Computer Vision Winter Workshop 2008, Janez Perš (ed.)
Moravske Toplice, Slovenia, February 4–6
Slovenian Pattern Recognition Society, Ljubljana, Slovenia

A Lower Bound by One-against-all Decomposition for Potts Model Energy Minimization

Alexander Shekhovtsov and Václav Hlaváč

Czech Technical University, Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
shekhovt@cmp.felk.cvut.cz, hlavac@fel.cvut.cz

Abstract *Energy minimization in computer vision can be formulated as minimizing a separable function of discrete variables. Such formulation arises, in particular, in Gibbs probabilistic models, which are suitable for many low-level image processing tasks. A special type of this probabilistic model, the Potts model, was successfully applied in segmentation, stereo and image denoising. Importantly, the Potts model is general enough to incorporate difficult problems: the energy minimization remains NP-hard.*

In this paper, we investigate a new linear relaxation for the Potts model. The relaxation is obtained by considering a family of decompositions of the energy function in the fashion of one label vs. all other. The tightest lower bound in the family constitutes a linear maximization program, for which we propose a coordinate ascent method. We show that our bound is dual to the known LP-relaxation of energy minimization, for which several dedicated solvers are available. Our new formulation could lead to more efficient algorithms (this was not confirmed yet) and to new per-variable-label bounds. The letter should allow to reject some selections which are not optimal. These bounds may be used efficiently, e.g., in branch and bound algorithm.

1 Introduction

We are interested in solving energy minimization problems in computer vision. There exist well-established approaches to the related hard discrete minimization problems, like the MAX-CUT problem. However, the scale of the tasks arising in computer vision is such that these techniques do not apply. In particular, one cannot imagine solving a semidefinite relaxation of a discrete minimization problem involving a million of variables. Even the Linear Programming (LP) relaxation is not within common computational resources yet. There were many efforts in designing dedicated algorithms for solving the LP-relaxation, which would efficiently exploit the separable structure of the problem, such as: Augmenting DAG [10, 15], TRW-S [13, 8].

Our study of the Potts model was motivated also by the following, often applied, heuristic to the multi-class segmentation problem. Instead of solving the multi-class problem, a series of two-class problems is solved to decide for one of the classes versus all others at a time. A solution of the multi-class problem is then combined somehow from these simpler problems. A less straightforward heuristic was proposed by Lu *et al*. [9]. They showed that energy function of a multi-label segmentation problem can be represented as a sum of solvable auxiliary binary subproblems. If solutions of these auxiliary subproblems are consistent, then it is guaranteed that a global optimum is found. However if they are consistent only for a subset of pixels, the proposition in *et al*. [9] to fix the segmentation of this subset as an optimal choice seems to be wrong (we will present a counterexample). We note that the auxiliary subproblems may serve to compute exact lower bounds, on basis of which the optimality can be decided. We also show how weights of the subproblems should be selected to force the consistency of decisions. Finally, we show that optimal selection of these weights is equivalent to the well-known LP-relaxation approach.

The LP-relaxation [10, 6, 13] is obtained by reformulating the energy minimization as 0-1 integer program and by relaxing the integrality constraints. There are tight correspondences between this LP-relaxation, equivalent reparametrizations of the energy [10] and decompositions of the energy into a collection of tree-structured subproblems [13]. In this paper, we take the view of decompositions. The basic idea is to represent the energy function as a sum of auxiliary energy functions and then to lower bound the minimum of the energy by the sum of minima of auxiliary functions. The trick then is to be able to compute minima of the auxiliary functions and try to find the tightest bound. In the case of tree-structured subproblems, the tightest bound problem is a dual to the standard LP relaxation [13]. Another family of decompositions is into sum of a submodular and a supermodular problems [11], which yields a different bound. Interestingly, this bound, while generally being weaker, can be computed more easily and provides certain optimality guarantees.

We introduce another decomposition applicable for the Potts model (see, *e.g*., [5]). It uses auxiliary binary subproblems, which are solvable due to constraints of the Potts model. The tightest bound obtained by these decompositions is shown to be equivalent to the standard LP-relaxation. The useful outcomes of this construction are 1) the novel algorithm for the problem and 2) per-node bounds, which al-

low to reject many non-optimal nodes in the case when optimal solution cannot be obtained. We call a *node* a pair (variable, label). This latter property is very useful practically: (a) to use the part of variables for which optimal labels are found in further steps of a computer vision application, (b) to constrain the search space and apply some other energy minimization method, and (c) to use the per-node bounds in the branch and bound approach.

Outline. We introduce the energy minimization problem in Section 2 and review the decomposition approach of [13] in Section 3. Our new decomposition and the associated tightest lower bound is proposed in Section 4. Our algorithm is based on necessary conditions of the optimality, Sections 5,6. Per-node lower bounds and equivalence to the standard LP-relaxation are considered in Sections 7,8. Experiments are summarized in Section 9. Discussion and conclusions are in Section 10.

2 Energy Minimization

We consider the following problem illustrated in Fig. 1. Let $\mathcal{L} = \{1 \dots K\}$ be a set of labels. Let $G = (\mathcal{V}, \mathcal{E})$ be a graph with $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ antisymmetric and antireflexive, *i.e.*, $(s,t) \in \mathcal{E} \Rightarrow (t,s) \notin \mathcal{E}$. In what follows, we will denote by st an ordered pair $(s,t) \in \mathcal{E}$. Let also $\bar{\mathcal{E}} = \mathcal{E} \cup \{ts \,|\, st \in \mathcal{E}\}$. 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. Let all potentials be concatenated into a single vector $\theta \in \Omega = \mathbb{R}^{\mathcal{I}}$, where set of indices $\mathcal{I} = \{(s,i) \,|\, s \in \mathcal{V}, i \in \mathcal{L}\} \cup \{(st,ij) \,|\, st \in \mathcal{E}, i,j \in \mathcal{L}\}$. Let in addition θ_{const} be a *constant* term, and let a concatenated vector of potentials, including the constant term, be denoted as $\mathring{\theta} = (\theta, \theta_{\text{const}}) \in \mathring{\Omega} = \Omega \times \mathbb{R}$. Let *energy* of a configuration x be defined by:

$$E(\mathrm{x}|\mathring{\theta}) = \sum_{s \in \mathcal{V}} \theta_s(x_s) + \sum_{st \in \mathcal{E}} \theta_{st}(x_s, x_t) + \theta_{\text{const}}. \qquad (1)$$

It is conveniently written using scalar product in Ω as $E(\mathrm{x}|\mathring{\theta}) = \langle \mu(\mathrm{x}), \theta \rangle + \theta_{\text{const}}$, where $\mu \colon \mathcal{L}^{\mathcal{V}} \mapsto \Omega$ is a mapping defined by $[\mu(\mathrm{x})]_s(k) = \delta_{\{x_s = k\}}$, $s \in \mathcal{V}$, $k \in \mathcal{L}$; $[\mu(\mathrm{x})]_{st}(k,k') = \delta_{\{x_s=k\}}\delta_{\{x_t=k'\}}$, $st \in \mathcal{E}$, $k,k' \in \mathcal{L}$, where $\delta_{\{A\}}$ is 1 if A is true and 0 otherwise. Thus, the problem of energy minimization can be shortly written using scalar product in Ω as

$$\min_{\mathrm{x} \in \mathcal{L}^{\mathcal{V}}} \langle \mu(\mathrm{x}), \theta \rangle + \theta_{\text{const}}. \qquad (2)$$

We restrict our attention to energy functions of simpler form, known as Potts model. Vector θ will be called a Potts model if for all $st \in \mathcal{E}$

$$\begin{aligned} \theta_{st}(i,j) &= 0, \qquad i \neq j \\ \theta_{st}(i,j) &\leq 0, \qquad i = j. \end{aligned} \qquad (3)$$

Even with restriction to models of this type, the energy minimization (2) is still an NP-hard problem, closely related to multiway cut problem as was pointed by [5].

[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.

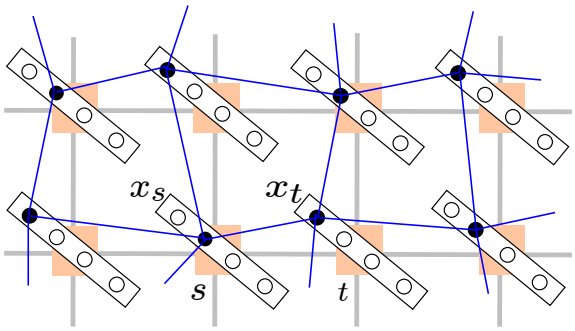

Figure 1: Energy minimization: each node s of the graph is assigned a discrete variable x_s, depicted by a box with labels. Labels in the box represent values which discrete variable may attain. A labeling x is shown by black circles and black solid lines.

Example 1 (A segmentation model)**.** Consider the following instance of the Potts model. Let set $\mathcal{V}$ be set of pixels and set $\mathcal{E}$ – pairs of neighboring pixels (*e.g.* horizontally and vertically neighboring). Each pixel is to be assigned a label x_s from the set of labels $\mathcal{L}$. Two neighboring pixels are expected to have identical labels, which is achieved by setting

$$\theta_{st}(x_s, x_t) = \begin{cases} 0, & x_s = x_t \\ c_{st}, & x_s \neq x_t, \end{cases} \qquad (4)$$

where c_{st} is a positive penalty paid when neighboring pixels s and t are assigned different labels. Univariate potentials $\theta_s(x_s)$ are set according to the local likelihood of assigning label x_s to pixel s. This model is easily convertible to the form satisfying (3) by subtracting a constant from all θ_{st} and adding the equivalent sum to θ_{const}.

3 Lower Bounds by Decompositions

Let us consider a collection of parameter vectors $\mathring{\boldsymbol{\theta}} = \{\mathring{\theta}^i \in \mathring{\Omega} \,|\, i \in I\}$, where I is a finite set. Collection $\mathring{\boldsymbol{\theta}}$ will be called a *decomposition* of θ, if

$$E(x|\mathring{\theta}) = \sum_{i \in I} E(x|\mathring{\theta}^i) \quad \forall \mathrm{x} \in \mathcal{L}^{\mathcal{V}}, \qquad (5)$$

i.e. E can be represented as a sum of the subproblems. In particular, $\mathring{\boldsymbol{\theta}}$ is a decomposition of $\mathring{\theta}$ if $\sum_i \mathring{\theta}^i = \mathring{\theta}$.

Statement 1 ([13])**.** For any decomposition $\mathring{\boldsymbol{\theta}}$ value $LB(\mathring{\boldsymbol{\theta}}) = \sum_i \min_{\mathrm{x}} \langle \theta^i, \mu(\mathrm{x}) \rangle + \theta_{\text{const}}$ constitutes a lower bound on the optimal energy:

$$LB(\mathring{\boldsymbol{\theta}}) \leq \min_{\mathrm{x}} \langle \theta, \mu(\mathrm{x}) \rangle + \theta_{\text{const}}. \qquad (6)$$

Proof. $\min_{\mathrm{x}} \langle \theta, \mu(\mathrm{x}) \rangle = \min_{\mathrm{x}} \langle \sum_i \theta^i, \mu(\mathrm{x}) \rangle = \min_{\mathrm{x}} \sum_i \langle \theta^i, \mu(\mathrm{x}) \rangle \geq \sum_i \min_{\mathrm{x}} \langle \theta^i, \mu(\mathrm{x}) \rangle$. Alternatively could be seen from the Jensen's inequality as in [13] and linearity of $\langle \cdot, \mu(\mathrm{x}) \rangle$. □

The computation of $LB(\mathring{\boldsymbol{\theta}})$ is tractable if all the problems $\min_{\mathrm{x}} \langle \theta^i, \mu(\mathrm{x}) \rangle$ are tractable. For this purpose, each θ^i can be chosen to define a tree-structured distribution [13]. For a chosen collection of trees covering the graph G there is

still a linear subspace of possible selections of values $\theta^i_s(\cdot)$, such that constraints $\sum_i \mathring{\theta}^i = \mathring{\theta}$ are satisfied. Finding the tightest lower bound over all possible decompositions from this subspace is a dual problem to the LP-relaxation [13]. However there are other possibilities.

Another family of lower bounds can be obtained by considering decompositions of θ into sum of submodular and supermodular problems [11]. For a bipartite graph G (like the grid graph in Fig. 1) both subproblems are solvable and the tightest lower bound of the family can be expressed as a specific linear program. Also this lower bound is generally much weaker than the one obtained from tree decompositions, it can be computed by a max-flow algorithm and provides certain theoretical guarantees.

We consider yet another family of decomposition, yielding easily computable lower bounds.

4 One-against-all Decomposition

In the one-against-all decomposition, each θ^i will represent a sub-problem with binary decision variables, where decision is made between label i and all other labels in each node. This is illustrated in Fig. 2.

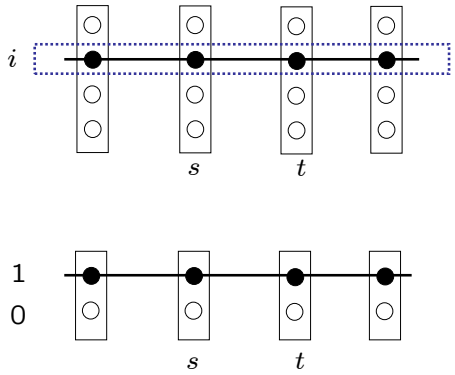

Figure 2: One vs. all decomposition. Top: in the multi-label problem, $E(\mathrm{x}|\theta)$, for each $i \in \mathcal{L}$ nodes and arcs associated with label i are used to construct auxiliary problem $E^i(\mathrm{z}|\theta^i)$. Bottom: The decision variables of the auxiliary problem $E^i(\mathrm{z}|\theta^i)$ are the indicators of whether to select the label i or some other. A reweightening of the auxiliary problems aims at achieving agreement of these decision variables.

Formally, let $I = \{1 \dots K\}$, let $\mathring{\theta}^i$ satisfy:

$$\begin{aligned} \theta^i_s(j) &= 0, \qquad \forall j \in \mathcal{L} : j \neq i \\ \theta^i_{st}(j,k) &= 0, \qquad \forall j,k \in \mathcal{L} : j \neq i \text{ or } k \neq i. \end{aligned} \tag{7}$$

In this case, energy function can be written as

$$E^i(\mathrm{x}|\mathring{\theta}^i) = \sum_{s\in\mathcal{V}} \theta^i_s(i)\delta_{\{x_s=i\}} + \sum_{st\in\mathcal{E}} \theta^i_{st}(i,i)\delta_{\{x_s=i\}}\delta_{\{x_t=i\}} + \theta^i_{\text{const}}, \tag{8}$$

or introducing auxiliary binary decision variables $\mathrm{z}_s = \delta_{\{x_s=i\}}$, as

$$E^i(\mathrm{z}|\mathring{\theta}^i) = \sum_{s\in\mathcal{V}} \theta^i_s(i)z_s + \sum_{st\in\mathcal{E}} \theta^i_{st}(i,i)z_s z_t + \theta^i_{\text{const}}. \tag{9}$$

Further on, let collection $\{\mathring{\theta}^i | i \in I\}$ satisfy:

$$\begin{aligned} \theta^i_{st}(i,i) &= \theta_{st}(i,i), && \forall st \in \mathcal{E},\ i \in \mathcal{L} \\ \theta^i_s(i) &= \theta_s(i) - q_s, && \forall s \in \mathcal{V},\ i \in \mathcal{L} \\ \theta^i_{\text{const}} &= \theta_{\text{const}} + \tfrac{1}{K}\textstyle\sum_{s\in\mathcal{V}} q_s \quad , \end{aligned} \tag{10}$$

where $q_s \in \mathbb{R} \mid s \in \mathcal{V}$ are free variables.

Statement 2. A collection $\mathring{\boldsymbol{\theta}}$ satisfying (7) and (10) forms a decomposition of $\mathring{\theta}$.

Indeed, substituting (7), (10) into $\sum_i E^i(\mathrm{x}|\mathring{\theta}^i)$ verifies that it equals $E(\mathrm{x}|\mathring{\theta})$.

For a given decomposition we are interested in computing the lower bound $LB(\mathring{\boldsymbol{\theta}}) = \sum_i \min_{\mathrm{x}} E^i(\mathrm{x}|\mathring{\theta}^i)$. Due to representation (9), auxiliary problem $\min_{\mathrm{x}} E^i(\mathrm{x}|\theta^i)$ is a binary minimization problem, and under constraints $\theta^i_{st}(i,i) \leq 0$, $st \in \mathcal{V}$ (which are valid in the Potts model) it is submodular and therefore can be minimized by a max-flow algorithm [2, 4].

We are interested in finding the tightest lower bound in the family, which means solving the following maximization problem:

$$LB = \max_{\{q_s | s\in\mathcal{V}\}} \sum_i \min_{\mathrm{x}} E^i(\mathrm{x}|\mathring{\theta}^i), \quad \text{s.t. (7) and (10).} \tag{11}$$

Our next step is to show that (11) can be written as a linear program and therefore it is practical to compute. To see this we first rewrite each subproblem $\min_{\mathrm{z}} E^i(\mathrm{z}|\mathring{\theta}^i)$ as a linear program [2, 1]:

$$\begin{aligned} &\min_{\nu} \sum_{s\in\mathcal{V}} (\theta^i_s(i) - q_s)\nu_s + \sum_{st\in\mathcal{E}} \theta^i_{st}(i,i)\nu_{st} + \theta^i_{\text{const}} \\ &\text{s.t.} \begin{cases} \nu_s \in [0,1], & s \in \mathcal{V} \\ \nu_{st} \in [0,1], & st \in \mathcal{E} \\ \nu_{st} \leq \min(\nu_s, \nu_t), & st \in \mathcal{E}. \end{cases} \end{aligned} \tag{12}$$

Problem (12) is a linear relaxation of the binary minimization $\min_{\mathrm{z}} E^i(\mathrm{z}|\mathring{\theta}^i)$, but because the latter problem is submodular the relaxation is tight (see,*e.g.*, [1]). Let us shortly denote problem (12) as

$$\min_{\nu\in\Lambda_2} E^i(\nu|\mathring{\theta}^i), \tag{13}$$

where $E^i(\nu|\mathring{\theta}^i)$ represents the objective of (12), which is linear in ν and $\mathring{\theta}^i$ and Λ_2 is the constraint polytope of (12). This allows us to rewrite (11) as follows

$$\begin{aligned} LB &= \max_q \sum_i \min_{\nu\in\Lambda_2} E^i(\nu|\mathring{\theta}^i) \\ &= \max_q \min_{\{\nu^i\in\Lambda_2 | i\in I\}} \sum_i E^i(\nu^i|\mathring{\theta}^i) \\ &= \inf_{\{\nu^i\in\Lambda_2 | i\in I\}} \sup_q \sum_i E^i(\nu^i|\mathring{\theta}^i), \end{aligned} \tag{14}$$

where the last equality holds due to linearity of objective and convexity of constraint sets. Expanding E^i, it is $\sum_i E^i(\nu^i|\mathring{\theta}^i) =$

$$\begin{aligned} &\sum_{s,i} (\theta_s(i) - q_s)\nu^i_s + \sum_{st,i} \theta_{st}(i,i)\nu^i_{st} + \sum_s q_s + \theta_{\text{const}} \\ &= \langle \nu, \theta \rangle + \sum_s q_s(1 - \sum_i \nu^i_s) + \theta_{\text{const}}. \end{aligned} \tag{15}$$

Finally, substituting (15) into the supremum in (14) it is seen, that the optimal value of (11) equals

$$\min \langle \nu, \theta \rangle + \theta_{\text{const}} \quad \text{s.t.} \begin{cases} \nu^i \in \Lambda_2, & i \in I \\ \sum_i \nu^i_s = 1, & \forall s \in \mathcal{V}. \end{cases} \tag{16}$$

In the linear program (16), the objective is the same as that of original problem (2). Only the constraint is a convex polytope with number of inequality constraints proportional to the number of variables. Unfortunately, state of the art linear programming solvers are too much memory and computation demanding for instances of energy minimization of any reasonably large size.

5 Conditions of the Optimality

We will propose an algorithm, to solve (16) suboptimally, so that it will satisfy certain necessary conditions of the optimality, derived in this section. As these conditions are not sufficient for the optimality, the outcome of the algorithm will be not the tightest lower bound, but a somewhat weaker lower bound.

We propose the following necessary conditions of the optimum of (11). Let $\Phi^i_{s,a} = \min_{\mathrm{z}|z_s=a} E^i(\mathrm{z}|\theta^i)$, $a \in \{0,1\}$, $s \in \mathcal{V}$ be a collection of *min-marginals* of subproblem θ^i. Let $\Delta\Phi^i_s = \Phi^i_{s,1} - \Phi^i_{s,0}$.

Statement 3. In the optimum of (11) the following conditions are satisfied

$$\begin{aligned} &\forall s \in \mathcal{V},\ \exists i \in \mathcal{L} \quad \Delta\Phi^i_s \le 0 \\ &\forall s \in \mathcal{V},\ \forall i,j \in \mathcal{L} \quad \Delta\Phi^i_s \ge 0 \text{ or } \Delta\Phi^j_s \ge 0. \end{aligned} \tag{17}$$

This conditions are naturally interpreted in terms of *agreement* of sets of optimal configurations for the subproblems. Namely, at least one of the subproblems θ^i must allow its label i as optimal. And if a subproblem i selects label i strictly in some s, then no other subproblem can select strictly a label $j \neq i$. Thus, subproblems are forced to share a solution x optimal for all of them simultaneously.

Proof.

- Assume the first part of (17) is violated. Then $\exists s \in \mathcal{V}$ such that $\varepsilon = \min_i \Delta\Phi^i_s > 0$. Then increasing q_s by ε will increase the value of $LB(\boldsymbol{\theta})$ by ε.

  Indeed, weight $\theta_s(i) - q_s$ will decrease only by ε, and the value $\min_{\mathrm{z}} E^i(\mathrm{z}|\theta^i)$ will increase by $\frac{\varepsilon}{K}$ for each $i \in \mathcal{L}$.

- Assume the second part of (17) is violated. Namely, $\exists s \in \mathcal{V}\ \exists i,j \in \mathcal{L}$ such that $\Delta\Phi^i_s < 0$ and $\Delta\Phi^j_s < 0$. Let $\varepsilon = \max(\Delta\Phi^i_s, \Delta\Phi^j_s) > 0$. Then decreasing q_s by ε will increase the $LB(\boldsymbol{\theta})$ by at least ε.

  Indeed, weight $\theta_s(i) - q_s$ will increase by ε, therefore values $\min_{\mathrm{z}} E^i(\mathrm{z}|\theta^i)$ will increase by ε. Thus $LB(\boldsymbol{\theta})$ will increase at least by $2\varepsilon - \varepsilon = \varepsilon$.

□

6 Algorithm

Our algorithm is designed to satisfy necessary conditions (17). Each time they are found violated, the correction is made and the $LB(\mathring{\boldsymbol{\theta}})$ is increased. Only one variable q_s is modified in each step, thus it is a coordinate ascent algorithm. As is usual for coordinate ascent maximization of a non-smooth function this algorithm may get stuck in a point fare from the optimum. Indeed, we observed it happens on practice. We will discuss this issue in the conclusion.

Algorithm 1: One-against-all Reweightening Coordinate Ascent

1. Initialize $q_s = 0, s \in \mathcal{V}$ and $\mathring{\theta}^i$ by (10).
2. For all $s \in \mathcal{V}$ do
   (a) compute values $\Phi^i_{s,a}$, $i \in \mathcal{L}$, $a \in \{0,1\}$ by fast reuse of residual networks in max-flow [7].
   (b) Check conditions (17). If not satisfied, increase the $LB(\boldsymbol{\theta})$ by appropriate change of q_s.
3. Repeat step 2 while $LB(\boldsymbol{\theta})$ is improving.

Note, that when run in the problems with integer weights θ, each step of the algorithm is guaranteed to increase the lower bound at least by one. So in this case it is finitely converging.

7 Per-node Bounds

Aside from the value LB computed by (11), which is a lower bound on the quality of the optimal solution, we are interested in obtaining per-node bounds $LB_{s,i}$ on the quality of the optimal solution under fixation of $x_s = i$.

We construct these per-node bounds as:

$$LB_{s,i}(\mathring{\boldsymbol{\theta}}) = \sum_j \min_{\mathrm{x}|x_s=i} E^j(\mathrm{x}|\mathring{\theta}^j) \le \min_{\mathrm{x}|x_s=i} E(\mathrm{x}|\mathring{\theta}). \tag{18}$$

These bounds of course hold for any decomposition $\mathring{\boldsymbol{\theta}}$ but they are tighter, when $\mathring{\boldsymbol{\theta}}$ is a solution of (11). Indeed, it is seen that $LB(\mathring{\boldsymbol{\theta}}) \le \min_i LB_{s,i}(\mathring{\boldsymbol{\theta}})$ for all $s \in \mathcal{V}$, and thus the larger the bound $LB(\mathring{\boldsymbol{\theta}})$, the larger are per-node bounds $LB_{s,i}(\mathring{\boldsymbol{\theta}})$.

Recall that computation of $\min_{\mathrm{x}|x_s=i} E^j(\mathrm{x}|\mathring{\theta}^j)$ reduces to problem with binary variables, $\min_{\mathrm{z}|z_s=\delta_{\{i=j\}}} E^j(\mathrm{z}|\mathring{\theta}^j)$. As was noted [3, 7], solutions for all $s \in \mathcal{V}$ can be computed cheaply by reusing the residual network in the max-flow algorithm.

Suppose, we have a good guess of solution x, then by comparing $LB_{s,i}$ with $E(\mathrm{x}|\mathring{\theta})$ we can reject some nodes as being certainly not optimal. Indeed, let $LB_{s,i} > E(\mathrm{x}|\mathring{\theta})$, then no configuration x′ with fixation $x'_s = i$ can be better then x, and therefore node (s,i) can be safely rejected.

8 Equivalence of Relaxations

In this section, we show that our new bound is in fact equivalent to the well known LP-relaxation approach.

The LP-relaxation of (2) is constructed as follows [6, 13, 14]: for each discrete variable x_s a group of relaxed variables $\mu_s(i) \in [0,1]$, $i \in \mathcal{L}$ is introduced and required to satisfy *normalization* constraints

$$\sum_{i\in\mathcal{L}} \mu_s(i) = 1, \quad \forall s \in \mathcal{V} \tag{19}$$

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

$$\begin{aligned} \sum_{j'\in\mathcal{L}} \mu_{st}(i,j') &= \mu_s(i), \quad \forall st \in \mathcal{E}, \forall i \in \mathcal{L}; \\ \sum_{i'\in\mathcal{L}} \mu_{st}(i',j) &= \mu_t(j), \quad \forall st \in \mathcal{E}, \forall j \in \mathcal{L}. \end{aligned} \tag{20}$$

Concatenated vector $\mu \in \Omega$ satisfying these constraints is called a *relaxed* labeling. Whenever μ has all components integral, it uniquely represents a labeling x by the relation $\mu = \mu(\mathrm{x})$. By dropping the integrality constraints the following relaxation of (2) is obtained:

$$\min_{\mu\in\Lambda_{G,\mathcal{L}}} \langle \mu, \theta\rangle + \theta_{\text{const}}, \tag{21}$$

where $\Lambda_{G,\mathcal{L}} = \{\mu \in \Omega_+ \mid A\mu = \mathbf{0},\ B\mu = \mathbf{1}\}$ is called *local* [13] polytope of graph G. Here set Ω_+ denotes vectors from Ω with all components nonnegative, equalities $A\mu = \mathbf{0}$ express marginalization constraints (20) and equalities $B\mu = \mathbf{1}$ express normalization constraints (19).

Statement 4. Let θ be a parameter vector of Potts model. Then linear program (16) is equivalent to the standard LP-relaxation (21) in the sense that their optimum values coincide and optimal solutions are convertible.

Proof.

- Let μ be optimal to (21). A feasible solution ν to (16) preserving the objective can be constructed as follows.

  Objective of (21) for Potts model expresses as

  $$\langle \mu, \theta\rangle + \theta_{\text{const}} = \sum_{s,i} \theta_s(i)\mu_s(i) + \sum_{st,i} \theta_{st}(i,i)\mu_{st}(i,i) + \theta_{\text{const}}. \tag{22}$$

  Letting

  $$\begin{aligned} \nu^i_s &= \mu_s(i), & s \in \mathcal{V},\ i \in \mathcal{L} \\ \nu^i_{st} &= \mu_{st}(i,i), & st \in \mathcal{E},\ i \in \mathcal{L} \end{aligned} \tag{23}$$

  it is seen that ν is feasible to (16) and gives the same value of the objective. Feasibility constraints $\sum_i \nu^i_s = 1$ follow from (20) and constraints $\nu^i_{st} \le \min(\nu^i_s, \nu^i_t)$ from (19) and nonnegativity of μ_{st}.

- Let ν be optimal to (16). A feasible solution μ to (21) can be constructed as follows:

  $$\begin{aligned} \mu_s(i) &= \nu^i_s, & s \in \mathcal{V},\ i \in \mathcal{L} \\ \mu_{st}(i,i) &= \nu^i_{st}, & st \in \mathcal{E},\ i \in \mathcal{L} \\ \mu_{st}(i,j) &= \tfrac{\tilde\nu^i_{st}\tilde\nu^i_{ts}}{1-\Delta_{st}}, & st \in \mathcal{E},\ i,j \in \mathcal{L} : i \ne j, \end{aligned} \tag{24}$$

  where $\Delta_{st} = \sum_i \min(\nu^i_s, \nu^i_t)$ and

  $$\begin{aligned} \tilde\nu^i_{st} &= \nu^i_s - \min(\nu^i_s, \nu^i_t), & st \in \mathcal{E},\ i \in \mathcal{L}; \\ \tilde\nu^i_{ts} &= \nu^i_t - \min(\nu^i_s, \nu^i_t), & st \in \mathcal{E},\ i \in \mathcal{L}. \end{aligned} \tag{25}$$

  It its seen that values μ are nonnegative. Noting that $\tilde\nu^i_{st}\tilde\nu^i_{ts} = 0$ for all $i \in \mathcal{L}, st \in \mathcal{E}$, marginalization constraints are verified as:

  $$\begin{aligned} \sum_j \mu_{st}(i,j) &= \min(\nu^i_s, \nu^i_t) + \tilde\nu^i_{st}\frac{1}{1-\Delta_{st}} \sum_{j\ne i} \tilde\nu^j_{st} = \\ &\min(\nu^i_s, \nu^i_t) + \tilde\nu^i_{st}\frac{1}{1-\Delta_{st}} \sum_{j} \tilde\nu^j_{st} = \\ &\min(\nu^i_s, \nu^i_t) + \tilde\nu^i_{st}\frac{1-\Delta_{st}}{1-\Delta_{st}} = \\ &\min(\nu^i_s, \nu^i_t) + \nu^i_s - \min(\nu^i_s, \nu^i_t) = \nu^i_s. \end{aligned} \tag{26}$$

  Thus μ is feasible to (21) and the value of the objective is preserved.

□

Example 2 (Non-optimal agreement for Potts model)**.** As a common case in integer programs, even the variables which appear to be integral in the solution of the relaxed problem need not be a part of an optimal assignment. Consider the Potts model in Fig. 3. Parameters $\theta_s(\cdot)$ are shown in the top figure. Values $\theta_{st}(i,j) = -\delta_{\{i=j\}}$ for all edges st. Value $\theta_{\text{const}} = 0$.

Energy of the optimal relaxed labeling μ is $E(\mu|\mathring\theta) = -5$. Energy of optimal integral configuration x is $E(\mathrm{x}|\mathring\theta) = 0$. Energy of the best integral labeling passing through the node $(s_1, 4)$ is $+1$. Thus it is seen that node $(s_1, 4)$ can not be taken as optimal.

According to our equivalence result, this counterexample applies also to one-against-all decompositions, and it shows that even for the decomposition $\boldsymbol{\theta}$ which maximizes the lower bound, agreement on one pixel does imply the optimality.

9 Experiments

So far only experiments with artificial random problems are proposed. We generate random instances of a Potts model on a grid graph of size 10×10 with 5 labels. Weights $\theta_s(\cdot)$ are sampled uniformly in $\{0, 1 \dots 10\}$ (integral weights) and weights $\theta_{st}(\cdot,\cdot)$ uniformly in $\{0, -1 \cdots -5\}$ (because in this graph cost of a labeling include twice more costs of edges than nodes).

For such sampled problems the algorithm converges in a small number of steps. Necessary conditions are not enough to guarantee that it reaches the solution of (11) and a significant gap is observed in experiments (we compared against MATLAB's LP-solver). However it can be still efficient in the rejection of non-optimal labels.

For each sampled problem, θ, we measure the percentage of the nodes, $\alpha(\theta)$, which can be rejected as being non-optimal based on our per-node bounds,

$$\alpha(\theta) = \frac{n_{\text{rejected}}}{|\mathcal{V}|(K-1)} \cdot 100, \tag{27}$$

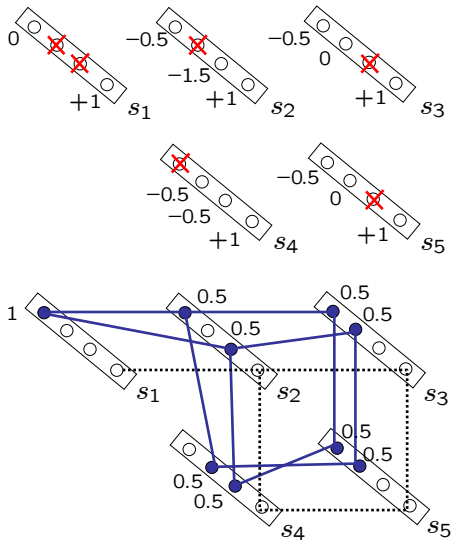

Figure 3: Example of Potts model, where optimal relaxed variables for node s_1 are integral, but do not correspond to the optimum of the discrete energy. Top: parameters $\theta_s(\cdot)$. Crosses indicate some large big numbers, say 10. Bottom: optimal relaxed labeling (solid lines) and optimal integral labeling (dashed). Numbers show values of non-zero relaxed variables $\mu_s(\cdot)$ associated to nodes.

where we divide over $(K-1)$, because if only one node in a vertex s is not rejected – then it is the only optimal solution in that vertex. Thus $\alpha(\theta) = 100$ means that an unique globally optimal solution was found. If there remain ambiguous nodes we say that we solved $\alpha(\theta)$ percents of the problem.

By sampling many random problems, we computed an empirical estimate of the probability of a random problem being solved up to a percents, $P\{\alpha(\theta) \geq a\}$. The plot of this empirical probability estimate from 10000 samples is shown in Fig. 4.

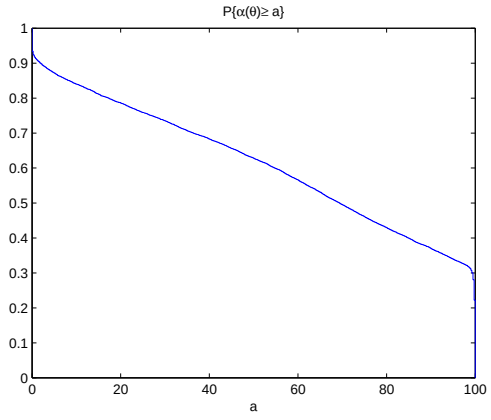

Figure 4: Empirical probability of solving a percents of a random problem. Estimated from 10000 uniform samples.

The estimate $P\{\alpha(\theta) \geq a\}$ is computed from the samples as follows. W calculate standard empirical pdf of the samples (with 100 bins), compute its cumulative distribution function, $P\{\alpha(\theta) < a\}$, and find $P\{\alpha(\theta) < a\} = 1 - P\{\alpha(\theta) \leq a\}$.

It is seen from the figure that there are about 10 percents of the problems, for which we cannot reject any of the nodes, and there are about 20 percents, for which we found the unique global minimum. Of course in practical applications the situation will be different, and we expect that real problems are much easier than random ones.

10 Conclusions

We showed that one vs. all label decompositions of Potts model gives rise to new family of lower bounds for energy minimization. The tightest lower bound is dual to a linear program, which we proved equivalent to the standard LP-relaxation. On the one hand, we cannot hope to achieve better results than a general LP-solver. On the other, we hope that this new formulation can yield a more efficient algorithm. So far, the algorithm we proposed is prone to certain stationary points which are not global minima of the relaxed problem. This behavior is similar to TRW-S and augmenting DAG algorithms: as their stationary points only satisfy necessary conditions of the optimum. Still, when general LP-solvers are not applicable, these suboptimal algorithms are of much use.

As a by-product of our algorithm, we obtain a collection of per-variable-label bounds, which can be used to reject some of non-optimal selections. This is of big advantage in applications: we often do not want to have very precise boundaries of the segmentation, *etc*. On the other hand, these bounds may be used to construct an efficient branch and bound algorithm to find the global optimum of the problem.

For each relaxation there are several ways it could be used in solving the initial discrete minimization problem. One of such possibilities is to construct a primal-dual integer method. This is likely to give another derivation of the popular expansion-move algorithm.

Acknowledgement

We thank Tomas Werner for discussions. We thank projects DIPLECS and CAK. And we thank reviewers for their useful comments.

References

[1] Authors. LP-relaxation of binarized energy minimization. Technical report, December 2007. To appear. 3

[2] E. Boros and P.L. Hammer. Pseudo-boolean optimization. Technical report, RUTCOR, October 2001. 3

[3] Yuri Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *International Conference on Computer Vision (ICCV)*, volume I, pages 105–112, 2001. 4

[4] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 26, pages 1124–1137, Sept 2004. 3

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, 23(11):1222–1239, Nov. 2001. 1, 2

[6] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM Journal on Discrete Mathematics*, 18(3):608–625, 2005. 1, 5

[7] P. Kohli and P.H.S. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV05*, volume 2, pages 922–929, 2005. 4

[8] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. In Robert G. Cowell and Zoubin Ghahramani, editors, *AI and Statistics*, pages 182–189. Society for Artificial Intelligence and Statistics, 2005. 1

[9] Fangfang Lu, Zhouyu Fu, and Antonio Robles-Kelly. Efficient graph cuts for multiclass interactive image segmentation. In Yasushi Yagi, Sing Bing Kang, In-So Kweon, and Hongbin Zha, editors, *ACCV (2)*, volume 4844 of *Lecture Notes in Computer Science*, pages 134–144. Springer, 2007. 1

[10] M.I. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika*, 4:113–130, 1976. In Russian. 1

[11] A. Shekhovtsov. Supermodular decomposition of structural labeling problem. *Control Systems and Computers*, 1:39–48, 2006. 1, 3

[12] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields. In *ECCV*, volume 1, 2006.

[13] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Exact MAP estimates by (hyper)tree agreement. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 809–816. MIT Press, 2003. 1, 2, 3, 5

[14] Tomáš Werner. A linear programming approach to max-sum problem: A review. Research Report CTU–CMP–2005–25, Center for Machine Perception, Czech Technical University, Dec. 2005. 5

[15] Tomáš Werner. A linear programming approach to max-sum problem: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, July 2007. 1