On Partial Optimality in Multi-label MRFs

Pushmeet Kohli PKOHLI@MICROSOFT.COM
Alexander Shekhovtsov SHEKHOVT@CMP.FELK.CVUT.CZ
Carsten Rother CARROT@MICROSOFT.COM
Vladimir Kolmogorov VNK@ADASTRAL.UCL.AC.UK
Philip Torr PHILIPTORR@BROOKES.AC.UK

Abstract

We consider the problem of optimizing multi-label MRFs, which is in general NP-hard and ubiquitous in low-level computer vision. One approach for its solution is to formulate it as an integer linear programming and relax the integrality constraints. The approach we consider in this paper is to first convert the multi-label MRF into an equivalent binary-label MRF and then to relax it. The resulting relaxation can be efficiently solved using a maximum flow algorithm. Its solution provides us with a partially optimal labelling of the binary variables. This partial labelling is then easily transferred to the multi-label problem. We study the theoretical properties of the new relaxation and compare it with the standard one. Specifically, we compare tightness, and characterize a subclass of problems where the two relaxations coincide. We propose several combined algorithms based on the technique and demonstrate their performance on challenging computer vision problems.

1. Introduction

One of the major advances in computer vision in the past few years has been the development of efficient deterministic algorithms for solving discrete labeling problems. Labeling problems occur in many places from dense stereo and image segmentation (Boykov et al., 2001; Szeliski et al., 2006) to the use of pictorial structures for object recognition (Felzenszwalb & Huttenlocher, 2000). They can be shown to be equivalent to the problem of estimating the maximum a posterior (MAP) solution in graphical models such as Markov Random Fields (MRF) and Conditional Random Fields (CRF), which is also often referred to as *energy minimization*.

A number of powerful algorithms are present in the literature which deal with the problem. For certain subclasses of the problem, it is possible to compute the exact solution in polynomial time: MRFs of bounded tree-width, *e.g.* (Lauritzen, 1998); with convex pairwise potentials (Ishikawa, 2003); with submodular potentials of binary (Hammer, 1965; Kolmogorov & Zabih, 2004) or multi-label (Schlesinger & Flach, 2000; Kovtun, 2004) variables; with permuted submodular potentials (Schlesinger, 2007). However, the problem of minimizing a general energy function is NP-hard. Nevertheless it is just these sorts of general energies that occur in many vision problems and making progress towards their solution is of paramount importance.

Energy Minimization as a Linear Program General discrete energy minimization can be seen as an integer programming problem. Dropping integrality constraints leads to an attractive linear programming relaxation (LP-1). Unfortunately, linear programs arising from this scheme in vision applications are of very large scale, and are not practical to solve with general methods. A number of algorithms have been developed (Schlesinger, 1976; Koval & Schlesinger, 1976; Wainwright et al., 2003; Kolmogorov, 2006; Werner, 2007) which attempt to solve this relaxation by exploiting the special structure of the problem. However, their common drawback is that they may converge to a suboptimal point. Other developed methods for energy minimization include local search algorithms (Boykov et al., 2001), primal-dual method (Komodakis & Tziritas, 2005), subgradient methods (Schlesinger & Giginyak, 2007; Komodakis et al., 2007). Some local search algorithms provide approximation ratio guarantee for (semi-)metric energies (Boykov et al., 2001; Komodakis & Tziritas, 2005). Finally, there are methods which output a partial assignment of labels with guaranteed optimality certificate for binary (Hammer et al., 1984;

Appearing in *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, 2008. Copyright 2008 by the author(s)/owner(s).

Boros et al., 1991; Boros et al., 2006; Rother et al., 2007) and multi-label (Kovtun, 2003; Kovtun, 2004) problems. Dead-end elimination is another related method for identifying non-optimal assignments based on local sufficient conditions, it was originally proposed (Desmet et al., 1992) for predicting and designing the structures of proteins. In this paper we develop a novel method for obtaining partial optimal solutions of functions of multi-label variables.

Our method works by first transforming the multi-label energy function to a function involving binary variables (Schlesinger & Flach, 2006). This binary energy is then minimized by applying the *roof dual* relaxation (Boros & Hammer, 2002; Hammer et al., 1984), which can be solved efficiently using a single graph cut. More importantly, solving the roof dual relaxation results in an assignment of a subset of the variables which is guaranteed to be valid for any optimal solution. This partially optimal solution can be used to constrain the state space of the original multi-label problem. As we show, this approach may be viewed as a different LP-relaxation of the multi-label energy minimization problem (LP-2).

Comparing LP-1 with LP-2 We present a number of theoretical results studying properties of LP-2 and relating LP-2 with LP-1. Our first result is that LP-1 is a tighter relaxation of the energy minimization problem compared to LP-2. We show in the paper that solutions of LP-1 necessarily satisfy the constraints derived by LP-2, so additional guarantees for methods based on LP-1 may follow. We also identify a subclass of problems for which LP-2 is as tight as LP-1. Thus, for problems of this subclass LP-1 can be solved exactly and efficiently using combinatorial methods.

It was recently demonstrated that the roof dual relaxation performs well for a number of computer vision applications (Kolmogorov & Rother, 2007; Rother et al., 2007) which are naturally formulated as a binary energy minimization. However, it turns out that when multi-label problems are formulated as binary energy minimization the roof dual relaxation leaves many variables unassigned. We therefore use recently proposed enhancements of the roof dual technique called "probing" (Boros et al., 2006; Rother et al., 2007) which can often help resolve these ambiguities. Our last contribution is providing an alternative formulation of LP-2: we prove that it is equivalent to computing a decomposition of the energy into submodular and supermodular parts so that the sum of the lower bounds for each part is maximized. Precise details of this theoretical result are given in (Shekhovtsov et al., 2007, section 10).

Although this is primarily a theoretical paper, we have performed a number of experiments with various energy models arising in vision applications, including image restoration, object-based segmentation, and image stitching. Our experiments show that the proposed method outperforms the competing method of (Kovtun, 2003) by labelling many more random variables. We also demonstrate that it may help solving difficult problems by reducing their state space and applying other methods to the reduced problem.

2. Energy Minimization

Let $\mathcal{L} = \{1 \dots K\}$ be a set of labels. Let $G = (\mathcal{V}, \mathcal{E})$ be a graph, where the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is antisymmetric and antireflexive, *i.e.* $(s,t) \in \mathcal{E} \Rightarrow (t,s) \notin \mathcal{E}$. We denote an ordered pair $(s,t) \in \mathcal{E}$ simply by st. Let each graph node $s \in \mathcal{V}$ be assigned a label $x_s \in \mathcal{L}$ and let a *labeling* (or *configuration*) be defined as $\mathrm{x} = \{x_s \,|\, s \in \mathcal{V}\}$[1]. Let $\{\theta_s(i) \in \mathbb{R} \,|\, i \in \mathcal{L},\, s \in \mathcal{V}\}$ be *univariate* potentials and $\{\theta_{st}(i,j) \in \mathbb{R} \,|\, i,j \in \mathcal{L},\, st \in \mathcal{E}\}$ be *pairwise* potentials of a random field. Let in addition θ_{const} be a *constant* term, and let a concatenated vector of potentials, including the constant term, be denoted as $\theta = (\theta, \theta_{\text{const}}) \in \Omega = \mathbb{R}^{\mathcal{I} \cup \{\text{const}\}}$, where set of indices

$$\mathcal{I} = \{(s,i) \,|\, s \in \mathcal{V},\, i \in \mathcal{L}\} \cup \{(st,ij) \,|\; st \in \mathcal{E},\, i,j \in \mathcal{L}\} \tag{1}$$

corresponds to all univariate and pairwise terms. Notation $\theta_{\mathcal{I}}$ will thus refer to all components of θ but the constant term. The *energy* of a configuration x of the random field is defined as:

$$E(\mathrm{x}|\theta) = \sum_{s \in \mathcal{V}} \theta_s(x_s) + \sum_{st \in \mathcal{E}} \theta_{st}(x_s, x_t) + \theta_{\text{const}}. \tag{2}$$

2.1. LP-relaxation

We will study two different relaxations of minimization of (2). Both relaxations can be written using the same formulation but will differ in the graph, number of labels and potential vector involved.

Energy function (2) can be conveniently written using a scalar product in Ω as $E(\mathrm{x}|\theta) = \langle \mu(\mathrm{x}), \theta \rangle$, where $\mu(\mathrm{x}) \in \{0,1\}^{\mathcal{I} \cup \{\text{const}\}}$ is defined by $\mu(\mathrm{x})_s(i) = \delta_{\{x_s = i\}}$, $\mu(\mathrm{x})_{st}(i,j) = \delta_{\{x_s = i\}} \delta_{\{x_t = j\}}$ and $\mu(\mathrm{x})_{\text{const}} = 1$. In this notation minimization of E is expressed as:

$$\min_{\mathrm{x} \in \mathcal{L}^{\mathcal{V}}} \langle \mu(\mathrm{x}), \theta \rangle. \tag{3}$$

The LP-relaxation of (3) reviewed in, *e.g.*, (Werner, 2007) is constructed as follows. For each variable x_s, a set of relaxed variables $\mu_s(i) \in [0,1]$, $i \in \mathcal{L}$ is introduced. These variables are required to satisfy the

[1] Notation $\{x_s | s \in S\}$ (bold brackets), where S is a finite set, will stand for the concatenated vector of variables x_s, rather than the set of their values.

normalization constraints

$$\sum_{i\in\mathcal{L}} \mu_s(i) = 1, \quad \forall s \in \mathcal{V}. \tag{4}$$

Further, for each pair (x_s, x_t), $st \in \mathcal{E}$ the relaxed variables $\mu_{st}(i,j) \in [0,1]$, $i,j \in \mathcal{L}$ are introduced which must satisfy the *marginalization* constraints:

$$\begin{array}{l} \sum\limits_{j'\in\mathcal{L}} \mu_{st}(i,j') = \mu_s(i), \quad \forall st \in \mathcal{E}, \forall i \in \mathcal{L}, \\ \sum\limits_{i'\in\mathcal{L}} \mu_{st}(i',j) = \mu_t(j), \quad \forall st \in \mathcal{E}, \forall j \in \mathcal{L}. \end{array} \tag{5}$$

The concatenated vector $\mu \in \Omega$ satisfying these constraints will be called a *relaxed* labeling. Indeed, a vector μ with all integral components uniquely represents a labeling x. When the integrality constraints are dropped we get the following relaxation of (3):

$$\min_{\mu\in\Lambda_{G,\mathcal{L}}} \langle \mu, \theta \rangle, \tag{6}$$

where $\Lambda_{G,\mathcal{L}} = \left\{ \mu \in \Omega_+ \,|\, A\mu_{\mathcal{I}} = \mathbf{0},\ B\mu_{\mathcal{I}} = \mathbf{1},\ \mu_{\text{const}} = 1 \right\}$ is called the *local* (Wainwright et al., 2003) polytope of graph G. Here set Ω_+ denotes vectors from Ω with all nonnegative components, equalities $A\mu_{\mathcal{I}} = \mathbf{0}$ express marginalization constraints (5), and equalities $B\mu_{\mathcal{I}} = \mathbf{1}$ express normalization constraints (4).

The dual of the linear program (6) can be written in the form (Shekhovtsov et al., 2007):

$$\max \theta'_{\text{const}} \quad \text{s.t.} \begin{cases} \theta' \equiv \theta \\ \theta'_{\mathcal{I}} \geq 0 \\ \theta'_{\text{const}} \in \mathbb{R}, \end{cases} \tag{LB}$$

where $\theta' \equiv \theta$ denotes that θ' is *equivalent* to (is a *reparametrization* of) θ i.e. $E(\mathrm{x}|\theta') = E(\mathrm{x}|\theta)$ for all $\mathrm{x} \in \mathcal{L}^{\mathcal{V}}$, see (Werner, 2007; Wainwright et al., 2003).

It is easy to see that θ'_{const} is a *lower bound* (denoted by *LB*) on $E(\cdot|\theta)$. This is because $\langle \mu, \theta' \rangle \geq 0, \forall \mu \in \Lambda$, and thus $E(\mathrm{x}|\theta) = \langle \mu(x), \theta \rangle + \theta_{\text{const}} = \langle \mu(x), \theta' \rangle + \theta'_{\text{const}} \geq \theta'_{\text{const}}$. There is a number of equivalent formulations of this lower bound (Wainwright et al., 2003; Werner, 2007). Our expression *LB* is in the form of maximizing the constant term of a *posiform* (Boros & Hammer, 2002). It can be easily seen that in the case $|\mathcal{L}| = 2$ it coincides with the *roof-dual* bound (Boros & Hammer, 2002).

2.2. Binary Energy Minimization

Energy minimization problems with 2 labels ($|\mathcal{L}| = 2$) are conveniently described in terms of *binary* (or Boolean) variables, *i.e.* with set of labels being $\mathbb{B} = \{0,1\}$. For clarity we will denote binary configurations by z. Univariate and pairwise terms of (2) can be written as:

$$\begin{array}{l} \theta_s(z_s) = \theta_s(1)z_s + \theta_s(0)(1-z_s), \\ \theta_{st}(z_s,z_t) = \theta_{st}(1,1)z_sz_t + \theta_{st}(0,1)(1-z_s)z_t \\ +\theta_{st}(1,0)z_s(1-z_t) + \theta_{st}(0,0)(1-z_s)(1-z_t). \end{array} \tag{7}$$

Expanding brackets in (7) it is clear that (2) can be transformed to the form:

$$E(\mathrm{z}|\eta) = \sum_s \eta_s z_s + \sum_{st} \eta_{st} z_s z_t + \eta_{\text{const}}, \tag{8}$$

which is a quadratic function of binary variables. Functions of the form $\mathbb{B}^{\mathcal{V}} \mapsto \mathbb{R}$ are called *pseudo-Boolean* (Boros & Hammer, 2002) and minimization (or maximization) of (8) is called quadratic Pseudo-boolean optimization.

Calculating coefficients η from θ is equivalent to choosing the reparametrization $\hat{\theta} \equiv \theta$ with non-zero elements being only $\hat{\theta}_s(1)$, $\hat{\theta}_{st}(1,1)$ and $\hat{\theta}_{\text{const}}$.

It is easy to see that pseudo-Boolean optimization, energy minimization with 2 labels and the MIN-CUT problem are all equivalent problems, and can be simply converted one into another. It is also known that polynomially solvable MIN-CUT problems (those having weights of all edges in the graph non-negative) correspond to quadratic pseudo-Boolean problems with all weights η_{st} being non-positive, which is equivalent to the condition of submodularity of $E(\cdot|\eta)$.

2.3. LP-relaxation of Binary Problems

As shown in (Hammer et al., 1984), the LP relaxation (6) for the case of binary variables has special properties, which in general do not hold in the multi-label case. First, there exists a *half-integral* optimal relaxed labeling μ^*, i.e. all components are 0, $\frac{1}{2}$ or 1. Second, if μ^* is integral for some node s (i.e. $\mu^*_s(\alpha) = 0$), then there exists a global minimum z of the original function in which $z_s = \alpha$. In other words, by solving the LP relaxation we can obtain constraints on the global minima of the binary energy. These constraints can be expressed as

$$\mathrm{z}^{\min} \leq \mathrm{z} \leq \mathrm{z}^{\max}, \tag{9}$$

where $\mathrm{z}^{\min}, \mathrm{z}^{\max} \in \mathbb{B}^{\mathcal{V}}$ and inequalities are componentwise. For instance, $0 \leq z_s \leq 1$ implies no constraints on z_s, while $0 \leq z_s \leq 0$ implies that z_s is constrained to be 0.

If (9) holds for all optimal solutions z, then the pair $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ is said to define *strong* persistency; if (9) holds for some optimal solution z then $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ defines *weak* persistency (Boros & Hammer, 2002). It is important to note that the LP relaxation can be solved very efficiently by computing minimum cut/maximum flow in an appropriately constructed graph. The technique in (Boros et al., 1991) is perhaps the most efficient; we will refer to it as the Qudratic Pseudo-Boolean Optimization (QPBO) method[2]. It yields a pair $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ defining strong persistency.

[2]Following (Rother et al., 2007) we use the abbreviation QPBO to denote the max-flow algorithm for computing the

Recently, two techniques were introduced which extend the QPBO method. The first one is *Probing* (Boros et al., 2006), or *QPBO-P*. It fixes a certain node s to a particular label α, runs QPBO thus obtaining some information about global minimizers of the energy. This information is incorporated into the energy (e.g. if we learn that $z_s = z_t$ for all optimal solutions then we contract nodes s and t), and the "probing" is performed for other nodes until convergence. An efficient implementation of QPBO-P was described in (Rother et al., 2007). The second method is *Improve*, or *QPBO-I* (Rother et al., 2007). It takes an input labeling z and tries to improve its energy by fixing a subset of nodes to labels specified by z and running QPBO. These operations guarantee not to increase the energy, and in practice do often decrease it.

3. Solving Multi-label Problems

We now address the problem of minimizing energy functions involving multi-label variables.

3.1. Converting Multi-label Problems Into Binary Ones

As was discussed in Sec. 2.2, there are simple transitions between energy minimization with two labels, MIN-CUT problem and the quadratic pseudo-Boolean optimization. Thus, it is not essentially important to which of those forms a multi-label problem will be reduced. The construction (Ishikawa, 2003; Kovtun, 2004; Schlesinger & Flach, 2006) adopted to our notation of binary energies is as follows.

We start the transformation procedure by obtaining a reparametrization $\hat{\theta} \equiv \theta$ which satisfies the following:

$$\begin{aligned} \hat{\theta}_{st}(1,j) = \hat{\theta}_{st}(i,1) &= 0 \quad st \in \mathcal{E},\ i,j \in \mathcal{L} \\ \hat{\theta}_{s}(1) &= 0 \quad s \in \mathcal{V}. \end{aligned} \tag{10}$$

This reparametrization is easy to construct. For example, to achieve $\hat{\theta}_{st}(i,1) = 0$ one needs to subtract the value $\gamma_c(i) = \theta_{st}(i,1)$ from $\theta_{st}(i,j)$ and add it to $\theta_s(i)$, which does not change the energy, and so on, see (Shekhovtsov et al., 2007) for details. Note that in the case of two labels, this reparametrization $\hat{\theta}$ immediately provides coefficients for writing a binary energy in the form (8).

Let the tuple $(\mathcal{L},\ G,\ \theta)$ define a multi-label energy minimization problem. Let $\tilde{\mathcal{L}}$ to refer to the set of the last $K-1$ labels of $\mathcal{L}$, i.e. $\tilde{\mathcal{L}} = \{2 \dots K\}$. The components of the equivalent binary energy minimization problem are given as follows (Fig 1):

- Graph $N = (V, A)$, where $V = \mathcal{V} \times \tilde{\mathcal{L}}$ and $A = A_{\mathcal{E}} \cup A_{\mathcal{V}}$. $A_{\mathcal{E}} = \big\{((s,i),(t,j)) \,|\, st \in \mathcal{E},\ i,j \in \tilde{\mathcal{L}}\big\}$. $A_{\mathcal{V}} = \big\{((s,i),(s,i-1)) \,|\, s \in \mathcal{V},\ i = 3 \dots K\big\}$.

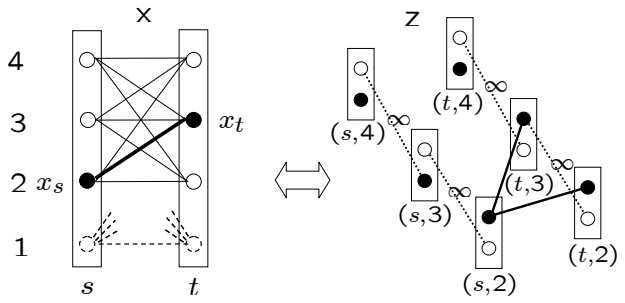


Figure 1. Converting multi-label problems into binary ones. Left: an interaction pair $st \in \mathcal{E}$ of the multi-label energy function; a labeling x is shown by black circles; lowest labels are dashed since all weights in $\hat{\theta}$ associated with them are 0. Right: binary variables $z_{(s,i)}, z_{(t,j)}$ used for encoding the multi-label problem; labeling z(x) is shown by black. Note that if the link $(x_s = 2, x_t = 3)$ is active (left) then two links $[(s,2),(t,2)]$ and $[(s,2),(t,3)]$ are active (right) .

- Binary configuration $\mathrm{z} \in \mathbb{B}^V$. For a configuration $\mathrm{x} \in \mathcal{L}^{\mathcal{V}}$ the corresponding binary configuration z(x) is defined by

$$\mathrm{z}(\mathrm{x})_{(s,i)} = \delta_{\{i \le x_s\}}, \quad (s,i) \in V. \tag{11}$$

- For a binary configuration z, the corresponding multi-label configuration x (denoted as x(z)) is found as:

$$x_s = 1 + \sum_{i \in \tilde{\mathcal{L}}} z_{(s,i)}. \tag{12}$$

- Binary energy function

$$E(\mathrm{z}|\eta) = H(\mathrm{z}) + \sum_{u \in V} \eta_u z_u + \sum_{uv \in A_{\mathcal{E}}} \eta_{uv} z_u z_v + \eta_{\text{const}}, \tag{13}$$

where weights η are set such that

$$E(\mathrm{z}(\mathrm{x})|\eta) = E(\mathrm{x}|\hat{\theta}) = E(\mathrm{x}|\theta),\ \forall \mathrm{x} \in \mathcal{L}^{\mathcal{V}}. \tag{14}$$

In particular pairwise terms η_{uv} are assigned as

$$\eta_{(s,i),(t,j)} = D_{ij}\hat{\theta}_{st} \quad st \in \mathcal{E},\ i,j \in \tilde{\mathcal{L}}, \tag{15}$$

where $D_{ij}\theta_{st} = \theta_{st}(i,j) + \theta_{st}(i-1,j-1) - \theta_{st}(i,j-1) - \theta_{st}(i-1,j)$. Hard constraints H are as follows:

$$H(\mathrm{z}) = \sum_{uv \in A_{\mathcal{V}}} h(z_u, z_v), \tag{16}$$

where $h(z_{(s,i)}, z_{(s,i-1)}) = 0$ if $z_{(s,i)} \le z_{(s,i-1)}$ and ∞ otherwise (see Fig. 1). Hard constraints ensure that any z with finite energy is in the form (11).

roof-dual (Boros et al., 1991) rather then the optimization problem itself.

It is already known that the above defined transformation can be used together with st-mincut algorithms to efficiently and exactly solve lattice-submodular multi-label problems. In this paper, we broaden its applicability by showing how it can be used in conjunction with roof-duality to obtain partial optimal solutions for general problems. An important aspect of the transformation is that (11) depends on the ordering of $\mathcal{L}$. This will lead to certain limitations in the sequel. Other reductions are possible which are order-independent, unfortunately they necessarily involve more of the hard constraints to ensure 1-1 correspondence with the original problem and this leads to a full loss of tightness in the LP-relaxation.

3.2. Multi-label QPBO

Let (G, K, θ) define a multi-label energy minimization problem, and $(N, \mathbb{B}, \eta)$ define the corresponding binary energy minimization problem. The LP-relaxation of the multi-label problem is defined as:

$$\min_{\mu \in \Lambda_{G,\mathcal{L}}} \langle \mu, \theta \rangle \qquad \text{(LP-1)}$$

while that of the binary problem defined as:

$$\min_{\nu \in \Lambda_{N,\mathbb{B}}} \langle \nu, \eta \rangle. \qquad \text{(LP-2)}$$

We attempt minimization of $E(\mathrm{x}|\theta)$ by applying QPBO and its extensions -P, -I (Boros et al., 2006; Rother et al., 2007) to the binary energy $E(\cdot|\eta)$. We call the new methods *multi-label* QPBO (-P,-I), or short MQPBO (-P,-I). As the original QPBO methods efficiently solves LP-2, it is important to see how LP-2 is related to the original problem $\min_{\mathrm{x}} E(\mathrm{x}|\theta)$ and to its relaxation LP-1. The following results apply.

Statement 1. Let $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ define strong persistency for $E(\cdot|\eta)$ such that $E(\mathrm{z}^{\min}) < \infty$ and $E(\mathrm{z}^{\max}) < \infty$. Then any optimal configuration $\mathrm{x} \in \operatorname{argmin}_{\mathcal{L}^{\mathcal{V}}} E(\cdot|\theta)$ must satisfy

$$\mathrm{x}^{\min} \leq \mathrm{x} \leq \mathrm{x}^{\max}, \qquad (17)$$

where $\mathrm{x}^{\min} = \mathrm{x}(\mathrm{z}^{\min})$, $\mathrm{x}^{\max} = \mathrm{x}(\mathrm{z}^{\max})$.

The proof (Shekhovtsov et al., 2007) is simply by using the relation (12).

Thus for each variable s there is an interval of labels $[x_s^{\min}, x_s^{\max}]$ outside of which no label may be selected in an optimal solution. All labels outside the interval may therefore be safely discarded. As is seen, the ordering of $\mathcal{L}$ turns to be very important. While in many practical application there is a natural ordering defined, generally, constraints in the form of intervals derived from arbitrary ordering may be very weak.

A pairwise term θ_{st} is called submodular (resp. supermodular) if

$$D_{ij}\theta_{st} \leq 0 \text{ (resp. } \geq 0 \text{)}, \quad i, j = 2 \dots K. \qquad (18)$$

Theorem 1. *If the term θ_{st} is either submodular or supermodular for each edge $st \in \mathcal{E}$, the relaxations LP-1 and LP-2 coincide, and there exist a mapping between their optimal solutions.*

Proof sketch. We have already used the mapping (11) to relate integral solutions of the two problems such that the energy is preserved. It is quite straightforward to extend it to an injective mapping Π of relaxed labelings μ to relaxed labelings ν, preserving the associated primal costs of LP-1 and LP-2. However, inverting this mapping is not always possible, so there might be a solution ν of LP-2, for which there is no corresponding solution μ of LP-1. Under conditions of the theorem a correction to a solution ν can be applied such that it remains optimal and has a preimage in the mapping Π feasible to LP-1. See details in (Shekhovtsov et al., 2007). □

Corollary 1. It is known that LP-2 can be solved using a network flow algorithm (Hammer et al., 1984). This implies that for a subclass of problems (defined by conditions of the above theorem) there exist **efficient** fully combinatorial algorithms to solve LP-1, which is an improvement over, *e.g.*, message passing algorithms such as TRW-S. But currently, we do not see applications for the case where a part of interactions is submodular and the other part is supermodular.

The next result shows that strong persistency constraints (17) can be also extended to relaxed labelings:

Theorem 2. *Let $\mathrm{x}^{\min} = \mathrm{x}(\mathrm{z}^{\min})$, $\mathrm{x}^{\max} = \mathrm{x}(\mathrm{z}^{\max})$ be the output of MQPBO, and let $\mu \in \Lambda_{G,\mathcal{L}}$ be an optimal solution of LP-1. Then $\mu_{s;i} = 0$ for labels i outside the interval $[x_s^{\min}, x_s^{\max}]$ for all $s \in \mathcal{V}$.*

Proof sketch. Assume μ violates constraints of the theorem. Let then $\nu = \Pi\mu$. For the binary problem it follows from the roof duality that a "truncated" solution, $\bar{\nu}$ may be constructed, which has non-zero weights $\bar{\nu}_u(a)$ only for labels $a \in \mathbb{B}$ such that $z_u^{\min} \leq a \leq z_u^{\max}$, $u \in V$ and such that the primal cost is decreased: $\langle \bar{\nu}, \eta \rangle < \langle \nu, \eta \rangle$. It can then be mapped back to a solution $\bar{\mu} = \Pi^{-1}\bar{\nu}$ of LP-1 of a strictly lower cost than μ. This contradicts to the optimality of μ. See details in (Shekhovtsov et al., 2007). □

The theorem shows that LP-1 never selects nodes which are rejected by LP-2, this may be useful in the analysis of the algorithms related to LP-1.

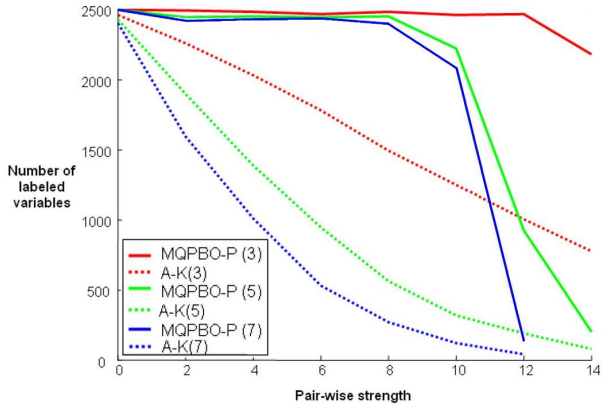


Figure 2. The number of labelled variables in the partially optimal solutions obtained by using MQPBO-P and the algorithm (Kovtun, 2003) (denoted by A-K). Results are shown for energy functions involving variables taking 3, 5 and 7 labels.

4. Enhanced Algorithms

In this section we discuss in more detail the "probing" and "improve" versions of MQPBO and show how they could be used in combination with other algorithms for minimization of multi-label energy functions.

MQPBO-P. This enhancement applies the probing technique (Boros et al., 2006; Rother et al., 2007) to binarized energy functions. It computes stronger constraints of the form (17) on the set of optimal configurations. Our results show that these constraints allow us to isolate the optimal labels for many random variables of the original energy function. In fact for certain energy functions we obtained the global minimum configuration. If latter is not the case then constraints (17) lead to a simplified minimization problem with a smaller or *restricted* solution space, which can be further approached by any other minimization algorithms.

As was mentioned above, MQPBO, and hence the probing extension, is not invariant to permutations of labels. While we are using a fixed ordering in all our experiments, the method could be potentially run under different orderings thus extracting multiple constraints.

MQPBO-P + X. Similar to QPBO+X (Rother et al., 2007), restriction of the energy function obtained by MQPBO(-P) is then minimized using any other minimization algorithm **X**. In our experiments we used max-product BP (Pearl, 1988), TRW-S (Kolmogorov, 2006) and α-expansion algorithm (Boykov et al., 2001).

MQPBO-I. By using QPBO-I (Rother et al., 2007) on the binarized problem any complete labelling of the multi-label MRF can be updated such that its energy never increases. This procedure can be seen as a local search algorithm.

	Pairwise Strength						
Truncation	λ=0	λ=2.4	λ=4.8	λ=7.2	λ=9.6	λ=12.0	λ=14.4
T = 6 (convex)	0	0	0	0	0	0	0
T = 5	0	0	0	0	0	0	0
T = 4	0	0	0	0	0	3	6
T = 3	0	3	4	14	22	26	30
T = 2	0	3	11	20	42	329	1971
T = 1 (Potts)	0	8	19	23	398	2336	2488

Figure 3. Effect of non-convexity: the number of unlabelled variables in the MQPBO-P solution for different values of pairwise strength λ and truncation T.

5. Experimental Results

We now provide the results of using our method to minimize energy functions encountered in computer vision problems. To get a good understanding of the performance of our method we also tested it on synthetic energy functions.

Synthetic Problems. The energy functions corresponding to the synthetic problems contained 50×50 multi-label variables which interacted under a 4-connected neighborhood. We used different numbers of labels and strengths of pairwise potentials in our experiments. Unary potentials $\theta_s(x_s)$ are sampled uniformly in $\{0,1\ldots100\}$. The pairwise potentials had the form of a linear truncated model $\theta_{st}(x_s, x_t) = \frac{\lambda}{T}\min(|x_s - x_t|, T)$, where T is the truncation and λ

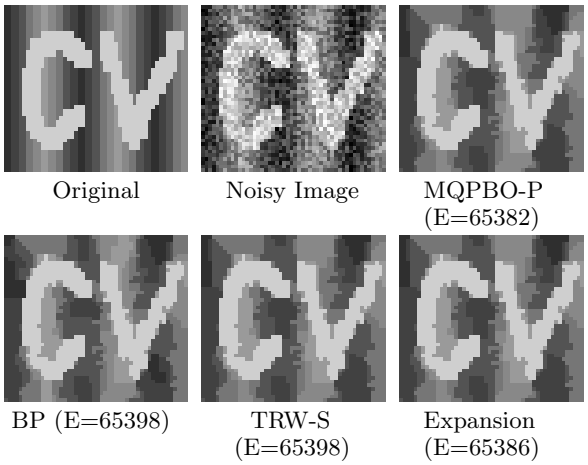


Figure 4. Image Denoising: results obtained by different energy minimization algorithms. The energy used for this experiment has $|\mathcal{L}_I| = 7$, $\gamma = 14$ and $T = 4$. Results are annotated with their respective energy costs. All algorithms were run untill convergence. Observe that MQPBO-P labels all variables and thus obtains the globally optimal solution for this energy.

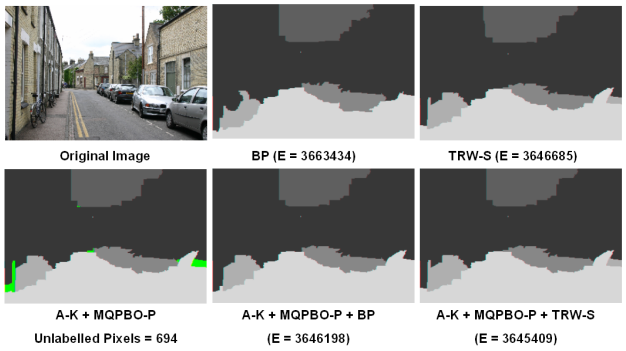


Figure 5. Object Segmentation and Recognition using (Shotton et al., 2006). The first row contains an image from the MSRC (Shotton et al., 2006) dataset and the results of running BP and TRW-S on the full energy. Different levels of gray denote different object classes such as "sky" and "road". The partially optimal solution obtained by running MQPBO-P and A-K (Kovtun, 2003) is shown in the bottom row, left image. Unlabeled pixels are shown in green. This solution is used to obtain a restricted energy. Running BP (bottom, middle) and TRW-S (bottom, right) on the restricted energy gives solutions with lower energy. Hence running A-K + MQPBO-P + "X" is better than running "X" alone. In all cases BP and TRW-S were run for 50 iterations.

is the pairwise strength. Fig. 2 shows comparison of MQPBO-P to (Kovtun, 2003), truncation T is fixed to 1. It is seen that the proposed method labels more variables for a range of parameters. To study the effect of non-convexity, we varied truncation T and strength λ of the pairwise terms (Fig. 3). The number of labels in this case is fixed to 7. Our experiments show that MQPBO-P finds the globally optimal solution of energy functions where the pairwise term is small in magnitude or is nearly *convex*. As expected, the number of variables labelled by the algorithm decrease with increase in the strength and non-convexity of the pairwise terms (Fig. 3).

Image Denoising. We now test the MQPBO-P algorithm on the problem of image denoising and restoration. There is a random variable for each pixel in the image. The label set for the problem is the set of intensities $\mathcal{L}_I$ the pixels can take. The unary cost for taking a particular label (intensity) is defined as $\theta_s(x_s) = |I_s - \phi(x_s)|$ where I_s is the intensity of the pixel s in the image, and the function ϕ maps the labels to their corresponding intensity levels. The pairwise terms of the energy are defined as: $\theta_{st}(x_s, x_t) = \gamma \min(|x_s - x_t|, T)$ where γ is a model parameter and T is the truncation used. Our experiments showed that the MQPBO-P algorithm performed quite well on the energy function and in some cases obtained the globally optimal solution which was not achieved by other energy minimization algorithms (see Fig. 4).

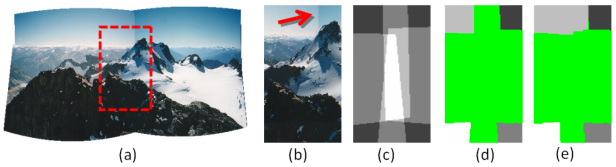


Figure 6. In this application we are stitching together four images (already rectified) to a panoramic image. Running alpha expansion gives result (a) and a zoom-in (dashed rectangle) is shown in (b). The red arrow indicates a visible cut. Image (c) shows the number of possible labels for this image area, where white means all four images overlap and very dark means only one image is possible for those pixels. When applying MQPBO and MQPBO-P to this image portion we obtain labelling (d) and (e) respectively, where green means unlabeled and the different gray scales represent different labels. We see that MQPBO-P is able to label more pixels than MQPBO. It is also worth noting that the visible cut in (b) is indeed the global minimum which can be seen from labelling (e).

Object based Segmentation. We now show the results of using the MQPBO-P method for restricting the energy functions. Our results show that the restriction significantly reduces the number of variables and makes them amenable for minimization using other algorithms. We test the algorithm on the problem of object segmentation and recognition. We use the energy function formulated in (Shotton et al., 2006). The binarized energy function corresponding to this problem has around 10^7 variables and running MQPBO-P on it directly is quite time consuming. To reduce the size of the problem we first run the partial optimality algorithm described in (Kovtun, 2003), referred to as A-K. MQPBO-P is run on the restriction obtained using A-K. This combined procedure leaves 694 variables unlabelled. The results are shown in Fig. 5.

Image Stitching. Finally, we investigated an application where MQPBO-P shows its limitations. For panoramic stitching the unary terms are either zero or infinity, depending on the presence or absence of an image, and consequently the pairwise terms dominate the energy. We use the panoramic stitching formulation from (Agarwala et al., 2004) where pairwise terms model the visibility of a transition, different for each pair of images. The results are discussed in fig 6.

6. Conclusions

This paper addressed the problem of minimizing non-submodular multi-label energy functions. These are used extensively in computer vision and are generally NP-hard to minimize. We present a method for obtaining partially optimal solutions of such energies derived from roof-dual based methods for binary energy functions. We give new theoretical insights in the un-

derlying LP relaxation, being efficiently solved by the method. Although this work is mainly theoretical in nature we hope to inspire people to use these ideas to develop better optimization algorithms. We also believe that this approach is useful for other important problems in computer vision such as MRF/CRF learning.

Acknowledgments

We would like to thank the anonymous reviewers for their useful comments. This research has been partially funded by Czech government grant MSM6840770038, EPSRC and PASCAL, EU.

References

Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S. M., Colburn, A., Curless, B., Salesin, D., & Cohen, M. F. (2004). Interactive digital photomontage. *ACM Trans. Graph.*, *23*, 294–302.

Boros, E., & Hammer, P. (2002). Pseudo-boolean optimization. *Discrete Applied Mathematics*, 155–225.

Boros, E., Hammer, P. L., & Sun, X. (1991). *Network flows and minimization of quadratic pseudo-Boolean functions* (Technical Report RRR 17-1991). RUTCOR.

Boros, E., Hammer, P. L., & Tavares, G. (2006). *Preprocessing of unconstrained quadratic binary optimization* (Technical Report RRR 10-2006). RUTCOR.

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *PAMI*, *23*, 1222–1239.

Desmet, J., Maeyer, M. D., Hazes, B., & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, *356*, 539–542.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2000). Efficient matching of pictorial structures. *CVPR* (pp. 66–73).

Hammer, P., Hansen, P., & Simeone, B. (1984). Roof duality, complementation and persistency in quadratic 0-1 optimization. *Math. Programming*, 121–155.

Hammer, P. L. (1965). Some network flow problems solved with pseudo-Boolean programming. *Operation Research*, *13*, 388–399.

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

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

Kolmogorov, V., & Rother, C. (2007). Minimizing non-submodular functions with graph cuts – a review. *PAMI*, *29*, 1274–1279.

Kolmogorov, V., & Zabih, R. (2004). What energy functions can be minimized via graph cuts? *PAMI*, *26*, 147–159.

Kolmogorov, V. N., & Wainwright, M. J. (2005). On optimality of tree-reweighted max-product message-passing. *Appeared in Uncertainty in Artificial Intelligence*.

Komodakis, N., Paragios, N., & Tziritas, G. (2007). MRF optimization via dual decomposition: Message-passing revisited. *ICCV* (pp. 1–8).

Komodakis, N., & Tziritas, G. (2005). A new framework for approximate labeling via graph cuts. *ICCV* (pp. 1018–1025).

Koval, V., & Schlesinger, M. (1976). Two-dimensional programming in image analysis problems. *Automatics and Telemechanics*, *2*, 149–168. In Russian.

Kovtun, I. (2003). Partial optimal labeling search for a NP-hard subclass of (max, +) problems. *DAGM-Symposium* (pp. 402–409).

Kovtun, I. (2004). *Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem*. Doctoral dissertation. In Ukrainian.

Lauritzen, S. L. (1998). *Graphical models*. No. 17 in Oxford Statistical Science Series. Oxford Science Publications.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Rother, C., Kolmogorov, V., Lempitsky, V., & Szummer, M. (2007). Optimizing binary MRFs via extended roof duality. *CVPR* (pp. 1–8).

Schlesinger, D. (2007). Exact solution of permuted submodular minsum problems. *EMMCVPR* (pp. 28–38). Springer.

Schlesinger, D., & Flach, B. (2006). *Transforming an arbitrary minsum problem into a binary one* (Technical Report TUD-FI06-01). Dresden University of Technology.

Schlesinger, M. (1976). Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika, Kiev*, *4*, 113–130. In Russian.

Schlesinger, M. I., & Flach, B. (2000). Some solvable subclasses of structural recognition problems. *Czech Pattern Recognition Workshop*.

Schlesinger, M. I., & Giginyak, V. V. (2007). Solution to structural recognition (MAX,+)-problems by their equivalent transformations. *Control Systems and Computers*, *1,2*.

Shekhovtsov, A., Kolmogorov, V., Kohli, P., Hlavac, V., Rother, C., & Torr, P. (2007). *LP-relaxation of binarized energy minimization* Research Report CTU–CMP–2007–27). Czech Technical University.

Shotton, J., Winn, J. M., Rother, C., & Criminisi, A. (2006). *TextonBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. *ECCV (1)* (pp. 1–15).

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. F., & Rother, C. (2006). A comparative study of energy minimization methods for Markov random fields. *ECCV (2)* (pp. 16–29).

Wainwright, M., Jaakkola, T., & Willsky, A. (2003). Exact MAP estimates by (hyper)tree agreement. In *Advances in neural information processing systems 15*, 809–816.

Werner, T. (2007). A linear programming approach to max-sum problem: A review. *PAMI*, *29*, 1165–1179.