Maximum Persistency in Energy Minimization

Research Report

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.

Contents

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*

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

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 [22] 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.* [10]. 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 [15, 16] 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 [12] proposed a polynomial time algorithm to find the tightest submodular relaxation. Lu and Williams [21], Ishikawa [11] and Fix *et al.* [6] obtained partial optimalities via different reductions to quadratic problems and subsequent application of the roof dual.

Multi-label energies The following methods were proposed for the pairwise model (1) with multi-label variables. Kohli *et al.* [14] 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 [18, 19] 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 [9]. Swoboda *et al.* [30] proposed a method for Potts model solving a series of LP relaxations approximately and generalized it recently to general and higher-order energies [31]. Unlike other approaches, methods [5, 18] 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, 19, 10, 14, 30] 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 [19] (including possible free choices in this method) and subsumes the method [30]. 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 §7.3.

In our previous work [27] a particular map $x \mapsto (x \vee y) \wedge z$ was extended to relaxed labelings, where $\vee$ and $\wedge$ are component-wise maximum and minimum, respectively. It allowed to relate Kovtun's methods to the standard LP-relaxation and the expansion move algorithm. In the previous work [26] a major part of the generalized approach was presented but with algorithms for a much more narrow class of mappings and without experiments.

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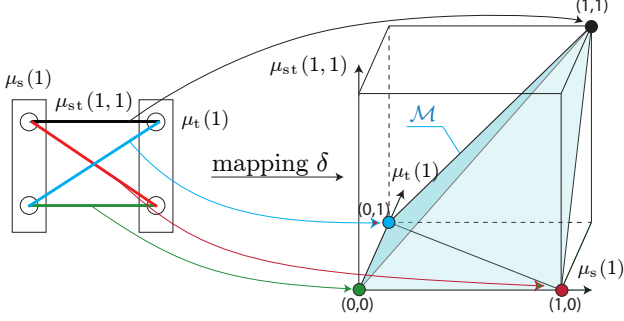
Figure 1. Mapping δ embeds discrete labelings as points in the space $\mathbb{R}^{\mathcal{I}}$. Projection onto components $\mu_s(1), \mu_t(1), \mu_{st}(1,1)$ is shown, the other components are dependent.

denote the scalar product in $\mathbb{R}^{\mathcal{I}}$. We can write the energy as

$$E_f(x) = \langle f, \delta(x) \rangle. \qquad (2)$$

The energy minimization can be expressed and relaxed as

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

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 [29], where the polytope Λ is given by the primal constraints in the following primal-dual pair:

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

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

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

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.*, [32] 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

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

Because a feasible dual solution satisfies $(\forall i')\, f_s^\varphi(i') \geqslant 0$, 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

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

and *strictly improving* if

$$(p(x) \neq x) \;\Rightarrow\; E_f(p(x)) < E_f(x), \qquad (6)$$

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

$$(\forall i) \quad p_s(i) \neq i \;\Rightarrow\; x_s^* \neq i. \qquad (7)$$

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 $E_f(p(x^*)) < E_f(x^*)$. $\qquad \square$

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.

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

$$(\forall x \in \mathcal{L}) \quad \delta(p(x)) = P\delta(x). \qquad (8)$$
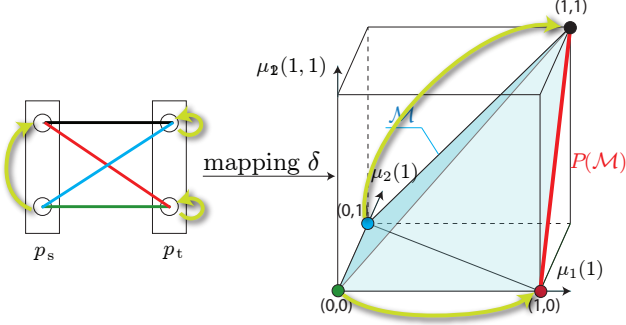
3

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 $\mathrm{aff}(\mathcal{M})$) wit this action – the oblique projection onto the red facet.

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

---

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

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

In order to show that other methods in the literature are special cases of condition (12) we first identify a pixel-wise idempotent mapping p they construct. Then we apply the following *trivial* sufficient condition for $[p] \in \mathbb{W}_f$.

Statement 2. Let $(\forall u \in \mathcal{V})\ (\forall st \in \mathcal{E})$

$$
\begin{aligned}
(\forall i \in \mathcal{L}_u) \quad & f_u(p_u(i)) \leqslant f_u(i), \\
(\forall ij \in \mathcal{L}_{st}) \quad & f_{st}(p_s(i), p_t(j)) \leqslant f_{st}(i,j).
\end{aligned}
\tag{15a}
\tag{15b}
$$

Then $[p] \in \mathbb{W}_f$. If additionally $(p_u(i) \neq i) \Rightarrow f_u(p_u(i)) < f_u(i)$ for all $s \in \mathcal{V}$, $i \in \mathcal{L}_s$, then $[p] \in \mathbb{S}_f$.

Proof. Let $\mu \in \Lambda$. By multiplying (15a) by $\mu_u(i)$ and summing over u and i and multiplying (15b) by $\mu_{st}(i,j)$ and summing over st and ij we get

$$
\langle [p]^{\mathsf{T}} f, \mu \rangle \leqslant \langle f, \mu \rangle,
\tag{16}
$$

which is equivalent to (12). Suppose $[p]\mu \neq \mu$. Then $\exists s \in \mathcal{V}, \exists i \in \mathcal{L}_s$ such that $\mu_s(i) > 0$ and $p_s(i) \neq i$. Therefore there will be at least one strict inequality in the sum (with positive coefficient) and (16) will hold strictly. $\square$

How this component-wise condition can be used to explain global methods? The trick is use it in combination with equivalent transformations. It turns out that this combination is very powerful and in fact characterizes $\mathbb{W}_f$ (see §5). We will consider mainly *weak* variants of all methods. We start with a simple local method.

4.1. DEE

There is a number of local sufficient conditions proposed that are generally referred to as *dead end elimination* (DEE) [5, 8, 20, 7, 24]. We will consider Goldstein's *simple DEE* [8]. This method for every pixel s and labels $\alpha, \beta \in \mathcal{L}_s$ verifies the condition $(\forall x \in \mathcal{L}_{\mathcal{N}(s)})$

$$
f_s(\alpha) - f_s(\beta) + \sum_{t \in \mathcal{N}(s)} [f_{st}(\alpha, x_t) - f_{st}(\beta, x_t)] \geqslant 0.
$$
$$
\tag{17}
$$

If the condition is satisfied it means that a (weakly) improving switch from α to β exists for an arbitrary labelling x. In that case, (s, α) can be eliminated, preserving at least one optimal assignment.

Let $p_s(\alpha) = \beta$, $p_s(i) = i$ for $i \neq \alpha$; and $p_t(i) = i$ for all $t \neq s$. Let $P = [p]$. We claim $P \in \mathbb{W}_f$.

Proof. The condition (17) can be written as

$$\min_{x \in \mathcal{L}}[f(x) - f(p(x))] \geqslant 0. \tag{18}$$

This minimization problem efficiently has a star structure (non-zero unary terms only for s and pairwise terms for neighbors of s). It is equivalent therefore to

$$\min_{\mu \in \Lambda}\langle f, \mu - P\mu \rangle \geqslant 0. \tag{19}$$

$\square$

Similarly, the strict inequality (17) implies $P \in \mathbb{S}_f$.

4.2. QPBO Weak Persistency

The *weak persistency* theorem [22, 10] states the following. Let $\mathcal{L}_s = \{0, 1\} = \mathbb{B}$. Let $\mu \in \operatorname{argmin}_{\mu \in \Lambda}\langle f, \mu \rangle$. Let $O_s = \{i \in \mathbb{B} \mid \mu_s(i) > 0\}$. Then

$$(\exists x \in \operatorname*{argmin}_{x} E_f(x)) \, (\forall s \in \mathcal{V}) \, x_s \in O_s. \tag{20}$$

In the case $|O_s| = 1$ vector μ_s is necessarily integer and the theorem states that for such integer pixels, x_s can be fixed accordingly.

We define $p_s(i) = 0$ if $O_s = \{0\}$, $p_s(i) = 1$ if $O_s = \{1\}$ and $p_s(i) = i$ otherwise. We claim $[p] \in \mathbb{W}_f$.

Proof. Let φ be a solution to LP-dual. We will show that the following component-wise inequalities hold:

$$f_s^\varphi(i) \geqslant f_s^\varphi(p_s(i)), \tag{21a}$$
$$f_{st}^\varphi(i, j) \geqslant f_{st}^\varphi(p_s(i), p_t(j)). \tag{21b}$$

Unary inequalities (21a) hold by construction of p and complementary slackness (4a). Let us show pairwise inequalities (21b). Let $y_{st} = p(x)_{st}$. Consider the following cases:
- $|O_s| = 1$, $|O_t| = 1$. Necessarily, $\mu_s(y_s) = 1$ and $\mu_t(y_t) = 1$. By feasibility, $\mu_{st}(y_{st}) = 1$. By complementary slackness, $f_{st}^\varphi(y_{st}) \leqslant f_{st}^\varphi(x_{st})$.
- $|O_s| = 1$, $|O_t| = 2$. Necessarily, $\mu_s(y_s) = 1$. By feasibility, $\mu_{st}(y_s, i) > 0$ for all $i \in \mathbb{B}$. By complementary slackness, $f_{st}^\varphi(y_s, x_t) \leqslant f_{st}^\varphi(x_s, x_t)$.
- $|O_s| = 2$, $|O_t| = 2$. In this case $y_{st} = x_{st}$ and hence $f_{st}^\varphi(y_{st}) = f_{st}^\varphi(x_{st})$.

Therefore, for every $st \in \mathcal{E}$ we have $f_{st}^\varphi(y_{st}) \leqslant f_{st}^\varphi(x_{st})$. By Statement 2, (21) implies $p \in \mathbb{W}_f$. $\square$

4.3. QPBO Strong Persistency

Let (μ, φ) be a feasible primal-dual pair for (LP). This pair is called *strictly complementary* if

$$\mu_s(i) > 0 \; \Leftrightarrow \; f_s^\varphi(i) = 0, \tag{22a}$$
$$\mu_{st}(i, j) > 0 \; \Leftrightarrow \; f_{st}^\varphi(i, j) = 0 \tag{22b}$$
$$\psi = f_0 + \sum_s \varphi_s. \tag{22c}$$

Clearly, strictly complementary pair is complementary and thus it is optimal. Such pair always exists and can be found by interior point algorithms. It is known that μ is a relative interior point of the primal optimal facet and φ is relative interior point of the dual optimal facet.

The *strong persistency* theorem [22, 10] considers pixels $s \in \mathcal{V}$ such that μ_s' is integer in *all* solutions μ' to the LP-relaxation. It is seen that $\mu_s'(i) > 0$ in some optimal solution μ' iff $\mu_s(i) > 0$ for the relative interior optimal solution μ. Clearly, the solution μ has the minimum number of integer components of all solutions. Let

$$O_s = \{i \in \mathcal{L}_s \mid \mu_s(i) > 0\}. \tag{23}$$

For a strictly complementary pair, (23) defines the same sets as

$$O_s = \operatorname*{argmin}_i f_s^\varphi(i). \tag{24}$$

So we need either primal or dual relative interior optimal point. The theorem can be formulated as follows. Let (μ, φ) be a strictly complementary primal-dual pair. Let O_s be defied by (24). Then

$$(\forall x \in \operatorname*{argmin}_{x} E_f(x)) \, (\forall s \in \mathcal{V}) \, x_s \in O_s. \tag{25}$$

Let us consider the pixel-wise mapping p:

$$p_s(i) = \begin{cases} 0, & O_s = \{0\}, \\ 1, & O_s = \{1\}, \\ i, & O_s = \{0, 1\}. \end{cases} \tag{26}$$

Statement 3. We claim that $[p] \in \mathbb{S}_f$.

Proof. By construction of p and O_s we have $(\forall i)$

$$f_s^\varphi(i) \geqslant f_s^\varphi(p_s(i)). \tag{27}$$

If $p_s(i) \neq i$, then $i \notin O_s$ and inequality (27) is strict. We also have the pairwise inequalities (21b) implied by non-strict complementary slackness as in the weak persistency case. By Statement 2, it follows that $p \in \mathbb{S}_f$. $\square$

We can verify that mapping p is the maximum because any mapping that is larger violates necessary conditions of $\mathbb{S}_f$ to be given in Lemma 1. Therefore it is the solution to MAX-SI.

4.4. MQPBO

MQPBO method [14] extends partial optimality properties of QPBO to multi-label problems via the reduction of the problem to 0-1 variables. The reduction is for a predefined ordering of labels. 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 the method constructs has the form $p(x) = (x \vee x^{\min}) \wedge x^{\max}$. Because the reduction is component-wise and we showed component-wise inequalities (21) for QPBO, it can be shown that component-wise conditions hold for p and therefore $[p] \in \mathbb{W}_f$.

Let f be a multi-label problem and g the equivalent binary (having $\{0, 1\}$-valued decision variables) energy minimization problem as defined in [14]. The mapping of multi-valued to binary labelings is given by $z_{s,i}(x) = [\![x_s > i]\!]$. The corresponding mapping of multi-label relaxed labelings μ to relaxed labelings ν of the binary problem is given [28] as follows. For index $i \in \mathcal{L}_s = \{0, 1, \dots, K-1\}$ introduce the following sets of labels:

$$L_s(i, 0) = \{0, \dots, i\}, \tag{28a}$$

$$L_s(i, 1) = \{i+1, \dots, K-1\}. \tag{28b}$$

The vector $\nu = \Pi\mu$ is defined as

$$\nu_{(s,i)}(\alpha) = \sum_{i' \in L_s(i,\alpha)} \mu_s(i'), \tag{29a}$$

$$\nu_{(s,i)(t,j)}(\alpha, \beta) = \sum_{\substack{i' \in L_s(i,\alpha) \\ j' \in L_t(j,\beta)}} \mu_{st}(i', j'), \tag{29b}$$

where i and j range in $\tilde{\mathcal{L}}_s = \tilde{\mathcal{L}}_t = \{0, 1, \dots, K-2\}$. The mapping Π is consistent with the mapping $z_{s,i}(x) = [\![x_s > i]\!]$ in the sense that $\Pi\delta(x) = \delta(z(x))$ for all x. Using the mapping Π, the equivalence of multi-label and binary problems $\forall x \in \mathcal{L}\ E_f(x) = E_g(z(x))$ is expressed as

$$(\forall \mu \in \Lambda) \quad \langle f, \mu \rangle = \langle g, \Pi\mu \rangle. \tag{30}$$

Let QPBO method for g construct labelings $z_s^{\min}, z_s^{\max}$ such that the mapping $q \colon z \to (z \vee z^{\min}) \wedge z^{\max}$ is strictly (resp. weakly) improving for g. Let

$$x_s^{\min} = \sum_{i' \in \tilde{\mathcal{L}}} z_{s,i'}^{\min}, \qquad x_s^{\max} = \sum_{i' \in \tilde{\mathcal{L}}} z_{s,i'}^{\max}. \tag{31}$$

It was shown [14] that the mapping $p \colon x \to (x \vee x^{\min}) \wedge x^{\max}$ is strictly (resp. weakly) improving for f.

Statement 4. We claim that the linear extension $[p]$ is in $\mathbb{S}_f$ (resp. in $\mathbb{W}_f$).

Proof. Let $\tilde{g}$ be the problem equivalent to g for which component-wise inequalities (15) hold (as proven to exist for QPBO). By equivalence (30), we have

$$(\forall \mu \in \Lambda) \quad \langle f, \mu \rangle = \langle g, \Pi\mu \rangle = \langle \tilde{g}, \Pi\mu \rangle. \tag{32}$$

Let $\tilde{f} = \Pi^{\mathsf{T}} \tilde{g}$. From (32) we conclude that $\tilde{f} \equiv f$. We expand now components of $\tilde{f}$ using (29) and component-wise conditions for $\tilde{g}$:

$$\tilde{f}_s(x_s) = \sum_{i'} \tilde{g}_{(s,i')}(z_{s,i'})$$
$$\geq \sum_{i'} \tilde{g}_{(s,i')}\big((z_{s,i'} \vee z_{s,i'}^{\min}) \wedge z_{s,i'}^{\max}\big) \tag{33}$$
$$= \tilde{f}_s((x_s \vee x^{\min})_s \wedge x_s^{\max}).$$

Similarly, for pairwise terms: $\tilde{f}_{st}(x_{st}) =$

$$\sum_{i'j'} \tilde{g}_{(s,i')(t,j')}(z_{(s,i')(t,j')}) \tag{34}$$
$$\geq \sum_{i',j'} \tilde{g}_{(s,i')(t,j')}\Big(\big(z_{(s,i')(t,j')} \vee z_{(s,i')(t,j')}^{\min}\big) \wedge z_{(s,i')(t,j')}^{\max}\Big)$$
$$= \tilde{f}_{st}((x_{st} \vee x^{\min})_{st} \wedge x_{st}^{\max}).$$

Therefore, component-wise inequalities hold for $\tilde{f}$. By the sufficient condition (15), we conclude that

$$(\forall \mu \in \Lambda) \quad \langle \tilde{f}, [p]\mu \rangle \leq \langle \tilde{f}, \mu \rangle, \tag{35}$$

therefore $[p]$ is in $\mathbb{W}_f$. In the case that q is strictly improving, from $p(x) \neq x$ follows $q(z) \neq z$ and one of the inequalities (33) holds strictly. In this case $[p] \in \mathbb{S}_f$. $\quad\square$

4.5. Auxiliary Submodular Problems

There were several methods proposed [18, 19] that differ in detail. All methods construct an *auxiliary* submodular (in a given ordering of labels) energy E_g. A minimizer y of E_g has the property that $E_g(x \vee y) \leq E_g(x)$, implied by submodularity. It follows that mapping $p(x) = x \vee y$ is improving for g. The construction of the auxiliary function (to be specified) ensures that improvement in f is at least as big as improvement in g for the full family of mappings $x \mapsto x \vee y$, assuming y is not known. It follows that p is improving for f and thus provides partial optimality.

Let $P = [p]$. We claim $P \in \mathbb{W}_f$.

Proof. First, we show that the auxiliary property of g implies

$$(\forall \mu \in \Lambda)\ \langle f, P\mu - \mu \rangle \leq \langle g, P\mu - \mu \rangle. \tag{36}$$

The auxiliary function E_g in [19] satisfies the following component-wise inequalities:

$$(\forall s \in \mathcal{V})\ (\forall i \in \mathcal{L}_s)\ (\forall i' \in \mathcal{K}_s)$$
$$(f - g)_s(i \vee i') \leq (f - g)_s(i), \tag{37a}$$

$$(\forall st \in \mathcal{E}), (\forall ij \in \mathcal{L}_{st})\ (\forall i'j' \in \mathcal{K}_s \times \mathcal{K}_t)$$
$$(f - g)_{st}(i \vee i', j \vee j') \leq (f - g)_{st}(i, j), \tag{37b}$$

where $\mathcal{K}_s \subset \mathcal{L}_s$ depend on a particular method. All methods ensure that $y_s \in \mathcal{K}_s$. Let $\mu \in \Lambda$. By multiplying inequalities (37a) for $i' = y_s$ with $\mu_s(i)$, inequalities (37b) for $i'j' = y_{st}$ with $\mu_{st}(ij)$ and adding we obtain (36).

Second, we show that $P \in \mathbb{W}_g$. Recall that E_g is submodular and y is a minimizer. LP-relaxation for g is tight, therefore there exists dual φ such that $(\delta(y), \varphi)$ satisfies complementary slackness (4). Let $\tilde{g} = g^\varphi$. By (4a), $\tilde{g}_s(x_s \vee y_s) \leqslant \tilde{g}_s(x_s)$. For the pairwise components we inspect the four cases in order to prove $\tilde{g}_{st}(x_{st} \vee y_{st}) \leqslant \tilde{g}_{st}(x_{st})$:

- $x_{st} \geqslant y_{st}$: in this case $x_{st} \vee y_{st} = x_{st}$.
- $x_{st} < y_{st}$: in this case $x_{st} \vee y_{st} = y_{st}$, which is minimal.
- $x_s < y_s, x_t \geqslant y_t$: in this case $x_{st} \vee y_{st} = (y_s, x_t)$. The submodularity inequality $\tilde{g}_{st}(x_{st}) + \tilde{g}_{st}(y_{st}) \geqslant \tilde{g}_{st}(x_s, y_t) + \tilde{g}_{st}(y_s, x_t)$ and minimality of y_{st} imply $\tilde{g}_{st}(x_{st}) \geqslant \tilde{g}_{st}(y_s, x_t)$.
- $x_s \geqslant y_s, x_t < y_t$: similar to the above.

By Statement 2, $p \in \mathbb{W}_{\tilde{g}} = \mathbb{W}_g$. From (36) follows $p \in \mathbb{W}_f$. $\qquad\square$

The *one-against-all-binary* method [19] restricts y_s to $\{0, \bar{y}_s\}$ for some fixed labeling $\bar{y}$ (*e.g.* $\bar{y}_s = \alpha$ for all s), let us call it the *test labeling*. The labels are then reordered such that $\bar{y}_s$ becomes the highest label and sets K_s are chosen to be $\{0, K-1\}$. Furthermore, g is additionally constrained to be equivalent to a problem with two labels. In this case the result of the method depends only on the choice of $\bar{y}$ and not on the actual ordering.

4.6. Iterative Pruning

Iterative Pruning method [30] was originally proposed for the Potts model: $f_{st}(i, j) = \gamma_{st}[\![i \neq j]\!]$. It constructs a subset $\mathcal{A} \subset \mathcal{V}$, a labeling y on $\mathcal{A}$ and an auxiliary energy E_g:

$$
\begin{aligned}
(\forall s \in \mathcal{A}) & \quad g_s = f_s, & (38) \\
(\forall st \in \mathcal{E},\ s \in \mathcal{A}, t \in \mathcal{A}) & \quad g_{st} = f_{st}, \\
(\forall st \in \mathcal{E},\ s \in \mathcal{A}, t \notin \mathcal{A},\ \forall ij) & \quad g_{st}(i, j) = \gamma_{st}[\![i = y_s]\!],
\end{aligned}
$$

with remaining terms set to zero. It can be seen that energy E_g depends on the assignment of y only on the *boundary* $\partial \mathcal{A} = \{s \in \mathcal{A} \mid \exists st \in \bar{\mathcal{E}}, t \notin \mathcal{A}\}$. Let us extend y to $\mathcal{V}$ in an arbitrary way, *e.g.*, by $y_{\mathcal{V} \backslash \mathcal{A}} = 0$. The sufficient condition of [30] imply that $\delta(y) \in \arg\min_{\mu \in \Lambda}\langle g, \mu\rangle$ (the relaxation is tight). We construct mapping p as

$$
p_s(i) = \begin{cases} y_s & \text{if } s \in \mathcal{A}, \\ i & \text{if } s \notin \mathcal{A}, \end{cases} \tag{39}
$$

i.e., p replaces part of labeling x on $\mathcal{A}$ with the labeling y. Let $P = [p]$. We claim that $P \in \mathbb{W}_f$.

Proof. We first show that g is auxiliary for f in the same sense as for the method [19]. We trivially have $f_s(p_s(i)) - f_s(i) = g_s(p_s(i)) - g_s(i)$. We also have equality of pairwise terms $f_{st}(p(x)_{st}) - f_{st}(x_{st}) = g_{st}(p(x)_{st}) - g_{st}(x_{st})$ for $st \in \mathcal{E}$ in all of the following cases: (a) $s \in \mathcal{A}$ and $t \in \mathcal{A}$; (b)

$s \notin \mathcal{A}$ and $t \notin \mathcal{A}$; (c) $s \in \mathcal{A}$ and $t \notin \mathcal{A}$, $x_s = y_s$. It remains to verify the inequality for boundary pairs $s \in \mathcal{A}$, $t \notin \mathcal{A}$ in the case $x_s \neq y_s$. We have

$$
\begin{aligned}
& f_{st}(x_{st}) - f_{st}(p(x)_{st}) \\
& \geqslant \min_{ij \in \mathcal{L}_{st}} \big(f_{st}(i, j) - f_{st}(p_s(i), p_t(j))\big) = -\gamma \quad (40) \\
& = g_{st}(x_{st}) - g_{st}(p(x)_{st}).
\end{aligned}
$$

It follows that (36) holds. The second step is to show that $P \in \mathbb{W}_g$. By assumption, we have $\delta(y) \in \arg\min_{\mu \in \Lambda}\langle g, \mu\rangle$. Given a labeling x, mapping p replaces part over $\mathcal{A}$ to the optimal labeling y. It follows that $(\forall \mu \in \Lambda)$ $\langle g, P\mu\rangle = \langle g, \delta(y)\rangle \leqslant \langle g, \mu\rangle$. Combined with (36), we obtain $P \in \mathbb{W}_f$. $\qquad\square$

5. Characterization

We introduced component-wise sufficient conditions (15) and observed while considering different methods that it was often possible to find a reparametrization of the problem such that these conditions hold. This is not a coincidence.

Theorem 1. Let $P = [p]$, p idempotent and $P \in \mathbb{W}_f$. Then exists φ such that

$$
P^\mathsf{T} f^\varphi \leqslant f^\varphi. \tag{41}
$$

Proof. Let $g = (I - P^\mathsf{T})f$. The steps of the proof are given by the following chain:

$$
\begin{aligned}
\min_{\substack{A\mu=0 \\ \mu \geqslant 0}} \langle f - P^\mathsf{T}f, \mu\rangle & \overset{(b)}{=} \min_{\substack{AP\mu=0 \\ A(I-P)\mu=0 \\ \mu \geqslant 0}} \langle f - P^\mathsf{T}f, \mu\rangle \\
& \overset{(c)}{=} \min_{\substack{A(I-P)\mu=0 \\ \mu \geqslant 0}} \langle f - P^\mathsf{T}f, \mu\rangle \overset{(d)}{=} \max_{\substack{\varphi \\ (I-P^\mathsf{T})(f-A^\mathsf{T}\varphi)\geqslant 0}} 0 .
\end{aligned}
\tag{42}
$$

On the LHS we have problem (52) which is bounded because $P \in \mathbb{W}_f$. Equality (b) is verified as follows. Inequality $\leqslant$ holds because $A\mu - AP\mu = 0$ and $AP\mu = 0$ implies $A\mu = 0$. On the other hand, P preserves all constraints of Λ and therefore $A\mu = 0 \Rightarrow AP\mu = 0$.

Equality (c) is the key step. We removed one constraint, therefore $\geqslant$ trivially holds. Let us prove $\leqslant$. Let μ be feasible to RHS of equality (c). Let $\mu = \mu_1 + \mu_2$ such that $\mu_1 \in \text{null}(I - P)$ and $\mu_2 \in \text{null}(P)$. Let us construct μ_1' as follows:

$$
\begin{aligned}
\gamma &= \max_{st, ij} |\mathcal{L}_{st}|(\mu_1)_{st}(i, j), \\
(\mu_1')_{st} &= \gamma/|\mathcal{L}_{st}|, \\
(\mu_1')_s &= \gamma/|\mathcal{L}_s|.
\end{aligned}
\tag{43}
$$

By construction,

$$
(\mu_1') \geqslant \mu_1 \qquad \text{and} \qquad A\mu_1' = 0. \tag{44}
$$

Let $\mu_1'' = P\mu_1'$. Because $P \geqslant 0$, we have

$$\mu_1'' = P\mu_1' \geqslant P\mu_1 = \mu_1 \,. \tag{45}$$

It also follows that $AP\mu_1'' = APP\mu_1' = AP\mu_1' = 0$ and $(I - P)\mu_1'' = (I - P)P\mu_1' = 0$. Let $\mu^* = \mu_1'' + \mu_2$. It preserves the objective,

$$\langle f - P^\mathsf{T} f, \mu^* \rangle = \langle f, (I - P)(\mu_1'' + \mu_2) \rangle \tag{46}$$
$$= \langle f, (I - P)\mu_2 \rangle = \langle f, (I - P)\mu \rangle \,.$$

We also have that

$$\mu^* = \mu_1'' + \mu_2 \geqslant \mu_1 + \mu_2 = \mu \geqslant 0 \,,$$
$$A(I - P)\mu^* = A(I - P)\mu_2 = A(I - P)\mu = 0 \,, \tag{47}$$
$$AP\mu^* = AP\mu_1'' = 0 \,.$$

Therefore, μ^* satisfies all constraints of the LHS of equality (c).

Equality (d) is the duality relation that asserts that the maximization problem on the RHS is feasible, which is the case iff

$$(\exists g \equiv f) \quad (I - P^\mathsf{T})g \geqslant 0 \,. \tag{48}$$

$\square$

6. 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:

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

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:

$$p_\xi(x)_s = \begin{cases} y_s & \text{if } \xi_{sx_s} = 1, \\ x_s & \text{otherwise.} \end{cases} \tag{49}$$

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.

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 §7.

- *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!

7. Algorithms

Case $K = 2$ 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], [17, §2.3]. Finding the relaxed solution with the maximum number of integer components was proven polynomial by Picard and Queyranne [23] in the context of vertex packing problem. We extend this proof to general quadratic pseudo-Boolean functions.

Statement 5. Let μ^1, μ^2 be two solutions to (LP). Let us denote sets where these solutions are integral as $U = \{s \in \mathcal{V} \mid \mu_s^1(i) \in \mathbb{B}\}$ and $V = \{s \in \mathcal{V} \mid \mu_s^2(i) \in \mathbb{B}\}$. Let x^1 and x^2 be corresponding partial labelings. Then there exists a solution μ such that its integral part $\mathcal{A} = \{s \in \mathcal{V} \mid \mu_s(i) \in \mathbb{B}\}$ is the union $U \cup V$.

Proof. We construct μ as follows

$$\mu_s = \begin{cases} \mu_s^1, & s \in U, \\ \mu_s^2, & s \notin U; \end{cases} \quad \mu_{st} = \begin{cases} \mu_{st}^1, & s \in U, \ t \in U, \\ \mu_{st}^2, & s \notin U, \ t \notin U, \\ \mu_s^1(\mu_t^2)^\mathsf{T}, & s \in U, \ t \notin U, \\ \mu_s^2(\mu_t^1)^\mathsf{T}, & s \notin U, \ t \in U. \end{cases} \tag{50}$$

First, we check that μ is feasible. We use feasibility of μ^1, μ^2 and verify that $\mathbf{1}^\mathsf{T}\mu_s^1(\mu_s^2)^\mathsf{T} = (\mu^2)^\mathsf{T}$.

Let us now show that μ is optimal. Let φ be relative interior dual solution. By complementarity slackness with

μ^1 and μ^2 it must be that $f_s^\varphi(i) = 0$ whenever $\mu_s^1(i) > 0$ or $\mu_s^2(i) > 0$ and the same holds for pairwise terms. We need to care only about the stitching, the pairwise terms in the case $s \in \mathcal{U}$, $t \notin \mathcal{U}$. Let $O_s^1 = \{i \mid \mu_s^1(i) > 0\}$. Since $|O_s^1| = 1$ and $|O_t^1| = 2$, by feasibility of μ we have that $\mu_{st}^1(x_s^1, 0) > 0$ and $\mu_{st}^1(x_s^1, 1) > 0$. By complementarity, $f_{st}^\varphi(x_s^1, 0) = f_{st}^\varphi(x_s^1, 1) = 0$. By construction, $\mu_{st}(1 - x_s^1, \cdot) = 0$ and we have that for any μ_t^2 the product $\mu_s^1(\mu_t^2)^\mathsf{T}$ satisfies complementarity with f^φ. The remaining case $s \notin \mathcal{U}$, $t \in \mathcal{U}$ is symmetric. Therefore μ is optimal. $\qquad\square$

It follows that the maximum can be found in polynomial time by trying to fix a variable and check whether there is a feasible solution with such fixation. This is trivial but inefficient. It can be done efficiently by analyzing connected components in the network flow model [17, §2.3].

Case $K \geqslant 3$ 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 $K \geqslant 3$.

7.1. General Properties

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

$$\{P \colon \mathbb{R}^\mathcal{I} \to \mathbb{R}^\mathcal{I} \mid (\exists \varphi \in \mathbb{R}^m) \ f^\varphi - P^\mathsf{T} f \geqslant 0\}. \quad (51)$$

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

$$\inf_{\mu \in \operatorname{coni}(\Lambda)} \langle g, \mu \rangle \geqslant 0. \quad (52)$$

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 (52) and its dual as

$$\begin{array}{ll} \inf\langle g, \mu \rangle & \max 0\,. \\ A\mu = 0 & \varphi \in \mathbb{R}^m \\ \mu \geqslant 0 & g - A^\mathsf{T}\varphi \geqslant 0 \end{array} \quad (53)$$

Inequality (52) holds iff the primal problem is bounded, and it is bounded iff the dual is feasible, which is the case iff $(\exists \varphi \in \mathbb{R}^m) \ (f - A^\mathsf{T}\varphi) - P^\mathsf{T} f \geqslant 0$. $\qquad\square$

With this reformulation we can write MAX-WI as

$$\max_{p,\varphi} \sum_{s,i} [\![p_s(i) \neq i]\!] \quad \text{s.t.:} \ (I - [p]^\mathsf{T})f - A^\mathsf{T}\varphi \geqslant 0. \quad (54)$$

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

$$(\forall s, \forall i) \quad f_s^\varphi(i) - f_s(p_s(i)) \geqslant \varepsilon[\![p_s(i) \neq i]\!], \quad (55\text{a})$$
$$(\forall st, \forall ij) \quad f_{st}^\varphi(i,j) - f_{st}(p_s(i), p_t(j)) \geqslant 0. \quad (55\text{b})$$

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

$$(\forall \mu \in \Lambda) \quad \langle (I - [p]^\mathsf{T})f, \mu \rangle \geqslant \varepsilon\langle h, \mu \rangle \quad (56)$$

for some $\varepsilon > 0$. We apply now the same inference as in Statement 6 for vector $g = f - P^\mathsf{T} f - \varepsilon h$. It follows that (56) is equivalent to $(\exists \varphi \in \mathbb{R}^m)\,(f - A^\mathsf{T}\varphi) - P^\mathsf{T} f - \varepsilon h \geqslant 0$. $\qquad\square$

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 $P \notin \mathbb{S}_f$. $\qquad\square$

7.2. Maximum Persistency by LP

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

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

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):

$$(P_\xi \mu)_s(i) = \sum_{i'} P_{s,ii'} \mu_s(i'), \tag{57a}$$

$$(P_\xi \mu)_{st}(i,j) = \sum_{i',j'} P_{st,ii',jj'} \mu_{st}(i',j'), \tag{57b}$$

$$P_{s,ii'} = [\![p_s(i')\!=\!i]\!] = \tag{58a}$$
$$[\![y_s\!=\!i]\!]\xi_{si'} + [\![i'\!=\!i]\!](1-\xi_{si'}),$$

$$P_{st,ii',jj'} = [\![y_s\!=\!i]\!][\![y_t\!=\!j]\!]\xi_{sti'j'} \tag{58b}$$
$$+ [\![i'\!=\!i]\!][\![y_t\!=\!j]\!](\xi_{tj'} - \xi_{sti'j'})$$
$$+ [\![y_s\!=\!i]\!][\![j'\!=\!j]\!](\xi_{si'} - \xi_{sti'j'})$$
$$+ [\![i'\!=\!i]\!][\![j'\!=\!j]\!](1-\xi_{si'}-\xi_{tj'}+\xi_{sti'j'}).$$

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:

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

By relaxing the integrality constraints we obtain the linear program

$$\max_{\xi,\varphi} \sum_{s,i} \xi_{si} \tag{L1}$$
$$(I - P_\xi^\mathsf{T})f - A^\mathsf{T}\varphi \geqslant 0$$
$$\xi \in \Sigma; \ \xi_{si} \in [0,1]; \xi_{sy_s} = 0.$$

We will prove in Theorem 2 that this relaxation is tight and then the program will be simplified by expanding the constraints and optimizing out variables ξ_{stij}. 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 (58a), (58b) are nonnegative, therefore $\mu' \geqslant 0$. Constraints $\mathbf{1}^\mathsf{T}\mu'_s = 1$ hold due to $\mathbf{1}^\mathsf{T}P_s = 1$. Constraints $\mathbf{1}^\mathsf{T}\mu'_{st} = (\mu'_t)^\mathsf{T}$ hold due to $\sum_{ii'} P_{st,ii',jj'} = P_{t,jj'}$. $\square$

Theorem 2. 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

$$(\forall \mu \in \Lambda) \quad \langle f, P_\xi P_\xi \mu \rangle \leqslant \langle f, P_\xi \mu \rangle \leqslant \langle f, \mu \rangle. \tag{61}$$

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':

$$\xi'_{si} = 1 - (1-\xi_{si})^2, \tag{62}$$
$$\xi'_{stij} = (1-\xi_{si}-\xi_{tj}+\xi_{stij})^2 - 1 + \xi'_{si} + \xi'_{tj}.$$

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

$$\xi^*_{si} = \lim_{n\to\infty} 1 - (1-\xi_{si})^{2^n} = [\![\xi_{si} > 0]\!]. \tag{63}$$

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 (63) 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. $\square$

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 (55b) are linearized exactly the same way as for the weak case, we can write them shortly as

$$((I - P_\xi^\mathsf{T})f - A^\mathsf{T}\varphi)_{st}(i,j) \geqslant 0. \tag{64}$$

The inequalities for univariate terms (55a), by substituting p_ξ can be expressed as

$$(f_s(i) - f_s(y_s))\xi_{si} - (A^\mathsf{T}\varphi)_s(i) \geqslant \varepsilon \xi_{s,i}[\![i \neq y_s]\!]. \tag{65}$$

Since we assume $\xi_{sy_s} = 0$, expression (65) is equivalent to

$$(f_s(i) - f_s(y_s) - \varepsilon)\xi_{si} - (A^\mathsf{T}\varphi)_s(i) \geqslant 0, \tag{66}$$

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 $\alpha = 0, \ldots, K-1$. This algorithm subsumes simple Goldstein's DEE [8] 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.

Reduced Linear Program We now detail the program (L1) in components and simplify it for the practical implementation. We will assume without loss of generality that $0 = f_s(y_s) = f_{st}(y_s, y_t) = f_{st}(i, y_t) = f_{st}(y_s, j)$. If

10

the problem $\hat{f}$ does not satisfy these constraints, we chose the *equivalent* problem f by letting

$$f_{st}(i,j) = \hat{f}_{st}(i,j) - \hat{f}_{st}(i,y_t) - \hat{f}_{st}(y_s,j) + \hat{f}_{st}(y_s,y_t),$$

$$f_s(i) = \hat{f}_s(i) - \hat{f}_s(y_s) + \sum_{t \in \mathcal{N}(s)} \left[\hat{f}_{st}(i,y_t) - \hat{f}_{st}(y_s,y_t) \right],$$

$$f_0 = \hat{f}_0 + \sum_{st \in \mathcal{E}} \hat{f}_{st}(y_s,y_t) + \sum_{s \in \mathcal{V}} \hat{f}_s(y_s). \tag{67}$$

It can be verified by substitution that $E_f = E_{\hat{f}}$ and that $(\forall \mu \in \Lambda) \langle \hat{f}, \mu \rangle = \langle f, \mu \rangle$. By construction, the optimal ξ for problem (MAX-WI) does not depend on this transformation. We have

$$\begin{aligned} (f - P^\mathsf{T} f)_s(i) &= f_s(i)\xi_{si}, \\ (f - P^\mathsf{T} f)_{st}(i,j) &= f_{st}(i,j)\big(\xi_{si} + \xi_{tj} - \xi_{stij}\big). \end{aligned} \tag{68}$$

With these expansions made, the problem (L1) expresses as

$$\max_{\xi,\varphi} \sum_{s,i} \xi_{si}$$

$$(\forall s,i) \ f_s(i)\xi_{s,i} + \sum_{t \in \mathcal{N}(s)} \varphi_{st}(i) - \varphi_s \geqslant 0, \tag{69a}$$

$$(\forall st, \ \forall ij)$$
$$f_{st}(i,j)(\xi_{si} + \xi_{tj} - \xi_{stij}) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0, \tag{69b}$$

$$\sum_s \varphi_s \geqslant 0, \tag{69c}$$

$$\xi \in \Sigma; \ \xi_{si} \in [0,1]; \xi_{sy_s} = 0. \tag{69d}$$

We next optimize out variables ξ_{stij}. For each ij variable ξ_{stij} is present only in the constraint (69b) and the constraint of the feasible set Σ, $\max(0, \xi_{si} + \xi_{tj} - 1) \leqslant \xi_{stij} \leqslant \min(\xi_{si}, \xi_{tj})$. Depending on whether $f_{st}(i,j)$ is positive or negative the optimal value for ξ_{stij}, which allows the maximum freedom for (69b) is either its lower or upper bound, respectively. Let $\mathcal{L}_{st}^+ = \{ij \mid f_{st}(i,j) > 0\}$ and $\mathcal{L}_{st}^- = \{ij \mid f_{st}(i,j) \leqslant 0\}$. Substituting the respective bounds into (69b) and using identities

$$\xi_{si} + \xi_{tj} - \max(0, \xi_{si} + \xi_{tj} - 1) = \min(\xi_{si} + \xi_{tj}, 1),$$
$$\xi_{si} + \xi_{tj} - \min(\xi_{si}, \xi_{tj}) = \max(\xi_{si}, \xi_{tj}) \tag{70}$$

we can rewrite constraints (69b) as

$$(\forall st, \ \forall ij \in \mathcal{L}_{st}^-)$$
$$f_{st}(i,j) \max(\xi_{si}, \xi_{tj}) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0, \tag{71}$$

$$(\forall st, \ \forall ij \in \mathcal{L}_{st}^+)$$
$$f_{st}(i,j) \min(\xi_{si} + \xi_{tj}, 1) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0. \tag{72}$$

Finally, by expressing $\min$ and $\max$ as two linear constraints each, we obtain the following representation of the problem (L1):

$$\max_{\xi,\varphi} \sum_{s,i} \xi_{si}$$

$$(\forall s,i) \ f_s(i)\xi_{s,i} + \sum_{t \in \mathcal{N}(s)} \varphi_{st}(i) - \varphi_s \geqslant 0, \tag{73a}$$

$$(\forall st, \ \forall ij \in \mathcal{L}_{st}^-)$$
$$f_{st}(i,j)\xi_{si} - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0, \tag{73b}$$
$$f_{st}(i,j)\xi_{tj} - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0; \tag{73c}$$

$$(\forall st, \ \forall ij \in \mathcal{L}_{st}^+)$$
$$f_{st}(i,j) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0, \tag{73d}$$
$$f_{st}(i,j)(\xi_{si} + \xi_{tj}) - \varphi_{st}(i) - \varphi_{ts}(j) \geqslant 0; \tag{73e}$$

$$\sum_s \varphi_s \geqslant 0,$$

$$\xi_{si} \in [0,1]; \xi_{sy_s} = 0.$$

In this form, only variables ξ_{si} remained. On the other hand, the number of constraints has doubled.

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. Necessary conditions in this case either provide the unique

---

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

1   $\mu \in \operatorname{argmin}_{\mu \in \Lambda} \langle f, \mu \rangle$;   /* solve (LP) */
2   For all s if exists $i \in \mathcal{L}_s$ such that $\mu_s(i) = 1$ then set $y_s = i$, otherwise set y_s arbitrarily;
3   Solve the problem (ε-L1) with y;

---

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 $\bar{y}$ and the iterative pruning method [30].

7.3. Windowing

In this section we would like to address large-scale problems, where solving (L1) for the full problem may be numerically intractable. We restrict consideration to a local *window* $\mathcal{W} \subset \mathcal{V}$, fix $p_s(i) = i$ for all $s \notin \mathcal{W}$ and solve for the part of p inside the window. This leads to a reduced problem (L1) with variables ξ_{si} and $\varphi_{st}(i)$ only inside the window. But how do we pick a good labeling y for (L1), without solving the full (LP)? We propose the following "local" necessary conditions.

Theorem 3. Let $P, Q\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$, $QP = Q$ and $P(\Lambda) \subset \Lambda$. Let

$$\mathcal{O} = \arg\min_{\mu \in \Lambda}\langle f, (I - Q)\mu\rangle \qquad (74)$$

Then

(a) $P \in \mathbb{W}_f \Rightarrow P(\mathcal{O}) \subseteq \mathcal{O}$.

(b) $P \in \mathbb{S}_f \Rightarrow (\forall \mu \in \mathcal{O})P\mu = \mu$.

Proof. **(a)** Assume $P(\mathcal{O}) \nsubseteq \mathcal{O}$. Then there exists $\mu \in \mathcal{O}$ such that $P\mu \in \Lambda\backslash\mathcal{O}$. Then $\langle f, (I - Q)\mu\rangle < \langle f, (I - Q)P\mu\rangle$. Equivalently, $0 > \langle f, ((I - Q) - (I - Q)P)\mu\rangle = \langle f, (I - P)\mu\rangle$. Hence $P \notin \mathbb{W}_f$.

(b) Assume for contradiction that $P\mu \neq \mu$. Then $\langle f, P\mu\rangle < \langle f, \mu\rangle$. Equivalently, $\langle f, ((I - Q) - (I - Q)P)\mu\rangle > 0$, which implies $\langle f, (I - Q)\mu\rangle > \langle f, (I - Q)P\mu\rangle$, which contradicts $\mu \in \mathcal{O}$. $\square$

As a corollary, we have the following result. Let $O_s = \{i \in \mathcal{L}_s \mid (\exists\mu \in \mathcal{O})\,\mu_s(i) > 0\}$. A pixel-wise map $p\colon \mathcal{L} \to \mathcal{L}$ is weakly Λ-improving only if $p_s(O_s) = O_s$ and strongly-Λ-improving only if $p_s(i) = i$ for all $i \in O_s$.

Instead of solving full (LP) we solve test problem (74) with $Q = [q]$, $q_s(i) = i$ for $s \notin \mathcal{W}$ and $q_s(i) = 0$ for $s \in \mathcal{W}$. Since for any p in the window there holds $q \circ p = q$, the solution to (74) identifies the subset of "immovable" labels and makes algorithms developed in the previous section applicable.

In order to better understand necessary conditions by Theorem 3 we give the next additional property. For a projection $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{I}$, its null space corresponds to the dimensions (variables) that become fixed. The larger the null space the more powerful the projection is, because the optimization domain reduces from $\mathcal{M}$ to $P(\mathcal{M})$. The next lemma clarifies why property $QP = Q$ was essential in Theorem 3.

Lemma 3. Let P be idempotent. Then $QP = Q$ iff $\mathrm{null}(P) \subseteq \mathrm{null}(Q)$.

Proof. Let $QP = Q$. Assume $Px = 0$, then $Qx = QPx = Q0 = 0$ and therefore $x \in \mathrm{null}(Q)$. In the other direction, let $\mathrm{null}(P) \subseteq \mathrm{null}(Q)$. Let x be arbitrary. Since $P^2 = P$, we can represent x with the orthogonal sum $x = x_1 + x_2$ with $x_1 \in \mathrm{null}(P)$ and $x_2 \in \mathrm{null}(I - P)$. We have $Px = Px_1 + Px_2 = Px_2 = x_2$. It follows that $QPx = Qx_2$. Since $\mathrm{null}(P) \subset \mathrm{null}(Q)$ we have $Qx_1 = 0$ and therefore $Qx = Qx_1 + Qx_2 = Qx_2$. It follows that $QPx = Qx_2 = Qx$ and therefore $QP = Q$. $\square$

8. Experiments

8.1. Random Instances

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
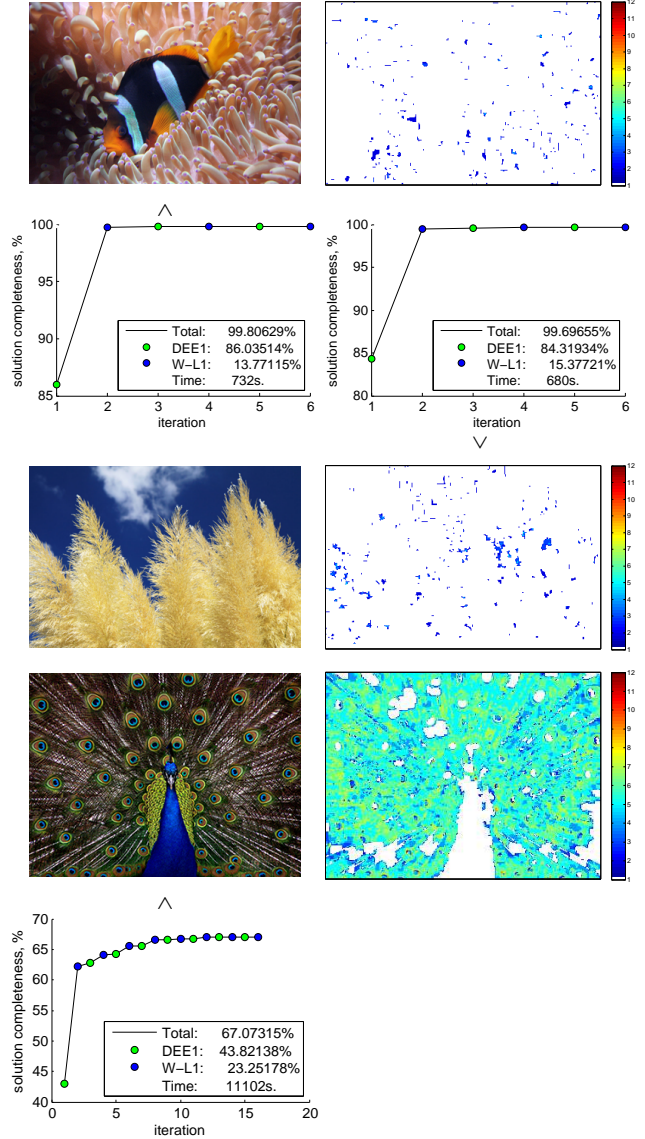

Figure 3. Other instances of `color-seg-n4`. For each instance shown: image, reminder of the problem (number of remaining labels), algorithm progress.

$\gamma_{st}(i) \sim U[0, 50]$. 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 $\frac{n_{\mathrm{elim}}}{|\mathcal{V}|(K-1)}100\%$, where n_{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 4.

Results of all methods that are covered by the proposed sufficient conditions were verified by solving the verification LP (14). For random problems, we also found a global minimum x^* with CPLEX mixed-integer solver (feasible for the size of the problems we used). Methods not verifiable with (14) we checked to satisfy $E_f(p(x^*)) = E_f(x^*)$.
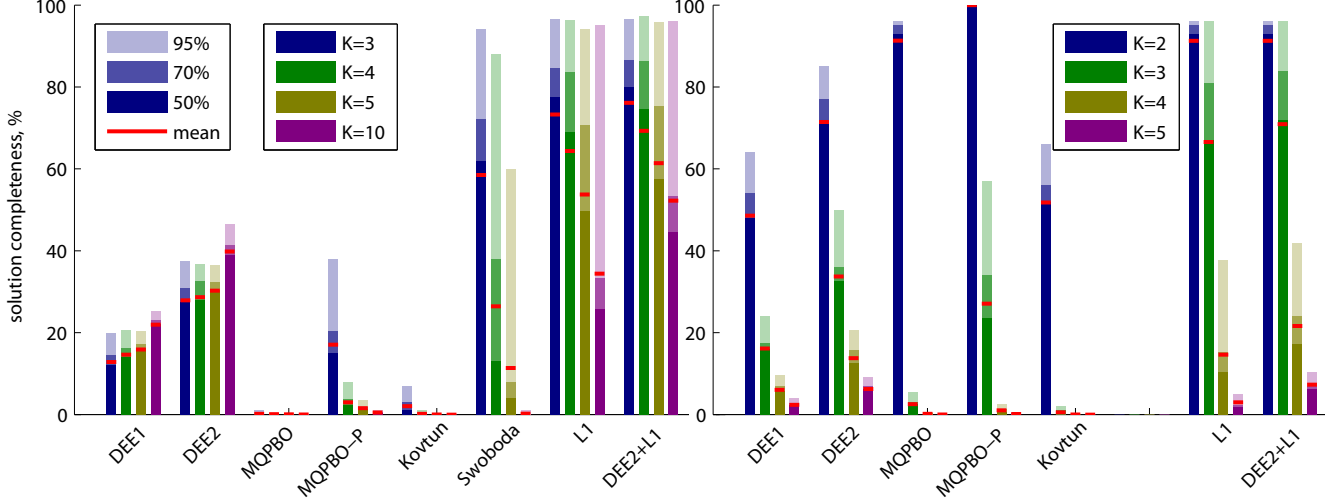
Figure 4. Solution completeness by different methods on random instances of size 10x10 pixels, 4-connected. Bars of different shades indicate the portion of the sample under the given solution completeness value (statistics over 100 instances). Left: Potts model, right: full model.

DEE1	Goldstein's Simple DEE [8]: 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: $f_s(\alpha_s) - f_s(\beta_s) + f_t(\alpha_t) - f_t(\beta_t) + f_{st}(\alpha_{st}) - f_{st}(\beta_{st}) + \sum_{t' \in \mathcal{N}(s)\backslash\{t\}} \min_{x_{t'}}[f_{st'}(\alpha_s, x_{t'}) - f_{st'}(\beta_s, x_{t'})] + \sum_{t' \in \mathcal{N}(t)\backslash\{s\}} \min_{x_{t'}}[f_{tt'}(\alpha_t, x_{t'}) - f_{tt'}(\beta_t, x_{t'})] \geqslant 0.$
MQPBO(-P)	The method of Kohli *et al*. [14]. The problem reduced to $\{0, 1\}$ variables is solved by QPBO(-P) [25], 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 [19]. We run a single pass over $\alpha = 1, \ldots K$ (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*. [30] 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.

In the case of Potts model, we see that performance of Swoboda *et al*. [30] drops quickly with the increase of the number of labels and ours decreases moderately. While the problem difficulty increases, the performance of DEE methods appears to benefit from more labels, which can be explained by the random nature of the problems. Increasing connectivity makes the problem more difficult for all methods, see Figure 5.

8.2. Large-Scale Segmentation

We propose experiments with multiclass image segmentation. We used `color-seg-n4` instances from [13], which have 4-12 labels and Potts pairwise interactions. Solving LP-relaxation for the whole problem is numerically intractable. We apply the technique described in §7.3. We maintain a global pixel-wise mapping $p\mathcal{L} \rightarrow \mathcal{L}$, which de-

fines the current problem reduction. At each step we select a window $\mathcal{W} \subset \mathcal{V}$ such that the problem (L1) over the window has no more that 10^4 variables or constraints (under the current reduction of label sets, $p_s(L_s)$). We find an improving mapping p' from the window subproblem and calculate the composition $p \circ p'$. We can process several overlapping windows in parallel, taking a composition of the mappings in the end. The result might depend on the order of composition, but any order corresponds to a correct weak partial optimality. An example of windows selected for processing instance `crops-small` are shown in Figure 6; Before each scan with local windows we perform simple DEE step, this step makes a big initial reduction for some of these problems, and our method works on the more difficult reminder. On some other problems DEE step is of almost of no help, (`fourcolors`, fourth in Figure 7). By
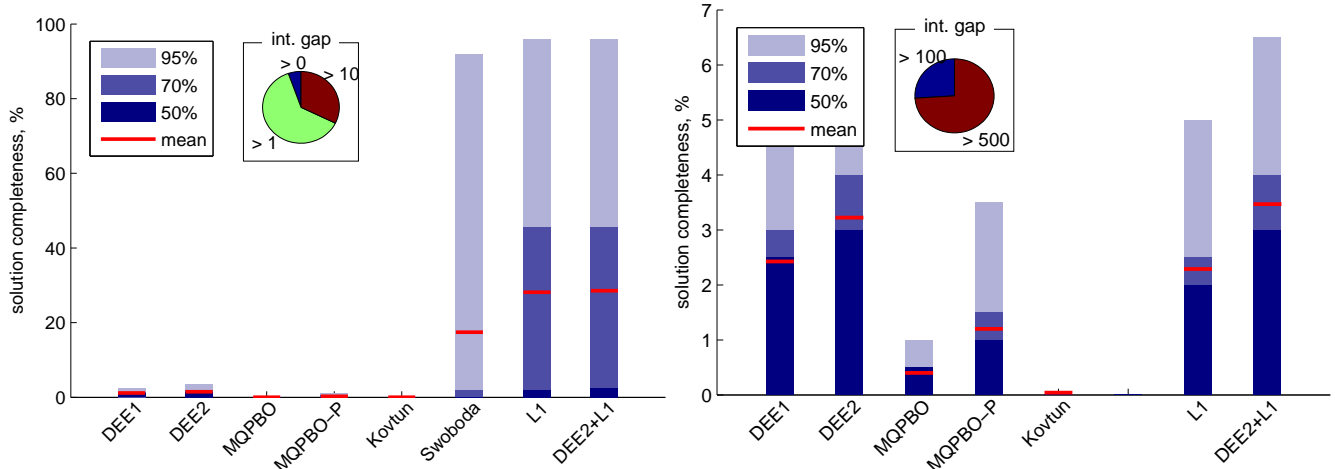
Figure 5. Results with 8-connected 10x10 grid, 3 labels. Left: Potts model, right: full model. The pie chart shows distribution of the integrality gap over 100 samples, which indicates problem difficulty. The "full" problem is much more difficult (note the different axis scale).
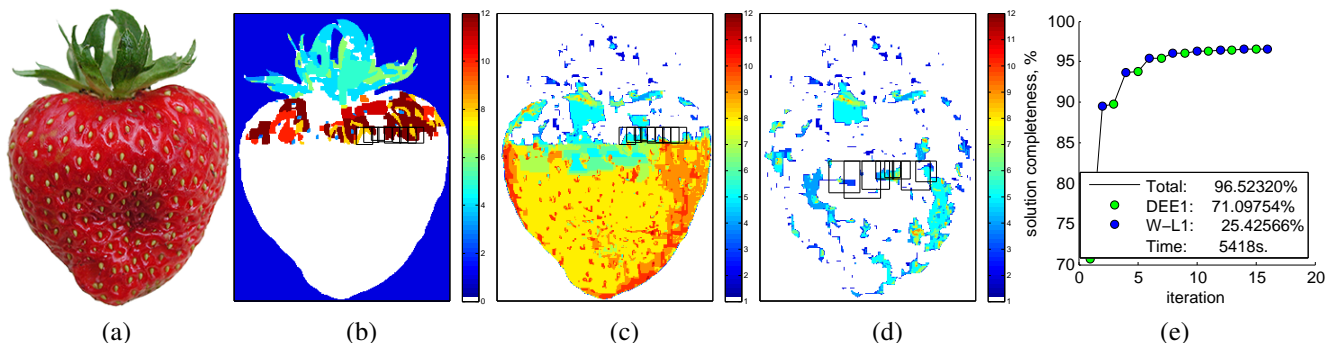

Figure 6. Windowing method in progress. (a) image corresponding to the instance `strawberry-glass-2-small`; (b) Partial labeling for the current reduction: only pixels with a single remaining label are assigned. Black boxes depict the current set of windows selected for application of L1 (processed in parallel). (c) The number of labels remaining in every pixel on the same iteration as (b). White color indicates fully resolved pixels. (d) The number of remained labels upon termination. Selected windows are larger, but cannot improve the reduction. (e) Algorithm progress during DEE1 iterations (green) and window-L1 iterations (blue).

this technique we demonstrate how a nearly complete solution can be found for large instances, by considering always only a part of problem at a time. We see that the *reminder of the problem* (the final reduced problem) often decouples in several small independent components, that are feasible to ,*e.g.*, ILP methods. These experiments are a proof-of-concept, we definitely need to develop methods further for a practical implementation. Results are shown in Figure 6-3. Note, for some of these instances method of Swoboda *et al.* [30] identifies a more complete solution, despite we claimed to generalize it. They are using a suboptimal LP solver, but applying it globally to the whole problem. It is likely that our results can be improved by picking the windows more accurately.

One interesting consequence of the windowing method is that it can be applied also with methods of Kovtun [19], Swoboda *et al.* [30], MQPBO [14] and, in fact, any other method that constructs a pixel-wise improving mapping.

Because MQPBO constructs a flow network on the graph with $K * |\mathcal{V}|$ nodes and $K^2 |\mathcal{E}|$ edges, it was reported as intractable for several vision problems [13]. The proposed windowing technique can remove this limitation.

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 nar-
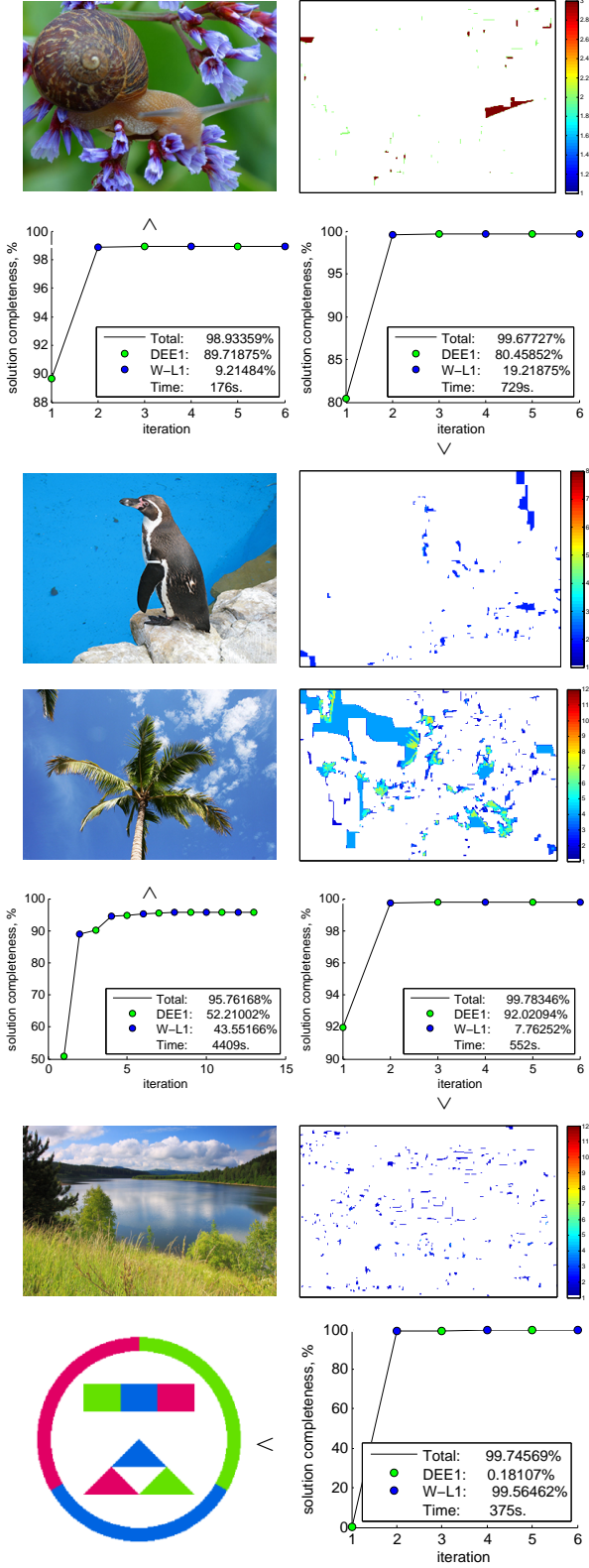
Figure 7. Other instances of `color-seg-n4`.

row 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 also proposed 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. The windowing technique allows to apply previous methods by parts as well.

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 (the local polytope Λ would be defined as $\mathrm{aff}(\mathcal{M}) \cap \mathbb{R}_+^{\mathcal{I}}$).

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] I. Georgiev, R. H. Lilien, and B. R. Donald. Improved pruning algorithms and divide-and-conquer strategies for dead-end elimination, with application to protein design. *Bioinformatics*, 22:174–183, July 2006.

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

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

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

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

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

[13] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, and C. Rother. A comparative study of modern inference techniques for discrete energy minimization problem. In *CVPR*, 2013.

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

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

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

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

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

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

[20] I. Lasters, M. De Maeyer, and J. Desmet. Enhanced deadend elimination in the search for the global minimum energy conformation of a collection of protein side chains. *Protein Engineering*, 8(8):815–22, 1995.

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

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

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

[24] N. A. Pierce, J. A. Spriet, J. Desmet, and S. L. Mayo. Conformational splitting: A more powerful criterion for dead-end elimination. *Journal of Computational Chemistry*, 21(11):999–1009, 2000.

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

[26] A. Shekhovtsov. *Exact and Partial Energy Minimization in Computer Vision*. PhD Thesis CTU–CMP–2013–24, Center for Machine Perception, K13133 FEE Czech Technical University in Prague, 2013.

[27] A. Shekhovtsov and V. Hlaváč. On partial opimality by auxiliary submodular problems. *Control Systems and Computers*, (2), 2011. Special issue.

[28] A. Shekhovtsov, V. Kolmogorov, P. Kohli, V. Hlavac, C. Rother, and P. Torr. LP-relaxation of binarized energy minimization. Technical Report CTU–CMP–2007–27, Czech Technical University, 2008.

[29] M. Shlezinger. Syntactic analysis of two-dimensional visual signals in the presence of noise. *Cybernetics and Systems Analysis*, 4:113–130, 1976. See review [32].

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

[31] 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.

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