CENTER FOR MACHINE PERCEPTION

CZECH TECHNICAL UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

LP-relaxation of Binarized Energy Minimization

(Version 1.50)

Alexander Shekhovtsov[1]
Vladimir Kolmogorov[2]
Pushmeet Kohli[3]
Václav Hlaváč[1]
Carsten Rother[3]
Philip Torr[4]

shekhovt@cmp.felk.cvut.cz vnk@adastral.ucl.ac.uk pkohli@microsoft.com
hlavac@fel.cvut.cz carrot@microsoft.com philiptorr@brookes.ac.uk

[1]Czech Technical University in Prague [2]University College London

[3]Microsoft Research Cambridge [4]Oxford Brookes University

Report number CTU–CMP–2007–27

September 13, 2008

A. Shekhovtsov and V. Hlavac were supported by EC project ICT-215078 DIPLECS and by Czech Government research program MSM6840770038. V. Kolmogorov thanks EPSRC for support. P. Torr thanks EPSRC and PASCAL, EU for support.

Research Reports of CMP, Czech Technical University in Prague, No. 27, 2007

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Abstract

We address the problem of energy minimization, which is (1) generally NP-complete and (2) involves many discrete variables – commonly a 2D array of them, arising from an MRF model.

One of the approaches to the problem is to formulate it as integer linear programming and relax integrality constraints. However this can be done in a number of possible ways. One, widely applied previously (LP-1) [19, 13, 4, 22, 9, 23], appears to lead to a large-scale linear program which is not practical to solve with general LP methods. A number of algorithms were developed which attempt to solve the problem exploiting its structure [14, 23, 22, 9], however their common drawback is that they may converge to a suboptimal point. The other LP relaxation we consider here is constructed by (1) reformulating the optimization problem in the form of a function of binary variables [18], and (2) applying the roof duality relaxation [6] to the reformulated problem. We refer to the resulting relaxation as LP-2. It is different from LP-1 in many respects, and this is the main question of our study. Most importantly, it is possible to apply an efficient, fully combinatorial, algorithm to solve the relaxed problem.

We also derive the following relations: a) LP-1 is generally a tighter relaxation than LP-2, b) LP-2 provides constraints on optimal integer configurations, which allows one to identify "a part" of an optimal solution, c) a subclass of problems can be identified for which LP-2 is as tight as LP-1 providing additional characterization of solutions of LP-1 for this subclass.

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.

1 Introduction

Outline. The energy minimization problems is introduced in Sect. 2, its natural linear programming relaxation is reviewed in Sect. 3. Transformation into energy minimization with binary variables is reviewed in Sect. 5. Then we show how the known results from operations research (in particular persistencies) can be interpreted in terms of the original multi-label problem in Sect. 6. The LP-relaxation of the binarized problem is studied in more detail in Sect. 8, and the subclass of problems is studied for which this relaxation coincides with LP-1. Sect. 9 relates persistency properties derived from LP-2 with active constraints in the solution of LP-1. Sect. 10 is devoted to the submodular-supermodular decomposition approach, and it shows that the approach is equivalent to the LP-2 relaxation. In appendix we consider an order-independent reduction to binary variables and show that its linear relaxation is degenerate.

Notation

$\mathbb{R}$ will denote set of reals, $\mathbb{R}_+$ set of non-negative reals, $\mathbb{B} = \{0, 1\}$ set of "booleans", where we adopt $0 =$*false* and $1 =$*true*.
$\delta_{\{R(x)\}}$ will denote the function:

$$\delta_{\{R(x)\}} := \begin{cases} 1, & R(x) = \textit{true} \\ 0, & R(x) = \textit{false}, \end{cases}$$

where R(x) is a boolean predicate (*e.g.* "$x \geq 0$").

Ordered pair (s,t) of entities will be shorthanded by st. In particular, in notation θ_{st} it is understood that θ is indexed by an ordered pair (s,t).

Euclidean scalar product of $x, y \in \mathbb{R}^d$ will be denoted by $\langle x, y\rangle$. In vector expressions $\mathbf{0}$ and $\mathbf{1}$ will denote vectors of appropriate length of zeros and ones respectively.

Notation $\boldsymbol{\{} x_s \,|\, s \in \mathcal{S} \boldsymbol{\}}$, where S is a finite set, will stand for the concatenated vector of variables x_s (rather that for the set of their values, which is denoted by $\{x_s \,|\, s \in \mathcal{S}\}$).

2 Energy Minimization Problem

We consider the following problem illustrated in Fig. 1. Let $\mathcal{L} = \{1 \dots K\}$ be 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 the ordered pair $(s,t) \in \mathcal{E}$. Let also $\bar{\mathcal{E}} = \mathcal{E} \cup \{ts \,|\, st \in \mathcal{E}\}$ denote set of all directed edges and their reverse. Let each graph node $s \in \mathcal{V}$ be assigned a label $x_s \in \mathcal{L}$ and let a *labeling* (or *configuration*) be denoted as $\mathrm{x} = \boldsymbol{\{} x_s \,|\, s \in \mathcal{V} \boldsymbol{\}}$. Let $\boldsymbol{\{} \theta_s(i) \in \mathbb{R} \,|\, i \in \mathcal{L}\ s \in \mathcal{V} \boldsymbol{\}}$ be *univariate* potentials and $\boldsymbol{\{} \theta_{st}(i,j) \in \mathbb{R} \,|\, i,j \in \mathcal{L}\ st \in \mathcal{E} \boldsymbol{\}}$ be *pairwise* potentials. Let in addition θ_{const} be a *constant* term (meaning it is a constant function of labeling).

Let all potentials, including the constant term, be concatenated into single vector $\theta \in \Omega = \mathbb{R}^{\mathcal{I}} \times \mathbb{R}$, where set of indices $\mathcal{I} = \{(s,i) \,|\, s \in \mathcal{V},\ i \in \mathcal{L}\} \cup \{(st, ij) \,|\, st \in \mathcal{E},\ i,j \in \mathcal{L}\}$ correspond to univariate and pairwise terms. Notation $\theta_{\mathcal{I}}$ will thus refer to all components of θ but the constant term.

Let *energy* of a configuration x be defined by:

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

It is conveniently written using scalar product in Ω as $E(\mathrm{x}|\theta) = \langle \mu(\mathrm{x}), \theta \rangle$, where $\mu(\mathrm{x}) \in \Omega$ is defined by $[\mu(\mathrm{x})]_s(i) = \delta_{\{x_s = i\}}$, $[\mu(\mathrm{x})]_{st}(i,j) = \delta_{\{x_s = i\}} \delta_{\{x_t = i\}}$ and $[\mu(\mathrm{x})]_{\text{const}} = 1$.

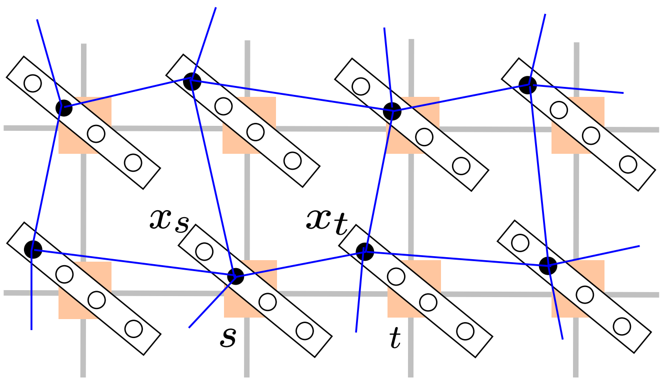

Figure 1: Energy minimization: each node s of the graph is assigned a discrete variable x_s, depicted by a box with labels. Labels in the box represent values which discrete variable may attain. A labeling x is shown by black circles and black solid lines.

3 LP-relaxation

The main subject of our analysis is the linear programming relaxation of energy minimization. The natural relaxation, studied *e.g.* by [19, 4, 22, 23] is obtained by reformulating the problem in terms of 0-1 integer variables and then relaxing integrality constraints. In this section we review this construction, the Lagrangian dual problem and the closely related notion of equivalent reparametrizations of the energy function.

Minimization of (1) can be written as:

$$\min_{\mu\in\{\mu(\mathrm{x})\,|\,\mathrm{x}\in\mathcal{L}^{\mathcal{V}}\}} \langle\mu,\theta\rangle, \tag{2}$$

where objective function is linear and therefore minimization can be performed over the convex closure of the constraint set:

$$(2) = \min_{\mu\in\mathcal{M}_{G,\mathcal{L}}} \langle\mu,\theta\rangle, \tag{3}$$

where $\mathcal{M}_{G,\mathcal{L}} = \mathrm{conv}\{\mu(\mathrm{x})\,|\,\mathrm{x}\in\mathcal{L}^{\mathcal{V}}\}$ is *marginal* polytope [22].

We consider the relaxation of (3) (see [4, 23, 22]), defined as:

$$\min_{\mu\in\Lambda_{G,\mathcal{L}}} \langle\mu,\theta\rangle, \tag{4}$$

where $\Lambda_{G,\mathcal{L}} = \big\{\mu\in\Omega_+ \mid A\mu_{\mathcal{I}} = \mathbf{0},\ B\mu_{\mathcal{I}} = \mathbf{1},\ \mu_{\mathrm{const}} = 1\big\}$ is the *local* polytope of graph G. Set Ω_+ denotes vectors from Ω with all components nonnegative. Equalities $A\mu_{\mathcal{I}} = \mathbf{0}$ express $m_1 = 2|\mathcal{E}||\mathcal{L}|$ *marginalization* constraints:

$$\begin{aligned} \sum_{j'\in\mathcal{L}} \mu_{st}(i,j') &= \mu_s(i), \quad \forall st\in\mathcal{E},\ \forall i\in\mathcal{L} \\ \sum_{i'\in\mathcal{L}} \mu_{st}(i',j) &= \mu_t(j), \quad \forall st\in\mathcal{E},\ \forall j\in\mathcal{L}, \end{aligned} \tag{5}$$

where A is $m_1\times|\mathcal{I}|$ matrix. Equalities $B\mu_{\mathcal{I}} = \mathbf{1}$ express $m_2 = |\mathcal{V}|$ *normalization* constraints:

$$\sum_{i\in\mathcal{L}} \mu_s(i) = 1, \quad \forall s\in\mathcal{V}, \tag{6}$$

where B is $m_2\times|\mathcal{I}|$ matrix. Polytope $\Lambda_{G,\mathcal{L}}$ inherits all linear equality constraints of $\mathcal{M}_{G,\mathcal{L}}$ but keeps only a small number of inequality constraints (only the constraint $\mu\in\Omega_+$), therefore it makes an outer approximation to $\mathcal{M}_{G,\mathcal{L}}$ [22].

Equivalent reparametrizations (see [23, 9] and references therein). We say that $\theta^1\in\Omega$ and $\theta^2\in\Omega$ are *equivalent*, which is denoted by $\theta^1\equiv\theta^2$, if $\langle\mu,\theta^1\rangle = \langle\mu,\theta^2\rangle$ holds for all $\mu\in\Lambda_{G,\mathcal{L}}$.

It can be shown that that the statement $\langle\mu,\phi\rangle = 0\ \forall\mu\in\Lambda_{G,\mathcal{L}}$ is equivalent to $\langle\mu,\phi\rangle = 0\ \forall\mu\in\mathcal{M}_{G,\mathcal{L}}$. Therefore $\theta^1\equiv\theta^2$ iff $E(\mathrm{x}|\theta^1) = E(\mathrm{x}|\theta^2)\ \forall\mathrm{x}\in\mathcal{L}^{\mathcal{V}}$, so our definition of equivalent problems coincides with the usual one [23, 22]. This fact follows from that $\mathrm{aff}\,\Lambda_{G,\mathcal{L}} = \mathrm{aff}\,\mathcal{M}_{G,\mathcal{L}}$ (in other words $\Lambda_{G,\mathcal{L}}$ is tight by equalities outer approximation to $\mathcal{M}_{G,\mathcal{L}}$).

Consider the set of *zero problems* defined as

$$\Omega^{\mathbf{0}} = \Big\{\phi\in\Omega \;\Big|\; \begin{matrix}\phi_{\mathcal{I}} = & A^{\mathsf{T}}y + B^{\mathsf{T}}z, \\ \phi_{\mathrm{const}} = & -\langle\mathbf{1},z\rangle\end{matrix},\ y\in\mathbb{R}^{m_1},\ z\in\mathbb{R}^{m_2}\Big\}. \tag{7}$$

Problems from this set and only they have the property $E(\mathrm{x}|\phi) = 0$ for all $x\in\mathcal{L}^{\mathcal{V}}$.

Statement 1. It is $\theta^1 \equiv \theta^2$ iff $\theta^1 - \theta^2 \in \Omega^{\mathbf{0}}$.

Proof. Clearly, the statement can be reformulated as $\phi \equiv \mathbf{0}$ iff $\phi \in \Omega^{\mathbf{0}}$.
- For any $\phi \in \Omega^{\mathbf{0}}$ and any $\mu \in \Lambda_{G,\mathcal{L}}$ it is $\langle \mu, \phi \rangle = \langle \mu_{\mathcal{I}}, A^{\mathsf{T}}y + B^{\mathsf{T}}z \rangle - \langle \mathbf{1}, z \rangle = \langle A\mu_{\mathcal{I}}, y \rangle + \langle B_{\mathcal{I}}\mu, z \rangle - \langle \mathbf{1}, z \rangle = \langle \mathbf{1}, z \rangle - \langle \mathbf{1}, z \rangle = 0$.
- Let $\langle \phi, \mu \rangle = 0$ holds for all $\mu \in \Lambda_{G,\mathcal{L}}$. Then it holds also for all $\mu \in \operatorname{aff} \Lambda_{G,\mathcal{L}} = \{\mu \in \Omega \mid A\mu_{\mathcal{I}} = \mathbf{0},\ B\mu_{\mathcal{I}} = \mathbf{1},\ \mu_{\text{const}} = 1\}$. Which means that the equation $\phi^{\mathsf{T}}\mu = 0$ is a linear combination of equations $A\mu_{\mathcal{I}} = \mathbf{0}$, $B\mu_{\mathcal{I}} = \mathbf{1}$ and $\mu_{\text{const}} = 1$. That is $\exists y \in \mathbb{R}^{m_1}$, $z \in \mathbb{R}^{m_2}$: $\phi_{\mathcal{I}} = A^{\mathsf{T}}y + B^{\mathsf{T}}z$, $\phi_{\text{const}} = -\langle \mathbf{1}, z \rangle$. $\square$

Often it is useful to consider the explicit form of elements $\phi \in \Omega^{\mathbf{0}}$. When matrices A and B are substituted, it takes the form [19, 23, 9]:

$$\begin{array}{rcll} \phi_{st}(i,j) & = & y_{st}(i) + y_{ts}(j), & \forall st \in \mathcal{E},\ i,j \in \mathcal{L} \\ \phi_s(i) & = & -\sum_{t|st \in \bar{\mathcal{E}}} y_{st}(i) + z_s, & \forall s \in \mathcal{V},\ i \in \mathcal{L} \\ \phi_{\text{const}} & = & -\sum_{s \in \mathcal{V}} z_s, & \end{array} \tag{8}$$

where it is understood that $y \in \mathbb{R}^{m_1}$ has components $\{y_{st}(i) \mid st \in \bar{\mathcal{E}},\ i \in \mathcal{L}\}$ and $z \in \mathbb{R}^{m_2}$ has components $\{z_s \mid s \in \mathcal{V}\}$. In [19, 23] it is shown that when graph G is connected and only problems with zero constant term are considered, the space $\Omega^{\mathbf{0}}$ can be parametrized without variables z. Indeed, in this case it is $\{B^{\mathsf{T}}z \mid z \in \mathbb{R}^{m_2},\ \langle \mathbf{1}, z \rangle = 0\} \subseteq \{A^{\mathsf{T}}y \mid y \in \mathbb{R}^{m_1}\}$.

LP-dual. We write dual of (4) as follows (standard LP dual with y, z being dual variables):

$$\begin{array}{ll} \min \langle \mu_{\mathcal{I}}, \theta_{\mathcal{I}} \rangle + \theta_{\text{const}} = & \max \langle \mathbf{1}, z \rangle + \theta_{\text{const}}. \\ A\mu_{\mathcal{I}} = \mathbf{0} & y \in \mathbb{R}^{m_1} \\ B\mu_{\mathcal{I}} = \mathbf{1} & z \in \mathbb{R}^{m_2} \\ \mu_{\mathcal{I}} \geq 0 & A^{\mathsf{T}}y + B^{\mathsf{T}}z \leq \theta_{\mathcal{I}} \end{array} \tag{9}$$

Introducing auxiliary variables $\theta' \in \Omega$, (9) can be written as:

$$\begin{array}{ll} = \max \theta'_{\text{const}} & = \max \theta'_{\text{const}}\ . \\ \theta'_{\mathcal{I}} = \theta_{\mathcal{I}} - A^{\mathsf{T}}y - B^{\mathsf{T}}z & \theta' \equiv \theta \\ \theta'_{\text{const}} = \theta_{\text{const}} + \langle \mathbf{1}, z \rangle & \theta'_{\text{const}} \in \mathbb{R} \\ \theta'_{\mathcal{I}} \geq 0 & \theta'_{\mathcal{I}} \geq 0 \end{array} \tag{LB}$$

The weak duality theorem implies that for any θ' feasible to the dual problem, the dual cost, θ'_{const} is a not grater then the primal cost, $\langle \mu, \theta \rangle$, $\forall \mu \in \Lambda_{G,\mathcal{L}}$. That is θ'_{const} is a *lower bound* on $\langle \mu, \theta \rangle$ and, in particular, on $E(\mathrm{x}|\theta) = \langle \mu(\mathrm{x}), \theta \rangle$.

The final expression LB can be seen to have similar form to the problem of maximizing the constant term of a *posiform* [1]. More precisely this relation can be expressed by the following statement.

Statement 2. When $|\mathcal{L}| = 2$, bound LB coincides with the *roof-dual* bound [1]. In this case optimization problem LB can be efficiently solved by a max-flow algorithm [2].

4 Binary Energy Minimization

Energy minimization problems with 2 labels are conveniently described in terms of 0-1 integer variables, which will be called *binary* throughout this paper.

Let $\mathcal{L} = \mathbb{B} = \{0, 1\}$, then each variable x_s is 0 or 1. Univariate and pairwise terms of (1) may be written as

$$\begin{array}{l}\theta_s(x_s) = \theta_{st}(1)x_s + \theta_{st}(0)(1 - x_s),\\ \theta_{st}(x_s, x_t) = \theta_{st}(1,1)x_s x_t + \theta_{st}(0,1)(1 - x_s)x_t\\ +\theta_{st}(1,0)x_s(1 - x_t) + \theta_{st}(0,0)(1 - x_s)(1 - x_t).\end{array} \tag{10}$$

Expanding braces in (10) it is clear that (1) may be written in the form:

$$E(\mathrm{x}|\eta) = \sum_s \eta_s x_s + \sum_{st} \eta_{st} x_s x_t + \eta_{\mathrm{const}}, \tag{11}$$

which is a quadratic polynomial in binary variables x_s. Functions of the form $\mathbb{B}^{\mathcal{V}} \mapsto \mathbb{R}$ are called *pseudo-Boolean* [1] and minimization (or maximization) of (11) is called quadratic pseudo-Boolean optimization.

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

5 Transformation to Binary Variables

Minimization of energy (1) can be always formulated as a MIN-CUT problem. Energies which correspond to polynomially solvable MIN-CUT (with all weights nonnegative) include energies with *convex* pairwise potentials [8] and, more generally, submodular pairwise potentials of binary [7, 11] or multi-label [20, 16, 10] variables. Reversely, any MIN-CUT problem can be formulated as energy minimization with 2 labels.

As there are simple transitions between MIN-CUT problem, quadratic pseudo-Boolean optimization and energy minimization with 2 labels it is not very important to which of them energy minimization with many labels will be reduced. The construction [8, 16, 18] adopted to our notation of binary energies is as follows.

Transformation $\mathcal{L} \to \mathbb{B}$. First, we construct $\hat{\theta} \equiv \theta$ satisfying the following:

$$\begin{array}{rll}\hat{\theta}_{st}(1,j) = \hat{\theta}_{st}(i,1) = & 0 & st \in \mathcal{E},\ i,j \in \mathcal{L}\\ \hat{\theta}_s(1) = & 0 & s \in \mathcal{V}.\end{array} \tag{12}$$

It is constructed as follows:

$$\begin{aligned}\hat{\theta}_{st}(i,j) &= \theta_{st}(i,j) - \theta_{st}(i,1) - \theta_{st}(1,j) + \theta_{st}(1,1), \quad st \in \mathcal{E},\ i,j \in \mathcal{L};\\ \hat{\theta}_s(i) &= \theta_s(i) + \sum_{t|\,st\in\mathcal{E}} \theta_{st}(i,1) + \sum_{t|\,ts\in\mathcal{E}} \theta_{ts}(1,i) - \theta_s(1) - \sum_{t|\,st\in\mathcal{E}} \theta_{st}(1,1) -\\ &\quad - \sum_{t|\,ts\in\mathcal{E}} \theta_{ts}(1,1), \quad s \in \mathcal{V},\ i \in \mathcal{L};\\ \hat{\theta}_{\mathrm{const}} &= \theta_{\mathrm{const}} + \sum_{st\in\mathcal{E}} \theta_{st}(1,1) + \sum_{s\in\mathcal{V}} \theta_s(1).\end{aligned} \tag{13}$$

It is easy to verify that condition (12) holds for $\hat{\theta}$ and that $\hat{\theta} \equiv \theta$. Note, that in the case of two labels, reparametrisation $\hat{\theta}$ immediately provide coefficients for (11). In more general case it is useful because of possibility to apply the next lemma.

Lemma 1 (Cumulative sum)**.** Let $f : \{1 \dots K\} \mapsto \mathbb{R}$, let $f(1) = 0$. Then the following representation holds for f:

$$f(i) = \sum_{2 \le i' \le i} D_{i'} f \quad i = 1 \dots K, \tag{14}$$

where $D_i f = f(i) - f(i-1), \quad i = 2 \dots K$, is the *discrete derivative* of f in i, and the result of the summation is assumed to be 0 if there is no summands (when $i = 1$).
Analogously, for a function $g : \{1 \dots K\}^2 \mapsto \mathbb{R}$ with $g(\cdot, 1) = g(1, \cdot) = 0$ it holds:

$$g(i,j) = \sum_{\substack{2 \le i' \le i \\ 2 \le j' \le j}} D_{i'j'} g \quad i, j = 1 \dots K, \tag{15}$$

where $D_{ij} g = g(i,j) + g(i-1, j-1) - g(i, j-1) - g(i-1, j), \quad i, j = 2 \dots K$.

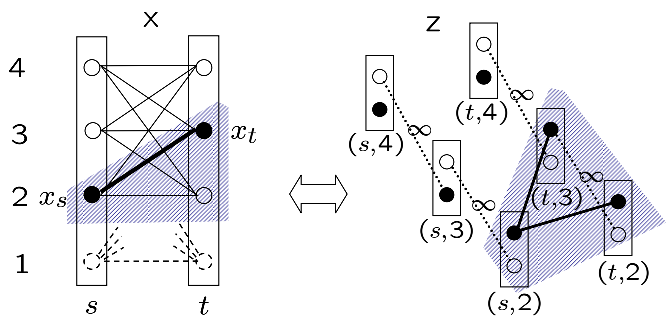

Figure 2: Transformation $\mathcal{L} \mapsto \mathbb{B}$. Left: an interaction pair $st \in \mathcal{E}$; a labeling x is shown by black circles; lowest labels are dashed since all weights in $\hat{\theta}$ associated with them are 0. Right: binary variables $z_{(s,i)}, z_{(t,j)}$; labeling $\mathrm{z}(\mathrm{x})$ is shown by black circles; dashed lines marked with ∞ show hard constraints. Shaded areas show the set $\{(s,i) \,|\, i \le x_s,\ s \in \mathcal{V},\ i \in \tilde{\mathcal{L}}\}$.

Let a tuple $(\mathcal{L}, G, \theta)$ define the energy minimization problem. Let $\tilde{\mathcal{L}} = \{2 \dots K\}$. Then components of the equivalent binary energy minimization problem are as follows (Fig 2):

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

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

$$\mathrm{z}(\mathrm{x})_{(s,i)} = \delta_{\{i \le x_s\}}, \quad (s,i) \in V. \tag{16}$$

III. Binary energy function

$$E(\mathrm{z}|\eta) = \sum_{u \in V} \eta_u z_u + \sum_{uv \in A_{\mathcal{E}}} \eta_{uv} z_u z_v + \eta_{\mathrm{const}} + H(\mathrm{z}), \tag{17}$$

where weights η are constructed as

$$\begin{aligned} \eta_{(s,i)(t,j)} &= D_{ij}\hat{\theta}_{st} = D_{ij}\theta_{st} && st \in \mathcal{E},\ i,j \in \tilde{\mathcal{L}} \\ \eta_{(s,i)} &= D_i\hat{\theta}_s && s \in \mathcal{V},\ i \in \tilde{\mathcal{L}} \\ \eta_{\text{const}} &= \hat{\theta}_{\text{const}}; \end{aligned} \tag{18}$$

and hard constraints read

$$H(\mathrm{z}) = \sum_{uv \in A_{\mathcal{V}}} h(z_u, z_v), \tag{19}$$

where $h(z_{(s,i)}, z_{(s,i-1)}) = 0$ if $z_{(s,i)} \leq z_{(s,i-1)}$ and ∞ otherwise, $i = 3 \ldots K$. Hard constraints ensure that any z with finite energy is in the form (16).

IV. For a binary configuration z of finite energy the corresponding configuration x, denoted as $\mathrm{x}(\mathrm{z})$, is found as

$$x_s = 1 + \sum_{i \in \tilde{\mathcal{L}}} z_{(s,i)}. \tag{20}$$

Statement 3. Constructed binary energy is equivalent to the original multi-label energy: For all $\mathrm{x} \in \mathcal{L}^{\mathcal{V}}$ it is $E(\mathrm{z}(\mathrm{x})|\eta) = E(\mathrm{x}|\theta)$.

Proof. Let $\mathrm{z} = \mathrm{z}(\mathrm{x})$. Using (16) and applying Lemma 1 pairwise terms in (17) expand as

$$\begin{aligned} \sum_{uv \in A_{\mathcal{E}}} \eta_{uv} z_u z_v &= \sum_{st \in \mathcal{E}} \sum_{i,j \in \tilde{\mathcal{L}}} \eta_{(s,i)(t,j)} z_{(s,i)} z_{(t,j)} = \\ &= \sum_{st \in \mathcal{E}} \sum_{\substack{2 \leq i' \leq x_s \\ 2 \leq j' \leq x_t}} D_{i'j'}\hat{\theta}_{st} = \sum_{st \in \mathcal{E}} \hat{\theta}_{st}(x_s, x_t), \end{aligned} \tag{21}$$

and univariate terms as:

$$\sum_{u \in V} \eta_u z_u = \sum_{s \in \mathcal{V}} \sum_{2 \leq i' \leq x_s} D_{i'}\hat{\theta}_s = \sum_{s \in \mathcal{V}} \hat{\theta}_s(x_s). \tag{22}$$

Therefore $E(\mathrm{z}(\mathrm{x})|\eta) = E(\mathrm{x}|\hat{\theta}) = E(\mathrm{x}|\theta)$. □

5.1 Dependence on the Label Order

The construction [8, 16, 18] outlined above depends on the ordering of the set of labels $\mathcal{L}$ (separate label sets $\mathcal{L}_s$ with different order for each node s can be considered). While for all discrete configurations the binarized problem is equivalent to the multi-label problem, the linear relaxation of it is not. Thus all the results derived in the sequel will depend on the selected order of labels, which is a significant limitation. Note that there is a combinatorial number of such orderings and solving all of the obtained in such way relaxations does not seem feasible.

An order-independent reduction to binary variables can be obtained by introducing binary variables as $\mathrm{z}_{(s,i)} = \delta_{\{i=x_s\}},\quad (s,i) \in V$. It is possible to construct weights of the binary problem in such a way that optimal binary configurations z corresponds to optimal configurations x of the multi-label problem.

It is achieved by enforcing constraint $\sum_i z_{(s,i)} \leq 1$ via $K(K-1)/2$ hard pairwise terms and constraints $\sum_i z_{(s,i)} >= 1$ by adding sufficiently large negative value to all unary terms. Unfortunately, this many hard constraints lead to full loss of tightness of the LP relaxation. We show in the appendix that this LP relaxation is degenerate.

6 Optimality Properties

Finding a minimizer of a function $E : \mathbb{B}^V \mapsto \mathbb{R}$ is generally NP-complete, however it is often possible to find certain constraints on the set of minimizers. In particular, if for some element $u \in V$ and some $\alpha \in \mathbb{B}$ it is known that any minimizer z must satisfy $z_u = \alpha$, then it is said that *strong persistency*[1, 3] holds for (u, α) or that (u, α) provides *part* of optimal solution. It is easy to see that all constraints of this form may be expressed as:

$$\mathrm{z}^{\min} \leq \mathrm{z} \leq \mathrm{z}^{\max}, \tag{23}$$

where $\mathrm{z}^{\min}, \mathrm{z}^{\max} \in \mathbb{B}^V$ and inequalities are component-wise (e.g., $0 \leq z_s \leq 1$ means no constraints on z_s, whereas $0 \leq z_s \leq 0$ means z_s is constrained to be 0). We will say that pair $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ defines *strong* (resp. *weak*) *persistency* if (23) holds for all (resp. for some) minimizers of E. Following [3] we will distinguish the notion of *autarkies*.

Let $\mathrm{z} \vee \mathrm{z}'$ denote component-wise maximum of z and z', let $\mathrm{z} \wedge \mathrm{z}'$ denote component-wise minimum of z and z'. We will say that pair $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ is *strong* (resp. *weak*) *autarky* for E if inequality

$$E((\mathrm{z} \vee \mathrm{z}^{\min}) \wedge \mathrm{z}^{\max}) < E(\mathrm{z}) \tag{24}$$

$$(\text{resp. } E((\mathrm{z} \vee \mathrm{z}^{\min}) \wedge \mathrm{z}^{\max}) \leq E(\mathrm{z}) \ ) \tag{25}$$

holds for all $\mathrm{z} \in \mathbb{B}^{\mathcal{V}}, \mathrm{z} \neq (\mathrm{z} \vee \mathrm{z}^{\min}) \wedge \mathrm{z}^{\max}$. It is easy to see that if $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ is strong (resp. weak) autarky for E then it also implies strong (resp. weak) persistency. Generally, not all persistencies can be derived via autarkies but we will consider only such type of persistencies since network flow model [2] efficiently computes strong (resp. weak) autarky $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ as a pair of "extreme" (resp. arbitrary) minimum cuts.

We will consider now the form which persistencies and autarkies take when transfered to multi-label setting by mapping $\mathrm{x}(\mathrm{z})$ defined by (20). For simplicity we will show only strong properties.

Statement 4. Let $(G, \mathcal{L}, \theta)$ define a multi-label energy minimization problem. Let $(N, \mathbb{B}, \eta)$ define the corresponding binary energy minimization problem and 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 $\mathrm{x} \in \mathrm{argmin}_{\mathcal{L}^{\mathcal{V}}} E(\cdot|\theta)$ must satisfy

$$\mathrm{x}^{\min} \leq \mathrm{x} \leq \mathrm{x}^{\max}, \tag{26}$$

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

Proof. Let us first note that conditions $E(\mathrm{z}^{\min}) < \infty$, $E(\mathrm{z}^{\max}) < \infty$ are needed in order that $\mathrm{x}^{\min}$ and $\mathrm{x}^{\max}$ are correctly defined. If they are not satisfied we can easily find tighter constraints $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ which does satisfy them.

Let us show that $\mathrm{x}^{\min} \leq \mathrm{x}$. Configuration $\mathrm{z}^{\min}$ satisfies hard constraints, therefore for $\mathrm{x}^{\min} = \mathrm{x}(\mathrm{z}^{\min})$ it holds $E(\mathrm{x}^{\min}|\theta) = E(\mathrm{z}^{\min}|\eta)$. Let $\mathrm{z} = \mathrm{z}(\mathrm{x})$, then z is optimal for $E(\cdot|\eta)$ and therefore (23) holds. We then verify that

$$\begin{aligned} z_{(s,i)} &\geq z^{\min}_{(s,i)}, \\ 1 + \sum_{i\in\tilde{\mathcal{L}}} z_{(s,i)} &\geq 1 + \sum_{i\in\tilde{\mathcal{L}}} z^{\min}_{(s,i)}, \\ \mathrm{x} &\geq \mathrm{x}(\mathrm{z}^{\min}). \end{aligned}$$

Constraint $\mathrm{x} \leq \mathrm{x}(\mathrm{z}^{\max})$ is verified similarly. □

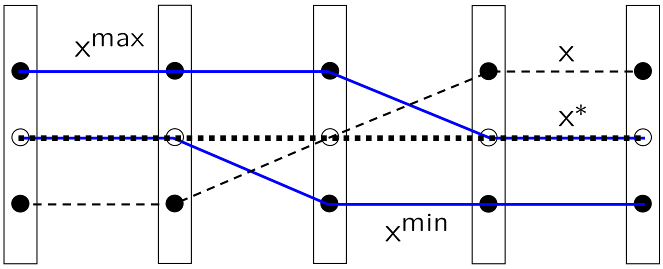

Figure 3: Strong autarky $(\mathrm{x}^{\min}, \mathrm{x}^{\max})$: if a labeling x (thin dashed) does not satisfy constraints (26) then labeling $\mathrm{x}^* = (\mathrm{x} \vee \mathrm{x}^{\min}) \wedge \mathrm{x}^{\max}$ (thick dashed) does satisfy them and it posseses a lower energy.

Statement 5. Let $(\mathrm{z}^{\min}, \mathrm{z}^{\max})$ be a strong autarky for $E(\cdot|\eta)$ such that $E(\mathrm{z}^{\min}) < \infty$ and $E(\mathrm{z}^{\max}) < \infty$. Then

$$\begin{aligned} &E((\mathrm{x} \vee \mathrm{x}^{\min}) \wedge \mathrm{x}^{\max}|\theta) < E(\mathrm{x}|\theta) \\ &\forall \mathrm{x} \in \mathcal{L}^{\mathcal{V}},\ \mathrm{x} \neq (\mathrm{x} \vee \mathrm{x}^{\min}) \wedge \mathrm{x}^{\max}, \end{aligned} \tag{27}$$

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

Proof. Let $\mathrm{x} \in \mathcal{L}^{\mathcal{V}}$ and let $\mathrm{z} = \mathrm{z}(\mathrm{x})$. We will show that $(\mathrm{x} \vee \mathrm{x}^{\min}) \wedge \mathrm{x}^{\max} = \mathrm{x}((\mathrm{z} \vee \mathrm{z}^{\min}) \wedge \mathrm{z}^{\max})$, then the statement will follow by equivalence of energies $E(\cdot|\theta)$ and $E(\cdot|\eta)$. Let us show that for any $\mathrm{z}, \mathrm{z}' \in \mathbb{B}^{\mathcal{V}}$ it is $\mathrm{x}(\mathrm{z} \vee \mathrm{z}') = \mathrm{x}(\mathrm{z}) \vee \mathrm{x}(\mathrm{z}')$, indeed

$$\begin{aligned} \mathrm{x}(\mathrm{z} \vee \mathrm{z}')_s = 1 + \sum_{i\in\tilde{\mathcal{L}}} \max(z_{(s,i)}, z'_{(s,i)}) = \\ 1 + \sum_{i\in\tilde{\mathcal{L}}} \delta_{\{i \leq \mathrm{x}(\mathrm{z})_s\}} \vee \delta_{\{i \leq \mathrm{x}(\mathrm{z}')_s\}} = \\ \max(\mathrm{x}(z)_s, \mathrm{x}(\mathrm{z}')_s). \end{aligned} \tag{28}$$

Similarly, we can verify that for any $\mathrm{z}, \mathrm{z}' \in \mathbb{B}^{\mathcal{V}}$ it is $\mathrm{x}(\mathrm{z} \wedge \mathrm{z}') = \mathrm{x}(\mathrm{z}) \wedge \mathrm{x}(\mathrm{z}')$. □

The statement shows how a strong autarky for binary energy is interpreted for multi-label energy, this interpretation is illustrated by Fig. 3.

7 Submodular Energy Minimization

Let $\mathrm{x} \vee \mathrm{x}'$ denote component-wise maximum and $\mathrm{x} \wedge \mathrm{x}'$ component-wise minimum of x and x′. Let us introduce the mapping $S : \mathcal{L}^{\mathcal{V}} \mapsto 2^V$ defined as

$$S(\mathrm{x}) = \{(s,i) \,|\, s \in \mathcal{V},\, 2 \leq i \leq x_s\}, \tag{29}$$

see Fig. 2. Union $S(x) \cup S(x')$ corresponds to $S(x \vee x')$ and intersection $S(x) \cap S(x')$ to $S(x \wedge x')$. Thus a set of subsets $U = \{S(\mathrm{x}) \,|\, \mathrm{x} \in \mathcal{L}^{\mathcal{V}}\}$ is closed under union and intersection. A function of subsets $E' : U \mapsto \mathbb{R}$ is called *submodular* if $E'(A \cup B) + E'(A \cap B) \leq E'(A) + E'(B)$ for all $A, B \subseteq U$.

It is seen therefore that certain energies $E(\mathrm{x}|\theta)$ may be identified with submodular set functions via $E'(S(\mathrm{x})) = E(\mathrm{x}|\theta)$. The submodularity of $E(\mathrm{x}|\theta)$ is then expressed as: $E(\mathrm{x} \vee \mathrm{x}') + E(\mathrm{x} \wedge \mathrm{x}') \leq E(\mathrm{x}) + E(\mathrm{x}')\ \forall \mathrm{x}, \mathrm{x}' \in \mathcal{L}^{\mathcal{V}}$. Vector θ in this case will be also called submodular and the condition may be expressed simpler (e.g. [23]) as:

$$\theta_{st}(x_{st}) + \theta_{st}(y_{st}) \geq \theta_{st}(x_{st} \wedge y_{st}) + \theta_{st}(x_{st} \vee y_{st}) \quad \forall x_{st}, y_{st} \in \mathcal{L}^2\ \forall st \in \mathcal{E}, \tag{30}$$

which means that all pairwise terms θ_{st} are submodular.

It can be seen (e.g. [5, 15]) that θ_{st} satisfies (30) iff

$$D_{ij}\theta_{st} \leq 0, \quad \forall i, j \in \{2 \ldots K\}. \tag{31}$$

We see now that when θ is submodular then pairwise coefficients of binary energy (17) calculated by (18) are not positive which is precisely the condition when binary energy is submodular and can be minimized exactly by max-flow.

Let Ω^{sub} denote the set of submodular vectors θ, let $\Omega^{\mathrm{sup}} = -\Omega^{\mathrm{sub}}$ denote set of supermodular vectors. Submodularity and supermodularity will play important role in our analysis. Note that for $\phi \in \Omega^{\mathbf{0}}$ it is $E(\cdot|\phi) \equiv 0$ therefore $\Omega^{\mathbf{0}} \subseteq \Omega^{\mathrm{sub}} \cap \Omega^{\mathrm{sup}}$.

8 LP-relaxation of Binary Energy

We will construct a linear relaxation of $\min_{\mathrm{z}} E(\mathrm{z}|\eta)$ explicitly. Also it will coincide with LP-relaxation as was defined by (4), a part of variables related by equality constraints will be excluded. In this form it is also easily identified with the *linearization* approach in quadratic pseudo-Boolean optimization [1]. Let us replace each variable z_u by a relaxed variable $\nu_u \in [0,\ 1]$ and each product $z_u z_v$, where $uv \in \mathcal{A}_{\mathcal{E}}$, by a relaxed variable $\nu_{uv} \in [0,\ 1]$. We restrict relaxed variables to satisfy the following constraints:

$$\begin{gathered} 0 \leq \nu_u \leq 1 \ \ \forall u \in V, \\ \max(0, \nu_u + \nu_v - 1) \leq \nu_{uv} \leq \min(\nu_u, \nu_v) \ \ \forall uv \in A_{\mathcal{E}}, \end{gathered} \tag{32}$$

which hold automatically for all 0-1 assignments $\nu_{uv} = \mathrm{z}_u z_v$, $\nu_u = z_u$, $\nu_v = z_v$. Moreover, for pairs $uv \in A_{\mathcal{V}}$ the inequality $z_u \leq z_v$ imposed by hard constraints (19) must hold, therefore corresponding relaxed variables ν_u, ν_v must satisfy:

$$\nu_{(s,i)} \leq \nu_{(s,i-1)} \quad \forall s \in \mathcal{V}\ i = 3 \ldots K. \tag{33}$$

Accordingly, we defile local polytope as

$$\hat{\Lambda}_{N,\mathbb{B}} = \{\nu \in \mathbb{R}_+^{V \cup A_\mathcal{E} \cup \{\text{const}\}} \mid (32) \text{ and } (33) \text{ hold for } \nu;\ \nu_{\text{const}} = 1\}. \tag{34}$$

LP-relaxation takes the form:

$$\min_{\nu \in \hat{\Lambda}_{N,\mathbb{B}}} \Big[\sum_{u\in V} \eta_u \nu_u + \sum_{uv \in A_\mathcal{E}} \eta_{uv}\nu_{uv} + \eta_{\text{const}}\nu_{\text{const}}\Big]. \tag{35}$$

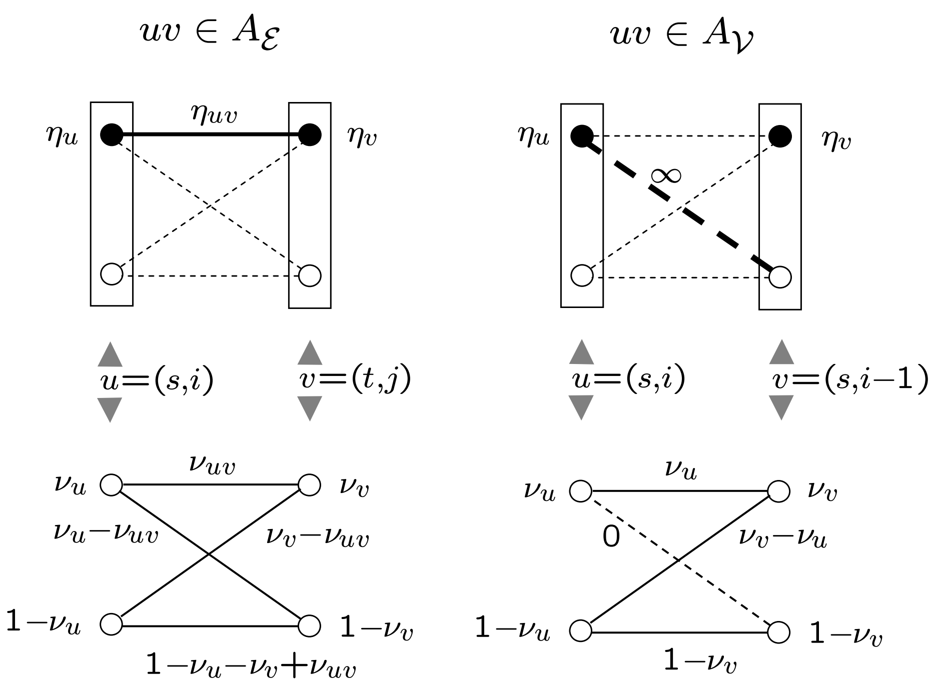

Figure 4: Simplifying LP relaxation of binary energy: equivalent assignment of $\tilde{\eta}$ and $\tilde{\nu}$ is shown for graph edges $uv \in A_\mathcal{E}$ (left) and $uv \in A_\mathcal{V}$ (right). Thin dashed lines have associated values zero.

Statement 6. Problem (35) can be equivalently written in the form (4) as $\min_{\tilde{\nu}\in\Lambda_{N,\mathbb{B}}} \langle \tilde{\nu}, \tilde{\eta} \rangle$ (see Fig. 4).

Proof. Let $\tilde{\eta}$ be defined as:[1]

$$\begin{aligned} &\tilde{\eta}_u = (0,\ \eta_u), \quad u \in V \\ &\tilde{\eta}_{uv} = \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & \eta_{uv} \\ \hline \end{array}, \quad uv \in A_\mathcal{E} \\ &\tilde{\eta}_{uv} = \begin{array}{|c|c|}\hline 0 & 0 \\ \hline \infty & 0 \\ \hline \end{array}, \quad uv \in A_\mathcal{V} \\ &\tilde{\eta}_{\text{const}} = \eta_{\text{const}} \end{aligned} \tag{36}$$

[1]We write functions $f : \mathbb{B} \mapsto \mathbb{R}$ as $(f(0), f(1))$ and functions $g : \mathbb{B}^2 \mapsto \mathbb{R}$ as $\begin{array}{|c|c|}\hline g(0,0) & g(0,1) \\ \hline g(1,0) & g(1,1) \\ \hline \end{array}$.

Let a relaxed labeling $\tilde{\nu} \in \Lambda_{N,\mathbb{B}}$ has finite energy: $\langle \tilde{\nu}, \tilde{\eta} \rangle < \infty$. By exploiting equality constraints of $\Lambda_{N,\mathbb{B}}$ we verify that any such $\tilde{\nu}$ can be written in the form:

$$
\begin{aligned}
&\tilde{\nu}_u = (1 - \nu_u,\ \nu_u), \quad u \in V \\
&\tilde{\nu}_{uv} = \begin{array}{|c|c|}\hline 1{-}\nu_u{-}\nu_v{+}\nu_{uv} & \nu_v - \nu_{uv} \\ \hline \nu_u{-}\nu_{uv} & \nu_{uv} \\ \hline \end{array}, \quad uv \in A_{\mathcal{E}} \\
&\tilde{\nu}_{uv} = \begin{array}{|c|c|}\hline 1{-}\nu_v & \nu_v{-}\nu_u \\ \hline 0 & \nu_u \\ \hline \end{array}, \quad uv \in A_{\mathcal{V}}
\end{aligned}
\tag{37}
$$

which provides a one-to-one mapping of relaxed labellings $\tilde{\nu} \in \Lambda_{N,\mathbb{B}}$ of finite energy and relaxed labellings $\nu \in \hat{\Lambda}_{N,\mathbb{B}}$. Note that constraints $\tilde{\nu}_{st} \geq 0$ are equivalent to (32). It is seen that under such defined mapping it holds $\langle \tilde{\nu}, \tilde{\eta} \rangle = \langle \nu, \eta \rangle$. □

It was observed that LP-relaxation of binary energy minimization always has a half-integral optimal solution, i.e. with $\nu \in \{0, \frac{1}{2}, 1\}^{V \cup A_{\mathcal{E}} \cup \{\text{const}\}}$ [1, 23].

Let $\nu \in \hat{\Lambda}_{N,\mathbb{B}}$ be an optimal half-integral relaxed labeling. Under constraints (33) it implies that for all $s \in \mathcal{V}$ components $\{\nu_{(s,i)} \mid i \in \tilde{\mathcal{L}}\}$ are in the form

$$
(\underbrace{1 \dots 1}_{n_1}, \underbrace{\tfrac{1}{2}, \dots, \tfrac{1}{2}}_{n_2}, \underbrace{0 \dots 0}_{n_3}), \tag{38}
$$

where $n_1, n_2, n_3 \geq 0$, $n_1{+}n_2{+}n_3 = |\tilde{\mathcal{L}}|$. Further on, under constraints (32) for all $st \in \mathcal{E}$ components $\{\nu_{(s,i)(t,j)} \mid i, j \in \tilde{\mathcal{L}}\}$ are in the form (i-row, j-column):

$$
\begin{array}{c|c|c|c|c}
 & 1 \dots 1 & \frac{1}{2} \dots \frac{1}{2} & 0 \dots 0 & \nu_{(t,j)} \\ \hline
1 \ \vdots \ 1 & \mathbf{1} & \mathbf{\tfrac{1}{2}} & \mathbf{0} & \\ \hline
\frac{1}{2} \ \vdots \ \frac{1}{2} & \mathbf{\tfrac{1}{2}} & \{\mathbf{0}, \mathbf{\tfrac{1}{2}}\} & \mathbf{0} & \\ \hline
0 \ \vdots \ 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \\ \hline
\nu_{(s,i)} & & & &
\end{array}
\tag{39}
$$

where corresponding values $\nu_{(s,i)}$ and $\nu_{(t,j)}$ appearing in constrains (32) are shown on the sides of the table. For example, constraints $\max(0, \frac{1}{2} + 1 - 1) \leq x \leq \min(\frac{1}{2}, 1)$ imply $x = \frac{1}{2}$. In the central part of the table values ν_{uv} with marginals $\nu_u = \nu_v = \frac{1}{2}$ are restricted to be in the set $\{0, \frac{1}{2}\}$.

Statement 7. Let θ_{st} be a submodular (resp. supermodular) pairwise term of the multi-label energy minimization problem. Let $\{\nu_{(s,i)(t,j)} \mid i, j \in \tilde{\mathcal{L}}\}$ be associated values of the relaxed labeling in the binarized problem. Then an increase (resp. decrease) of $\nu_{(s,i)(t,j)}$ does not lead to an increase of the associated part of objective (35), $\sum\limits_{i,j \in \tilde{\mathcal{L}}} \eta_{(s,i)(t,j)} \nu_{(s,i)(t,j)}$.

Proof. It is $\eta_{(s,i)(t,j)} = D_{ij}\theta_{st} \leq 0$ (resp. ≥ 0). Therefore an increase (resp. decrease) of any of $\nu_{(s,i)(t,j)}$ does not increase the objective. □

Statement 7 implies that for problems where each interaction term θ_{st} is either submodular or supermodular there exists an optimal solution ν^* of (35) with central part of ν^*_{uv} set to $\frac{1}{2}$ for submodular θ_{st} and to 0 for supermodular θ_{st}.

More generally, for arbitrary $v \in \hat{\Lambda}_{N,\mathbb{B}}$ let us introduce two mappings projecting pairwise components of ν on the corresponding side of the constraints (32):

$$U : \hat{\Lambda}_{N,\mathbb{B}} \mapsto \hat{\Lambda}_{N,\mathbb{B}}, \quad \text{defined by} \begin{cases} \mathrm{U}_u(\nu) = \nu_u, & u \in V \\ \mathrm{U}_{uv}(\nu) = \min(\nu_u, \nu_v), & uv \in A_{\mathcal{E}} \\ \mathrm{U}_{\text{const}}(\nu) = 1; \end{cases} \tag{40}$$

$$L : \hat{\Lambda}_{N,\mathbb{B}} \mapsto \hat{\Lambda}_{N,\mathbb{B}}, \quad \text{defined by} \begin{cases} \mathrm{L}_u(\nu) = \nu_u, & u \in V \\ \mathrm{L}_{uv}(\nu) = \max(0, \nu_u + \nu_v - 1), & uv \in A_{\mathcal{E}} \\ \mathrm{L}_{\text{const}}(\nu) = 1. \end{cases} \tag{41}$$

Clearly, mapping U preserves the optimality of the solution for submodular problems, while mapping L preserves optimality of solutions for supermodular problems. Another important property of these mappings is given by the following statement.

Statement 8. Result of the mapping $\nu^* = \mathrm{U}(\nu)$ (resp. $\nu^* = \mathrm{L}(\nu)$) satisfy supermodularity constraints:

$$\nu^*_{(s,i)(t,j)} + \nu^*_{(s,i-1)(t,j-1)} - \nu^*_{(s,i)(t,j-1)} - \nu^*_{(s,i-1)(t,j)} \geq 0 \quad \forall st \in \mathcal{E},\ \forall i,j = [3 \dots K]. \tag{42}$$

Proof. □

Let us fix arbitrary $st \in \mathcal{E}$. Under constraints (33) on ν it is $\nu_{s,i} \leq \nu_{s,i-1}$ and $\nu_{t,j} \leq \nu_{t,j-1}$.

U: Without loss of generality, let $\nu_{s,i} \leq \nu_{t,j}$. Then it is also $\nu_{s,i} \leq \nu_{t,j-1}$ and (42) reduces to

$$\nu_{si} + \min(\nu_{s,i-1}, \nu_{t,j-1}) \geq \nu_{si} + \min(\nu_{s,i-1}, \nu_{t,j}), \tag{43}$$

which follows now from $\nu_{t,j-1} \geq \nu_{t,j}$ and that $\min(a, \cdot)$ is a monotonous non-decreasing function.

L: Inequality (42), when mapping L is substituted, can be expressed as

$$\max(1-\nu_{t,j}, \nu_{s,i}) + \max(1-\nu_{t,j-1}, \nu_{s,i-1}) \geq \max(1-\nu_{t,j-1}, \nu_{s,i}) + \max(1-\nu_{t,j}, \nu_{s,i-1}). \tag{44}$$

Under constraints (33) it is $\nu_{s,i} \leq \nu_{s,i-1}$ and $1 - \nu_{t,j} \geq 1 - \nu_{t,i-1}$. Consider now three possible cases:

1. $1 - \nu_{t,j} \leq \nu_{s,i} \leq \nu_{s,i-1}$, then (44) reduces to

$$\nu_{s,i} + \nu_{s,i-1} \geq \nu_{s,i} + \nu_{s,i-1}. \tag{45}$$

2. $\nu_{s,i} \leq \nu_{s,i-1} \leq 1 - \nu_{t,j}$, then (44) reduces to

$$1 - \nu_{t,j} + \max(1 - \nu_{t,j-1}, \nu_{s,i-1}) \geq \max(1 - \nu_{t,j-1}, \nu_{s,i}) + 1 - \nu_{t,j}, \tag{46}$$

which follows now from $\nu_{s,i-1} \geq \nu_{s,i}$ and that $\max(c, \cdot)$ is a monotonous non-decreasing function.

3. $\nu_{s,i} \leq 1 - \nu_{t,j} \leq \nu_{s,i-1}$, then (44) reduces to

$$1 - \nu_{t,j} + \nu_{s,i-1} \geq \max(1 - \nu_{t,j-1}, \nu_{s,i}) + \nu_{s,i-1}, \tag{47}$$

which follows from $1 - \nu_{t,j} \geq 1 - \nu_{t,j-1}$ and the assumption.

Statement 9 (A reduced submodular linear program)**.** Let $(G, \mathcal{L}, \theta)$ define a submodular multi-label energy minimization problem. Let $(N, \mathbb{B}, \eta)$ define a corresponding binary energy minimization problem. Then LP-relaxation (35) can be reduced to

$$\begin{aligned} &\min\langle \eta, \nu \rangle \\ &\nu_u \in [0, 1], \quad \forall u \in V \\ &\nu_{uv} \leq \min(\nu_u, \nu_v), \quad \forall uv \in A_{\mathcal{E}} \end{aligned} \tag{48}$$

Proof. Let ν be optimal to (48). Then making a correction of ν by setting $\nu_{uv} := \min(\nu_u, \nu_v), \quad \forall uv \in A_{\mathcal{E}}$ will not increase the objective (as implied by Statement 7) and gives a solution which is feasible to (35). □

While it is always possible to map a relaxed labeling $\mu \in \Lambda_{G,\mathcal{L}}$ to a relaxed binary labeling $\nu \in \hat{\Lambda}_{N,\mathbb{B}}$ such that costs $\langle \mu, \theta \rangle$ and $\langle \nu, \eta \rangle$ are equal, the reverse mapping is not always possible. This is why LP-2 is weaker than LP-1. To construct the reverse mapping we will need the following extended supermodularity constraints on ν:

$$D_{ij} \left(\begin{array}{c|c|c} 1 & \nu_t & 0 \\ \hline \nu_s & \nu_{st} & 0 \\ \hline 0 & 0 & 0 \end{array} \right) \geq 0, \qquad i, j = 2 \dots K + 1; \tag{49}$$

Accordingly we define additionally constrained polytope

$$\hat{\Lambda}^{\sup}_{N,\mathbb{B}} = \{\nu \in \hat{\Lambda}_{N,\mathcal{L}} \mid \nu \text{ satisfies (49)}\}. \tag{50}$$

Proposition 1 (Mapping of relaxed labelings)**.** Let mapping $\Pi : \Lambda_{G,\mathcal{L}} \mapsto \hat{\Lambda}^{\sup}_{N,\mathbb{B}}$ defined as follows:

$$\begin{aligned} [\Pi\mu]_{si'} &= \Pi_{i'}\mu_s = \sum_{i' \leq i \leq K} \mu_s(i), && i' \in \tilde{\mathcal{L}},\ s \in \mathcal{V}; \\ [\Pi\mu]_{si',tj'} &= \Pi_{i'j'}\mu_{st} = \sum_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} \mu_{st}(i, j), && i'j' \in \tilde{\mathcal{L}},\ st \in \mathcal{E}; \\ [\Pi\mu]_{\text{const}} &= 1. \end{aligned} \tag{51}$$

Proof. We need to verify that $\nu = \Pi\mu$ obeys constraints of $\hat{\Lambda}^{\sup}_{N,\mathbb{B}}$.

- Constraint $\nu_{s,i} \leq \nu_{s,i-1}$, $i \in \tilde{\mathcal{L}}$ follows from nonnegativity of μ.
- Constraint $\nu_{s,i} \in [0,1]$ follows from nonnegativity of μ and normalization constraints (6).
- Constraints $\nu_{u,v} \leq \min(\nu_u, \nu_v)$, $uv \in \mathcal{A}_\mathcal{E}$ are verified as follows. Let us show *e.g.* $\nu_u - \nu_{uv} \geq 0$. Substituting (51) and using marginalization constraints (5), it is

$$\nu_{s,i'} - \nu_{si',tj'} = \sum_{\substack{i' \leq i \leq K \\ 1 \leq j \leq K}} \mu_{st}(i,j) - \sum_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} \mu_{st}(i,j) = \sum_{\substack{i' \leq i \leq K \\ 1 \leq j < j'}} \mu_{st}(i,j) \geq 0. \tag{52}$$

- Constraints $0 \leq \nu_{uv}$ follow from nonnegativity of μ.
- Constraints $\nu_{uv} \geq \nu_u + \nu_v - 1$ are verified as follows. Substituting (51) they read

$$\sum_{\substack{i' < i \leq K \\ j' < j \leq K}} \mu_{st}(i,j) - \sum_{i' \leq i \leq K} \mu_s(i) - \sum_{j' \leq j \leq K} \mu_t(j) + 1 \geq 0. \tag{53}$$

Using marginalization constraints (5), the LHS equals to

$$\begin{aligned} &\sum_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} \mu_{st}(i,j) - \sum_{\substack{i' \leq i \leq K \\ 1 \leq j \leq K}} \mu_{st}(i,j) - \sum_{\substack{1 \leq i \leq K \\ j' \leq j \leq K}} \mu_{st}(i,j) + \sum_{\substack{1 \leq i \leq K \\ 1 \leq j \leq K}} \mu_{st}(i,j) = \\ &= \sum_{\substack{1 \leq i < i' \\ 1 \leq j < j'}} \mu_{st}(i,j) \geq 0. \end{aligned} \tag{54}$$

- Supermodularity constraints follow from that when the mapping Π is substituted into the expression (49) it yields exactly matrix of non-negative values μ_{st}.

$\square$

Statement 10. The inverse mapping $\Pi^{-1} : \hat{\Lambda}^{\text{sup}}_{N,\mathbb{B}} \mapsto \Lambda_{G,\mathcal{L}}$ is given by

$$\begin{aligned} [\Pi^{-1}\nu]_s(i) &= -D_{i+1} \begin{pmatrix} 1 & \nu_s & 0 \end{pmatrix}, & i = 1 \dots K; \\ [\Pi^{-1}\nu]_{st}(i,j) &= D_{i+1\,j+1} \left(\begin{array}{c|c|c} 1 & \nu_t & 0 \\ \hline \nu_s & \nu_{st} & 0 \\ \hline 0 & 0 & 0 \end{array}\right) & i,j = 1 \dots K; \\ [\Pi^{-1}\nu]_{\text{const}} &= 1. \end{aligned} \tag{55}$$

Proof. Let $\mu = \Pi^{-1}\nu$. We need to verify that $\mu \in \Lambda_{G,\mathcal{L}}$ for all $\nu \in \hat{\Lambda}^{\text{sup}}_{G,\mathbb{B}}$ and that $\Pi^{-1}\Pi\mu = \mu$ for all $\mu \in \Lambda_{G,\mathcal{L}}$. First claim is verified as follows. Expanding expression (55) it is

$$\begin{aligned} \mu_s(1) &= -(\nu_{s,2} - 1) = 1 - \nu_{s,2} \\ \mu_s(i) &= -(\nu_{s,i+1} - \nu_{s,i}) = \nu_{s,i} - \nu_{s,i+1}, \quad i = 2 \dots K-1 \\ \mu_s(K) &= -(0 - \nu_{s,K}) = \nu_{s,K}. \end{aligned} \tag{56}$$

- Constraints $\mu_s(i) \geq 0$ follow from that ν_s are monotonous non-increasing (33) and from constraints $\nu_{s,i'} \leq 1$.
- It is also $\sum_{i\in\mathcal{L}} \mu_s(i) = 1 - \nu_{s,2} + \nu_{s,2} - \nu_{s,3} + \cdots + \nu_{s,K-1} - \nu_{s,K} + \nu_{s,K} = 1$.
- It may be verified that

$$\sum_{j\in\mathcal{L}} \mu_{st}(i,j) = \begin{cases} 1 - \nu_{s,2} & i = 1 \\ \nu_{s,i} - \nu_{s,i+1} & i = 2 \dots K-1 \\ \nu_{s,K} & i = K, \end{cases} \tag{57}$$

  which coincides with (56), and therefore marginalization constraints (5) are satisfied.
- Non-negativity of $\mu_{st}(i,j)$ is implied by the supermodularity (49) of ν.

The claim $\Pi^{-1}\Pi\mu = \mu$ is verified as follows. Let $\mu' = \Pi^{-1}\Pi\mu$.

- Substituting mapping Π into (56) it is

$$\begin{aligned} \mu'_s(1) &= 1 - \sum_{2\leq i''\leq K} \mu_s(i'') = \mu_s(1) \\ \mu'_s(i) &= \sum_{i\leq i''\leq K} \mu_s(i'') - \sum_{i+1\leq i''\leq K} \mu_s(i'') = \mu_s(i), \qquad i \in \mathcal{L} \\ \mu'_s(K) &= \sum_{K\leq i''\leq K} \mu_s(i'') = \mu_s(K). \end{aligned} \tag{58}$$

- Similarly it is verified for pairwise terms. Let us expand, *e.g.*, the term $\mu'_{st}(i,1)$:

$$\begin{aligned} \mu'_{st}(i,1) = \nu_{s,\,i} - \nu_{s,\,i+1} - \nu_{s,\,i;t,2} + \nu_{s,\,i+1;t,2} = \mu_s(i) - \nu_{s,\,i;t,2} + \nu_{s,\,i+1;t,2} = \\ \sum_{1\leq j''\leq K} \mu_{st}(i,j'') - \sum_{\substack{i\leq i''\leq K \\ 2\leq j''\leq K}} \mu_{st}(i'',j'') + \sum_{\substack{i+1\leq i''\leq K \\ 2\leq j''\leq K}} \mu_{st}(i'',j'') = \mu_{st}(i,1). \end{aligned} \tag{59}$$

□

Thus we proved that Π is a bijective mapping between $\Lambda_{G,\mathcal{L}}$ and $\hat{\Lambda}^{\text{sup}}_{N,\mathbb{B}}$.

Statement 11. Let $(G,\mathcal{L},\theta)$ and $(N,\mathbb{B},\eta)$ define the pair of corresponding multi-label and binary energy minimization problems. Mapping Π preserves the associated primal cost:

$$\langle \Pi\mu, \eta\rangle = \langle \mu, \theta\rangle. \tag{60}$$

Proof. Without loss of generality we assume that θ satisfies (12), so $\eta_{(s,i)} = D_i\theta_s$, $\eta_{(s,i)(t,j)} = D_{ij}\theta_{st}$ and $\eta_{\text{const}} = \theta_{\text{const}}$.

- Univariate terms of $\langle \mu, \theta\rangle$ are expressed as:

$$\sum_{i\in\mathcal{L}} \mu_s(i)\theta_s(i) = \sum_{i\in\mathcal{L}} \mu_s(i) \sum_{2\leq i'\leq i} D_{i'}\theta_s = \sum_{i'\in\tilde{\mathcal{L}}} \eta_{(s,i')} \sum_{i'\leq i\leq K} \mu_s(i) = \sum_{i'\in\tilde{\mathcal{L}}} \eta_{(s,i')}[\Pi\mu]_{s,i'}. \tag{61}$$

- Analogously, pairwise terms of $\langle\mu,\theta\rangle$ are expressed as:

$$\begin{aligned}&\sum_{i,j\in\mathcal{L}}\mu_{st}(i,j)\theta_{st}(i,j)=\sum_{i,j\in\mathcal{L}}\mu_{st}(i,j)\sum_{\substack{2\leq i'\leq i\\2\leq j'\leq j}}D_{i'j'}\theta_{st}=\\&\sum_{\substack{2\leq i'\leq i\\2\leq j'\leq j}}\eta_{si';tj'}\sum_{\substack{i'\leq i\leq K\\j'\leq j\leq K}}\mu_{st}(i,j)=\sum_{\substack{2\leq i'\leq i\\2\leq j'\leq j}}\eta_{si';tj'}[\Pi\mu]_{si';tj'}.\end{aligned}\tag{62}$$

□

Theorem 1. *Let $(G,\mathcal{L},\theta)$ define a multi-label energy minimization problem. Let for each edge $st\in\mathcal{E}$ term θ_{st} is either submodular or supermodular. Let $(N,\mathbb{B},\eta)$ define the corresponding binary energy minimization problem. Then*

(a) LP-relaxations of $E(\mathrm{x}|\theta)$ and $E(\mathrm{z}|\eta)$ coincide:

$$\min_{\mu\in\Lambda_{G,\mathcal{L}}}\langle\mu,\theta\rangle=\min_{\nu\in\Lambda_{N,\mathbb{B}}}\langle\nu,\eta\rangle;\tag{63}$$

(b) There exists half-integral optimal solution of LHS of (63).

Proof. Using mappings Π, U and L the claim is proved as follows:

Let μ^* be optimal to the LHS of (63). Then $\langle\Pi\mu^*,\eta\rangle=\langle\mu^*,\theta\rangle$ and therefore LHS $\geq$ RHS.

Let ν^* be optimal solution of the RHS of (63). Let ν^{**} be constructed from ν^* using Stat. (7), with $\nu^{**}_{(si)(tj)}=\min(\nu^*_{si},\nu^*_{tj})$ for submodular θ_{st} and with $\nu^{**}_{(si)(tj)}=\max(0,\nu^*_{si}+\nu^*_{tj}-1)$ for supermodular θ_{st}. In this case ν^{**}_{st} is supermodular (see Statment 8) and it is $\langle\nu^{**},\eta\rangle=\langle\Pi\Pi^{-1}\nu^{**},\eta\rangle=\langle\Pi^{-1}\nu^{**},\theta\rangle$ and therefore LHS $\leq$ RHS.

Claim in (b) follows trivially from the existence of half-integral optimal ν^* by considering $\mu=\Pi^{-1}\nu^{**}$. □

It worse noting that under conditions of the theorem half-integrality of optimal μ implies that for each $s\in\mathcal{V}$ there could be at most two labels assigned non-zero weights by μ, i.e. μ_s is either integral: $(0\dots0,1,0\dots0)$, either half-integral: $(0\dots0,\frac{1}{2},0\dots0,\frac{1}{2},0\dots0)$.

Corollary 1. For a subclass of problems defined by conditions of the theorem there exist efficient fully combinatorial algorithm to maximize the LB, which is an improvement over e.g. TRW-S algorithm [22, 9]. It is the network flow algorithm [2] applied to binary energy $E(\cdot|\eta)$.

9 Persistency for Relaxed Labelings

Theorem 2. *Let $(G,\mathcal{L},\theta)$ define a multi-label energy minimization problem. Let $(N,\mathbb{B},\eta)$ define the corresponding binary energy minimization problem. Let $(\mathrm{z}^{\min},\mathrm{z}^{\max})$ provide strong persistency for $E(\cdot|\eta)$. Let $\mu\in\Lambda_{G,\mathcal{L}}$ be an optimal solution of the LP-1 relaxation.*

Then it is $\mu_{s;i}=0$ for all labels i outside the interval $[x^{\min}_s,x^{\max}_s]$, where $\mathbf{x}^{\min}=\mathbf{x}(\mathbf{z}^{\min})$ and $\mathbf{x}^{\max}=\mathbf{x}(\mathbf{z}^{\max})$.

Proof. Let us denote components of vector $\nu \in \Lambda_{N,\mathbb{B}}$ as $\nu_{u;\alpha}$, $\nu_{uv;\alpha\beta}$ where $\alpha, \beta \in \{0,1\}$ are binary labels. If $u = (s,i)$ then for brevity we will use notation like $\nu_{s;i} = \nu_{s;i;1}$ and $\nu_{st;ij} = \nu_{st;ij;11}$.

Earlier we defined a mapping $\mathbf{z} = \mathbf{z}(\mathbf{x})$ from multi-valued configurations to binary configurations. Let us extend it to a mapping $\Pi : \Lambda_{G;\mathcal{L}} \to \Lambda_{N,\mathbb{B}}$ for relaxed labelings. For index $i \in \{2, \ldots, K\}$ define the following sets of labels:

$$L_s(i,0) = \{1, \ldots, i-1\} \quad L_s(i,1) = \{i, \ldots, K\}$$

Then we can define vector $\nu = \Pi\mu$ as

$$\nu_{s;i;\alpha} = \sum_{i' \in L_s(i,\alpha)} \mu_{s;i'} \qquad \nu_{st;ij;\alpha\beta} = \sum_{\substack{i' \in L_s(i,\alpha) \\ j' \in L_t(j,\beta)}} \mu_{st;i'j'}$$

It is easy to see that mapping Π is precisely the mapping $\mu \mapsto \nu$ defined by (51) only extended to polytope $\Lambda_{N,\mathbb{B}}$ by relations (37). When proving Theorem 1 we already showed that under this mapping objective is preserved, which may be written as:

$$\langle \mu, \theta \rangle = \langle \nu, \eta \rangle, \tag{64}$$

where $\nu = \Pi\mu$ and η is understood to have components $\eta_{u;\alpha}$, $\eta_{uv;\alpha\beta}$, where $\alpha, \beta \in \{0,1\}$ are binary labels and only non-zero components are $\eta_{u;1}$, $\eta_{uv;11}$.

Next we define "truncated" vector $\bar{\mu}$. Loosely speaking, values of μ associated with labels $i < x_s^{\min}$ are reassigned to $x_s^{\min}$, and values associated with labels $i > x_s^{\max}$ are reassigned to $x_s^{\max}$. Formally, if $x_s^{\min} < x_s^{\max}$ then for label i let us define the set of labels $T_s(i)$ associated with i as

$$T_s(i) = \begin{cases} \varnothing & \text{if } i < x_s^{\min} \\ \{1, 2, \ldots, i\} & \text{if } i = x_s^{\min} \\ \{i\} & \text{if } x_s^{\min} < i < x_s^{\max} \\ \{i, i+1, \ldots, K\} & \text{if } i = x_s^{\max} \\ \varnothing & \text{if } i > x_s^{\max} \end{cases}$$

If $x_s^{\min} = x_s^{\max}$ then $T_s(i) = \{1, \ldots, K\}$ if $i = x_s^{\min}$ and $T_s(i) = \varnothing$ otherwise. Using these sets, "truncated" vector is defined as

$$\bar{\mu}_{s;i} = \sum_{i' \in T_s(i)} \mu_{s;i'} \qquad \bar{\mu}_{st;ij} = \sum_{\substack{i' \in T_s(i) \\ j' \in T_t(j)}} \mu_{st;i'j'}$$

It is trivial to check that $\bar{\mu} \in \Lambda_{G,\mathcal{L}}$.

We define "truncated" vector $\bar{\nu}$ in a similar way:

$$\begin{cases} T_{s;i}(0) = \{0,1\} \\ T_{s;i}(1) = \varnothing \end{cases} \qquad \text{if } z_{s;i}^{\min} = z_{s;i}^{\max} = 0$$

$$\begin{cases} T_{s;i}(0) = \{0\} \\ T_{s;i}(1) = \{1\} \end{cases} \qquad \text{if } z_{s;i}^{\min} = 0, z_{s;i}^{\max} = 1$$

$$\begin{cases} T_{s;i}(0) = \varnothing \\ T_{s;i}(1) = \{0,1\} \end{cases} \qquad \text{if } z_{s;i}^{\min} = z_{s;i}^{\max} = 1$$

$$\bar{\nu}_{s;i;\alpha} = \sum_{\alpha' \in \hat{T}_{s;i}(\alpha)} \nu_{s;i;\alpha'} \qquad \bar{\nu}_{st;ij;\alpha\beta} = \sum_{\substack{\alpha' \in \hat{T}_{s;i}(\alpha) \\ \beta' \in \hat{T}_{t;j}(\beta)}} \nu_{st;ij;\alpha'\beta'}$$

We prove below that $\Pi\bar{\mu} = \bar{\nu}$. This will imply the desired result. Indeed, suppose that the condition in theorem 2 does not hold, then $\nu_{s;i;\alpha} > 0$ for some index $(s;i)$ and binary label α outside the interval $[z_s^{\min}, z_s^{\max}]$. Therefore,

$$\begin{aligned} \langle \bar{\mu}, \theta \rangle &= \langle \bar{\nu}, \eta \rangle \\ &< \langle \nu, \eta \rangle = \langle \mu, \theta \rangle \end{aligned}$$

(The inequality follows from the the properties of roof duality). This contradicts to the optimality of μ.

In order to prove that $\Pi\bar{\mu} = \bar{\nu}$, let us show the following

Lemma 2. For any node s, index $i \in \{2, \ldots, K\}$ and label $\alpha \in \{0, 1\}$ there holds

$$\bigcup_{i' \in L_s(i,\alpha)} T_s(i') = \bigcup_{\alpha' \in \hat{T}_{s;i}(\alpha)} L_s(i, \alpha')$$

Furthermore, all unions are disjoint.

Proof. We prove the lemma only for $\alpha = 1$; the other case is similar. (Actually, the case $\alpha = 1$ would be sufficient for showing that $\Pi\bar{\mu} = \bar{\nu}$ since it is enough to establish the latter equality only for indexes $(u; 1)$ and $(uv; 11)$.) Also, we consider only the case when $x_s^{\min} < x_s^{\max}$; the case $x_s^{\min} = x_s^{\max}$ can be analyzed in the same way.

Three cases are possible; in each case it is easy to see that all unions are disjoint:

- $i \in \{2, \ldots, x_s^{\min}\}$. Then the LHS is $\{1, \ldots, K\}$ since $\{x_s^{\min}, \ldots, x_s^{\max}\}$ is a subset of $L_s(i, \alpha)$. The RHS equals the same set since $z_{s;i}^{\min} = z_{s;i}^{\max} = 1$ and $\hat{T}_{s;i}(1) = \{0, 1\}$.
- $i \in \{x_s^{\min} + 1, \ldots, x_s^{\max}\}$. Then the LHS is $\{i, \ldots, K\}$. The RHS equals the same set since $z_{s;i}^{\min} = 0$, $z_{s;i}^{\max} = 1$ and $\hat{T}_{s;i}(1) = \{1\}$.
- $i \in \{x_s^{\max} + 1, \ldots, K\}$. Then the LHS is empty. The RHS is empty as well since $z_{s;i}^{\min} = z_{s;i}^{\max} = 0$ and $\hat{T}_{s;i}(1) = \varnothing$.

□

Using the lemma, we can write

$$\begin{aligned} (\Pi\bar{\mu})_{s;i;\alpha} &= \sum_{i' \in L_s(i,\alpha)} \bar{\mu}_{s;i'} \\ &= \sum_{i' \in L_s(i,\alpha)} \sum_{i'' \in T_s(i')} \mu_{s;i''} \\ &= \sum_{\alpha' \in \hat{T}_{s;i}(\alpha)} \sum_{i' \in L_s(i,\alpha')} \mu_{s;i'} \\ &= \sum_{\alpha' \in \hat{T}_{s;i}(\alpha)} \nu_{s;i;\alpha'} = \bar{\nu}_{s;i;\alpha} \end{aligned}$$

$$
\begin{aligned}
(\Pi\bar{\mu})_{st;ij;\alpha\beta} &= \sum_{\substack{i'\in L_s(i,\alpha)\\ j'\in L_t(j,\beta)}} \bar{\mu}_{st;i'j'} \\
&= \sum_{\substack{i'\in L_s(i,\alpha)\\ j'\in L_t(j,\beta)}} \sum_{\substack{i''\in T_s(i')\\ j''\in T_t(j')}} \mu_{st;i''j''} \\
&= \sum_{\substack{\alpha'\in \hat{T}_{s;i}(\alpha)\\ \beta'\in \hat{T}_{t;j}(\beta)}} \sum_{\substack{i'\in L_s(i,\alpha')\\ j'\in L_t(j,\beta')}} \mu_{st;i'j'} \\
&= \sum_{\substack{\alpha'\in \hat{T}_{s;i}(\alpha)\\ \beta'\in \hat{T}_{t;j}(\beta)}} \nu_{s;i;\alpha'} \;=\; \bar{\nu}_{st;ij;\alpha\beta}
\end{aligned}
$$

□

10 Submodular-Supermodular Decomposition

LP-relaxation of binary energy can be written directly in terms of the multi-label problem via additionally relaxed LP (4). The construction is motivated by considering a lower bound obtained by decomposing θ into sum of submodular and supermodular problems.

Lemma 3 (Decomposition, [21])**.** Any $\theta \in \Omega$ can be decomposed as $\theta = \theta^1 + \theta^2$, with $\theta^1 \in \Omega^{\rm sub}$, $\theta^2 \in \Omega^{\rm sup}$.

Proof. For any $st \in \mathcal{E}$ let $c^1_{st}(i,j) = \min(D_{ij}\theta_{st}, 0)$ and $c^2_{st}(i,j) = \max(D_{ij}\theta_{st}, 0)$, $i,j = 2\dots K$ and let $q_{st}(i,j) = \theta_{st}(i,1) + \theta_{st}(1,j) - \theta_{st}(1,1)$, $i,j \in \mathcal{L}$. Then choosing

$$
\begin{aligned}
\hat{\theta}^1_{st}(i,j) &= \sum_{\substack{2\le i'\le i\\ 2\le j'\le j}} c^1_{st}(i',j') + q_{st}(i,j) \\
\hat{\theta}^2_{st}(i,j) &= \sum_{\substack{2\le i'\le i\\ 2\le j'\le j}} c^2_{st}(i',j') \\
\hat{\theta}^1_s &= \hat{\theta}^2_s = \tfrac{1}{2}\theta_s, \quad \hat{\theta}^1_{\rm const} = \hat{\theta}^2_{\rm const} = \tfrac{1}{2}\theta_{\rm const}
\end{aligned}
\tag{65}
$$

provides such a decomposition. Note it is not unique. □

It was noticed [21] that the family of all decompositions of this kind provide an alternative to the tree-structured decomposition [22] and may be used to construct a lower bound on the energy. The tightest lower bound by all possible decompositions from the family may be written in the form, similar to our expression of LB as:

$$
\max_{\theta^1,\,\theta^2}(\theta^1_{\rm conts} + \theta^2_{\rm conts}) \quad \text{s.t.} \begin{cases} \theta^1 \in \Omega^{\rm sub} \\ \theta^2 \in \Omega^{\rm sup} \\ \theta^1 + \theta^2 \equiv \theta \\ \theta^1_{\mathcal{I}} \ge 0, \quad \theta^2_{\mathcal{I}} \ge 0 \\ \theta^1_{\rm const}, \theta^2_{\rm const} \in \mathbb{R}. \end{cases} \tag{$LB2$}
$$

It is clear that $LB2 \leq LB$ since there are additional constraints in the optimization problem $LB2$ compared to LB (the submodularity of θ^1 and supermodularity of θ^2), which means that $LB2$ bound is weaker. However, as we show here, $LB2$ is equivalent to the LP-relaxation of the binarized energy minimization (LP-2), considered in Sect. 8. Thus the bound $LB2$ can be efficiently computed via a max-flow algorithm. The optimal solution (θ^1, θ^2) can be shown to provide constraints in the form of autarky on the set of minimizers of (1), completely analogous to those considered in Sect. 6.

To show the stated equivalence we first derive a non-trivial dual of $LB2$ and then show it is equivalent to LP-2. The construction is accomplished in several steps: characterize the set of constraints in the problem $LB2$ and use it to simplify the standard Lagrangian dual.

The following lemma describes the (convex) set of all possible submodular-supermodular decompositions. Decomposition $\hat{\theta}^1 + \hat{\theta}^2$ by Lemma 3 plays a role of an extreme point in this set.

Lemma 4 (All decompositions)**.** Let $\theta^1 + \theta^2 = \theta$, $\theta^1 \in \Omega^{\text{sub}}$, $\theta^2 \in \Omega^{\text{sup}}$. Then there exist $\varphi \in \Omega^{\text{sub}}$ such that

$$\theta^1 = \hat{\theta}^1 + \varphi, \quad \theta^2 = \hat{\theta}^2 - \varphi, \tag{66}$$

where $\hat{\theta}^1$ and $\hat{\theta}^1$ are defined by decomposition (65).

Proof. Let us construct φ as follows

$$\begin{aligned} \varphi_{st}(i,j) &= \sum_{\substack{2 \leq i' \leq i \\ 2 \leq j' \leq j}} D_{i'j'}\theta^1_{st} - D_{i'j'}\theta^1_{st}, \qquad && st \in \mathcal{E},\ i,j \in \mathcal{L} \\ \varphi_s(i) &= \theta^1_s(i) - \hat{\theta}^1_s(i), && s \in \mathcal{V},\ i \in \mathcal{L} \\ \varphi_{\text{const}} &= \theta^1_{\text{const}} - \hat{\theta}^1_{\text{const}}. \end{aligned} \tag{67}$$

The definition of φ ensures equation (66) holds. We will prove now φ is submodular, i.e. $D\varphi_{st} \leq 0$. From $D\theta^1_{st} + D\theta^2_{st} = D\theta_{st}$ it is $D\theta^2_{st} = D\theta_{st} - D\theta^1_{st} \geq 0$, since θ^2 is supermodular. And it is $D\theta^1_{st} \leq 0$ (since θ^1 is submodular). Therefore $D\theta^1_{st} \leq \min(0, D\theta_{st}) = D\hat{\theta}^1_{st}$. □

It is also easily seen that for any $\varphi \in \Omega^{\text{sub}}$ equation (66) provides a submodular-supermodular decomposition. It follows that the family of the decompositions can be constructively defined as follows:

$$\{(\theta^1, \theta^2) \mid \theta^1 + \theta^2 = \theta,\ \theta^1 \in \Omega^{\text{sub}},\ \theta^2 \in \Omega^{\text{sup}}\} = \{(\hat{\theta}^1 + \varphi, \hat{\theta}^2 - \varphi) \mid \varphi \in \Omega^{\text{sub}}\}. \tag{68}$$

Lemma 5. Space Ω^{sub}, up to equivalent transformations, can be parametrized as follows:

$$\varphi \in \Omega^{\text{sub}} \quad \Leftrightarrow \quad \exists c \in \mathbb{R}_-^{\mathcal{E} \times \tilde{\mathcal{L}}^2},\ \exists q \in \mathbb{R}^{\mathcal{V} \times \mathcal{L}} :\ \varphi \equiv \varphi(c,q), \tag{69}$$

where

$$\begin{aligned} &[\varphi(c,q)]_s(i) = q_s(i), && s \in \mathcal{V},\ i \in \mathcal{L} \\ &[\varphi(c,q)]_{st}(i,j) = \sum_{\substack{2 \leq i' \leq i \\ 2 \leq j' \leq j}} c_{st}(i',j'), \qquad && st \in \mathcal{E},\ i,j \in \mathcal{L} \\ &[\varphi(c,q)]_{\text{const}} = 0. \end{aligned} \tag{70}$$

Proof. Let $\hat{\varphi} = \varphi$ satisfy (12). Then, setting $c_{st}(i',j') = D_{i'j'}\hat{\varphi}_{st}$ and $q_s(i) = \hat{\varphi}_s(i) + \frac{1}{|\mathcal{V}|}\hat{\varphi}_{\text{const}}$ provides the necessary result. □

Statement 12 ($LB2$-dual)**.** The dual of $LB2$ can be written as:

$$\min_{\mu^1,\,\mu^2} \langle \mu^1, \hat{\theta}^1 \rangle + \langle \mu^2, \hat{\theta}^2 \rangle \quad \text{s.t.} \begin{cases} \mu^1, \mu^2 \in \Lambda \\ \mu^1_s = \mu^2_s \quad \forall s \in \mathcal{V}, \end{cases} \tag{71}$$

where $\hat{\theta}^1$ and $\hat{\theta}^2$ are defined by decomposition (65).

Proof. Using Lemma 4, we rewrite $LB2$ as

$$\max_{\theta^1,\,\theta^2,\varphi} (\theta^1_{\text{conts}} + \theta^2_{\text{conts}}) \quad \text{s.t.} \begin{cases} \theta^1 \equiv \hat{\theta}^1 + \varphi \\ \theta^2 \equiv \hat{\theta}^2 - \varphi \\ \theta^1_{\mathcal{I}} \geq 0 \\ \theta^2_{\mathcal{I}} \geq 0 \\ \theta^1_{\text{const}}, \theta^2_{\text{const}} \in \mathbb{R} \\ \varphi \in \Omega^{\text{sub}}, \end{cases} \tag{72}$$

or, changing the order of maximization, as:

$$= \max_{\varphi \in \Omega^{\text{sub}}} \Big(\max_{\substack{\theta^1_{\mathcal{I}} \geq 0 \\ \theta^1 \equiv \hat{\theta}^1 + \varphi}} \theta^1_{\text{const}} + \max_{\substack{\theta^2_{\mathcal{I}} \geq 0 \\ \theta^2 \equiv \hat{\theta}^2 - \varphi}} \theta^2_{\text{const}} \Big), \tag{73}$$

applying the duality relation (4)=LB to the two maximization problems in the brackets it is

$$= \max_{\varphi \in \Omega^{\text{sub}}} \Big(\min_{\mu^1 \in \Lambda_{G,\mathcal{L}}} \langle \mu^1, \hat{\theta}^1 + \varphi \rangle + \min_{\mu^2 \in \Lambda_{G,\mathcal{L}}} \langle \mu^2, \hat{\theta}^2 - \varphi \rangle \Big) \tag{74}$$

$$= \inf_{\mu^1, \mu^2 \in \Lambda_{G,\mathcal{L}}} \sup_{\varphi \in \Omega^{\text{sub}}} \Big(\langle \mu^1, \hat{\theta}^1 + \varphi \rangle + \langle \mu^2, \hat{\theta}^2 - \varphi \rangle \Big) \tag{75}$$

$$= \inf_{\mu^1, \mu^2 \in \Lambda_{G,\mathcal{L}}} \Big(\langle \mu^1, \hat{\theta}^1 \rangle + \langle \mu^2, \hat{\theta}^2 \rangle + \sup_{\varphi \in \Omega^{\text{sub}}} \langle \varphi, \mu^1 - \mu^2 \rangle \Big). \tag{76}$$

The supremum in the expression (76) can be expanded using Lemma (5) as follows:

$$\sup_{\varphi \in \Omega^{\text{sub}}} \langle \varphi, \mu^1 - \mu^2 \rangle = \sup_{\substack{q \in \mathbb{R}^{\mathcal{V} \times \mathcal{L}} \\ c \in \mathbb{R}_-^{\mathcal{E} \times \tilde{\mathcal{L}}^2}}} \sum_{s \in \mathcal{V},\, i \in \mathcal{L}} (\mu^1_s(i) - \mu^2_s(i)) q_s(i) + \sum_{st \in \mathcal{E},\, i,j \in \mathcal{L}} (\mu^1_{st}(i,j) - \mu^2_{st}(i,j)) \sum_{\substack{2 \leq i' \leq i \\ 2 \leq j' \leq j}} c_{st}(i',j')$$

$$= \sup_{q \in \mathbb{R}^{\mathcal{V} \times \mathcal{L}}} \sum_{s \in \mathcal{V},\, i \in \mathcal{L}} (\mu^1_s(i) - \mu^2_s(i)) q_s(i) + \sup_{-c \in \mathbb{R}_-^{\mathcal{E} \times \tilde{\mathcal{L}}^2}} \sum_{st \in \mathcal{E},\, i',j' \in \tilde{\mathcal{L}}} c_{st}(i',j') \sum_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} (\mu^1_{st}(i,j) - \mu^2_{st}(i,j))$$

$$= \begin{cases} 0, & \mu^1_s(i) = \mu^2_s(i) \quad \forall s \in \mathcal{V},\, \forall i \in \mathcal{L} \quad \wedge \quad \sum\limits_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} (\mu^1_{st}(i,j) - \mu^2_{st}(i,j)) \geq 0, \quad \forall st \in \mathcal{E},\, \forall i,j \in \tilde{\mathcal{L}} \\ \infty, & \text{otherwise.} \end{cases} \tag{77}$$

Thus, (76) may be continued as:

$$= \min \Big(\langle \mu^1, \hat{\theta}^1 \rangle + \langle \mu^2, \hat{\theta}^2 \rangle \Big). \quad \text{s.t.} \begin{cases} \mu^1, \mu^2 \in \Lambda_{G,\mathcal{L}} \\ \mu^1_s(i) = \mu^2_s(i) \quad \forall s \in \mathcal{V},\, \forall i \in \mathcal{L} \\ \sum\limits_{\substack{i' \leq i \leq K \\ j' \leq j \leq K}} (\mu^1_{st}(i,j) - \mu^2_{st}(i,j)) \geq 0, \quad \forall st \in \mathcal{E},\, \forall i,j \in \tilde{\mathcal{L}} \end{cases} \tag{78}$$

where the inequality constraints may be discarded (consider proofs of Statement 9 and Theorem 1), and it is

$$= \min \left(\langle \mu^1, \hat\theta^1 \rangle + \langle \mu^2, \hat\theta^2 \rangle \right). \\ \text{s.t.} \begin{cases} \mu^1, \mu^2 \in \Lambda_{G,\mathcal{L}} \\ \mu^1_s(i) = \mu^2_s(i) \quad \forall s \in \mathcal{V}, \, \forall i \in \mathcal{L} \end{cases} \tag{79}$$

□

Theorem 3. *Let* $(G, \mathcal{L}, \theta)$ *define a multi-label energy minimization problem. Let* $(N, \mathbb{B}, \eta)$ *define a corresponding binary energy minimization problem. Then* (71) *is equivalent to* $\min_{\nu \in \hat\Lambda_{N,\mathbb{B}}} \langle \nu, \eta \rangle$.

Proof. Let ν be optimal solution of the relaxed binarized problem. Then the corresponding (μ^1, μ^2) is constructed as follows:

- Values $\mu^1_s(i) = \mu^2_s(i) = \Pi^{-1}_i \nu_s, \quad \forall s \in \mathcal{V}, \, \forall i \in \mathcal{L}$.
- Values $\mu^1_{st}(i,j) = \Pi^{-1}_{ij} \mathrm{U}_{st}(\nu), \quad \forall st \in \mathcal{E}, \, \forall i,j \in \mathcal{L}$.
- Values $\mu^2_{st}(i,j) = \Pi^{-1}_{ij} \mathrm{L}_{st}(\nu), \quad \forall st \in \mathcal{E}, \, \forall i,j \in \mathcal{L}$.

The pair (μ^1, μ^2) is feasible to (71) as followed by properties of Π^{-1}. To see that objective is preserved we need to verify that pairwise terms preserve the associated parts of the objective. It is seen as follows:

$$\sum_{i,j \in \mathcal{L}} \left[\mu^1_{st;ij} \hat\theta^1_{st;ij} + \mu^2_{st;ij} \hat\theta^2_{st;ij} \right] = \\ \sum_{i',j' \in \tilde{\mathcal{L}}} \left[\mathrm{U}_{st}(\nu)_{i'j'} D_{i'j'} \hat\theta^1_{st} + \mathrm{L}_{st}(\nu)_{i'j'} D_{i'j'} \hat\theta^2_{st} \right] = \sum_{i',j' \in \tilde{\mathcal{L}}} \nu_{st;i'j'} D_{i'j'} \theta_{st} = \sum_{i',j' \in \tilde{\mathcal{L}}} \nu_{st;i'j'} \eta_{st;i'j'}. \tag{80}$$

Let (μ^1, μ^2) be optimal to (71), then the corresponding ν is constructed as follows:

- Values $\nu_{(s,i')} = \Pi_{i'} \mu^1_s = \Pi_{i'} \mu^2_s, \quad \forall t \in \mathcal{V}, \, \forall i' \in \tilde{\mathcal{L}}$.
- Values $\nu_{(s,i')(t,j')} = \begin{cases} \Pi_{i'j'} \mu^1_{st}, & D_{i'j'} \theta \le 0 \\ \Pi_{i'j'} \mu^2_{st}, & D_{i'j'} \theta > 0 \end{cases}, \quad \forall st \in \mathcal{E}, \, \forall i', j' \in \tilde{\mathcal{L}}$.

Marginalization constraints hold for ν since $\mu^1_s = \mu^2_s$ for all $s \in \mathcal{V}$. Pairwise terms of the objective are expressed as:

$$\sum_{i',j' \in \tilde{\mathcal{L}}} \nu_{st;i'j'} \eta_{st;i'j'} = \sum_{i',j' \in \tilde{\mathcal{L}}} \nu_{st;i'j'} \left[D_{i'j'} \hat\theta^1_{st} + D_{i'j'} \hat\theta^2_{st} \right] = \\ \sum_{i',j' \in \tilde{\mathcal{L}}} (\Pi_{i'j'} \mu^1_{st})(D_{i'j'} \hat\theta^1_{st}) + \sum_{i',j' \in \tilde{\mathcal{L}}} (\Pi_{i'j'} \mu^2_{st})(D_{i'j'} \hat\theta^2_{st}) = \\ \sum_{i,j \in \mathcal{L}} \mu^1_{st;ij} \hat\theta^1_{st;ij} + \sum_{i,j \in \mathcal{L}} \mu^2_{st;ij} \hat\theta^2_{st;ij}, \tag{81}$$

where we used the following rearrangement of sums:

$$\sum_{i',j'\in\tilde{\mathcal{L}}}(\Pi_{i'j'}\mu^1_{st})(D_{i'j'}\hat{\theta}^1_{st}) = \sum_{i',j'\in\tilde{\mathcal{L}}}\sum_{\substack{i'\le i\le K\\ j'\le j\le K}}\mu^1_{st;ij}D_{i'j'}\hat{\theta}^1_{st} = \sum_{i,j\in\mathcal{L}}\mu^1_{st;ij}\sum_{\substack{2\le i'\le i\\ 2\le j'\le j}}D_{i'j'}\hat{\theta}^1_{st} = \sum_{i,j\in\mathcal{L}}\mu^1_{st;ij}\hat{\theta}^1_{st;ij} \tag{82}$$

and analogous equality for μ^2. □

Appendix: Order-independent Reduction to Binary Variables

In section 5 we mentioned an order-independent reduction of multi-label energy 1 minimization to the following binary energy minimization:

$$\begin{aligned} E(z) &= \sum_{(s,i)}\theta_s(i)z_{(s,i)} + \sum_{(s,i),(t,j)}\theta_{st}(i,j)z_{(s,i)}z_{(tj)} \\ &\quad + H(z) + C\sum_s(1-\sum_i z_{(s,i)}). \end{aligned} \tag{83}$$

Here $z_{(s,i)} = \delta_{\{x_s=i\}}$ are binary indicator variables. $H(z)$ is a hard constraint prohibiting two variables $z_{(s,i)}, z_{(s,j)}$ with $i \neq j$ to be 1 simultaneously; it can be written as

$$H(z) = \sum_s\sum_{i,j:i\neq j}h(z_{(s,i)}, z_{(s,j)}), \tag{84}$$

where $h(z_u, z_v)$ is $+\infty$ if $z_u = z_v = 1$, and 0 otherwise. Finally, C is a sufficiently large constant which ensures that at least one of the indicator variables for node s is 1, e.g. $C > 2C_0$ where

$$C_0 = \sum_s\max_i|\theta_s(i)| + \sum_{(s,t)}\max_{i,j}|\theta_{st}(i,j)|. \tag{85}$$

Let us show that the roof duality relaxation applied to (83) is degenerate if $K \geq 3$, i.e. it does not label any nodes. Let $\nu_{(s,i)} \in [0,1]$ be the fractional variable which is the relaxation of binary variable $z_{(s,i)} \in \{0,1\}$. (The relaxation also uses variables for pairwise terms; below we always assume that given variables $\nu = \{\nu_{(s,i)}\}$, variables for pairwise terms are chosen so that the objective of the roof duality relaxation is minimized.) It is known (see *e.g.* [1]) that extreme points of the polytope in the roof duality relaxation are half-integral, i.e. $\nu_{(s,i)} \in \{0, 0.5, 1\}$. (Note, variables for pairwise terms are half-integral as well). Thus, there exists a half-integral vector ν^* which is an optimal solution of the roof duality relaxation.

Lemma 6. There holds $\nu^*_{(s,i)} = 0.5$ for all (s,i).

Proof. Let Λ be the set of half-integral vectors $\nu \in \{0, 0.5, 1\}^{\mathcal{V}\times\mathcal{L}}$ whose cost in the roof duality relaxation is finite. It is,

$$\Lambda = \{\nu \in \{0,0.5,1\}^{\mathcal{V}\times\mathcal{L}} \mid \forall s\ \big(\exists i\ \nu_{(s,i)}{=}1\big) \Rightarrow \big(\forall j{\neq}i\ \nu_{(s,j)}{=}0\big)\ \}.$$

Indeed, suppose $i \neq j$ and $\nu_{(s,i)}=1$ and $\nu_{(s,j)} > 0$. By constraints (32) it must be that $\nu_{(s,i)(s,j)} \geq \max(0, \nu_{(s,i)}+\nu_{(s,j)}-1) = \nu_{(s,j)} > 0$, and because of the hard terms h the relaxed cost is infinite. Under other cases $\nu_{(s,i)(s,j)}$ can be 0 and the relaxed cost is finite.

For vectors $\nu \in \Lambda$ the relaxed cost can be written as

$$E(\nu) = E_0(\nu) + C\sum_s (1 - \sum_i \nu_{(s,i)}),$$

where $E_0(\nu)$ corresponds to the relaxation of the first two terms in (83). Without loss of generality we can assume that $E_0(\nu) \geq 0 \ \forall \nu$. We will not need an explicit form of $E_0(\nu)$; however, we will use the property $-C_0 \leq E_0(\nu') - E_0(\nu) \leq C_0$ for $\nu, \nu' \in \Lambda$, which is easy to check.

For $\nu^* \in \Lambda$ for each s consider the following hypothetical cases:

- Suppose $\exists i \ \nu^*_{(s,i)} = 1$. Then it is necessary $\nu^*_s = (0, \dots, 0, 1, 0, \dots, 0)$. Let ν be the vector obtained from ν^* by setting $\nu_{(s,i)}$ for all i to 0.5. We have $\nu \in \Lambda$ and

$$E(\nu) - E(\nu^*) = E_0(\nu) - E_0(\nu^*) + \frac{1}{2}CK - C(K-1) \leq C_0 - \frac{K-2}{2}C < 0,$$

which contradicts to the optimality of ν^*.

- Suppose $\forall j \ \nu^*_{(s,j)} < 1$ and $\nu^*_{(s,i)} = 0$ for some i. Let ν be the vector obtained from ν^* by modifying $\nu_{(s,i)}$ from 0 to 0.5. We have $\nu \in \Lambda$ and

$$E(\nu) - E(\nu^*) = E_0(\nu) - E_0(\nu^*) - \frac{C}{2} \leq C_0 - \frac{1}{2}C < 0,$$

which again contradicts to the optimality of ν^*.

These contradictions imply the lemma. □

References

[1] E. Boros and P. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, (123(1-3)):155–225, 2002. 4, 5, 8, 10, 12, 24

[2] E. Boros, P. L. Hammer, and X. Sun. Network flows and minimization of quadratic pseudo-Boolean functions. Technical Report RRR 17-1991, RUTCOR, May 1991. 4, 8, 17

[3] E. Boros, P. L. Hammer, and G. Tavares. Preprocessing of unconstrained quadratic binary optimization. Technical Report RRR 10-2006, RUTCOR, Apr. 2006. 8

[4] 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, 3

[5] M. Fisher, G. Nemhauser, and L. Wolsey. An analysis of approximations for maximizing submodular setfunctionsi. *Math. Programming*, pages 265–294, 1978. 10

[6] P. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Math. Programming*, pages 121–155, 1984. 1

[7] P. L. Hammer. Some network flow problems solved with pseudo-Boolean programming. *Operation Research*, 13:388–399, 1965. 5

[8] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, 2003. 5, 7

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, october 2006. 1, 3, 4, 17

[10] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. In *ECCV 2002*, number 2352 in LNCS, pages 65–81, 2002. 5

[11] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, February 2004. 5

[12] V. N. Kolmogorov and M. J. Wainwright. On optimality of tree-reweighted max-product message-passing. In *Appeared in Uncertainty in Artificial Intelligence*, July 2005.

[13] A. M. Koster, S. Hoesel, and A. Kolen. The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, 23:89–97(9), October 1998. 1

[14] V. Koval and M. Schlesinger. Two-dimensional programming in image analysis problems. *Automatics and Telemechanics*, 2:149–168, 1976. In Russian. 1

[15] I. Kovtun. Partial optimal labeling search for a NP-hard subclass of (max, +) problems. In *DAGM-Symposium*, pages 402–409, 2003. 10

[16] I. Kovtun. *Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem*. PhD thesis, 2004. In Ukrainian. 5, 7

[17] D. Schlesinger. Exact solution of permuted submodular minsum problems. In A. L. Yuille, S. C. Zhu, D. Cremers, and Y. Wang, editors, *International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 4679 of *Lecture Notes in Computer Science*, pages 28–38. Springer, 2007.

[18] D. Schlesinger and B. Flach. Transforming an arbitrary minsum problem into a binary one. Research Report TUD-FI06-01, Dresden University of Technology, April 2006. 1, 5, 7

[19] M. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika, Kiev*, 4:113–130, 1976. In Russian. 1, 3, 4

[20] M. I. Schlesinger and B. Flach. Some solvable subclasses of structural recognition problems. In *Czech Pattern Recognition Workshop*, February 2000. 5

[21] A. Shekhovtsov. Supermodular decomposition of structural labeling problem. *Control Systems and Computers*, 1:39–48, 2006. 20

[22] M. Wainwright, T. Jaakkola, and A. Willsky. Exact MAP estimates by (hyper)tree agreement. In S. T. S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 809–816. MIT Press, 2003. 1, 3, 17, 20

[23] T. 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, 3, 4, 10, 12