Maximum Persistency in Energy Minimization

Alexander Shekhovtsov
Graz University of Technology
shekhovtsov@icg.tugraz.at*

Abstract

We consider discrete pairwise energy minimization problem (weighted constraint satisfaction, max-sum labeling) and methods that identify a globally optimal partial assignment of variables. When finding a complete optimal assignment is intractable, determining optimal values for a part of variables is an interesting possibility. Existing methods are based on different sufficient conditions. We propose a new sufficient condition for partial optimality which is: (1) verifiable in polynomial time (2) invariant to reparametrization of the problem and permutation of labels and (3) includes many existing sufficient conditions as special cases. We pose the problem of finding the maximum optimal partial assignment identifiable by the new sufficient condition. A polynomial method is proposed which is guaranteed to assign same or larger part of variables than several existing approaches. The core of the method is a specially constructed linear program that identifies persistent assignments in an arbitrary multi-label setting.

1. Introduction

Energy Minimization Given a graph $(\mathcal{V}, \mathcal{E})$ and functions $f_s \colon \mathcal{L}_s \to \mathbb{R}$ for all $s \in \mathcal{V}$ and $f_{st} \colon \mathcal{L}_s \times \mathcal{L}_t \to \mathbb{R}$ for all $st \in \mathcal{E}$, where $\mathcal{L}_s$ are finite sets of *labels*, the problem is to minimize the *energy*

$$E_f(x) = f_0 + \sum_{s\in\mathcal{V}} f_s(x_s) + \sum_{st\in\mathcal{E}} f_{st}(x_s, x_t), \qquad (1)$$

over all assignments $x \in \mathcal{L} = \prod_s \mathcal{L}_s$ (Cartesian product). Notation st denotes the ordered pair (s,t) for $s,t \in \mathcal{V}$. The general energy minimization problem is APX-hard.

Partial Optimality Let $\mathcal{A} \subset \mathcal{V}$. By $x_\mathcal{A}$ we denote the restriction of x to $\mathcal{A}$. An assignment y with domain $\mathcal{A}$ is a *partial assignment* denoted $(\mathcal{A}, y)$. The pair $(\mathcal{A}, y)$ is called *strong optimal partial assignment* if there holds $x^*_\mathcal{A} = y$ for *any* minimizer x^* of E_f. And *weak optimal partial assignment* if *there exists* a minimizer x^* of E_f such that $x^*_\mathcal{A} = y$.

Related Work Several fundamental results identifying optimal partial assignments are obtained from the properties of linear relaxations of some discrete problems. An optimal solution to continuous relaxation of a mixed-integer 0-1 programming problem is defined by Adams *et al*. [2] to be *persistent* if the set of $[0,1]$ relaxed variables realizing binary values retains the same binary values in at least one integer optimum. A mixed-integer program is said to be *persistent* (or possess the *persistency* property) if *every* solution to its continuous relaxation is persistent. Nemhauser & Trotter [19] proved that the vertex packing problem is persistent. This result was later generalized to optimization of quadratic pseudo-Boolean functions (equivalent to energy minimization with two labels) by Hammer *et al*. [9]. The relaxed problem in this case is known as the *roof dual*. *Strong persistency* was also proven, stating that if a variable takes the same binary value in *all* optimal solutions to the relaxation, then *all* optimal solutions to the original 0-1 problem take this value. However, it is a rare case that a relaxation of a particular problem is persistent.

Several works considered generalization of persistency to higher-order pseudo-Boolean functions. Adams *et al*. [2] considered a hierarchy of continuous relaxations of 0-1 polynomial programming problems. Given an optimal relaxed solution, they derive sufficient conditions on the dual multipliers which ensure that the solution is persistent. This result generalizes the roof duality approach, coinciding with it in the case of quadratic polynomials in binary variables. Kolmogorov [13, 14] studied submodular and bisubmodular relaxations and showed that they provide a natural generalization of the quadratic pseudo-Boolean case to higher-order terms and possess the persistency property. Kahl and Strandmar [11] proposed a polynomial time algorithm to find the tightest submodular relaxation. Lu and Williams [18], Ishikawa [10] and Fix *et al*. [6] obtained partial optimalities via different reductions to quadratic problems and subsequent application of the roof dual.

*This work was supported by the EU project FP7-ICT-247870 NIFTi and the Austrian Science Fund (FWF) under the START project BIVISION, No. Y729.

Multi-label energies The following methods were proposed for the pairwise model (1) with multi-label variables. Kohli *et al*. [12] reduced multi-label energy to quadratic pseudo-Boolean and applied roof dual. The family of local methods known as *dead end elimination* (DEE), originally proposed by Desmet *et al*. [5], uses simple sufficient conditions that consider a variable and its immediate neighbors in the graph. Kovtun [16, 17] proposed to construct an auxiliary submodular problem whose solution provides a partial optimal assignment for the original problem. For the Potts model it was shown that K auxiliary problems can be solved in time $O(\log(K)F)$, where F is the time to solve a single auxiliary problem [8]. Swoboda *et al*. [24] proposed a method for Potts model solving a series of LP relaxations approximately and generalized it recently to general and higher-order energies [25]. Unlike other approaches, methods [5, 16] are not directly related to relaxation techniques.

Contribution We observed that in many methods there is an underlying mapping of labelings $p\colon \mathcal{L} \to \mathcal{L}$ that *improves* the energy of any given labeling: $E_f(p(x)) \leqslant E_f(x)$. It follows that there exists a minimizer in the reduced search space $p(\mathcal{L})$. However, even in the case that such mapping is given, the verification of the improving property is NP-hard (see below). We propose instead to verify that a suitable linear *extension* of this mapping improves the energy of all *relaxed* labelings. This constitutes a sufficient condition which is polynomial to verify. It includes sufficient conditions used in methods [5, 17, 9, 12, 24] as special cases.

We pose the problem of finding the *maximum* weak/strong optimal partial assignment identifiable by the new sufficient conditions (denoted MAX-WI / MAX-SI, respectively). We propose polynomial algorithms for several classes of mappings p, which include many of previously proposed constructions. The algorithms involve solving the LP-relaxation and an additional linear program of a comparable size. We give a method that improves over one-against-all method of Kovtun [17] (including possible free choices in this method) and subsumes the method [24]. In the case of two labels, our method reduces to known QPBO results. Experimental verification of correctness and quantification of achieved improvement is performed on difficult random instances. Preliminary experiments with large-scale vision problems are reported in [22].

2. Background

We will assume that $st \in \mathcal{E} \Rightarrow ts \notin \mathcal{E}$. Let us denote the set $\mathcal{L}_s \times \mathcal{L}_t$ as $\mathcal{L}_{st}$ and the pair of labels $(i,j) \in \mathcal{L}_{st}$ as ij. The following set of indices is associated with the graph $(\mathcal{V},\mathcal{E})$ and the set of labelings: $\mathcal{I} = \{0\} \cup \{(s,i) \mid s \in \mathcal{V},\ i \in \mathcal{L}_s\} \cup \{(st,ij) \mid st \in \mathcal{E},\ ij \in \mathcal{L}_{st}\}$. A vector

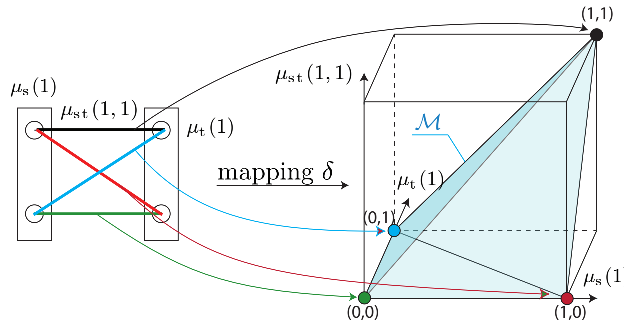

Figure 1. Mapping δ embeds discrete labelings as points in the space $\mathbb{R}^{\mathcal{I}}$. Projection onto components $\mu_s(1), \mu_t(1), \mu_{st}(1,1)$ is shown, the other components are dependent.

$f \in \mathbb{R}^{\mathcal{I}}$ has components (coordinates) f_0, $f_u(l)$, $f_{st}(i,j)$ for all $u \in \mathcal{V}$, $l \in \mathcal{L}_u$, $st \in \mathcal{E}$, $ij \in \mathcal{L}_{st}$. We further define that $f_{ts}(j,i) = f_{st}(i,j)$. Let $\tilde{\mathcal{E}} = \mathcal{E} \cup \{ts \mid st \in \mathcal{E}\}$, the symmetric closure of $\mathcal{E}$. The *neighbors* of a pixel s are pixels in the set $\mathcal{N}(s) = \{t \mid st \in \tilde{\mathcal{E}}\}$.

LP Relaxation Let $\delta(x) \in \mathbb{R}^{\mathcal{I}}$ be the vector with components $\delta(x)_0 = 1$, $\delta(x)_s(i) = [\![x_s{=}i]\!]$ and $\delta(x)_{st}(i,j) = [\![(x_s,x_t){=}ij]\!]$, where $[\![\,]\!]$ is the Iverson bracket. Let $\langle\cdot,\cdot\rangle$ denote the scalar product in $\mathbb{R}^{\mathcal{I}}$. We can write the energy as

$$E_f(x) = \langle f, \delta(x)\rangle. \tag{2}$$

The energy minimization can be expressed and relaxed as

$$\min_{x\in\mathcal{L}}\langle f,\delta(x)\rangle = \min_{\mu\in\delta(\mathcal{L})}\langle f,\mu\rangle = \min_{\mu\in\mathcal{M}}\langle f,\mu\rangle \geqslant \min_{\mu\in\Lambda}\langle f,\mu\rangle, \tag{3}$$

where $\mathcal{M} = \operatorname{conv}\delta(\mathcal{L})$ and Λ is the *local* polytope that makes an outer approximation of $\mathcal{M}$. We consider the standard Schlesinger's LP relaxation [23], where the polytope Λ is given by the primal constraints in the following primal-dual pair:

(LP-primal)		(LP-dual)
$\min\langle f,\mu\rangle$	$=$	$\max\psi$
$\sum_j \mu_{st}(i,j) - \mu_s(i) = 0,$		$\varphi_{st}(i) \in \mathbb{R},$
$\sum_i \mu_{st}(i,j) - \mu_t(j) = 0,$		$\varphi_{ts}(j) \in \mathbb{R},$
$\sum_i \mu_s(i) - \mu_0 = 0,$		$\varphi_s \in \mathbb{R},$
$\mu_0 = 1,$		$\psi \in \mathbb{R},$
$\mu_s(i) \geqslant 0,$		$f_s(i) + \sum_{t\in\mathcal{N}(s)}\varphi_{st}(i) - \varphi_s \geqslant 0,$
$\mu_{st}(i,j) \geqslant 0,$		$f_{st}(i,j) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0,$
$\mu_0 \geqslant 0;$		$f_0 + \sum_s \varphi_s - \psi \geqslant 0.$

This relaxation is illustrated in Figure 1. We write it compactly as

$$\begin{aligned} \min\langle f,\mu\rangle &= & \max\psi\,, \\ A\mu &= 0 & \varphi \in \mathbb{R}^m \\ \mu_0 &= 1 & \psi \in \mathbb{R} \\ \mu &\geqslant 0 & f - A^{\mathsf T}\varphi - e_0\psi \geqslant 0 \end{aligned} \tag{LP}$$

where A is $m \times |I|$ and $e_0 \in \mathbb{R}^{\mathcal{I}}$ is the basis vector for component 0. Vector $f^{\varphi} := f - A^{\mathsf{T}}\varphi$ is called an *equivalent transformation* (*reparametrization*) of f. There holds $\langle f^{\varphi}, \mu \rangle = \langle f, \mu \rangle - \langle \varphi, A\mu \rangle = \langle f, \mu \rangle$ for all $\mu \in \Lambda$. Because $\Lambda \supset \delta(\mathcal{L})$, it follows that $E_f(x) = E_{f^{\varphi}}(x)$ for all $x \in \mathcal{L}$. If there exists φ such that $g = f^{\varphi}$ we write $g \equiv f$. In this case vectors f and g are different but they define equal energy functions $E_f = E_g$. See, *e.g.*, [26] for more detail.

Let $(\mu, (\varphi, \psi))$ be a feasible primal-dual pair. Complementary slackness for (LP) states that μ is optimal to the primal and (φ, ψ) to the dual iff

$$\mu_s(i) > 0 \;\Rightarrow\; f^{\varphi}_s(i) = 0, \tag{4a}$$
$$\mu_{st}(i,j) > 0 \;\Rightarrow\; f^{\varphi}_{st}(i,j) = 0, \tag{4b}$$
$$\mu_0 > 0 \;\Rightarrow\; \psi = f_0 + \sum_s \varphi_s. \tag{4c}$$

Because a feasible dual solution satisfies $(\forall i')\ f^{\varphi}_s(i') \geqslant 0$, condition on the RHS[1] of (4a) imply that label i is *minimal* for f^{φ}. Similarly, in case of (4b) we say that ij is a minimal pair. Implication (4c) has its premise always satisfied.

3. Improving Mapping

Definition 1. A mapping $p\colon \mathcal{L} \to \mathcal{L}$ is called *(weakly) improving* for f if

$$(\forall x \in \mathcal{L}) \quad E_f(p(x)) \leqslant E_f(x), \tag{5}$$

and *strictly improving* if

$$(p(x) \neq x) \;\Rightarrow\; E_f(p(x)) < E_f(x), \tag{6}$$

We will consider *pixel-wise* mappings, of the form $p(x)_s = p_s(x_s)$, where $(\forall s \in \mathcal{V})\ p_s\colon \mathcal{L}_s \to \mathcal{L}_s$. Furthermore, we restrict to idempotent mappings, *i.e.*, satisfying $p \circ p = p$, where $\circ$ denotes composition.

Statement 1. Let p be an improving pixel-wise idempotent mapping. Then there exists an optimal solution x^* such that

$$(\forall i) \quad p_s(i) \neq i \;\Rightarrow\; x^*_s \neq i. \tag{7}$$

In case p is strictly improving any optimal solution x^* satisfies (7).

Proof. Let x be optimal. Then $x^* = p(x)$ is optimal as well. By idempotency, x^* satisfies $p(x^*) = x^*$. Condition (7) is equivalent to $(\forall i)\ x^*_s = i \Rightarrow p_s(i) = i$. If p is strictly improving, for any optimal solution x^* there must hold $p(x^*) = x^*$, otherwise $E_f(p(x^*)) < E_f(x^*)$. □

It follows that knowing an improving mapping, we can eliminate labels (s,i) for which $p_s(i) \neq i$ as non-optimal. Given a mapping p, the verification of the improving property is NP-hard: in case of binary variables it includes NP-hard decision problem of whether a partial assignment is an autarky [4]. We need a simpler sufficient condition. It will be constructed by embedding the mapping into the linear space and applying a relaxation there.

[1]RHS = Right-hand side of an equation.

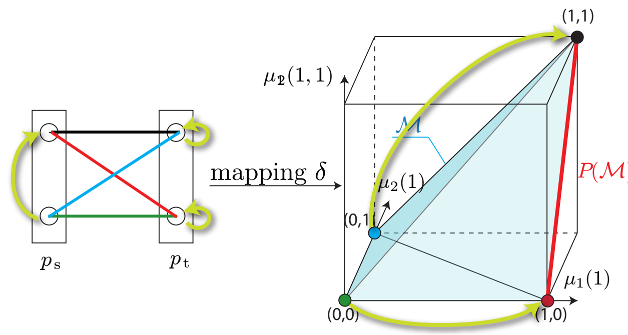

Figure 2. Discrete map p sends some labelings to other (the green labeling to red and the blue one to black). There is a corresponding linear map $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ (unique on aff$(\mathcal{M})$) wit this action – the oblique projection onto the red facet.

3.1. Relaxed Improving Mapping

Definition 2. A *linear extension* of $p\colon \mathcal{L} \to \mathcal{L}$ is a linear mapping $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ that satisfies

$$(\forall x \in \mathcal{L}) \quad \delta(p(x)) = P\delta(x). \tag{8}$$

See Figure 2 for illustration. We will only use the following particular linear extension for a pixel-wise mapping $p\colon \mathcal{L} \to \mathcal{L}$, which will be denoted $[p]$. For each p_s define matrix $P_s \in \mathbb{R}^{\mathcal{L}_s \times \mathcal{L}_s}$ as $P_{s,ii'} = [\![p_s(i') = i]\!]$. The linear extension $P = [p]$ is given by

$$\begin{aligned} (P\mu)_0 &= 1, \\ (P\mu)_s(i) &= P_s\mu_s, \\ (P\mu)_{st}(ij) &= P_s\mu_{st}P_t^{\mathsf{T}}. \end{aligned} \tag{9}$$

Linear maps of the form (9) with general matrices P_s satisfying $P_s \geqslant 0$ and $1^{\mathsf{T}}P_s = 1$ will be called *pixel-wise*. To verify that (8) holds true we expand the components as follows. $(P\delta(x))_s(i) = \sum_{i' \in \mathcal{L}_s} [\![p_s(i'){=}i]\!][\![x_s{=}i']\!] = [\![p_s(x_s){=}i]\!] = \delta(p(x))_s(i)$. Similarly, for pairwise components, $(P\delta(x))_{st}(i,j) = [\![p_s(x_s){=}i]\!][\![p_t(x_t){=}j]\!] = \delta(p(x))_{st}(i,j)$.

Using the linear extension P of p we can write

$$E_f(p(x)) = \langle f, \delta(p(x)) \rangle = \langle f, P\delta(x) \rangle. \tag{10}$$

This allows to express condition (5) as

$$(\forall x \in \mathcal{L}) \quad \langle f, P\delta(x) \rangle \leqslant \langle f, \delta(x) \rangle. \tag{11}$$

We introduce a sufficient condition by requiring that this inequality is satisfied over a larger subset Λ.

Definition 3. A linear mapping $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ is a (weak) Λ-*improving* mapping for f if

$$(\forall \mu \in \Lambda) \quad \langle f, P\mu \rangle \leqslant \langle f, \mu \rangle; \tag{12}$$

and is a *strict* Λ-*improving* mapping for f if

$$(\forall \mu \in \Lambda,\ P\mu \neq \mu) \quad \langle f, P\mu \rangle < \langle f, \mu \rangle. \tag{13}$$

The set of mappings for which (12) (resp. (13)) is satisfied will be denoted $\mathbb{W}_f$ (resp. $\mathbb{S}_f$). For convenience, we will use the term *relaxed improving* map, meaning it w.r.t. polytope Λ. Note, this definition and some theorems are given for arbitrary linear maps, at the same time for the purpose of this paper it would be sufficient to assume pixel-wise maps of the form (9). Clearly, (12) implies (11) because $\delta(\mathcal{L}) \subset \Lambda$ and for the linear extension $[p]$ it implies that p is improving. Sets $\mathbb{W}_f$ and $\mathbb{S}_f$ are convex as they are intersections of half-spaces (respectively, closed and open). *Verification* of (12) for a given P can be performed via solving

$$\min_{\mu \in \Lambda} \langle (I - P^{\mathsf{T}}) f, \mu \rangle \tag{14}$$

and checking that the result is non-negative, *i.e*. can be decided in polynomial time.

4. Special Cases

The new sufficient condition generalizes several previously proposed sufficient conditions. We will show here improving mappings corresponding to each method. Proofs that the mapping is actually relaxed improving in each case (which is a stronger requirement) is given in [22].

DEE There is a number of local sufficient conditions proposed [5, 7], *etc*. We can show to include Goldstain's simple DEE condition. The condition allows to eliminate label $\alpha \in \mathcal{L}_s$ if there is label $\beta \in \mathcal{L}_s$ such that replacing α with β does not increase the energy for all configurations of neighboring pixels. Clearly, we have the map p which is pixel-wise, it's component p_t is identity for all $t \neq s$ and component p_s has $p_s(\alpha) = \beta$ and $p_s(i) = i$ for $i \neq \alpha$.

QPBO [19, 9] As mentioned above, in the case of 2 lables the LP relaxation has persistency property. The partial assignment of the integral part $(y, \mathcal{A})$ is globally optimal. Moreover, it has the *autarky* property: for any labeling x there holds $E_f(x[\mathcal{A} \leftarrow y]) \leqslant E_f(x)$, where $x[\mathcal{A} \leftarrow y]$ is the labeling obtained from x by switching its $\mathcal{A}$ components to y [4]. Clearly, the map $p\colon x \mapsto x[\mathcal{A} \leftarrow y]$ is improving. The relaxed improving property does not hold for autarkies in general, but for the solution of the LP relaxation.

MQPBO [12] This method extends partial optimality properties of QPBO to multi-label problems via the reduction of the problem to 0-1 variables. The method outputs two labelings $x^{\min}$ and $x^{\max}$ with the guarantee that there exists optimal labeling x that satisfy $x_s \in [x_s^{\min}, x_s^{\max}]$. The improving mapping for this method is given by $p\colon x \mapsto (x \vee x^{\min}) \wedge x^{\max}$.

Auxiliary Submodular Problems [16, 17] These methods construct an *auxiliary* submodular (in a given ordering of labels) energy E_g. A minimizer y of E_g has the property that $E_g(x \vee y) \leqslant E_g(x)$, implied by submodularity. It follows that mapping $p\colon x \mapsto x \vee y$ is improving for g. The construction of the auxiliary function g ensures the inequality $E_f(x \vee y) - E_f(x) \leqslant E_g(x \vee y) - E_g(x)$. It follows immediately that p is improving for f.

Iterative Pruning [24] This method is applicable to the Potts model and constructs a partial optimal assignment $(\mathcal{A}, y)$. It turns out that the mapping $p\colon x \mapsto x[\mathcal{A} \leftarrow y]$ is relaxed improving. A recent generalization of this method to arbitrary energies [25] in the pairwise model satisfies our sufficient condition as well.

5. Maximum Improving Mapping

Having a more powerful sufficient condition, which can be verified in polynomial time, how do we find a map that satisfies it? How do we find the map that delivers *the largest* partial optimal assignment, or, equivalently, eliminates the maximum number of labels as non-optimal? Recall that the label (s, i) is eliminated by pixel-wise mapping p if $[\![p_s(i) {\neq} i]\!]$. We therefore formulate the following *maximum persistency* problem:

$$\max_p \sum_{s,i} [\![p_s(i) \neq i]\!] \quad \text{s.t. } [p] \in \mathbb{W}_f. \tag{MAX-WI}$$

The strict variant, with constraint $[p] \in \mathbb{S}_f$, will be denoted MAX-SI. The problem may look difficult, however, we will be able to solve it in polynomial time for some types of maps covering nearly all types that appeared in the previous section:

- *all-to-one maps*. Set $\mathcal{P}^{1,y}$ of maps of the form $p\colon x \mapsto x[\mathcal{A} \leftarrow y]$ for all $\mathcal{A} \subset \mathcal{V}$ and fixed $y \in \mathcal{L}$.
- *subset-to-one maps*. Let $V = \{(s, i) \mid s \in \mathcal{V}, i \in \mathcal{L}_s\}$. Let $\xi \in \{0, 1\}^V$. Mapping p_ξ in every pixel either preserves the label or switches it to y_s:
$$p_\xi(x)_s = \begin{cases} y_s & \text{if } \xi_{s x_s} = 1, \\ x_s & \text{otherwise.} \end{cases} \tag{15}$$
Vector $(\xi_{si} \mid i \in \mathcal{L}_s)$ serves as the indicator of a subset of labels in pixel s that are mapped to y_s. The set $\mathcal{P}^{2,y}$ of all such maps is considered.
- *all-to-one-unknown* maps. Set $\mathcal{P}^1 = \bigcup_{y \in \mathcal{L}} \mathcal{P}^{1,y}$.

Additionally, we define *subset-to-one-unknown* maps as the set $\mathcal{P}^2 = \bigcup_{y \in \mathcal{L}} \mathcal{P}^{2,y}$. This set is considered merely to draw the boundary between solvable and unsolvable cases of maximum persistency problem. All complexity results are summarized in Table 1. We see that as soon as $K > 3$ the problem with unconstrained maps becomes intractable. We also see that the complexity jumps with the number of possible destinations for each label increased. Note, in case of all-to-one-unknown maps the difference between strict and weak conditions results in a different complexity class!

problem type	MAX-SI	MAX-WI
$K=2$	P (QPBO)	P (QPBO)
$K=3$	?	NP-hard
$K>3$	NP-hard	NP-hard
$\mathcal{P}^{1,y}$	P (ε-L1)	P (L1)
$\mathcal{P}^{2,y}$	P (ε-L1)	P (L1)
$\mathcal{P}^{1}$	P (nec. cond. + ε-L1)	NP-hard
$\mathcal{P}^{2}$	NP-hard	NP-hard

Table 1. Complexity of maximum persistency problem. Notation $K = 2$ means the class of problems with 2 labels and arbitrary maps. In brackets we denote the respective polynomial method, see §6.

6. Algorithms

For the case of two labels ($K = 2$), problem MAX-SI (resp. MAX-WI) can be solved by finding solution to (LP) with the minimum (resp. maximum) number of integer components. This corresponds to finding specific cuts in the network flow model [3], [15, §2.3]. Finding the relaxed solution with the maximum number of integer components was proven polynomial by Picard and Queyranne [20] in the context of vertex packing problem. We give more detail in [22].

To show that for $K \geqslant 3$ problem MAX-WI is NP-hard we notice that (LP) is tight iff there exists $y \in \mathcal{L}$ such that mapping $p\colon \mathcal{L} \mapsto y$ is relaxed-improving. Clearly, this mapping is a (non-unique) solution to MAX-WI. Verifying tightness of (LP) is a pairwise constraint satisfaction problem which is NP-hard for $K \geqslant 3$.

We will now derive some properties of MAX-WI/SI problem that will enable our main result – reduction to a single linear program for subset-to-one maps. The problem will be gradually reformulated in terms of linear extension $P = [p]$ only. The constraint $P \in \mathbb{W}_f$ is complicating because set $\mathbb{W}_f$ is defined with quantifier $(\forall x \in \Lambda)$, see (12). However, since Λ is polyhedral, this set can be reformulated as a projection of a higher-dimensional polytope:

Statement 2 (Dual $\mathbb{W}$)**.** Set $\mathbb{W}_f$ can be expressed as

$$\{P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}} \mid (\exists \varphi \in \mathbb{R}^m)\ \ f^{\varphi} - P^{\mathsf{T}} f \geqslant 0\}. \tag{16}$$

Proof. Denote $g = (I - P^{\mathsf{T}})f$. Condition (14), equivalent to (12), can be stated for the conic hull of Λ:

$$\inf_{\mu \in \operatorname{coni}(\Lambda)} \langle g, \mu \rangle \geqslant 0. \tag{17}$$

This is because for any $\mu \in \Lambda$ and any $\alpha \geqslant 0$ vector $\alpha\mu$ will satisfy RHS of (12) as well. Observe that $\operatorname{coni}(\Lambda) = \{\mu \mid A\mu = 0,\ \mu \geqslant 0\}$ (in the specific representation of the polytope we used we just have to drop the constraint $\mu_0 = 1$). We can write minimization problem in (17) and its dual as

$$\begin{array}{ll} \inf \langle g, \mu \rangle & \max 0\,. \\ A\mu = 0 & \varphi \in \mathbb{R}^m \\ \mu \geqslant 0 & g - A^{\mathsf{T}}\varphi \geqslant 0 \end{array} \tag{18}$$

Inequality (17) holds iff the primal problem is bounded, and it is bounded iff the dual is feasible, which is the case iff $(\exists \varphi \in \mathbb{R}^m)\ (f - A^{\mathsf{T}}\varphi) - P^{\mathsf{T}} f \geqslant 0$. □

With this reformulation we can write MAX-WI as

$$\max_{p,\varphi} \sum_{s,i} [\![p_s(i) \neq i]\!] \quad \text{s.t.: } (I - [p]^{\mathsf{T}})f - A^{\mathsf{T}}\varphi \geqslant 0. \tag{19}$$

Notice, quantifier $(\exists \varphi)$ turned into an extra minimization variable. To handle the strict case, we would need a similar dual reformulation for the set $\mathbb{S}_f$. This set has a more complicated quantifier $(\forall \mu \in \Lambda, P\mu \neq \mu)$. Fortunately, the following reformulation holds for pixel-wise maps:

Statement 3 (Dual $\mathbb{S}$)**.** Let $p\colon \mathcal{L} \to \mathcal{L}$ be pixel-wise. Then $[p] \in \mathbb{S}_f$ iff $(\exists \varepsilon > 0)\ (\exists \varphi \in \mathbb{R}^m)$

$$(\forall s,\ \forall i) \quad f^{\varphi}_s(i) - f_s(p_s(i)) \geqslant \varepsilon [\![p_s(i) \neq i]\!], \tag{20a}$$

$$(\forall st,\ \forall ij) \quad f^{\varphi}_{st}(i,j) - f_{st}(p_s(i), p_t(j)) \geqslant 0. \tag{20b}$$

Proof. Let $h \in \mathbb{R}^{\mathcal{I}}$ with components $h_s(i) = [\![p_s(i) \neq i]\!]$, $h_{st}(i,j) = 0$. For $\mu \in \Lambda$ there holds $\langle h, \mu \rangle = 0$ iff $[p]\mu = \mu$. Conditions (13) are equivalent to

$$(\forall \mu \in \Lambda) \quad \langle (I - [p]^{\mathsf{T}})f, \mu \rangle \geqslant \varepsilon \langle h, \mu \rangle \tag{21}$$

for some $\varepsilon > 0$. We apply now the same inference as in Statement 2 for vector $g = f - P^{\mathsf{T}} f - \varepsilon h$. It follows that (21) is equivalent to $(\exists \varphi \in \mathbb{R}^m)\ (f - A^{\mathsf{T}}\varphi) - P^{\mathsf{T}} f - \varepsilon h \geqslant 0$. □

Additionally, the following lemma provides necessary conditions for sets $\mathbb{W}_f$, $\mathbb{S}_f$. It will help to narrow down the set of maps over which the optimization is carried out.

Lemma 1 (Necessary Conditions)**.** Let $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$, $P(\Lambda) \subset \Lambda$ and $\mathcal{O} = \operatorname{argmin}_{\mu \in \Lambda} \langle f, \mu \rangle$. Then
 (a) $P \in \mathbb{W}_f \Rightarrow P(\mathcal{O}) \subset \mathcal{O}$.
 (b) $P \in \mathbb{S}_f \Rightarrow (\forall \mu \in \mathcal{O})\ P(\mu) = \mu$.

Proof. **(a)** Assume $(\exists \mu \in \mathcal{O})\ P\mu \in \Lambda \backslash \mathcal{O}$. Then $\langle f, P\mu \rangle > \langle f, \mu \rangle$, therefore $P \notin \mathbb{W}_f$. **(b)** Assume $(\exists \mu \in \mathcal{O})\ P\mu \neq \mu$. Then $\langle f, P\mu \rangle \geqslant \langle f, \mu \rangle$ and therefore $P \notin \mathbb{S}_f$. □

Subset-to-One Maps Let us consider the class of maps $\mathcal{P}^{2,y}$, in which mapping p_ξ is defined by the indicator variable $\xi \in \{0,1\}^V$. We will first consider problem (MAX-WI). The constraint $[p_\xi] \in \mathbb{W}_f$ in the dual form is still complicated by that $[p_\xi]$ defined by (9) involves products $\xi_{si}\xi_{tj}$. We are going to linearize these terms by introducing additional variables ξ_{stij}. Let Σ be set the of vectors ξ with components ξ_{si}, ξ_{stij} such that

$$\begin{aligned} 0 \leqslant{}& \xi_{si} \leqslant 1, \\ \max(0, \xi_{si} + \xi_{tj} - 1) \leqslant{}& \xi_{stij} \leqslant \min(\xi_{si}, \xi_{tj}). \end{aligned} \tag{Σ}$$

If $\xi \in \Sigma$ and all ξ_{si} are integral, there holds $\xi_{stij} = \xi_{si}\xi_{tj}$. Set Σ is convex, polyhedral. For $\xi \in \Sigma$ we introduce the following corresponding mapping P_ξ by replacing products $\xi_{si}\xi_{tj}$ with ξ_{stij} in (9):

$$(P_\xi \mu)_s(i) = \sum_{i'} P_{s,ii'} \mu_s(i'), \tag{22a}$$

$$(P_\xi \mu)_{st}(i,j) = \sum_{i',j'} P_{st,ii',jj'} \mu_{st}(i',j'), \tag{22b}$$

$$\begin{aligned} P_{s,ii'} =& [\![p_s(i'){=}i]\!] = \\ & [\![y_s{=}i]\!]\xi_{si'} + [\![i'{=}i]\!](1-\xi_{si'}), \end{aligned} \tag{23a}$$

$$\begin{aligned} P_{st,ii',jj'} =& [\![y_s{=}i]\!][\![y_t{=}j]\!]\xi_{sti'j'} \\ &+[\![i'{=}i]\!][\![y_t{=}j]\!](\xi_{tj'} - \xi_{sti'j'}) \\ &+[\![y_s{=}i]\!][\![j'{=}j]\!](\xi_{si'} - \xi_{sti'j'}) \\ &+[\![i'{=}i]\!][\![j'{=}j]\!](1-\xi_{si'} - \xi_{tj'} + \xi_{sti'j'}). \end{aligned} \tag{23b}$$

Mapping P_ξ is linear in ξ and for integer ξ it coincides with $[p_\xi]$. We can now formulate (MAX-WI) as the following mixed integer linear program:

$$\begin{aligned} &\max_{\xi,\varphi} \sum_{s,i} \xi_{si} \\ &(I - P_\xi^{\mathsf{T}})f - A^{\mathsf{T}}\varphi \geqslant 0 \\ &\xi \in \Sigma;\ \xi_{si} \in \{0,1\}; \xi_{sy_s} = 0. \end{aligned} \tag{IL1}$$

By relaxing the integrality constraints we obtain linear program (L1). We will prove in Theorem 1 that this relaxation is tight[2]. We first need the following lemma.

Lemma 2. Polytope Λ is closed under mapping P_ξ, $\xi \in \Sigma$.

Proof. We verify that $(\forall \mu \in \Lambda)\ P_\xi \mu \in \Lambda$. Denote $\mu' = P_\xi \mu$. By constraints of Σ, all numbers (23a), (23b) are nonnegative, therefore $\mu' \geqslant 0$. Constraints $1^{\mathsf{T}}\mu'_s = 1$ hold due to $1^{\mathsf{T}} P_s = 1$. Constraints $1^{\mathsf{T}}\mu'_{st} = (\mu'_t)^{\mathsf{T}}$ hold due to $\sum_{ii'} P_{st,ii',jj'} = P_{t,jj'}$. □

Theorem 1. In a solution (ξ,φ) to (L1) vector ξ is integer.

Proof. We will show that rounding ξ up results in a feasible solution with equal or better objective. Because ξ is feasible to (L1), the mapping P_ξ is Λ-improving for f. Note, at this point, unless ξ is integer it is not guaranteed that $P_\xi(\mathcal{M}) \subset \mathcal{M}$ and we cannot draw any partial optimalities from it, neither P_ξ is guaranteed to be idempotent. By Lemma 2, $P_\xi(\Lambda) \subset \Lambda$. Therefore

$$(\forall \mu \in \Lambda) \quad \langle f, P_\xi P_\xi \mu\rangle \leqslant \langle f, P_\xi \mu\rangle \leqslant \langle f, \mu\rangle. \tag{25}$$

It follows that $P_\xi^2 = P_\xi P_\xi$ is Λ-improving. Since $P_\xi(\Lambda) \subset \Lambda$, it is also $P_\xi^2(\Lambda) \subset P_\xi(\Lambda) \subset \Lambda$. Moreover, $P_\xi^2 = P_{\xi'}$ with the following coefficients ξ':

$$\begin{aligned} \xi'_{si} &= 1 - (1-\xi_{si})^2, \\ \xi'_{stij} &= (1 - \xi_{si} - \xi_{tj} + \xi_{stij})^2 - 1 + \xi'_{si} + \xi'_{tj}. \end{aligned} \tag{26}$$

It can be verified that $\xi' \in \Sigma$. Let $P_{\xi*} = \lim_{n\to\infty}(P_\xi)^{2^n}$. Then

$$\xi^*_{si} = \lim_{n\to\infty} 1 - (1-\xi_{si})^{2^n} = [\![\xi_{si} > 0]\!]. \tag{27}$$

Since $P_{\xi*}$ is Λ-improving, it is feasible to (L1). Assume for contradiction that there exist (s',i') such that $0 < \xi_{s'i'} < 1$. From (27) we have $\xi^*_{si} \geqslant \xi_{si}$ for all si and $\xi^*_{s'i'} > \xi_{s'i'}$. It follows that ξ^* achieves a better objective value, which contradicts the optimality of ξ. Therefore ξ is integer. □

Since the optimal solution to (L1) is integer and unique (as seen from the objective), it is the solution to (MAX-WI).

Problem (MAX-SI) can be approached similarly, using the dual definition of $\mathbb{S}$. The inequalities for pairwise terms (20b) are linearized exactly the same way as for the weak case, we can write them shortly as

$$((I - P_\xi^{\mathsf{T}})f - A^{\mathsf{T}}\varphi)_{st}(i,j) \geqslant 0. \tag{28}$$

The inequalities for univariate terms (20a), by substituting p_ξ can be expressed as

$$(f_s(i) - f_s(y_s))\xi_{si} - (A^{\mathsf{T}}\varphi)_s(i) \geqslant \varepsilon\xi_{s,i}[\![i \neq y_s]\!]. \tag{29}$$

Since we assume $\xi_{sy_s} = 0$, expression (29) is equivalent to

$$(f_s(i) - f_s(y_s) - \varepsilon)\xi_{si} - (A^{\mathsf{T}}\varphi)_s(i) \geqslant 0, \tag{30}$$

i.e., we obtained the same form of constraints as for the weak case, but with slightly modified vector f. Namely, components $f_s(y_s)$ are increased by ε for all s. Let us denote the problem (L1) with ε-modified vector f as (ε-L1). Since the solution ξ to (ε-L1) is integer it solves MAX-SI.

These solutions can be applied for one or more test labelings y. A polynomial algorithm, for example, can iterate over labelings $(y^\alpha \mid \forall s\ y_s = \alpha)$ for $\alpha = 0,\dots,K-1$. This algorithm subsumes simple Goldstein's DEE [7] and the series of Kovtun's weak one-against-all subproblems for candidate labelings y^α. Most efficient in practice seems to set y_s to one of the immovable labels by the necessary conditions by Lemma 1. This approach in fact allows to solve optimally MAX-SI problem for the next class of mappings.

All-to-One-Unknown Let us consider the class $\mathcal{P}^1$, in which map p_ξ is defined by $\xi \in \{0,1\}^{\mathcal{V}}$ and labeling $y \in \mathcal{L}$. Problem (MAX-WI) is NP-hard by our argument above for $K \geqslant 3$, valid for this class as well. However, we can solve the MAX-SI problem combining necessary conditions by Lemma 1 and (ε-L1) problem as proposed in Algorithm 1.

[2]Problem (L1) can be further simplified by expanding the constraints and optimizing out variables ξ_{stij}, this however would occlude the proof.

Algorithm 1: Max Strong all-to-one-unknown

```
μ ∈ argmin_{μ∈Λ}⟨f, μ⟩;            /* solve (LP) */
For all s if exists i ∈ L_s such that μ_s(i) = 1 then set
  y_s = i, otherwise set y_s arbitrarily;
Solve the problem (ε-L1) with y;
```

Necessary conditions in this case either provide the unique labeling y_s or prove that p_s must be identity. The optimality of the method follows. This algorithm subsumes strict variant of Kovtun's one-against-all auxiliary problem, under an arbitrary choice of a test labeling y and the iterative pruning method [24].

7. Experiments

We report results on random problems with Potts interactions and full interactions. Both types have unary weights $f_s(i) \sim U[0,\, 100]$ (uniformly distributed). Full random energies have pairwise terms $f_{st}(i,j) \sim U[0,\, 100]$ and Potts energies have $f_{st}(i,j) = -\gamma_{st}(i)[\![i{=}j]\!]$, where $\gamma_{st}(i) \sim U[0,\, 50]$. All costs are integer to allow for exact verification of correctness. Only instances with non-zero integrality gap w.r.t. standard LP-relaxation are considered. For each of the methods in Table 2, we measure *solution completeness* as $\frac{n_{\rm elim}}{|\mathcal{V}|(K-1)}100\%$, where $n_{\rm elim}$ is the total number of pairs $(s \in \mathcal{V}, i \in \mathcal{L}_s)$ eliminated by the method as non-optimal. The results are shown in Figure 3.

8. Conclusion

We have identified a common mechanism of improving mappings that works in different methods for partial optimality and proposed how to obtain more general optimality guarantees from a given linear relaxation. It leads to a coherent and short description of several methods and analysis of their common properties. From necessary conditions by Lemma 1 it follows that all the methods reviewed in §4 (as well as the proposed method) cannot be used to tighten the LP-relaxation, they can only simplify it in some cases. While our algorithms work for a restricted class of mappings, many previous methods are based on more narrow classes and use less powerful sufficient conditions. We therefore have a theoretical guarantee to improve over these methods and we have verified on difficult random problems that the improvement is significant.

The difference between week and strict conditions may seem unimportant in practice and was often neglected in the previous work. However, the class of mappings for which the maximum persistency problem is polynomially solvable is larger for strict conditions. Therefore, the difference is important for developing algorithms and for the theoretical comparison of different methods. We believe it is also essential for clarity and completeness to keep track of both. Moreover, it may be useful in practice to have a threshold ε, below which (*e.g.*, due to limited numerical or data accuracy) the optimal assignment is not reliable, *cf*. our strict conditions.

We can also propose how our method can be applied to large-scale problems on sparse graphs, where solving full-size (L1) is numerically intractable. We can solve constrained variants of MAX-WI/MAX-SI, where the mapping is chosen only inside a window $\mathcal{W} \subset \mathcal{V}$. This leads to linear programs of a smaller size and allows to test the method on vision problems (details in [22]).

Our approach is readily generalizable to higher order energies. It would be sufficient to augment the embedding δ with more components in order to obtain a tighter relaxation and a tighter partial optimality condition.

References

[1] IBM ILOG CPLEX Optimization Studio V12.5.

[2] W. P. Adams, J. B. Lassiter, and H. D. Sherali. Persistency in 0-1 polynomial programming. *Mathematics of Operations Research*, 23(2):359–389, Feb. 1998.

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

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

[6] A. Fix, A. Gruber, E. Boros, and R. Zabih. A graph cut algorithm for higher-order Markov random fields. In *ICCV*, pages 1020 –1027, 2011.

[7] R. F. Goldstein. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophysical Journal*, 66(5):1335–1340, May 1994.

[8] I. Gridchyn and V. Kolmogorov. Potts model, parametric maxflow and k-submodular functions. In *ICCV*, 2013.

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

[10] H. Ishikawa. Transformation of general binary MRF minimization to the first-order case. *PAMI*, 33(6):1234–1249, 2011.

[11] F. Kahl and P. Strandmark. Generalized roof duality. *Discrete Applied Mathematics*, 160(16-17):2419–2434, 2012.

[12] P. Kohli, A. Shekhovtsov, C. Rother, V. Kolmogorov, and P. Torr. On partial optimality in multi-label MRFs. In *ICML*, pages 480–487, 2008.

[13] V. Kolmogorov. Generalized roof duality and bisubmodular functions. In *NIPS*, pages 1144–1152, 2010.

[14] V. Kolmogorov. Generalized roof duality and bisubmodular functions. *Discrete Applied Mathematics*, 160(4-5):416–426, 2012.

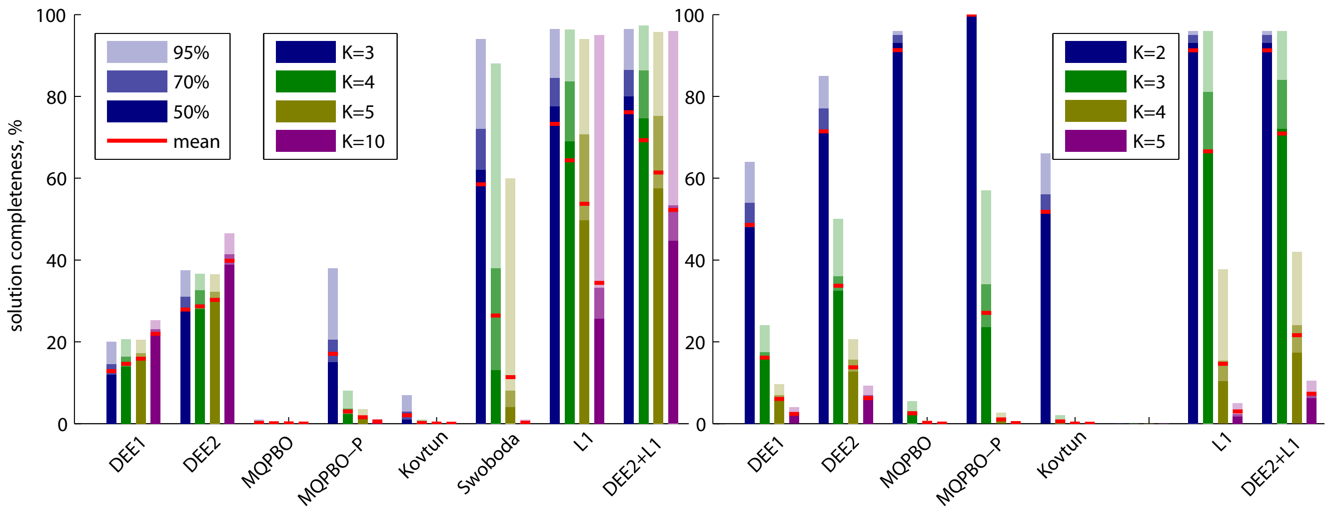

Figure 3. Solution completeness by different methods on random instances of size 10x10 pixels, 4-connected. Bars of different shades indicate the portion of the sample under the given solution completeness value (statistics over 100 instances). Left: Potts model, right: full model.

DEE1	Goldstein's Simple DEE [7]: If $f_s(\alpha) - f_s(\beta) + \sum_{t\in\mathcal{N}(s)} \min_{x_t}[f_{st}(\alpha, x_t) - f_{st}(\beta, x_t)] \geqslant 0$ eliminate α. Iterate until no elimination possible.
DEE2	Similar to DEE1, but including also the pairwise condition: $f_s(\alpha_s) - f_s(\beta_s) + f_t(\alpha_t) - f_t(\beta_t) + f_{st}(\alpha_{st}) - f_{st}(\beta_{st}) + \sum\limits_{t'\in\mathcal{N}(s)\backslash\{t\}} \min_{x_{t'}}[f_{st'}(\alpha_s, x_{t'}) - f_{st'}(\beta_s, x_{t'})] + \sum\limits_{t'\in\mathcal{N}(t)\backslash\{s\}} \min_{x_{t'}}[f_{tt'}(\alpha_t, x_{t'}) - f_{tt'}(\beta_t, x_{t'})] \geqslant 0$.
MQPBO(-P)	The method of Kohli *et al*. [12]. The problem reduced to $\{0, 1\}$ variables is solved by QPBO(-P) [21], where "-P" is the variant with probing [4]. In the options for probing we chose: use weak persistencies, allow all possible directed constraints and dilation=1.
Kovtun	One-against-all Kovtun's method [17]. We run a single pass over $\alpha = 1, \dots K$ (test labelings are $(y_s = \alpha \mid s \in \mathcal{V})$). Labels eliminated in earlier steps are taken correctly into account in the subsequent steps.
Swoboda	Iterative Pruning method of Swoboda *et al*. [24] using CPLEX [1] for each iteration. This version is applicable only to Potts model.
L1	The proposed method solving (L1) with CPLEX. The test labeling y is selected from the necessary conditions.
DEE2+L1	Sequential application of DEE2 and L1. Note, DEE2 uses condition on pairs which is not covered by the proposed sufficient condition under standard relaxation polytope Λ.

Table 2. List of tested methods.

[15] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts – a review. *PAMI*, 29(7):1274–1279, 2007.

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

[17] I. Kovtun. Sufficient condition for partial optimality for (max, +) labeling problems and its usage. *Control Systems and Computers*, (2), 2011. Special issue.

[18] S. H. Lu and A. C. Williams. Roof duality for polynomial 0-1 optimization. *Mathematical Programming*, 37(3):357–360, May 1987.

[19] G. L. Nemhauser and L. E. Trotter, Jr. Vertex packings: Structural properties and algorithms. *Mathematical Programming*, 8:232–248, 1975.

[20] J.-C. Picard and M. Queyranne. On the integer-valued variables in the linear vertex packing problem. *Mathematical Programming*, 12(1):97–101, 1977.

[21] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.

[22] A. Shekhovtsov. Maximum Persistency in Energy Minimization. Technical report, Graz University of Technology, 2014.

[23] M. Shlezinger. Syntactic analysis of two-dimensional visual signals in the presence of noise. *Cybernetics and Systems Analysis*, 4:113–130, 1976. See review [26].

[24] P. Swoboda, B. Savchynskyy, J. Kappes, and C. Schnörr. Partial optimality via iterative pruning for the Potts model. In *SSVM*, 2013.

[25] P. Swoboda, B. Savchynskyy, J. H. Kappes, and C. Schnörr. Partial optimality by pruning for MAP-inference with general graphical models. In *CVPR*, page 8, 2014.

[26] T. Werner. A linear programming approach to max-sum problem: A review. *PAMI*, 29(7):1165–1179, July 2007.