the Algorithm

Table 1 describes the main steps of our approach which relies on solving an st-MINCUT problem at each iteration. Recall that, given a directed, non-negatively weighted graph with two terminal vertices s (the source) and t (the sink), an st-cut is defined as a partitioning of the vertices of the graph into two disjoint sets such that the first partition contains s while the second partition contains t. The st-MINCUT problem is to find the minimum cost st-cut, where the cost of a cut is measured as the sum of the weights of the edges whose starting point belongs to the first partition and ending point belongs to the second partition. The st-MINCUT problem has several efficient, provably polynomial-time solvers [14] and is used as a building block for several approximate MAP estimation techniques [5, 9, 18, 29].

Unlike the methods described in [5, 29] we will not be able to obtain the optimal move at each iteration. In other words, if in the m^{th} iteration we move from label f_m to f_{m+1} then it is possible that there exists another labelling f'_{m+1} such that

$$\begin{aligned} f'_{m+1}(a) = f_m(a) \text{ or } f'_{m+1}(a) \in I_m, \forall v_a \in \mathbf{v} \\ Q(f'_{m+1}, \mathbf{D}; \boldsymbol{\theta}) < Q(f_{m+1}, \mathbf{D}; \boldsymbol{\theta}). \end{aligned} \tag{7}$$

However, our analysis in the next section shows that we will still be able to reduce the Gibbs energy sufficiently at each iteration so as to obtain the guarantees of the LP relaxation.

We now turn our attention to designing a method of moving from labelling f_m to f_{m+1}. Our approach relies on constructing a graph such that every st-cut on the graph corresponds to a labelling f' of the random variables which satisfies:

$$f'(a) = f_m(a) \text{ or } f'(a) \in I_m, \forall v_a \in \mathbf{v}. \tag{8}$$

The new labelling f_{m+1} is obtained in two steps: (i) we obtain a labelling f' which corresponds to the st-MINCUT on our graph; and (ii) we choose the new labelling f_{m+1} as

$$f_{m+1} = \begin{cases} f' & \text{if} \quad Q(f', \mathbf{D}; \boldsymbol{\theta}) \leq Q(f_m, \mathbf{D}; \boldsymbol{\theta}), \\ f_m & \text{otherwise.} \end{cases} \tag{9}$$

Below, we provide the details of the graph construction.

2.1 Graph Construction

At each iteration of our algorithm, we are given an interval $I_m = [i_m + 1, j_m]$ of L labels (i.e. $(j_m - i_m) = L$) where $d(L) \geq M$. We also have the current labelling f_m for all the random variables. We construct a directed weighted graph (with non-negative weights) $\mathcal{G}_m = \{\mathcal{V}_m, \mathcal{E}_m, c_m(\cdot, \cdot)\}$ such that for each $v_a \in \mathbf{v}$, we define vertices $\{a_{i_m+1}, a_{i_m+2}, \cdots, a_{j_m}\} \in \mathcal{V}_m$. In addition, as is the case with every st-MINCUT problem, there are two additional vertices called terminals which we denote by s (the source) and t (the sink). The edges $e \in \mathcal{E}_m$ with capacity (i.e. weight) $c_m(e)$ are of two types: (i) those that represent the unary potentials of a labelling corresponding to an st-cut in the graph and; (ii) those that represent the pairwise potentials of the labelling.

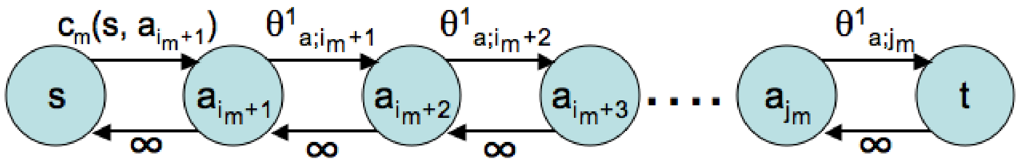


Figure 1: *Part of the graph $\mathcal{G}_m$ containing the terminals and the vertices corresponding to the variable v_a. The edges which represent the unary potential of the new labelling are also shown. The term $c_m(s, a_{i_m+1})$ is shown in equation (10)*
.

Representing Unary Potentials For all random variables $v_a \in \mathbf{v}$, we define the following edges which belong to the set $\mathcal{E}_m$:

1. For all $k \in [i_m + 1, j_m)$, edges (a_k, a_{k+1}) have capacity $c_m(a_k, a_{k+1}) = \theta^1_{a;k}$.
2. For all $k \in [i_m + 1, j_m)$, edges (a_{k+1}, a_k) have capacity $c_m(a_{k+1}, a_k) = \infty$.
3. Edges (a_{j_m}, t) have capacity $c_m(a_{j_m}, t) = \theta^1_{a;j_m}$.
4. Edges (t, a_{j_m}) have capacity $c_m(t, a_{j_m}) = \infty$.
5. Edges (s, a_{i_m+1}) have capacity
$$c_m(s, a_{i_m+1}) = \begin{cases} \theta^1_{a;f_m(a)} & \text{if} \quad f_m(a) \notin I_m \\ \infty & \text{otherwise.} \end{cases} \tag{10}$$
6. Edges (a_{i_m+1}, s) have capacity $c_m(a_{i_m+1}, s) = \infty$.

Fig. 1 shows the above edges together with their capacities for one random variable v_a. Note that there are two types of edges in the above set: (i) with finite capacity; and (ii) with infinite capacity. Any st-cut with finite cost contains only one of the finite capacity edges for each random variable v_a. This is because if an st-cut included more than one finite capacity edge, then by construction it must include at least one infinite capacity edge thereby making its cost infinite [10, 29]. We interpret a finite cost st-cut as a relabelling of the random variables as follows:

$$f'(a) = \begin{cases} k & \text{if st-cut includes edge } (a_k, a_{k+1}) \text{ where } k \in [i_m + 1, j_m), \\ j_m & \text{if st-cut includes edge } (a_{j_m}, t), \\ f_m(a) & \text{if st-cut includes edge } (s, a_{i_m+1}). \end{cases} \tag{11}$$

Note that the sum of the unary potentials for the labelling f' is exactly equal to the cost of the st-cut over the edges defined above. However, the Gibbs energy of the labelling also includes the sum of the pairwise potentials (as shown in equation (3)). Unlike the unary potentials we will not be able to model the sum of pairwise potentials exactly. However, we will be able to obtain its upper bound using the cost of the st-cut over the following edges.

Representing Pairwise Potentials For all neighbouring random variables v_a and v_b, i.e. $(a, b) \in \mathcal{E}$, we define edges $(a_k, b_{k'}) \in \mathcal{E}_m$ where either one or both of k and k' belong to the set $(i_m + 1, j_m]$ (i.e. at least one of them is different from $i_m + 1$). The capacity of these edges is given by

$$c_m(a_k, b_{k'}) = \frac{w_{ab}}{2} \left(d(k - k' + 1) - 2d(k - k') + d(k - k' - 1) \right). \tag{12}$$

The above capacity is non-negative due to the fact that $w_{ab} \geq 0$ and $d(\cdot)$ is convex. Furthermore, we also add the following edges:

$$\begin{aligned} c_m(a_k, a_{k+1}) &= \tfrac{w_{ab}}{2} \left(d(L - k + i_m) + d(k - i_m) \right), \forall (a, b) \in \mathcal{E}, k \in [i_m + 1, j_m) \\ c_m(b_{k'}, b_{k'+1}) &= \tfrac{w_{ab}}{2} \left(d(L - k' + i_m) + d(k' - i_m) \right), \forall (a, b) \in \mathcal{E}, k' \in [i_m + 1, j_m) \\ c_m(a_{j_m}, t) = c_m(b_{j_m}, t) &= \tfrac{w_{ab}}{2} d(L), \forall (a, b) \in \mathcal{E}. \end{aligned} \tag{13}$$

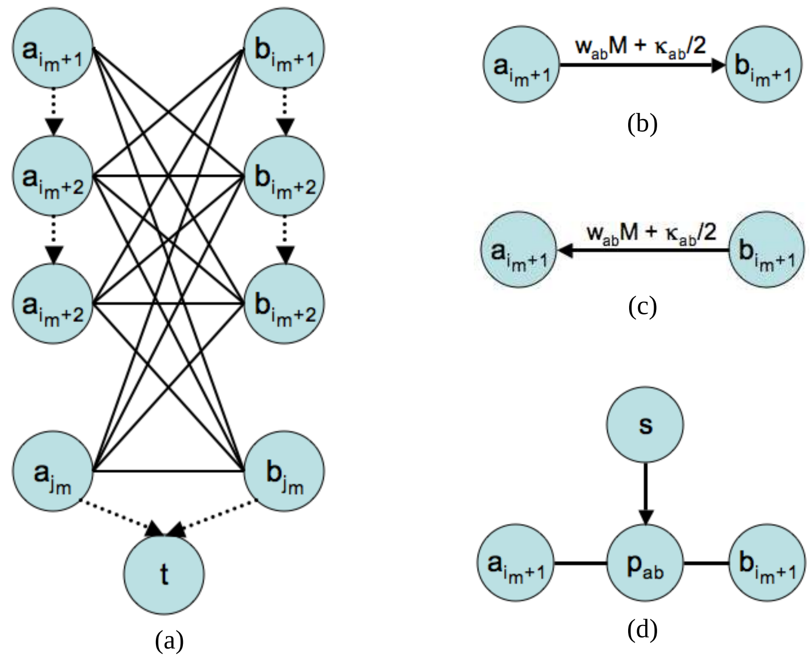


Figure 2: **(a)** *Edges that are used to represent the pairwise potentials of two neighbouring random variables v_a and v_b are shown. Undirected edges indicate that there are opposing edges in both directions with equal capacity (as given by equation 12). Directed dashed edges, with capacities shown in equation (13), are added to ensure that the graph models the convex pairwise potentials correctly.* **(b)** *An additional edge is added when $f_m(a) \in I_m$ and $f_m(b) \notin I_m$. Here, $\kappa_{ab} = w_{ab}d(L)$.* **(c)** *A similar additional edge is added when $f_m(a) \notin I_m$ and $f_m(b) \in I_m$.* **(d)** *Five edges, with capacities as shown in equation (20), are added when $f_m(a) \notin I_m$ and $f_m(b) \notin I_m$. Undirected edges indicate the presence of opposing edges with equal capacity.*

Fig. 2(a) provides an illustration of the above edges. The following Lemma shows that we are now able to model convex pairwise potentials exactly (up to an additive constant).

Lemma 1: For the capacities defined in equations (12) and (13), the cost of the st-cut which includes the edges (a_k, a_{k+1}) and $(b_{k'}, b_{k'+1})$ (i.e. v_a and v_b take labels l_k and $l_{k'}$ respectively) is given by $w_{ab}d(k-k') + \kappa_{ab}$, where $\kappa_{ab} = w_{ab}d(L)$.

Proof: The following proof is due to [29] and is included here for the sake of completeness only. We start by observing that due to the presence of the infinite capacity edges representing unary potentials, the st-cut will consist of only the following edges:

$$\begin{aligned}(a_k, a_{k+1}) \cup (b_{k'}, b_{k'+1}) \cup \{(a_{i'}, b_{j'}), i_m + 1 \leq i' \leq k, k'+1 \leq j' \leq j_m\} \\ \cup \{(a_{i'}, b_{j'}), k+1 \leq i' \leq k, i_m + 1 \leq j' \leq k'\}.\end{aligned} \tag{14}$$

Using equations (12) and (13) to sum the capacities of the above edges, we obtain the following expression:

$$\begin{aligned}&\tfrac{w_{ab}}{2}\left(d(L-k+i_m) + d(k-i_m)\right) + \tfrac{w_{ab}}{2}\left(d(L-k'+i_m) + d(k'-i_m)\right) \\ &+ \textstyle\sum_{i'=i_m+1}^{k} \sum_{j'=k'+1}^{j_m} \tfrac{w_{ab}}{2}\left(d(i'-j'+1) - 2d(i'-j') + d(i'-j'-1)\right) \\ &+ \textstyle\sum_{i'=k+1}^{j_m} \sum_{j'=i_m+1}^{k'} \tfrac{w_{ab}}{2}\left(d(i'-j'+1) - 2d(i'-j') + d(i'-j'-1)\right).\end{aligned} \tag{15}$$

In order to simplify this expression, consider

$$\begin{array}{rc} & \sum_{j'=k'+1}^{j_m} \left(d(i'-j'+1) - 2d(i'-j') + d(i'-j'-1)\right) \\ = & d(i'-k') - 2d(i'-k'-1) + d(i'-k'-2) \\ + & d(i'-k'-1) - 2d(i'-k'-2) + d(i'-k'-3) \\ & \vdots \\ + & d(i'-j_m+2) - 2d(i'-j_m+1) + d(i'-j_m) \\ + & d(i'-j_m+1) - 2d(i'-j_m) + d(i'-j_m-1) \\ = & d(i'-k') - d(i'-k'-1) - d(i'-j_m) + d(i'-j_m+1). \end{array} \tag{16}$$

Hence, it follows that

$$\begin{array}{rc} & \sum_{i'=i_m+1}^{k} \sum_{j'=k'+1}^{j_m} \left(d(i'-j'+1) - 2d(i'-j') + d(i'-j'-1)\right) \\ = & d(i_m+1-k') - d(i_m-k') - d(i_m-j_m+1) + d(i_m-j_m) \\ + & d(i_m+2-k') - d(i_m+1-k') - d(i_m-j_m+2) + d(i_m-j_m+1) \\ & \vdots \\ + & d(k-k'-1) - d(k-k'-2) - d(k-j_m-1) + d(k-j_m-2) \\ + & d(k-k') - d(k-k'-1) - d(k-j_m) + d(k-j_m-1) \\ = & d(k-k') - d(j_m-k) - d(k'-i_m) + d(j_m-i_m) \\ = & d(k-k') - d(L-k+i_m) - d(i_m-k') + d(L), \end{array} \tag{17}$$

where the last expression is obtained using the fact that $L = j_m - i_m$. Note that we also use the fact that $d(x) = d(-x)$. As noted before, if this is not the case then $d(x)$ can be replaced by $(d(x) + d(-x))/2$ to obtain an equivalent MAP estimation problem. Similarly, it can be shown that

$$\begin{array}{rc} & \sum_{i'=k+1}^{j_m} \sum_{j'=i_m+1}^{k'} \left(d(i'-j'+1) - 2d(i'-j') + d(i'-j'-1)\right) \\ = & d(k-k') - d(L-k'+i_m) - d(i_m-k) + d(L), \end{array} \tag{18}$$

Substituting equations (17) and (18) into expression (15), we obtain the cost of the st-cut as

$$\begin{array}{rc} & \frac{w_{ab}}{2}\left(d(L-k+i_m) + d(k-i_m)\right) + \frac{w_{ab}}{2}\left(d(L-k'+i_m) + d(k'-i_m)\right) \\ + & \frac{w_{ab}}{2}\left(d(k-k') - d(L-k+i_m) - d(i_m-k') + d(L)\right) \\ + & \frac{w_{ab}}{2}\left(d(k-k') - d(L-k'+i_m) - d(i_m-k) + d(L)\right) \\ = & w_{ab} d(k-k') + \kappa_{ab}. \end{array} \tag{19}$$

This proves that the capacities in equations (12) and (13) model convex pairwise potentials exactly up to an additive constant. ∎

Since the cost of the st-cut exactly models the convex pairwise potential plus a constant, it follows that the above graph (together with the edges representing unary potentials) can be used to find the exact MAP estimate of the random field with convex pairwise potentials. However, we are concerned with the NP-hard case where the pairwise potentials are truncated. In order to model this case, we incorporate some additional edges to the above set. These additional edges are best described by considering the following three cases for all $(a,b) \in \mathcal{E}$.

1. If $f_m(a) \in I_m$ and $f_m(b) \in I_m$ then we do not add any more edges in the graph (see Fig. 2(a)).
2. If $f_m(a) \in I_m$ and $f_m(b) \notin I_m$ then we add an edge (a_{i_m+1}, b_{i_m+1}) with capacity $w_{ab}M + \kappa_{ab}/2$ (see Fig. 2(b)). Similarly, if $f_m(a) \notin I_m$ and $f_m(b) \in I_m$ then we add an edge (b_{i_m+1}, a_{i_m+1}) with capacity $w_{ab}M + \kappa_{ab}/2$ (see Fig. 2(c)).
3. If $f_m(a) \notin I_m$ and $f_m(b) \notin I_m$, we introduce a new vertex p_{ab}[4]. Using this vertex p_{ab}, five edges are defined with the following capacities (see Fig. 2(d)):

$$\begin{array}{r} c_m(a_{i_m+1}, p_{ab}) = c_m(p_{ab}, a_{i_m+1}) = w_{ab}M + \kappa_{ab}/2, \\ c_m(b_{i_m+1}, p_{ab}) = c_m(p_{ab}, b_{i_m+1}) = w_{ab}M + \kappa_{ab}/2, \\ c_m(s, p_{ab}) = \theta^2_{ab;f_m(a),f_m(b)} + \kappa_{ab}. \end{array} \tag{20}$$

[4] We note here that an equivalent graph can be constructed without adding the vertex p_{ab} using the method of [24]. However, the vertex p_{ab} helps make the analysis easier.

This completes our graph construction. Given the graph $\mathcal{G}_m$ we solve the st-MINCUT problem which provides us with a labelling f' as described in equation (11). The new labelling f_{m+1} is obtained using equation (9).

2.2 Properties of the Graph

We now describe the properties of the above graph construction, with the aim of facilitating the analysis of our algorithm for the case of truncated linear and truncated quadratic models.

Property 1 As mentioned above, the cost of the st-cut includes exactly the sum of the unary potentials associated with the labelling f', i.e. $\sum_{v_a \in \mathbf{v}} \theta^1_{a;f'(a)}$.

Property 2 For $(a,b) \in \mathcal{E}$, if $f'(a) = f_m(a) \notin I_m$ and $f'(b) = f_m(b) \notin I_m$ then the cost of the st-cut includes exactly the pairwise potential $\theta^2_{ab;f'(a)f'(b)}$ plus a constant κ_{ab}. This is due to the fact that the st-cut contains the edge (s, p_{ab}) whose capacity is $\theta^2_{ab;f_m(a)f_m(b)} + \kappa_{ab}$. Note that in this case p_{ab} belongs to the partition containing the sink t. This can be easily verified by observing that the cost of the st-cut would increase if p_{ab} belonged to the partition containing the source s (since this would include edge (p_{ab}, a_{i_m+1}) and (p_{ab}, b_{i_m+1}) in the st-cut).

Property 3 For $(a,b) \in \mathcal{E}$, if $f'(a) \in I_m$ and $f'(b) \in I_m$ such that

$$d(f'(a) - f'(b)) \leq M, \tag{21}$$

then the cost of the st-cut includes exactly the pairwise potential $\theta^2_{ab;f'(a)f'(b)}$ plus a constant κ_{ab}, i.e

$$w_{ab} d(f'(a) - f'(b)) + \kappa_{ab}. \tag{22}$$

This follows from the fact that in this case the pairwise potential lies in the 'convex' part of the truncated convex model. For the convex part, our graph construction is exactly the same as that of [29] which models the pairwise potentials exactly up to the constant κ_{ab} (see Lemma 1 in § 2.1).

Property 4 For $(a,b) \in \mathcal{E}$, if $f'(a) \in I_m$ and $f'(b) \in I_m$ such that

$$d(f'(a) - f'(b)) > M, \tag{23}$$

then the cost of the st-cut overestimates the pairwise potential $\theta^2_{ab;f'(a)f'(b)}$ as

$$w_{ab} d(f'(a) - f'(b)) + \kappa_{ab}. \tag{24}$$

This again follows from the fact that our graph construction boils down to that of [29] where the 'truncation' part of the truncated convex model has been overestimated by the convex function $w_{ab} d(\cdot)$ (see Lemma 1 in § 2.1).

Property 5 For $(a,b) \in \mathcal{E}$, if $f'(a) \in I_m$ and $f'(b) = f_m(b) \notin I_m$ then the cost of the st-cut overestimates the pairwise potential $\theta^2_{ab;f'(a)f'(b)}$ as

$$w_{ab} d(f'(a) - (i_m + 1)) + w_{ab} d'(f'(a) - (i_m + 1)) + w_{ab} M + \kappa_{ab}, \tag{25}$$

where $d'(\cdot)$ denotes the following function:

$$d'(x) = d(x+1) - d(x) - d(1) + \frac{d(0)}{2}, \forall x \geq 0. \tag{26}$$

Note that $d'(\cdot)$ is only defined for a non-negative argument. Clearly, the argument of $d'(\cdot)$ in equation (25) is non-negative since $f'(a) \in [i_m + 1, j_m]$. For example, $d'(x) = 0$ when $d(\cdot)$ is a linear metric and $d'(x) = 2x$ when $d(\cdot)$ is the quadratic semi-metric. Similarly, if $f'(a) = f_m(a) \notin I_m$ and $f'(b) \in I_m$ then the cost of the st-cut overestimates the pairwise potential $\theta^2_{ab;f'(a)f'(b)}$ as

$$w_{ab} d(f'(b) - (i_m + 1)) + w_{ab} d'(f'(b) - (i_m + 1)) + w_{ab} M + \kappa_{ab}. \tag{27}$$

The above property can be shown to be true using the following Lemma.

Lemma 2: For the graph described in § 2.1, property 5 holds true.

Proof: We will show the proof for $f'(a) \in I_m$ and $f'(b) = f_m(b) \notin I_m$. The proof for $f'(a) = f_m(a) \notin I_m$ and $f'(b) \in I_m$ can be obtained from the following arguments trivially.

There are two possible cases to be considered: (i) $f_m(a) \in I_m$; and (ii) $f_m(a) \notin I_m$. In the first case, the edges that specify the st-cut are given by

$$\begin{array}{c} (a_{f'(a)}, a_{f'(a)+1}) \cup \{(a_{i'}, b_{j'}), i_m + 2 \leq i' \leq f'(a), i_m + 1 \leq j' \leq j_m\} \\ \cup\{(a_{i_m+1}, b_{j'}), i_m + 2 \leq j' \leq j_m\} \cup (a_{i_m+1}, b_{i_m+1}). \end{array} \tag{28}$$

In the second case, the st-cut is specified by

$$\begin{array}{c} (a_{f'(a)}, a_{f'(a)+1}) \cup \{(a_{i'}, b_{j'}), i_m + 2 \leq i' \leq f'(a), i_m + 1 \leq j' \leq j_m\} \\ \cup\{(a_{i_m+1}, b_{j'}), i_m + 2 \leq j' \leq j_m\} \cup (p_{ab}, b_{i_m+1}). \end{array} \tag{29}$$

Note that in this case p_{ab} belongs to the same partition as the source s. This can be shown easily by observing that the cost of the st-cut increases if p_{ab} belongs to the partition containing the sink t (since this would include edges (a_{i_m+1}, p_{ab}) and (s, p_{ab}) in the st-cut). The two cases differ only in that the first includes the edge (a_{i_m+1}, b_{i_m+1}) and the second includes the edge (p_{ab}, b_{i_m+1}). However, the capacity of both these edges is equal to $w_{ab}M + \kappa_{ab}/2$. Hence it follows that the cost of the st-cut in both the cases is the same. Therefore it is sufficient to show that the Lemma holds true for the first case.

The cost of the st-cut for the edges in equation (28) is given by

$$\begin{array}{cc} & \frac{w_{ab}}{2}\left(d(L - f'(a) + i_m) + d(f'(a) - i_m)\right) \\ + & \sum_{i'=i_m+2}^{f'(a)} \sum_{j'=i_m+1}^{j_m} \frac{w_{ab}}{2}\left(d(i' - j' + 1) - 2d(i' - j') + d(i' - j' - 1)\right) \\ + & \sum_{j'=i_m+2}^{j_m} \frac{w_{ab}}{2}\left(d(i_m - j' + 2) - 2d(i_m - j' + 1) + d(i_m - j')\right) \\ + & w_{ab}M + \frac{\kappa_{ab}}{2}. \end{array} \tag{30}$$

In order to simplify the above expression, we begin by observing that

$$\begin{array}{cc} & \sum_{j'=i_m+1}^{j_m}\left(d(i' - j' + 1) - 2d(i' - j') + d(i' - j' - 1)\right) \\ = & d(i' - i_m) - d(i' - i_m - 1) - d(i' - j_m) + d(i' - j_m - 1). \end{array} \tag{31}$$

The above equation is obtained by substituting $k' = i_m$ in equation (16). It follows that

$$\begin{array}{cc} & \sum_{i'=i_m+2}^{f'(a)} \sum_{j'=i_m+1}^{j_m} \frac{w_{ab}}{2}\left(d(i' - j' + 1) - 2d(i' - j') + d(i' - j' - 1)\right) \\ = & d(2) - d(1) - d(i_m - j_m + 2) + d(i_m - j_m + 1) \\ + & d(3) - d(2) - d(i_m - j_m + 3) + d(i_m - j_m + 2) \\ & \vdots \\ + & d(f'(a) - i_m - 1) - d(f'(a) - i_m - 2) - d(f'(a) - j_m - 1) + d(f'(a) - j_m - 2) \\ + & d(f'(a) - i_m) - d(f'(a) - i_m - 1) - d(f'(a) - j_m) + d(f'(a) - j_m - 1) \\ = & d(f'(a) - i_m) - d(j_m - f'(a)) - d(1) + d(j_m - i_m - 1) \\ = & d(f'(a) - i_m) - d(L - f'(a) + i_m) - d(1) + d(L - 1), \end{array} \tag{32}$$

where the last expression is obtained using $L = j_m - i_m$. Once again, we use the property $d(x) = d(-x)$. Similarly, by substituting $k' = i_m + 1$ in equation (16), we get

$$\begin{array}{cc} & \sum_{j'=i_m+2}^{j_m} \frac{w_{ab}}{2}\left(d(i_m - j' + 2) - 2d(i_m - j' + 1) + d(i_m - j')\right) \\ = & d(0) - d(1) - d(j_m - i_m - 1) + d(j_m - i_m) \\ = & d(0) - d(1) - d(L - 1) + d(L). \end{array} \tag{33}$$

By simplifying expression (30) using equations (32) and (33), the cost of the st-cut is given by

$$\begin{array}{cc} & \frac{w_{ab}}{2}\left(d(L - f'(a) + i_m) + d(f'(a) - i_m)\right) \\ + & \frac{w_{ab}}{2}\left(d(f'(a) - i_m) - d(L - f'(a) + i_m) - d(1) + d(L - 1)\right) \\ + & \frac{w_{ab}}{2}\left(d(0) - d(1) - d(L - 1) + d(L)\right) \\ + & w_{ab}M + \frac{\kappa_{ab}}{2} \\ = & w_{ab}d(f'(a) - (i_m + 1)) + w_{ab}d'(f'(a) - (i_m + 1)) + w_{ab}M + \kappa_{ab}, \end{array} \tag{34}$$

where the last expression is obtained using the definition of $d'(\cdot)$ in equation (26) and the fact that $\kappa_{ab} = w_{ab}d(L)$. This proves the Lemma. ∎

In summary, property 1 tells us that the cost of the st-cut exactly models the sum of the unary potentials. Properties 2 and 3 specify the cases where the cost of the st-cut exactly models the pairwise potentials, while properties 4 and 5 specify the remaining cases where the cost of the st-cut overestimates the pairwise potentials. In other words, the Gibbs energy of the labelling f', and hence the Gibbs energy of f_{m+1}, is at most equal to the cost of the st-MINCUT on $\mathcal{G}_m$.

Note that our graph construction is similar to that of Gupta and Tardos [9] with two notable exceptions: (i) we can handle any general truncated convex model and not just truncated linear as in the case of [9]. This is achieved in part by using the graph construction of [29] which generalizes Ishikawa's previous work on linear metric [11]; and (ii) we have the freedom to choose the value of L, while [9] fixed this value to M. A logical choice would be to use that value of L which minimizes the worst case multiplicative bound for a particular class of problems. The following analysis obtains such a value of L for both the truncated linear and the truncated quadratic models. Our worst case multiplicative bounds are exactly those achieved by the LP relaxation (see [6]).

3 Analysis of the Algorithm

Before we begin our analysis, we require the following definitions. Let $r \in [0, L-1]$ be a uniformly distributed random integer. Using r we define the following set of intervals

$$\mathcal{S}_r = \{[0,r],[r+1,r+L],[r+L+1,r+2L],\cdots,[.,h-1]\}, \tag{35}$$

where $h = |\mathbf{l}|$ is the total number of labels associated with the MRF. We denote an optimal labelling of the MRF by f^*. Given such a labelling f^* and an interval $I_m = [i_m+1, j_m] \in \mathcal{S}_r$, we define the following five sets:

1. $\mathbf{v}(I_m) \subseteq \mathbf{v}$ such that $v_a \in \mathbf{v}(I_m)$ if, and only if, $f^*(a) \in I_m$.
2. $\mathcal{E}(I_m) \subseteq \mathcal{E}$ such that $(a,b) \in \mathcal{E}(I_m)$ if, and only if, $f^*(a) \in I_m$ and $f^*(b) \in I_m$.
3. $\mathcal{D}_1(I_m) \subseteq \mathcal{E}$ such that $(a,b) \in \mathcal{E}(I_m)$ if, and only if, $f^*(a) \in I_m$ and $f^*(b) \notin I_m$.
4. $\mathcal{D}_2(I_m) \subseteq \mathcal{E}$ such that $(a,b) \in \mathcal{E}(I_m)$ if, and only if, $f^*(a) \notin I_m$ and $f^*(b) \in I_m$.
5. $\mathcal{D}(I_m) = \mathcal{D}_1(I_m) \bigcup \mathcal{D}_2(I_m)$.

In other words, $\mathbf{v}(I_m)$ contains all the random variables which take an optimal labelling in I_m, $\mathcal{E}(I_m)$ contains the set of all edges in the graphical model of the MRF whose endpoints take an optimal labelling in the interval I_m, and $\mathcal{D}(I_m)$ contains edges where only one endpoint takes an optimal labelling in I_m.

Clearly, the following equation holds true:

$$\sum_{v_a \in \mathbf{v}} \theta^1_{a;f^*(a)} = \sum_{I_m \in \mathcal{S}_r} \sum_{v_a \in \mathbf{v}(I_m)} \theta^1_{a;f^*(a)}, \tag{36}$$

since $f^*(a)$ belongs to one and only one interval in $\mathcal{S}_r$ for all $v_a \in \mathbf{v}$. In order to make the analysis less cluttered, we introduce the following shorthand notation for some terms:

1. For $(a,b) \in \mathcal{E}(I_m)$, we denote $w_{ab}d(f^*(a) - f^*(b))$ by e^m_{ab}.
2. For $(a,b) \in \mathcal{D}_1(I_m)$, we denote $w_{ab}d(f^*(a)-(i_m+1))+w_{ab}d'(f^*(a)-(i_m+1))+w_{ab}M$ by e^m_a.
3. For $(a,b) \in \mathcal{D}_2(I_m)$, we denote $w_{ab}d(f^*(b)-(i_m+1))+w_{ab}d'(f^*(b)-(i_m+1))+w_{ab}M$ by e^m_b.

We are now ready to prove our main results, starting with the following Lemma.

Lemma 3: At an iteration of our algorithm, given the current labelling f_m and an interval $I_m = [i_m+1, j_m]$, the new labelling f_{m+1} obtained by solving the st-MINCUT problem reduces the Gibbs energy by at least the following:

$$\sum_{v_a \in \mathbf{v}(I_m)} \theta^1_{a;f_m(a)} + \sum_{(a,b)\in\mathcal{E}(I_m)\bigcup\mathcal{D}(I_m)} \theta^2_{ab;f_m(a)f_m(b)} - \left(\sum_{v_a\in\mathbf{v}(I_m)} \theta^1_{a;f^*(a)} + \sum_{(a,b)\in\mathcal{E}(I_m)} e^m_{ab} + \sum_{(a,b)\in\mathcal{D}_1(I_m)} e^m_a + \sum_{(a,b)\in\mathcal{D}_2(I_m)} e^m_b\right). \tag{37}$$

Proof: From the arguments in § 2.2, it is clear that the Gibbs energy of the new labelling f_{m+1} is bounded from above by the cost of the st-MINCUT. The cost of the st-MINCUT itself is bounded from above by the cost of any other st-cut in the graph $\mathcal{G}_m$. Consider one such st-cut which results in the following labelling:

$$f'(a) = \begin{cases} f^*(a) & \text{if} \quad v_a \in \mathbf{v}(I_m) \\ f_m(a) & \text{otherwise.} \end{cases} \tag{38}$$

We will now derive the cost of this st-cut using the properties in § 2.2. We consider the following six cases:

1. For random variables $v_a \notin \mathbf{v}(I_m)$ it follows from Property 1 that the cost of the st-cut will include the unary potentials associated with such variables exactly, i.e.
$$\sum_{v_a\notin\mathbf{v}(I_m)} \theta^1_{a;f_m(a)}. \tag{39}$$
2. For neighbouring random variables $(a,b) \notin \mathcal{E}(I_m)\bigcup\mathcal{D}(I_m)$ it follows from Property 2 that the cost of the st-cut will include the pairwise potentials associated with such neighbouring variables exactly up to a constant κ_{ab}, i.e.
$$\sum_{(a,b)\notin\mathcal{E}(I_m)\bigcup\mathcal{D}(I_m)} \left(\theta^2_{ab;f_m(a)f_m(b)} + \kappa_{ab}\right). \tag{40}$$
3. For random variables $v_a \in \mathbf{v}(I_m)$, it follows from Property 1 that the cost of the st-cut will include the unary potentials associated with such variables exactly, i.e.
$$\sum_{v_a\in\mathbf{v}(I_m)} \theta^1_{a;f^*(a)}. \tag{41}$$
4. For neighbouring random variables $(a,b) \in \mathcal{E}(I_m)$ it follows from Properties 3 and 4 that the cost of the st-cut will include the following:
$$\sum_{(a,b)\in\mathcal{E}(I_m)} \left(e^m_{ab} + \kappa_{ab}\right). \tag{42}$$
5. For neighbouring random variables $(a,b) \in \mathcal{D}_1(I_m)$ it follows from Property 5 that the cost of the st-cut will include the following:
$$\sum_{(a,b)\in\mathcal{D}_1(I_m)} \left(e^m_{a} + \kappa_{ab}\right). \tag{43}$$
6. For neighbouring random variables $(a,b) \in \mathcal{D}_2(I_m)$ it follows from Property 5 that the cost of the st-cut will include the following:
$$\sum_{(a,b)\in\mathcal{D}_2(I_m)} \left(e^m_{b} + \kappa_{ab}\right). \tag{44}$$

The Gibbs energy of f' (i.e. $Q(f',\mathbf{D};\boldsymbol{\theta})$), and hence $Q(f_{m+1},\mathbf{D};\boldsymbol{\theta})$, is at most the sum of terms (39)-(44) minus $\sum_{(a,b)\in\mathcal{E}} \kappa_{ab}$. It follows that the difference between the Gibbs energy of the current labelling f_m and the new labelling f_{m+1}, i.e $Q(f_m,\mathbf{D};\boldsymbol{\theta}) - Q(f_{m+1},\mathbf{D};\boldsymbol{\theta})$, is at least

$$\sum_{v_a \in \mathbf{v}(I_m)} \theta^1_{a;f_m(a)} + \sum_{(a,b)\in\mathcal{E}(I_m)\bigcup\mathcal{D}(I_m)} \theta^2_{ab;f_m(a)f_m(b)} - \left(\sum_{v_a\in\mathbf{v}(I_m)} \theta^1_{a;f^*(a)} + \sum_{(a,b)\in\mathcal{E}(I_m)} e^m_{ab} + \sum_{(a,b)\in\mathcal{D}_1(I_m)} e^m_a + \sum_{(a,b)\in\mathcal{D}_2(I_m)} e^m_b\right). \tag{45}$$

This proves the Lemma. ∎

Let f be the final labelling obtained using our algorithm. Since f is a local optimum with respect to all intervals I_m, it follows that the term (37) should be non-positive for all I_m (otherwise the Gibbs energy could be further reduced thereby contradicting the fact that f is the local optimum labelling). In other words,

$$\begin{aligned}\sum_{v_a\in\mathbf{v}(I_m)}\theta^1_{a;f(a)}+\sum_{(a,b)\in\mathcal{E}(I_m)\bigcup\mathcal{D}(I_m)}\theta^2_{ab;f(a)f(b)}\\ \leq\left(\sum_{v_a\in\mathbf{v}(I_m)}\theta^1_{a;f^*(a)}+\sum_{(a,b)\in\mathcal{E}(I_m)}e^m_{ab}\right.\\ \left.+\sum_{(a,b)\in\mathcal{D}_1(I_m)}e^m_a+\sum_{(a,b)\in\mathcal{D}_2(I_m)}e^m_b\right),\forall I_m.\end{aligned}\tag{46}$$

We sum the above inequality over all $I_m\in\mathcal{S}_r$. The summation of the LHS is at least $Q(f,\mathbf{D};\boldsymbol{\theta})$. Furthermore, using equation (36), the summation of the above inequality can be written as

$$\begin{aligned}Q(f,\mathbf{D};\boldsymbol{\theta})\leq\sum_{v_a\in\mathbf{v}}\theta^1_{a;f^*(a)}+\\ \sum_{I_m\in\mathcal{S}_r}\left(\sum_{(a,b)\in\mathcal{E}(I_m)}e^m_{ab}+\sum_{(a,b)\in\mathcal{D}_1(I_m)}e^m_a+\sum_{(a,b)\in\mathcal{D}_2(I_m)}e^m_b\right).\end{aligned}\tag{47}$$

We now take the expectation of the above inequality over the uniformly distributed random integer $r\in[0,L-1]$. The LHS of the inequality and the first term on the RHS (i.e. $\sum\theta^1_{a;f^*(a)}$) are constants with respect to r. Hence, we get

$$\begin{aligned}Q(f,\mathbf{D};\boldsymbol{\theta})\leq\sum_{v_a\in\mathbf{v}}\theta^1_{a;f^*(a)}+\\ \frac{1}{L}\sum_r\sum_{I_m\in\mathcal{S}_r}\left(\sum_{(a,b)\in\mathcal{E}(I_m)}e^m_{ab}+\sum_{(a,b)\in\mathcal{D}_1(I_m)}e^m_a+\sum_{(a,b)\in\mathcal{D}_2(I_m)}e^m_b\right).\end{aligned}\tag{48}$$

We conclude by observing that this is the same bound that is obtained by the LP relaxation. Thus, using the analysis of [6] we obtain the following results.

Lemma 4: When $d(\cdot)$ is linear, i.e. $d(x)=|x|$, the following inequality holds true:

$$\begin{aligned}\frac{1}{L}\sum_r\sum_{I_m\in\mathcal{S}_r}\left(\sum_{(a,b)\in\mathcal{E}(I_m)}e^m_{ab}+\sum_{(a,b)\in\mathcal{D}_1(I_m)}e^m_a+\sum_{(a,b)\in\mathcal{D}_2(I_m)}e^m_b\right)\\ \leq\left(2+\max\left\{\frac{2M}{L},\frac{L}{M}\right\}\right)\sum_{(a,b)\in\mathcal{E}}\theta^2_{ab;f^*(a)f^*(b)}.\end{aligned}\tag{49}$$

Proof: The following is a slight modification of the proof of Lemma 4.5 of [6] and is presented here for the sake of completeness. Since we are dealing with the truncated linear metric, the terms e^m_{ab}, e^m_a and e^m_b can be simplified as

$$\begin{aligned}e^m_{ab}&=w_{ab}|f^*(a)-f^*(b)|,\\ e^m_a&=w_{ab}(f^*(a)-i_m-1+M),\\ e^m_b&=w_{ab}(f^*(b)-i_m-1+M).\end{aligned}\tag{50}$$

We begin by observing that the LHS of inequality (49) can be rewritten as

$$\frac{1}{L}\sum_{(a,b)\in\mathcal{E}}\left(\sum_{\mathcal{E}(I_m)\ni(a,b)}e^m_{ab}+\sum_{\mathcal{D}_1(I_m)\ni(a,b)}e^m_a+\sum_{\mathcal{D}_2(I_m)\ni(a,b)}e^m_b.\right)\tag{51}$$

In order to prove the Lemma, we consider the following three cases for two neighbouring random variables $(a,b)\in\mathcal{E}$.

Case I: $d(f^*(a),f^*(b))=|f^*(a)-f^*(b)|\leq L$ and hence, $\theta^2_{ab;f^*(a)f^*(b)}=w_{ab}M$.

In this case, it is clear that $(a,b)\notin\mathcal{E}(I_m)$ for all intervals I_m since the length of each interval is L. Furthermore, the conditions for $(a,b)\in\mathcal{D}_1(I_m)$ and $(a,b)\in\mathcal{D}_2(I_m)$ are given by

$$\begin{aligned}(a,b)\in\mathcal{D}_1(I_m)&\iff i_m\in[f^*(a)-L,f^*(a)-1],\\ (a,b)\in\mathcal{D}_2(I_m)&\iff i_m\in[f^*(b)-L,f^*(b)-1].\end{aligned}\tag{52}$$

In order to prove inequality (49), we observe that

$$\begin{array}{cc} & \sum_{\mathcal{E}(I_m)\ni(a,b)} e_{ab}^m + \sum_{\mathcal{D}_1(I_m)\ni(a,b)} e_a^m + \sum_{\mathcal{D}_2(I_m)\ni(a,b)} e_b^m \\ = & w_{ab}\left(\sum_{i_m=f^*(a)-L}^{f^*(a)-1}(M+f^*(a)-i_m-1) + \sum_{i_m=f^*(b)-L}^{f^*(b)-1}(M+f^*(b)-i_m-1)\right) \\ = & w_{ab}\left(2LM + \sum_{i_m=f^*(a)-L}^{f^*(a)-1}(f^*(a)-i_m-1) + \sum_{i_m=f^*(a)-L}^{f^*(a)-1}(f^*(a)-i_m-1)\right) \\ \leq & w_{ab}\left(2LM+2L^2\right) \\ = & L\left(2+\frac{L}{M}\right)\theta^2_{ab;f^*(a)f^*(b)}, \end{array} \tag{53}$$

where the last expression is obtained using the fact that $\theta^2_{ab;f^*(a)f^*(b)} = w_{ab}M$.

Case II: $M \leq d(f^*(a), f^*(b)) = |f^*(a) - f^*(b)| < L$ and hence, $\theta^2_{ab;f^*(a)f^*(b)} = w_{ab}M$.

We will assume, without loss of generality, that $f^*(a) \leq f^*(b)$. In this case, the conditions for $(a,b) \in \mathcal{E}(I_m)$, $(a,b) \in \mathcal{D}_1(I_m)$ and $(a,b) \in \mathcal{D}_2(I_m)$ are given by

$$\begin{array}{c} (a,b) \in \mathcal{E}(I_m) \iff i_m \in [f^*(b)-L, f^*(a)-1], \\ (a,b) \in \mathcal{D}_1(I_m) \iff i_m \in [f^*(a)-L, f^*(b)-L-1], \\ (a,b) \in \mathcal{D}_2(I_m) \iff i_m \in [f^*(a), f^*(b)-1]. \end{array} \tag{54}$$

Again, in order to prove inequality (49), we observe that

$$\begin{array}{cc} & \sum_{\mathcal{E}(I_m)\ni(a,b)} e_{ab}^m + \sum_{\mathcal{D}_1(I_m)\ni(a,b)} e_a^m + \sum_{\mathcal{D}_2(I_m)\ni(a,b)} e_b^m \\ = & w_{ab}\left(\sum_{i_m=f^*(b)-L}^{f^*(a)-1}(f^*(b)-f^*(a)) + \sum_{i_m=f^*(a)-L}^{f^*(b)-L-1}(M+f^*(a)-i_m-1)\right. \\ + & \left.\sum_{i_m=f^*(a)}^{f^*(b)-1}(M+f^*(b)-i_m-1)\right) \\ \leq & w_{ab}\left(2L+2M-(f^*(b)-f^*(a))\right)(f^*(b)-f^*(a)) \\ \leq & w_{ab}L(2M+L) \\ = & L\left(2+\frac{L}{M}\right)\theta^2_{ab;f^*(a)f^*(b)}, \end{array} \tag{55}$$

where the last expression is obtained using the fact that $\theta^2_{ab;f^*(a)f^*(b)} = w_{ab}M$.

Case III: $d(f^*(a), f^*(b)) = |f^*(a) - f^*(b)| \leq M$ and hence, $\theta^2_{ab;f^*(a)f^*(b)} = w_{ab}|f^*(a) - f^*(b)|$.

We will assume, without loss of generality, that $f^*(a) \leq f^*(b)$. Similar to case II, the conditions for $(a,b) \in \mathcal{E}(I_m)$, $(a,b) \in \mathcal{D}_1(I_m)$ and $(a,b) \in \mathcal{D}_2(I_m)$ are given by

$$\begin{array}{c} (a,b) \in \mathcal{E}(I_m) \iff i_m \in [f^*(b)-L, f^*(a)-1], \\ (a,b) \in \mathcal{D}_1(I_m) \iff i_m \in [f^*(a)-L, f^*(b)-L-1], \\ (a,b) \in \mathcal{D}_2(I_m) \iff i_m \in [f^*(a), f^*(b)-1]. \end{array} \tag{56}$$

Once again, we consider

$$\begin{array}{cc} & \sum_{\mathcal{E}(I_m)\ni(a,b)} e_{ab}^m + \sum_{\mathcal{D}_1(I_m)\ni(a,b)} e_a^m + \sum_{\mathcal{D}_2(I_m)\ni(a,b)} e_b^m \\ = & w_{ab}\left(\sum_{i_m=f^*(b)-L}^{f^*(a)-1}(f^*(b)-f^*(a)) + \sum_{i_m=f^*(a)-L}^{f^*(b)-L-1}(M+f^*(a)-i_m-1)\right. \\ + & \left.\sum_{i_m=f^*(a)}^{f^*(b)-1}(M+f^*(b)-i_m-1)\right) \\ \leq & w_{ab}\left(2L+2M-(f^*(b)-f^*(a))\right)(f^*(b)-f^*(a)) \\ \leq & w_{ab}(2L+2M)(f^*(b)-f^*(a)) \\ = & L\left(2+\frac{2M}{L}\right)\theta^2_{ab;f^*(a)f^*(b)}, \end{array} \tag{57}$$

where the last expression is obtained using the fact that $\theta^2_{ab;f^*(a)f^*(b)} = w_{ab}(f^*(b) - f^*(a))$.

Substituting inequalities (53), (55) and (57) in expression (51) and dividing both sides by L for all $(a,b) \in \mathcal{E}$, we obtain inequality (49). This proves the Lemma. ∎

Theorem 1: For the truncated linear metric, our algorithm obtains a multiplicative bound of $2 + \sqrt{2}$ using $L = \sqrt{2}M$.

The proof of the above theorem follows by substituting $L = \sqrt{2}M$ in inequality (49) and simplifying inequality (48). Note that this bound is better than those obtained by α-expansion [5] (i.e. $2M$) and its generalization [9] (i.e. 4). In fact, the bound of [9] can be obtained directly from the above analysis by using the non-optimal assignment of $L = M$.

Similarly, using Theorem 4 of [6], we obtain the following multiplicative bound for the truncated quadratic semi-metric.

Theorem 2: For the truncated quadratic semi-metric, our algorithm obtains a multiplicative bound of $O(\sqrt{M})$ using $L = \sqrt{M}$.

Note that both α-expansion and the approach of Gupta and Tardos provide no bounds for the above case. The primal-dual method of [18] obtains a bound of $2M$ which is clearly inferior to our guarantees.

4 Experiments

We tested our approach using both synthetic and standard real data. Below, we describe the experimental setup and the results obtained in detail.

4.1 Synthetic Data

Experimental Setup We used 100 random fields for both the truncated linear and truncated quadratic models. The variables $\mathbf{v}$ and neighbourhood relationship $\mathcal{E}$ of the random fields described a 4-connected grid graph of size 50×50. Note that 4-connected grid graphs are widely used to model several problems in Computer Vision [28]. Each variable was allowed to take one of 20 possible labels, i.e. $\mathbf{l} = \{l_0, l_1, \cdots, l_{19}\}$. The parameters of the random field were generated randomly. Specifically, the unary potentials $\theta^1_{a;i}$ were sampled uniformly from the interval $[0, 10]$ while the weights w_{ab}, which determine the pairwise potentials, were sampled uniformly from $[0, 5]$. The parameter M was also chosen randomly while taking care that $d(5) \leq M \leq d(10)$.

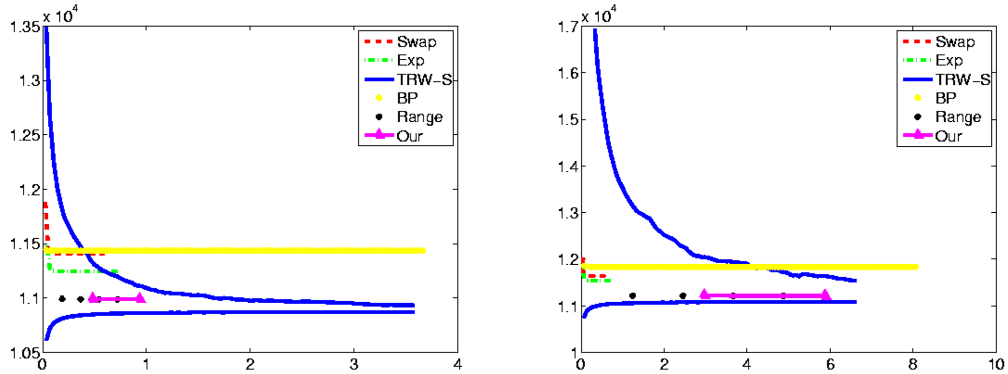
Results Fig. 3 shows the results obtained by our approach and five other state of the art algorithms: $\alpha\beta$-swap, α-expansion, BP, TRW-S and the range move algorithm of [29]. We used publicly available code for all previously proposed approaches with the exception of the range move algorithm[5]. As can be seen from the figure, the most accurate approach is the method proposed in this paper, followed closely by the range move algorithm. Recall that, unlike range move, our algorithm is guaranteed to provide the same worst case multiplicative bounds as the LP relaxation. As expected, both the range move algorithm and our method are slower than $\alpha\beta$-swap and α-expansion (since each iteration computes an st-MINCUT on a larger graph). However, they are faster than TRW-S, which attempts to minimize the LP relaxation, and BP. We note here that our implementation does not use any clever tricks to speed up the max-flow algorithm (such as those described in [1]) which can potentially decrease the running time by orders of magnitude.

4.2 Real Data - Stereo Reconstruction

Given two *epipolar rectified* images $\mathbf{D}_1$ and $\mathbf{D}_2$ of the same scene, the problem of stereo reconstruction is to obtain a correspondence between the pixels of the images. This problem can be modelled using a random field whose variables correspond to pixels of one image (say $\mathbf{D}_1$) and take labels from a set of *disparities* $\mathbf{l} = \{0, 1, \cdots, h-1\}$. A disparity value i for a random variable a denoting pixel (x, y) in $\mathbf{D}_1$ indicates that its corresponding pixel lies in location $(x + i, y)$ in the second image.

For the above random field formulation, the unary potentials were defined as in [3] and were truncated at 15. As is typically the case, we chose the neighbourhood relationship $\mathcal{E}$ to define a 4-

[5]When using α-expansion with the truncated quadratic semi-metric, all edges with negative capacities in the graph construction were removed, similar to the experiments in [28].



(a) (b)

Figure 3: *Results of the synthetic experiment.* **(a)** *Truncated linear metric.* **(b)** *Truncated quadratic semi-metric. The x-axis shows the time taken in seconds. The y-axis shows the average Gibbs energy obtained over all 100 random fields using the six algorithms. The lower blue curve is the value of the dual obtained by* TRW-S. *In both the cases, our method and the range move algorithm provide the most accurate solution and are faster than* TRW-S *and* BP.

neighbourhood grid graph. The number of disparities h was set to 20. We experimented using the following truncated convex potentials:

$$\theta^2_{ab;ij} = 50\min\{|i-j|, 10\},$$
$$\theta^2_{ab;ij} = 50\min\{(i-j)^2, 100\}. \tag{58}$$

The above form of pairwise potentials encourage neighbouring pixels to take similar disparity values which corresponds to our expectations of finding smooth surfaces in natural images. Truncation of pairwise potentials is essential to avoid oversmoothing, as observed in [5, 29]. Note that using spatially varying weights w_{ab} provides better results. However, the main aim of this experiment is to demonstrate the accuracy and speed of our approach and not to design the best possible Gibbs energy. Fig. 4 shows the results obtained using various algorithms when using the truncated linear metric on a standard stereo pair (Tsukuba). Table 2 provides the value of the Gibbs energy and the total time taken by all the approaches for three stereo pairs. Similar to the synthetic experiments, the range move algorithm and our method provide the most accurate solutions while taking less time than TRW-S and BP. Our method does marginally better than range move. However, we would again like to emphasize that unlike our method the range move algorithm provides no theoretical guarantees about the quality of the solution.

5 Discussion

We have presented an st-MINCUT based algorithm for obtaining the approximate MAP estimate of discrete random fields with truncated convex pairwise potentials. Our method improves the multiplicative bound for the truncated linear metric compared to [5, 9] and provides the best known bound for the truncated quadratic semi-metric. Due to the use of only the st-MINCUT problem in its design, it is faster than previous approaches based on the LP relaxation. In fact, its speed can be further improved by a large factor using clever techniques such as those described in [16] (for convex unary potentials) and/or [1] (for general unary potentials). Furthermore, it overcomes the well-known deficiencies of TRW and its variants. Experiments on synthetic and real data problems demonstrate its effectiveness compared to several state of the art algorithms.

Our method can easily be extended to handle truncated submodular models by using the graph construction of [24] instead of [29]. However, the resulting multiplicative bounds would start to depend on the value of the pairwise potentials thereby making the analysis cluttered. For this reason, we have restricted our discussion to the truncated convex model.

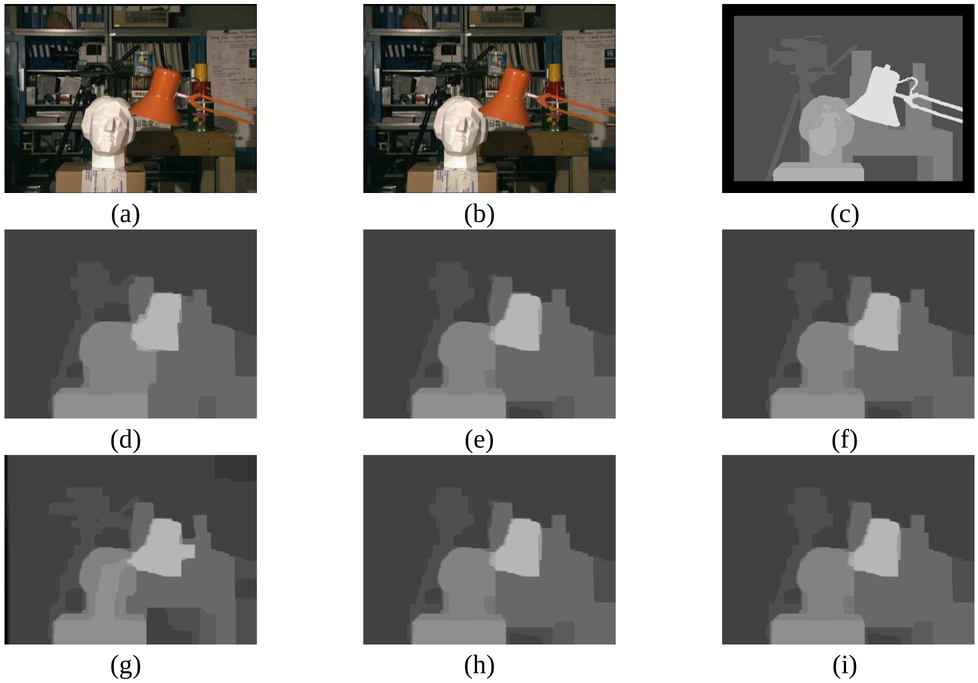


Figure 4: *Tsukuba stereo pair.* **(a)** *First image.* **(b)** *Second image.* **(c)** *Ground truth disparity map.* **(d)-(i)** *Results obtained using various algorithms: in the above order, $\alpha\beta$-swap algorithm, α-expansion,* TRW-S, BP, *the range move algorithm of [29] and our approach.*

A slight modification of Theorem 3.7 of [9] also proves that if our algorithm is run for $O(h/L)(\log Q(f_1, \mathbf{D}; \boldsymbol{\theta}) + \log \epsilon^{-1})$ iterations (where f_1 is the initial labelling, and $\epsilon > 0$), then the expected value of the Gibbs energy would be at most $(2 + \sqrt{2} + \epsilon)Q(f^*, \mathbf{D}; \boldsymbol{\theta})$ for the truncated linear metric and $(O(\sqrt{M}) + \epsilon)Q(f^*, \mathbf{D}; \boldsymbol{\theta})$ for the truncated quadratic semi-metric (where f^* is an optimal labelling). In other words, our method provides the same guarantees as the LP relaxation in polynomial time. Although theoretically interesting, the practical implications of this result are minimal, since in most scenarios we will be able to run our algorithm for a sufficient number of iterations so as to end up in the local minimum over all intervals I_m. In fact, in all our experiments we reached the local minimum in less than 5 iterations.

The analysis in section 3 shows that, for the truncated linear and truncated quadratic models, the bound achieved by our move making algorithm over intervals of any length L is equal to that of rounding the LP relaxation's optimal solution using the same intervals [6]. This equivalence also extends to the Potts model (in which case α-expansion provides the same bound as the LP relaxation when using the rounding scheme of [12]). A natural question would be to ask about the relationship between move making algorithms and the rounding schemes used in convex relaxations. Note that despite recent efforts [18] which analyze certain move making algorithms in the context of primal-dual approaches for the LP relaxation, not many results are known about their connection with randomize rounding schemes. Although the discussion in section 3 cannot be trivially generalized to all random fields, it offers a first step towards answering this question. We believe that further exploration in this direction would help improve the understanding of the nature of the MAP estimation problem, e.g. how to derandomize approaches based on convex relaxations. Furthermore, it would also help design efficient move making algorithms for more complex relaxations such as those described in [20].

Acknowledgments The first author was supported by the EU CLASS project and EPSRC grant EP/C006631/1(P). The second author is in receipt of a Royal Society Wolfson Research Merit Award, and would like to acknowledge support from the Royal Society and Wolfson foundation.

Algorithm	Energy-1	Time-1(s)	Energy-2	Time-2(s)
$\alpha\beta$-swap	645227	28.86	709120	20.04
α-expansion	634931	9.52	723360	9.78
TRW-S	**634720**	94.86	**651696**	226.07
BP	662108	170.67	2155759	244.71
Range Move	**634720**	39.75	**651696**	80.40
Our Approach	**634720**	66.13	**651696**	80.70

(a)

Algorithm	Energy-1	Time-1(s)	Energy-2	Time-2(s)
$\alpha\beta$-swap	1056109	35.00	1198029	52.98
α-expansion	1052860	15.16	1320088	11.95
TRW-S	1053341	142.19	1057371	339.02
BP	1117782	180.65	2443796	368.14
Range Move	**1052762**	100.49	**1057041**	168.28
Our Approach	**1052762**	129.30	**1057041**	155.98

(b)

Algorithm	Energy-1	Time-1(s)	Energy-2	Time-2(s)
$\alpha\beta$-swap	3678200	18.48	3707268	20.25
α-expansion	3677950	11.73	3687874	8.79
TRW-S	3677578	131.65	3679563	332.94
BP	3789486	272.06	5180705	331.36
Range Move	3686844	97.23	**3679552**	141.78
Our Approach	**3613003**	120.14	**3679552**	191.20

(c)

Table 2: *The energy obtained and the time taken by the algorithms used in the stereo reconstruction experiment. Columns 2 and 3 : truncated linear metric. Columns 4 and 5: truncated quadratic semi-metric.* **(a)** *Tsukuba.* **(b)** *Venus.* **(c)** *Teddy. The lowest energy obtained in each case is indicated using bold font.*

We thank the reviewers for careful reading of the submitted manuscript and for helpful comments which improved the clarity of the paper.

References

[1] K. Alahari, P. Kohli, and P. H. S. Torr. Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In *CVPR*, 2008.

[2] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.

[3] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. In *ICCV*, pages 1073–1080, 1998.

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

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

[6] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. A linear programming formulation and approximation algorithms for the metric labelling problem. *SIAM Journal on Discrete Mathematics*, 18(3):606–635, 2005.

[7] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient belief propagation for early vision. In *CVPR*, pages I: 261–268, 2004.

[8] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing for MAP LP-relaxations. In *NIPS*, 2007.

[9] A. Gupta and E. Tardos. A constant factor approximation algorithm for a class of classification problems. In *STOC*, 2000.

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

[11] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *CVPR*, pages 125–131, 1998.

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

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

[14] V. Kolmogorov and Y. Boykov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26(9):1124–1137, 2004.

[15] V. Kolmogorov and C. Rother. Comparison of energy minimization algorithms for highly connected graphs. In *ECCV*, pages II: 1–15, 2006.

[16] V. Kolmogorov and A. Shioura. New algorithms for the dual of the convex cost network flow problem with applications to computer vision. Technical report, University College London, 2007.

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

[18] N. Komodakis and G. Tziritas. Approximate labeling via graph-cuts based on linear programming. *PAMI*, 2007.

[19] 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.

[20] M. P. Kumar, V. Kolmogorov, and P. H. S. Torr. An analysis of convex relaxations for MAP estimation. In *NIPS*, 2007.

[21] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labelling sequence data. In *ICML*, 2001.

[22] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman, 1988.

[23] P. Ravikumar, A. Agarwal, and M. Wainwright. Message-passing for graph-structured linear programs: Proximal projections, convergence and rounding schemes. In *ICML*, 2008.

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

[27] 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.

[28] 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 with smoothness-based priors. *PAMI*, 2008.

[29] O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *CVPR*, 2007.

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

[31] Y. Weiss, C. Yanover, and T. Meltzer. MAP estimation, linear programming and belief propagation with convex free energies. In *UAI*, 2007.