

Maximum Persistency via Iterative Relaxed Inference in Graphical Models

Alexander Shekhovtsov, Paul Swoboda, and Bogdan Savchynskyy

Abstract—We consider the NP-hard problem of MAP-inference for undirected discrete graphical models. We propose a polynomial time and practically efficient algorithm for finding a part of its optimal solution. Specifically, our algorithm marks some labels of the considered graphical model either as (i) *optimal*, meaning that they belong to all optimal solutions of the inference problem; (ii) *non-optimal* if they provably do not belong to any solution. With access to an exact solver of a linear programming relaxation to the MAP-inference problem, our algorithm marks the maximal possible (in a specified sense) number of labels. We also present a version of the algorithm, which has access to a suboptimal dual solver only and still can ensure the (non-)optimality for the marked labels, although the overall number of the marked labels may decrease. We propose an efficient implementation, which runs in time comparable to a single run of a suboptimal dual solver. Our method is well-scalable and shows state-of-the-art results on computational benchmarks from machine learning and computer vision.

Index Terms—Persistency, partial optimality, LP relaxation, discrete optimization, WCSP, graphical models, energy minimization

1 INTRODUCTION

WE consider the energy minimization or maximum a posteriori (MAP) inference problem for discrete graphical models. In the most common pairwise case it has the form

$$\min_{x \in \mathcal{X}} E_f(x) := f_\emptyset + \sum_{v \in \mathcal{V}} f_v(x_v) + \sum_{uv \in \mathcal{E}} f_{uv}(x_u, x_v), \quad (1)$$

where minimization is performed over vectors x , containing the discrete-valued components x_v . Further notation is to be detailed in § 2. The problem has numerous applications in computer vision, machine learning, communication theory, signal processing, information retrieval and statistical physics, see [18, 47, 29] for an overview of applications. Even in the binary case, when each coordinate of x can be assigned two values only, the problem is known to be NP-hard and is also hard to approximate [27].

Hardness of the problem justifies a number of existing approximate methods addressing it [18]. Among them, solvers addressing its linear programming (LP) relaxations and in particular, the LP dual [41, 49, 20], count among the most versatile and efficient ones. However, apart from some notable exceptions (see the overview of related work below), approximate methods can not guarantee neither optimality of their solutions as a whole, nor even optimality of any individual solution coordinates. That is, if x is a solution returned by an approximate method and x^* is an optimal one, there is no guarantee that $x_v^* = x_v$ for any coordinate v .

In contrast, our method provides such guarantees for some coordinates. More precisely, for each component x_v it eliminates

those of its values (henceforth called *labels*), which *provably* can not belong to any optimal solution. We call these eliminated labels *persistent non-optimal*. Should a single label a remain non-eliminated, it implies that for all optimal solutions x^* it holds $x_v^* = a$ and the label a is called *persistent optimal*. Our elimination method is polynomial and is applicable with *any* (approximate) solver for a dual of a linear programming relaxation of the problem, employed as a subroutine.

1.1 Related Work

A trivial but essential observation is that any method identifying persistency has to be based on tractable sufficient conditions in order to avoid solving the NP-hard problem (1).

Dead-end elimination methods (DEE) [10] verify local sufficient conditions by inspecting a given node and its immediate neighbors at a time. When a label in the node can be substituted with another one such that the energy for all configurations of the neighbors does not increase, this label can be eliminated without loss of optimality.

A similar principle for eliminating interchangeable labels was proposed in constraint programming [11]. It’s generalization to a related problem of Weighted Constraint Satisfaction (WCS) is known as *dominance rules* or *soft neighborhood substitutability*. However, because general WCS considers bounded addition $a \oplus b = \min(k, a+b)$, instead of ordinary addition, the condition appears to be intractable and therefore weaker sufficient local conditions were introduced, e.g., [26, 9]. The way [9] selects a local substitute label using equivalence preserving transforms is related to our method, in which we use an approximate solution based on the dual of the LP relaxation as a tentative substitute (or test) labeling.

Although the local character of the DEE methods allows for an efficient implementation, it also significantly limits their quality, i.e., the number of persistencies found. As shown in [35, 43, 48], considering more global criteria may significantly increase the algorithm’s quality.

- Alexander Shekhovtsov is with the Institute for Computer Graphics and Vision (ICG), Graz University of Technology, Inffeldgasse 16, Graz 8010, Austria.
E-mail: shekhovtsov@icg.tugraz.at.
- Paul Swoboda is with the Discrete Optimization Group, Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria.
E-mail: pswoboda@ist.ac.at.
- Bogdan Savchynskyy is with Computer Vision Lab, Faculty of Computer Science, Institute for Artificial Intelligence, Dresden University of Technology, 01062 Dresden, Germany.
E-mail: bogdan.savchynskyy@tu-dresden.de.

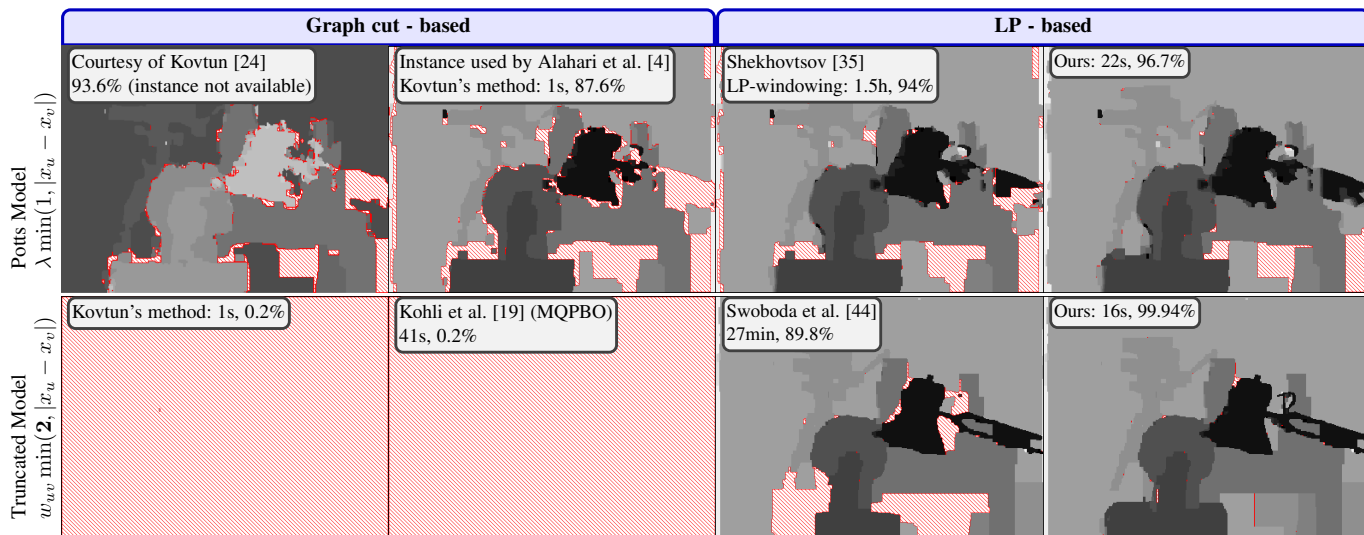


Fig. 1: Progress of partial optimality methods. The top row corresponds to a stereo model with Potts interactions and large aggregating windows for unary costs used in [24, 4] (instance published by [4]). The bottom row is a more refined stereo model with truncated linear terms [45] (instance from [1]). The hashed red area indicates that the optimal persistent label in the pixel is not found (but some non-optimal labels might have been eliminated). Solution completeness is given by the percentage of persistent labels. Graph cut based methods are fast but only efficient for strong unary terms. LP-based methods are able to determine larger persistent assignments but are extremely slow prior to this work.

The roof dual relaxation in quadratic pseudo-Boolean Optimization (QPBO) [5, 30] (equivalent to pairwise energy minimization with binary variables) has the property that all variables that are integer in the relaxed solution are persistent. Several generalizations of roof duality to higher-order energies were proposed (e.g., [2, 21]). The MQPBO method [19] and the generalized roof duality [50] extend roof duality to the multi-label case by reducing the problem to binary variables and generalizing the concept of submodular relaxation [21], respectively. Although for binary pairwise energies these methods provide a very good trade-off between computational efficiency and a number of found persistencies, their efficacy drops as the number of label grows.

Auxiliary submodular problems were proposed in [24, 25] as a sufficient persistency condition for multilabel energy minimization. In the case of Potts model, the method has a very efficient specialized algorithm [14]. Although these methods have shown very good efficacy for certain problem classes appearing in computer vision, the number of persistencies they find drastically decreases when the energy does not have strong unary terms (see Fig. 1).

In contrast to the above methods that technically rely either on local conditions or on computing a maximum flow (min-cut), the works [42, 43, 44] and [35] proposed **persistency approaches relying on a general linear programming relaxation**. Authors of [42, 43, 44] demonstrated applicability of their approach to large-scale problems by utilizing existing efficient approximate MAP-inference algorithms, while in [35] the large-scale problems are addressed using a windowing technique. Despite the superior persistency results, the running time of the approximate-LP-based methods remained prohibitively slow for practical applications as illustrated by an example in Fig. 1.

Not only LP-based methods can achieve superior results in practice, but they are even theoretically guaranteed to do so, as proven for the method [35, 37]. In this method, the problem of determining the *maximum number of persistencies* is formulated

as a polynomially solvable linear program. It is guaranteed to find a *provably larger* persistency assignment than most of the above mentioned approaches. However, solving this linear program for large scale instances is numerically unstable/intractable and applying it to multiple local windows is prohibitively slow. This poses a challenge of designing an LP-based method that would be indeed practical.

1.2 Contribution

In this work we propose a method which solves the same *maximum persistency problem* as in [35] and therefore delivers *provably* better results than other methods. Similar to [44], our method requires to iteratively (approximately) solve the linear programming relaxation of (1) as a subroutine. However, our method is significantly faster than [35, 44] due to a substantial theoretical and algorithmic elaboration of this subroutine.

We demonstrate the efficiency of our approach on benchmark problems from machine learning and computer vision. We outperform *all* competing methods in terms of the number of persistent labels and method [35] in speed and scalability. On randomly generated small problems, we show that the set of persistent labels found using approximate LP solver is close to the maximal one as established by the (costly and not scalable) method [35].

The present paper is a revised version of [39]. Besides reworked explanations, shortened and clarified proofs, one new technical extension is a more general dual algorithm, with termination guarantees for a larger class of approximate solvers.

2 WORK OVERVIEW

This section serves as an overview of our method, where we give the most general definitions, formulate the maximum persistency problem and briefly describe a generic method to solve it. This description, equipped with references to subsequent sections, should serve as a road map for the rest of the paper.

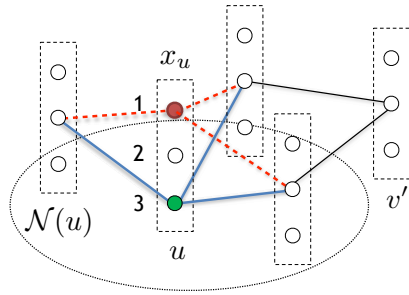


Fig. 2: Dead end elimination / dominance. Variables are shown as boxes and their possible labels as circles. Label $x_u = 1$ is substituted with label $x_u = 3$. If for any configuration of neighbors $x_{\mathcal{N}(u)}$ the energy does not increase (only the terms inside $\{u\} \cup \mathcal{N}(u)$ contribute to the difference), label $x_u = 1$ can be eliminated without loss of optimality.

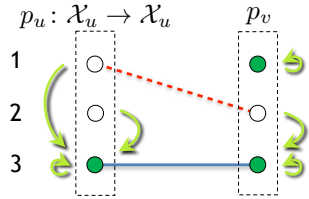


Fig. 3: Simultaneous substitution of labels in two variables. $\mathcal{X}_u = \mathcal{X}_v = \{1, 2, 3\}$, labels at arrow tails are substituted with labels at arrow heads. So the joint configuration (1, 2) (dashed) is substituted with the configuration (3, 3) (solid).

2.1 Notation

In the MAP-Inference Problem (1) we assume $(\mathcal{V}, \mathcal{E})$ to be a directed graph with the set of *nodes* \mathcal{V} and the set of *edges* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Let uv denote an ordered pair (u, v) and $\mathcal{N}(u) = \{v \mid uv \in \mathcal{E} \vee vu \in \mathcal{E}\}$ stands for the set of *neighbors* of u . Each node $v \in \mathcal{V}$ is associated with a variable x_v taking its values in a finite *set of labels* \mathcal{X}_v . *Cost functions or potentials* $f_v: \mathcal{X}_v \rightarrow \mathbb{R}$, $f_{uv}: \mathcal{X}_u \times \mathcal{X}_v \rightarrow \mathbb{R}$ are associated with nodes and edges respectively. Let $f_\emptyset \in \mathbb{R}$ be a *constant term*, which we introduce for the sake of notation. Finally, \mathcal{X} stands for the Cartesian product $\prod_{v \in \mathcal{V}} \mathcal{X}_v$ and its elements $x \in \mathcal{X}$ are called *labelings*.

We represent all potentials of energy (1) by a single *cost vector* $f \in \mathbb{R}^{\mathcal{I}}$, where the set \mathcal{I} enumerates all components of all terms: $\mathcal{I} = \{\emptyset\} \cup \{(u, i) \mid u \in \mathcal{V}, i \in \mathcal{X}_u\} \cup \{(uv, ij) \mid uv \in \mathcal{E}, i \in \mathcal{X}_u, j \in \mathcal{X}_v\}$.

2.2 Improving Substitutions

We formulate our persistency method in the framework of (strictly) improving substitutions, called *improving mappings* in our previous works [35, 37]. It was shown in [35] that most existing persistency techniques can be expressed as improving substitutions. A mapping $p: \mathcal{X} \rightarrow \mathcal{X}$ is called a *substitution*, if it is idempotent, *i.e.*, $p(x) = p(p(x))$.

Definition 2.1. A substitution $p: \mathcal{X} \rightarrow \mathcal{X}$ is called *strictly improving* for the cost vector f if

$$(\forall x \mid p(x) \neq x) \quad E_f(p(x)) < E_f(x). \quad (2)$$

When a strictly improving substitution is applied to any labeling x , it is guaranteed that $p(x)$ has equal or better energy. In particular, strictly improving substitutions generalize the strong

autarky property [5]. When applied to the whole search space \mathcal{X} we obtain its image $p(\mathcal{X})$ – a potentially smaller search space containing all optimal labelings.

In what follows we will restrict ourselves to *node-wise* substitutions, *i.e.*, those defined locally for each node: $p(x)_u = p_u(x_u)$, where $p_u: \mathcal{X}_u \rightarrow \mathcal{X}_u$. Indeed, already this class of substitutions covers most existing persistency methods.

Example 2.1. Let us consider the dead-end elimination (DEE) [10, 13]. It is a test whether a given label in a single node, *e.g.*, $x_u = 1$ in Fig. 2 can be substituted with another one, *e.g.*, $x_u = 3$ in Fig. 2. The change of the energy under this substitution depends only on the configuration of neighbors $x_{\mathcal{N}(u)}$, and the value of the change is additive in neighbors, so that it can be verified for all $x_{\mathcal{N}(u)}$ whether the substitution always improves the energy. If it is so, the label $x_u = 1$ can be eliminated and the test is repeated for a different label in the reduced problem.

A general substitution we consider is applied to labels in all nodes simultaneously, as illustrated in Fig. 3 for two variables. We obtain the following principle for identifying persistencies.

Proposition 2.2. If p is a strictly improving substitution, then any optimal solution x^* of (1) must satisfy $(\forall v \in \mathcal{V}) p_v(x_v^*) = x_v^*$.

Indeed, otherwise $E_f(p(x^*)) < E_f(x^*)$, which is a contradiction. If $p_v(i) \neq i$, then idempotency implies that label (v, i) is non-optimal persistent and can be excluded from consideration.

2.3 Verification Problem

Verifying whether a given substitution is strictly improving is an NP hard decision problem [35]. In order to obtain a polynomial sufficient condition we will first rewrite (2) as an energy minimization problem and then relax it. To this end we reformulate Definition 2.1 in an optimization form:

Proposition 2.3. Substitution p is strictly improving *iff*

$$\min_{x \in \mathcal{X}} (E_f(x) - E_f(p(x))) \geq 0, \quad (3a)$$

$$p(x) = x \text{ for all minimizers.} \quad (3b)$$

Proof. Indeed, condition (3a) is equivalent to $(\forall x) E_f(x) \geq E_f(p(x))$. Sufficiency: if $x \neq p(x)$, then x is not a minimizer and $E_f(x) > E_f(p(x))$. Necessity: for $x = p(x)$ we have that $E_f(x) = E_f(p(x))$, therefore from Definition 2.1 it follows that condition (3a) holds and any $x = p(x)$ is a minimizer, moreover, for any minimizer x it must be $E_f(x) - E_f(p(x)) = 0$ and from Definition 2.1 it follows $x = p(x)$. \square

In § 3 we will show that the difference of the energies in (3a) can be represented as a pairwise energy with an appropriately constructed cost vector g so that there holds

$$E_f(x) - E_f(p(x)) = E_g(x). \quad (4)$$

Therefore, according to Proposition 2.3 the verification of the strictly improving property reduces to minimizing the energy (4) and checking that (3b) is fulfilled. To make the verification problem *tractable*, we relax it as

$$\min_{\mu \in \Lambda} E_g(\mu) \geq 0, \quad (5a)$$

$$p(\mu) = \mu \text{ for all minimizers,} \quad (5b)$$

where Λ is a tractable polytope such that its integer vertices correspond to labelings (the standard LP relaxation that we use

will be defined in § 5), μ is a relaxed labeling and $E_g(\mu)$ and $p(\mu)$ are appropriately defined extensions of discrete functions $E_g(x)$ and $p(x)$, defined in § 3. By construction, the objective value (5) matches exactly to that of (2) for all integer labelings $\mu \in \Lambda \cap \{0, 1\}^{\mathcal{X}}$, which is sufficient for (2) to hold. The sufficient condition (5) (made precise in Definition 3.2) means that p strictly improves not only integer labelings but also all relaxed labelings and therefore such substitutions will be called *strictly relaxed-improving* for a cost vector f . Assuming Λ is fixed by the context, let \mathbb{S}_f denotes the set of all substitutions p satisfying (5).

2.4 Maximum Persistency and Subset-to-One Substitutions

The *maximum persistency* approach [35] consists of finding a relaxed-improving substitution $p \in \mathbb{S}_f$ that eliminates the *maximal* number of labels:

$$\max_{p \in \mathcal{P}} \sum_{v \in \mathcal{V}} |\{i \in \mathcal{X}_v \mid p(i) \neq i\}|, \text{ s.t. } p \in \mathbb{S}_f, \quad (6)$$

where \mathcal{P} is a class of substitutions.

While maximizing over all substitutions is not tractable, maximizing over the following restricted class is [35]. Assume we are given a *test* labeling y , which in our case will be an approximate solution of the MAP inference (1). We then consider substituting in each node v a subset of labels with y_v .

Definition 2.4 ([35]). A substitution p is in the class of *subset-to-one* substitutions $\mathcal{P}^{2,y}$, where $y \in \mathcal{X}$, if there exist subsets $\mathcal{Y}_v \subseteq \mathcal{X}_v \setminus \{y_v\}$ for all v such that

$$p_v(i) = \begin{cases} y_v, & \text{if } i \in \mathcal{Y}_v; \\ i, & \text{if } i \notin \mathcal{Y}_v. \end{cases} \quad (7)$$

See Figs. 3 and 4 for examples. Note, this class is rather large: there are $2^{|\mathcal{X}_v|-1}$ possible choices for p_v and most of the existing methods for partial optimality still can be represented using it [35] (in particular, methods [25, 14] can be represented using a constant test labeling, $y_v = \alpha$ for all v).

The restriction to the class $\mathcal{P}^{2,y}$ allows to represent the search of the substitution that eliminates the maximum number of labels as the one with the largest (by inclusion) sets \mathcal{Y}_v of substituted labels. This allows to propose a relatively simple algorithm.

2.5 Cutting Plane Algorithm

The algorithm is a cutting plane method in a general sense: we maintain a substitution p^t which is in all iterations better or equal than the solution to (6) and achieve feasibility by iteratively constraining it.

Initialization: Define the substitution p^0 by the sets $\mathcal{Y}_v^0 = \mathcal{X}_v \setminus \{y_v\}$. It substitutes everything with y and clearly maximizes the objective (6).

Verification: Check whether current p^t is strictly relaxed-improving for f by solving relaxed problem (5). If yes, return p^t . If not, the optimal relaxed solution μ^* corresponds to the most violated constraint.

Cutting plane: Assign p^{t+1} to the substitution defined by the largest sets \mathcal{Y}_v^{t+1} such that $\mathcal{Y}_v^{t+1} \subseteq \mathcal{Y}_v^t$ and the constraints $E_g(\mu^*) \geq 0$, $p(\mu^*) = \mu^*$, are satisfied. Repeat the verification step.

The steps of this meta-algorithm are illustrated in Fig. 4. It is clear that when the algorithm stops the substitution p^t is strictly improving, although it could be the identity map that does not

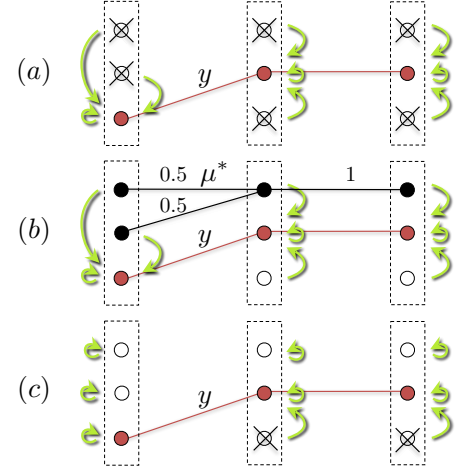


Fig. 4: Steps of the discrete cutting-plane algorithm. (a) Starting from substitution that maps everything to the test labeling y (red), crossed labels would be eliminated if p passes the sufficient condition. (b) A relaxed solution μ^* violating the sufficient condition is found (black). (c) Substitution p is pruned.

eliminate any labels. The exact specification of the cutting plane step will be derived in § 4 and it will be shown that this algorithm solves the maximum persistency problem (6) over $\mathcal{P}^{2,y}$ optimally.

2.6 Work Outline

In § 3 we give a precise formulation of the relaxed condition (5) and its components. In § 4 we specify details of the algorithm and prove its optimality. These results hold for a general relaxation $\Lambda \supseteq \mathcal{M}$ but require to solve linear programs (5a) precisely.

The rest of the paper is devoted to an approximate solution of the problem (6), *i.e.* finding a relaxed improving mapping, which is *almost* maximum. We consider specifically the standard LP relaxation and reformulate the algorithm to use a dual solver for the problem (5a), § 5. We then gradually relax requirements on the optimality of the dual solver while keeping persistency guarantees, §§ 5 to 7, and propose several theoretical and algorithmic tools to solve the series of verification problems incrementally and overall efficiently, § 8. Finally, we provide an exhaustive experimental evaluation in § 9, which clearly demonstrates efficacy of the developed method.

3 RELAXED-IMPROVING SUBSTITUTIONS

3.1 Overcomplete Representation

In this section we formally derive the strictly relaxed-improving sufficient condition (5). To obtain the relaxation we use the standard lifting approach (a.k.a. overcomplete representation [47]), in which a labeling is represented using the 1-hot encoding. This lifting allows to linearize the energy function, the substitution and consequently both the non-relaxed (3) and relaxed (5) improving substitution criteria.

The lifting is defined by the mapping $\delta: \mathcal{X} \rightarrow \mathbb{R}^{\mathcal{I}}$:

$$\delta(x)_{\emptyset} = 1, \quad (8a)$$

$$\delta(x)_u(i) = \llbracket x_u = i \rrbracket, \quad (8b)$$

$$\delta(x)_{uv}(i, j) = \llbracket x_u = i \rrbracket \llbracket x_v = j \rrbracket, \quad (8c)$$

where $\llbracket \cdot \rrbracket$ is the Iverson bracket, *i.e.*, $\llbracket A \rrbracket$ equals 1 if A is true and 0 otherwise. Using this lifting, we can linearize unary

terms as $f_v(x_v) = \sum_k f_v(k) \mathbb{1}[x_v=k] = \sum_k f_v(k) \delta(x)_v(k)$ and similarly for the pairwise terms. This allows to linearize the energy function E_f and write it as a scalar product $E_f(x) = \langle f, \delta(x) \rangle$ in $\mathbb{R}^{\mathcal{X}}$. The energy minimization problem (1) can then be written as

$$\min_{x \in \mathcal{X}} E_f(x) = \min_{x \in \mathcal{X}} \langle f, \delta(x) \rangle = \min_{\mu \in \mathcal{M}} \langle f, \mu \rangle, \quad (9)$$

where $\mathcal{M} = \text{conv } \delta(\mathcal{X})$ is the convex hull of all labelings in the lifted space, also known as marginal polytope [47]. The last equality in (9) uses the fact that the minimum of a linear function on a finite set equals the minimum on its convex hull. Expression (9) is an equivalent reformulation of the energy minimization problem as a linear program, however over a generally intractable polytope \mathcal{M} .

3.2 Lifting of Substitutions

Next we show how a substitution $p: \mathcal{X} \rightarrow \mathcal{X}$ can be represented as a linear map in the lifted space $\mathbb{R}^{\mathcal{X}}$. This will allow to express the term $E_f(p(x))$ as a linear function of $\delta(x)$ and hence also to represent the non-relaxed criterion (3).

Proposition 3.1. Given a substitution p , let $P^\top: \mathbb{R}^{\mathcal{X}} \rightarrow \mathbb{R}^{\mathcal{X}}$ be defined by its action on a cost vector $f \in \mathbb{R}^{\mathcal{X}}$ as follows:

$$(P^\top f)_\emptyset = f_\emptyset, \quad (10a)$$

$$(P^\top f)_u(i) = f_u(p_u(i)), \quad (10b)$$

$$(P^\top f)_{uv}(i, j) = f_{uv}(p_u(i), p_v(j)) \quad (10c)$$

$\forall u \in \mathcal{V}, uv \in \mathcal{E}, ij \in \mathcal{X}_{uv}$. Then P satisfies

$$(\forall x \in \mathcal{X}) \delta(p(x)) = P\delta(x). \quad (11)$$

Proof. Let $x \in \mathcal{X}$. From (10) it follows that $E_{P^\top f}(x) = E_f(p(x))$, which can be expressed as a scalar product

$$\langle P^\top f, \delta(x) \rangle = \langle f, P\delta(x) \rangle = \langle f, \delta(p(x)) \rangle.$$

Since this equality holds for all $f \in \mathbb{R}^{\mathcal{X}}$ it follows that $P\delta(x) = \delta(p(x))$. \square

The expression (11) allows to write the energy of the substituted labeling $p(x)$ as

$$E_f(p(x)) = \langle f, \delta(p(x)) \rangle = \langle f, P\delta(x) \rangle = E_{P^\top f}(x). \quad (12)$$

For this reason, the mapping P is called the *linear extension* of p and will be denoted with the symbol $[p]$. The following example illustrates how $[p]$ looks in coordinates.

Example 3.1. Consider the substitution p depicted in Fig. 3 and defined by $p_u: 1, 2, 3 \mapsto 3, 3, 3$; $p_v: 1, 2, 3 \mapsto 1, 3, 3$. The relaxed labeling $\mu \in \mathbb{R}^{\mathcal{I}}$ has the structure $(\mu_\emptyset, \mu_u, \mu_v, \mu_{uv})$. The linear extension $[p]: \mathbb{R}^{\mathcal{X}} \rightarrow \mathbb{R}^{\mathcal{X}}$ can be written as a block-diagonal matrix

$$\begin{pmatrix} 1 & & & \\ & P_u & & \\ & & P_v & \\ & & & P_{uv} \end{pmatrix}, \text{ where } P_u = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, P_v = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad (13)$$

and P_{uv} is defined by $P_{uv}\mu_{uv} = P_u\mu_u P_v$, where μ_{uv} is shaped as a 3×3 matrix. The action of the block P_u expresses as

$$P_u \begin{pmatrix} \mu_u(1) \\ \mu_u(2) \\ \mu_u(3) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (14)$$

i.e. all relaxed labels are mapped to the indicator of the label $x_u = 3$. And the adjoint operator P^\top acts as follows (*cf.* (10b)):

$$P_u^\top (f_u(1) f_u(2) f_u(3))^\top = (f_u(3) f_u(3) f_u(3))^\top, \quad (15)$$

$$P_v^\top (f_u(1) f_u(2) f_u(3))^\top = (f_u(1) f_u(3) f_u(3))^\top. \quad (16)$$

Similarly, due to (10c) we have, *e.g.*, $(P_{uv}^\top f_{uv})(1, 2) = f_{uv}(3, 3)$, $(P_{uv}^\top f_{uv})(1, 1) = f_{uv}(3, 1)$ and so on. \square

3.3 Strictly improving substitutions

Let I denote the identity mapping $\mathbb{R}^{\mathcal{X}} \rightarrow \mathbb{R}^{\mathcal{X}}$. Using Proposition 2.3 and the linear extension $[p]$, we obtain that substitution p is strictly improving iff the value of

$$\begin{aligned} \min_{x \in \mathcal{X}} \langle f, \delta(x) - \delta(p(x)) \rangle &= \min_{x \in \mathcal{X}} \langle f, (I - [p])\delta(x) \rangle \\ &= \min_{x \in \mathcal{X}} \langle (I - [p])^\top f, \delta(x) \rangle = \min_{\mu \in \mathcal{M}} \langle (I - [p])^\top f, \mu \rangle \end{aligned} \quad (17)$$

is zero and $[p]\mu = \mu$ for all minimizers. Note that problem (17) is of the same form as the energy minimization (9) with the cost vector $g = (I - [p])^\top f$, as introduced in (4).

The sufficient condition for persistency (5) is obtained by relaxing the intractable marginal polytope \mathcal{M} in (17) to a tractable outer approximation $\Lambda \supseteq \mathcal{M}$.

Definition 3.2 ([35]). Substitution p is *strictly Λ -improving* for the cost vector $f \in \mathbb{R}^{\mathcal{X}}$ (shortly, *strictly relaxed-improving*, or $p \in \mathbb{S}_f$) if

$$\min_{\mu \in \Lambda} \langle (I - [p])^\top f, \mu \rangle = 0, \quad (18a)$$

$$[p]\mu^* = \mu^* \text{ for all minimizers.} \quad (18b)$$

In § 5, Λ will be defined as the polytope of the standard LP relaxation but until then the arguments are general and require only that $\Lambda \supseteq \mathcal{M}$. Since Λ includes all integer labelings, it is a sufficient condition for improving substitution and hence persistency.

Corollary 3.3. If substitution p is strictly Λ -improving for f and $\Lambda \supseteq \mathcal{M}$, then p is strictly improving for f .

The problem (18a) will be called the *verification LP* and the decision problem to test for $p \in \mathbb{S}_f$, *i.e.* to verify conditions (18), will be called the *verification problem*.

4 GENERIC PERSISTENCE ALGORITHM

4.1 Structure of $P^{2,y}$

In [35] it was shown that the maximum persistency problem (6) over the class of substitutions $\mathcal{P}^{2,y}$ can be formulated as a single linear program, where the substitution is represented using auxiliary (continuous) variables. Here we take a different approach based on observing a lattice-like structure of improving substitutions.

Throughout this section we will assume that the test labeling $y \in \mathcal{X}$ is fixed. Let us compare two substitutions p and q by the sets of the labels they eliminate. A substitution $p \in \mathcal{P}^{2,y}$ eliminates all labels in \mathcal{Y}_v , or equivalently all labels not in $p_v(\mathcal{X}_v)$.

Definition 4.1. A substitution $p \in \mathcal{P}^{2,y}$ is *better equal* than a substitution $q \in \mathcal{P}^{2,y}$, denoted by $p \geq q$, if $(\forall v \in \mathcal{V})$ it holds $p_v(\mathcal{X}_v) \subseteq q_v(\mathcal{X}_v)$.

Proposition 4.2. Let the partially ordered set $(\mathbb{S}_f \cap \mathcal{P}^{2,y}, \geq)$ of subset-to-one strictly relaxed-improving substitutions has the maximum and let it be denoted r . Then r is the unique solution of (6) with $\mathcal{P} = \mathcal{P}^{2,y}$.

Proof. Since r is the maximum, it holds $r \geq q$ for all q . From Definition 4.1 we have $r_v(\mathcal{X}_v) \subseteq q_v(\mathcal{X}_v)$ for all v and thus $\sum_v |r_v(\mathcal{X}_v)| \leq \sum_v |q_v(\mathcal{X}_v)|$. Therefore r is optimal to (6). Additionally, if $r \neq q$, then it holds $\sum_v |r_v(\mathcal{X}_v)| < \sum_v |q_v(\mathcal{X}_v)|$ and therefore r is the unique solution to (6). \square

The existence of the maximum will formally follow from the correctness proof of the algorithm, Theorem 4.6. A stronger claim, which is not necessary for our analysis, but which may provide a better insight is that $(\mathbb{S}_f \cap \mathcal{P}^{2,y}, \geq)$ is a lattice isomorphic to the lattice of sets with union and intersection operations. This is seen as follows. If both p and q are strictly-improving then so is their composition $r(x) = p(q(x))$, as can be verified by chaining inequalities (2). In $\mathcal{P}^{2,y}$ the composition satisfies the property $r_u(\mathcal{X}_u) = p_u(\mathcal{X}_u) \cap q_u(\mathcal{X}_u)$, and can be identified with the *join* of p and q (the least r such that $r \geq p$ and $r \geq q$). This can be shown to hold also for $\mathbb{S}_f \cap \mathcal{P}^{2,y}$. It is this structure that allows to find the maximum in $\mathbb{S}_f \cap \mathcal{P}^{2,y}$ by a relatively simple algorithm.

4.2 Generic Algorithm

Our generic primal algorithm, displayed in Algorithm 1, represents a substitution $p \in \mathcal{P}^{2,y}$ by the sets \mathcal{Y}_v of labels to be substituted with y , via (7). Line 1 initializes these sets to all labels but y . Line 3 constructs the cost vector of the verification LP in condition (18). Lines 4 and 6 solve the verification LP and test whether sufficient conditions (18) are satisfied via the following reformulation.

Proposition 4.3. For a given substitution p , let $\mathcal{O}^* \subseteq \Lambda$ denote the set of minimizers of the verification LP (18) and

$$\mathcal{O}_v^* := \{i \in \mathcal{X}_v \mid (\exists \mu \in \mathcal{O}^*) \mu_v(i) > 0\}, \quad (19)$$

which is the support set of all optimal solutions in node v . Then $p \in \mathbb{S}_f$ iff $(\forall v \in \mathcal{V} \forall i \in \mathcal{O}_v^*) p_v(i) = i$. Proof in § A.

Corollary 4.4. For substitution $p \in \mathcal{P}^{2,y}$ defined by (7) it holds $p \in \mathbb{S}_f$ iff $(\forall v \in \mathcal{V}) \mathcal{O}_v^* \cap \mathcal{Y}_v = \emptyset$.

In the remainder of the paper we will relate the notation \mathcal{O}_v^* to \mathcal{O}^* as in (19). The set of optimal solutions \mathcal{O}^* can, in general, be a d -dimensional face of Λ . We need to determine in (19) whether an optimal solution $\mu \in \mathcal{O}^*$ exists such that the coordinate $\mu_v(i)$ is strictly positive. From the theory of linear programming [46], it is known that if one takes an optimal solution μ in the relative interior of \mathcal{O}^* (i.e., if \mathcal{O}^* is a 2D face the relative interior excludes vertices and edges of \mathcal{O}^*) then its support set $\{i \mid \mu_v(i) > 0\}$ is the same for all such points and it matches \mathcal{O}_v^* (19). Therefore, it is practically feasible to find the support sets \mathcal{O}_v^* by using a single solution found by an interior point/barrier method (which are known to converge to a “central” point of the optimal face) or methods based on smoothing [31, 32]. Obtaining an *exact* solution by these methods may become computationally expensive as the size of the inference problem (1) grows. Despite that, Algorithm 1 is implementable and defines the baseline for its practically efficient variants solving (6) approximately. These are developed further in the paper.

Since Line 6 of Algorithm 1 verifies precisely the condition of Corollary 4.4, the algorithm terminates as soon as $p \in \mathbb{S}_f$ and

Algorithm 1: Iterative Pruning LP-Primal

Input: Cost vector $f \in \mathbb{R}^{\mathcal{I}}$, test labeling $y \in \mathcal{X}$;
Output: Maximum strictly improving substitution p ;

- 1 $(\forall v \in \mathcal{V}) \mathcal{Y}_v := \mathcal{X}_v \setminus \{y_v\}$;
- 2 **while true**
- 3 Construct the verification problem potentials
 $g := (I - [p])^\top f$ with p defined by (7);
- 4 $\mathcal{O}^* = \operatorname{argmin}_{\mu \in \Lambda} \langle g, \mu \rangle$;
- 5 $(\forall v \in \mathcal{V}) \mathcal{O}_v^* = \{i \in \mathcal{X}_v \mid (\exists \mu \in \mathcal{O}^*) \mu_v(i) > 0\}$;
- 6 **if** $(\forall v \in \mathcal{V}) \mathcal{O}_v^* \cap \mathcal{Y}_v = \emptyset$ **then return** p ;
- 7 **for** $v \in \mathcal{V}$ **do**
- 8 Pruning of substitutions: $\mathcal{Y}_v := \mathcal{Y}_v \setminus \mathcal{O}_v^*$;

hence p is strictly improving. In the opposite case, Line 8 *prunes* the sets \mathcal{Y}_v by removing labels corresponding to the support set \mathcal{O}_v^* of all optimal solutions of the verification LP, which have been identified now to violate the sufficient condition. These labels may be a part of some optimal solution to (1) and will not be eliminated.

To complete the analysis of Algorithm 1 it remains to answer two questions: i) does it terminate and ii) is it optimal for the maximum persistency problem (6)?

Proposition 4.5. Algorithm 1 runs in polynomial time and returns a substitution $p \in \mathbb{S}_f \cap \mathcal{P}^{2,y}$.

Proof. As we discussed above, sets \mathcal{O}_v^* in Line 5 can be found in polynomial time. At every iteration, if the algorithm has not terminated yet, at least one of the sets \mathcal{Y}_v strictly shrinks, as can be seen by comparing termination condition in Line 6 with pruning in Line 8. Therefore the algorithm terminates in at most $\sum_v (|\mathcal{X}_v| - 1)$ iterations. On termination, $p \in \mathbb{S}_f$ by Corollary 4.4. \square

Theorem 4.6. Substitution p returned by Algorithm 1 is the maximum of $\mathbb{S}_f \cap \mathcal{P}^{2,y}$ and thus it solves (6). Proof in § A.

It is noteworthy that Algorithm 1 can be used to solve problem (6) with *any* polytope Λ satisfying $\mathcal{M} \subseteq \Lambda$, i.e., with any LP relaxation of (1) that can be expressed in the lifted space $\mathbb{R}^{\mathcal{I}}$. Moreover, in order to use the algorithm with higher order models one needs merely to (straightforwardly) generalize the linear extension (10) as done in [37].

The test labeling y can itself be chosen using the approximate solution of the LP-relaxation, e.g., via the zeroth iteration of the algorithm with $g = f$ and picking y_v from \mathcal{O}_v^* . This choice is motivated by the fact that a strict relaxed-improving substitution cannot eliminate the labels from the support set of optimal solutions of the LP relaxation [35], and thus these labels may not be substituted with anything else.

4.3 Comparison to Previous Work

Substitutions in $\mathcal{P}^{2,y}$ are related to the expansion move algorithm [6] in the following sense. While [6] seeks to improve a single current labeling x by calculating an optimized crossover (fusion) with a candidate labeling y , we seek which labels can be moved with a guaranteed improvement to y for all possible labelings x .

Algorithm 1 is similar in structure to [43]. The later finds an improving substitution in a small class $\mathcal{P}^{1,y}$ by incrementally shrinking the set of potentially persistent variables. More

specifically, given a test labeling $y \in \mathcal{X}$, the *all-to-one* class of substitutions $\mathcal{P}^{1,y}$ contains substitutions p , which in every node v either replace *all* labels with y_v or leaves all labels unchanged. There are only two possible choices for p_v : either $i \mapsto y_v$ for all $i \in \mathcal{X}_v$ or the identity $i \mapsto i$. Methods [24, 42, 43] can be explained as finding an improving mapping in this class [35]. We generalize the method [43] to substitutions in $\mathcal{P}^{2,y}$. The original sufficient condition for persistency in [43] does not extend to such substitutions. Even for substitutions in $\mathcal{P}^{1,y}$ it is generally weaker than condition (18) unless a special reparametrization is applied [44]. Criterion (18) extends to general substitutions and does not depend on reparametrization. Similarly to [43], we will use approximate dual solvers in this more general setting.

In [35, (ε -L1)] problem (6) is formulated as one big linear program. We solve problem (6), and hence also the LP problem [35] in a more combinatorial fashion w.r.t. to the variables defining the substitution.

It may seem that solving a series of linear programs rather than a single one is a disadvantage of the proposed approach. However, as we show further, the proposed iterative algorithm can be implemented using a dual, possibly suboptimal, solver for the relaxed verification problem (18). This turns out to be much more beneficial in practice since the verification problems can be incrementally updated from iteration to iteration and solved overall very efficiently. This approach achieves scalability by exploiting available specialized approximate solvers for the relaxed MAP inference. Essentially, any dual (approximate) solver can be used as a black box in our method.

5 PERSISTENCY WITH DUAL SOLVERS

Though Algorithm 1 is quite general, its practical use is limited by the strict requirements on the solver, which must be able to determine the exact support set of all optimal solutions. However, finding even a single solution of the relaxed problem with standard methods like simplex or interior point can be practically infeasible and one has to switch to specialized solvers developed for this problem. Although there are scalable algorithms based on smoothing techniques [31, 32], which converge to an optimal solution, waiting until convergence in *each* iteration of Algorithm 1 can make the whole procedure impractical. In general, we would like to avoid restricting ourselves to certain selected solvers to be able to choose the most efficient one for a given problem. In the standard LP relaxation (introduced below in (20)), the number of primal variables grows quadratically with the number of labels, while the number of dual variables grows only linearly. It is therefore desirable to use solvers working in the dual domain, including suboptimal ones, (e.g. [20, 32, 12, 33], performing block-coordinate descent) as they offer the most performance for a limited time budget. Furthermore, fast parallel versions of such methods have been developed to run on GPU/FPGA [38, 7, 16], making the LP approach feasible for more vision applications.

We will switch to the dual verification LP and gradually relax our requirements on the solution returned by a dual solver. This is done in the following steps:

1. an optimal dual solution;
2. an arc consistent dual point;
3. any dual point.

Our main objective is to ensure in each of these cases that the found substitution p is strictly improving, while possibly compromising its maximality. The final practical algorithm operating in

the mode 3 relies on the persistency problem reduction introduced in § 6. Intermediate steps 1 and 2 are considered right after defining the standard LP relaxation and its dual.

5.1 LP Relaxation

We consider the standard local polytope relaxation [41, 49, 47] of the energy minimization problem (1) given by the following primal-dual pair:

$$\begin{aligned}
 \text{(primal)} \quad & \min \langle f, \mu \rangle \\
 & \sum_j \mu_{uv}(i, j) = \mu_u(i), \\
 & \sum_i \mu_{uv}(i, j) = \mu_v(j), \\
 & \sum_i \mu_u(i) = \mu_\emptyset, \\
 & \mu_u(i) \geq 0, \\
 & \mu_{uv}(i, j) \geq 0, \\
 & \mu_\emptyset = 1
 \end{aligned}
 \quad = \quad
 \begin{aligned}
 \text{(dual)} \quad & \max f_\emptyset^\varphi \\
 & \varphi_{uv}(i) \in \mathbb{R}, \\
 & \varphi_{vu}(j) \in \mathbb{R}, \\
 & \varphi_u \in \mathbb{R}, \\
 & f_u^\varphi(i) \geq 0, \\
 & f_{uv}^\varphi(i, j) \geq 0,
 \end{aligned}
 \quad (20)$$

where f^φ abbreviates

$$f_u^\varphi(i) = f_u(i) + \sum_{v \in \mathcal{N}(u)} \varphi_{uv}(i) - \varphi_u, \quad (21a)$$

$$f_{uv}^\varphi(i, j) = f_{uv}(i, j) - \varphi_{uv}(i) - \varphi_{vu}(j), \quad (21b)$$

$$f_\emptyset^\varphi = f_\emptyset + \sum_u \varphi_u. \quad (21c)$$

The constraints of the primal problem (20) define the *local polytope* Λ . The cost vector f^φ is called a *reparametrization* of f . There holds cost equivalence: $\langle f^\varphi, \mu \rangle = \langle f, \mu \rangle$ for all $\mu \in \Lambda$ (as well as $E_f = E_{f^\varphi}$), see [49]. Using the reparametrization, the dual problem (20) can be briefly expressed as

$$\max_{\varphi} f_\emptyset^\varphi \quad \text{s.t.} \quad (\forall \omega \in \mathcal{V} \cup \mathcal{E}) f_\omega^\varphi \geq 0. \quad (22)$$

Note that for a feasible φ the value f_\emptyset^φ is a lower bound on the primal problem (20). In what follows we will assume that φ in (22) additionally satisfies the following *normalization*: $\min_i f_u^\varphi(i) = 0$ and $\min_{ij} f_{uv}^\varphi(i, j) = 0$ for all u, v , which is automatically satisfied for any optimal solution.

5.2 Expressing \mathcal{O}_v^* in the Dual Domain

Let (μ, φ) be a pair of primal and dual optimal solutions to (20). From complementary slackness we know that if $\mu_v(i) > 0$ then the respective dual constraint holds with equality:

$$\mu_v(i) > 0 \Rightarrow f_v^\varphi(i) = 0, \quad (23)$$

in this case we say that $f_v^\varphi(i)$ is *active*. The set of such active dual constraints matches the sets of local minimizers of the reparametrized problem,

$$\mathcal{O}_v(\varphi) := \{i \in \mathcal{X}_v \mid f_v^\varphi(i) = 0\} = \underset{i}{\operatorname{argmin}} f_v^\varphi(i). \quad (24)$$

From complementary slackness (23) we obtain that

$$\mathcal{O}_v^* \subseteq \mathcal{O}_v(\varphi). \quad (25)$$

This inclusion is insufficient for an exact reformulation of Algorithm 1, however it is sufficient for correctness if we make sure that $\mathcal{Y}_v \cap \mathcal{O}_v(\varphi) = \emptyset$ on termination, i.e., that the substitution p does not displace labels in $\mathcal{O}_v(\varphi)$. Then, by Corollary 4.4, $p \in \mathbb{S}_f$ follows.

There always exists an optimal primal solution μ and dual φ satisfying *strict complementarity* [46], in which case relation (23) becomes an equivalence:

$$\mu_v(i) > 0 \Leftrightarrow f_v^\varphi(i) = 0. \quad (26)$$

Algorithm 2: Iterative Pruning Arc Consistency

Input: Cost vector $f \in \mathbb{R}^{\mathcal{I}}$, test labeling $y \in \mathcal{X}$;
Output: Strictly improving substitution p ;
1 $(\forall v \in \mathcal{V}) \mathcal{Y}_v := \mathcal{X}_v \setminus \{y_v\}$;
2 **while true**
3 Construct verification problem $g := (I - [p])^\top f$ with p defined by (7);
4 Use dual solver to find φ such that g^φ is arc consistent;
5 $\mathcal{O}_v(\varphi) := \{i \in \mathcal{X}_v \mid g_v^\varphi(i) = 0\}$;
6 **if** $(\forall v \in \mathcal{V}) \mathcal{O}_v(\varphi) \cap \mathcal{Y}_v = \emptyset$ **then return** p ;
7 **for** $v \in \mathcal{V}$ **do**
8 Pruning of substitutions: $\mathcal{Y}_v := \mathcal{Y}_v \setminus \mathcal{O}_v(\varphi)$;

It is the case when both μ and φ are relative interior points of the optimal primal, resp. optimal dual, faces [46]. For a relative interior optimal φ , the set of constraints that are satisfied as equalities $f^\varphi(i) = 0$ is the smallest and does not depend on the specific choice of such φ . Under strict complementarity, (25) turns into equality $\mathcal{O}_v^* = \mathcal{O}_v(\varphi)$, which allows to compute the exact maximum persistency using a dual algorithm without reconstructing a primal solution. However finding such φ appears more difficult: e.g. the most efficient dual block-coordinate ascent solvers [20,12, 8, 33] only have convergence guarantees (see [20, 33]) allowing to find a sub-optimal solution, satisfying arc consistency.

Definition 5.1 ([49]). A reparametrized problem f^φ is called *arc consistent* if: (i) for all $uv \in \mathcal{E}$ from $f_{uv}^\varphi(i, j)$ being active follows that $f_u^\varphi(i)$ and $f_v^\varphi(j)$ are active; (ii) for all $u \in \mathcal{V}$ from $f_u^\varphi(i)$ active follows that for all $v \in \mathcal{N}(u)$ there exists a $j \in \mathcal{X}_v$ such that $f_{uv}^\varphi(i, j)$ is active.

An optimal dual solution need not be arc consistent, but it can be reparametrized without loss of optimality to enforce arc consistency [49]. Labels that become inactive during this procedure are not in the support set of primal solutions. In general the following holds.

Proposition 5.2. Arc consistency is a necessary condition for relative interior optimality: if $\mathcal{O}_v(\varphi) = \mathcal{O}_v^*$ for all $v \in \mathcal{V}$ then f^φ is arc consistent. Proof in § A.

This property is in our favor, since we are ideally interested in the equality $\mathcal{O}_v(\varphi) = \mathcal{O}_v^*$. Next, we propose an algorithm utilizing an arc consistent solver and prove that it is guaranteed to output $p \in \mathbb{S}_f$.

5.3 Persistency with an Arc Consistency Solver

We propose Algorithm 2 which is based on a dual solver attaining the arc consistency condition (differences to Algorithm 1 underlined). If the dual solver (in line 4) finds a relative interior optimal solution, Algorithm 2 solves (6) exactly. Otherwise it is suboptimal and we need to reestablish its correctness and termination.

Lemma 5.3 (Termination of Algorithm 2). Algorithm 2 terminates in at most $\sum_v (|\mathcal{X}_v| - 1)$ iterations.

Proof. In case the return condition in line 6 is not satisfied, $\mathcal{O}_v \cap \mathcal{Y}_v \neq \emptyset$ for some v and the pruning in line 8 excludes at least one label from \mathcal{Y}_v . □

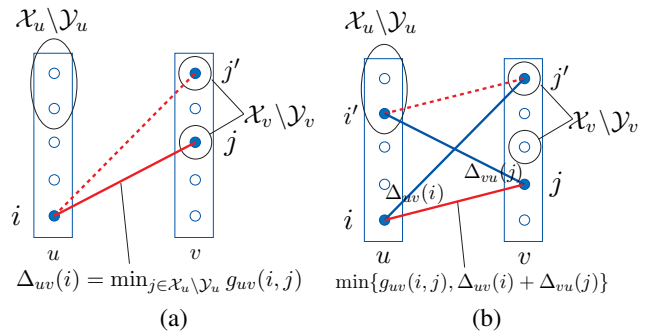


Fig. 5: Illustration for the reduction. Labels $\mathcal{X}_u \setminus \mathcal{Y}_u$ are not displaced by p hence their associated unary and pairwise costs are zero in $g = (I - [p])^\top f$. In case (a) the indicated pairwise costs are replaced with their minimum. In case (b) the value of $g_{uv}(i, j)$ can be decreased, assuming all reductions of type (a) and their symmetric counterparts are already performed. The amount of decrease matches the value of the mixed derivative (non-submodularity) associated to i, i' paired with j, j' .

Lemma 5.4 (Correctness of Algorithm 2). If $(\forall v \in \mathcal{V}) \mathcal{O}_v(\varphi) \cap \mathcal{Y}_v = \emptyset$ holds for an arc consistent dual vector φ , then φ is optimal. Proof in § A.

It follows that when Algorithm 2 terminates, the found arc consistent solution φ is optimal, in which case inclusion (25) is satisfied and the found substitution p is guaranteed to be in \mathbb{S}_f .

5.4 Solvers Converging to Arc Consistency

One can see that arc consistency is only required on termination of Algorithm 2. In the intermediate iterations we may as well perform the pruning step, line 8, without waiting for the solver to converge. This motivates the following practical strategy:

- Perform a number of iteration towards finding an arc-consistent dual point φ ;
- Check whether there are some labels to prune, *i.e.*, $(\exists u) \mathcal{O}_u(\varphi) \cap \mathcal{Y}_u \neq \emptyset$;
- Terminate if φ is arc consistent and there is nothing to prune; otherwise, perform more iterations towards arc consistency.

If the solver is guaranteed to eventually find an arc consistent solution, the overall algorithm will either terminate with an arc consistent and (by Lemma 5.4) optimal φ or there will be some labels to prune. However, we have to face the question what happens if the dual solver does not find an arc consistent solution in finite time. In this case the algorithm can be iterating infinitely with no pruning available. At the same time there is no guarantee that a pruning step will not occur at some point and thus if we simply terminate the algorithm we get no persistency guarantees. Even if the dual solver was guaranteed to converge in a finite number of iterations, it is in principle possible that the time needed for a pruning to succeed would be proportional to the time of convergence, making the whole algorithm very slow. Instead, it is desirable to guarantee a valid result while allowing only a fixed time budget for the dual solver. We will overcome this difficulty with the help of the reduced verification LP presented next.

6 VERIFICATION PROBLEM REDUCTION

Algorithms 1 and 2 iteratively solve verification problems. We can replace the verification LP solved in step 4 by a simpler, reduced one, *without loss of optimality* of the algorithms.

Definition 6.1. Let $g := (I - [p])^\top f$ be the cost vector of the verification LP. The *reduced cost vector* \bar{g} is defined as

$$\bar{g}_v(i) := g_v(i), \quad v \in \mathcal{V}; \quad \bar{g}_\emptyset = 0; \quad (27a)$$

$$\bar{g}_{uv}(i, j) := \begin{cases} 0, & i \notin \mathcal{Y}_u, j \notin \mathcal{Y}_v, \\ \Delta_{vu}(j) := \min_{i' \notin \mathcal{Y}_u} g_{uv}(i', j), & i \notin \mathcal{Y}_u, j \in \mathcal{Y}_v, \\ \Delta_{uv}(i) := \min_{j' \notin \mathcal{Y}_v} g_{uv}(i, j'), & i \in \mathcal{Y}_u, j \notin \mathcal{Y}_v, \\ \min\{\Delta_{vu}(j) + \Delta_{uv}(i), g_{uv}(i, j)\}, & i \in \mathcal{Y}_u, j \in \mathcal{Y}_v. \end{cases} \quad (27b)$$

The reduction is illustrated in Fig. 5. Taking into account that $g_{uv}(i', j') = 0$ for $i' \in \mathcal{X}_u \setminus \mathcal{Y}_u, j' \in \mathcal{X}_v \setminus \mathcal{Y}_v$, the reduction can be interpreted as forcing the inequality

$$g_{uv}(i, j') + g_{uv}(i', j) - g_{uv}(i, j) - g_{uv}(i', j') \geq 0, \quad (28)$$

i.e., the non-negativity of mixed discrete derivatives, for all four-tuples $i \in \mathcal{Y}_u, j \in \mathcal{Y}_v, i' \notin \mathcal{Y}_u, j' \notin \mathcal{Y}_v$. The cost vector \bar{g} is therefore a *partial submodular truncation* of g .

Recall that Algorithm 1 on each iterations prunes all substitutions $q \leq p$ that do not belong to \mathbb{S}_f based on the solutions of the verification LP. The following theorem reestablishes optimality of this step with the above reduction.

Theorem 6.2 (Reduction). Let $p \in \mathcal{P}^{2,y}$ and \bar{g} be the corresponding reduced cost vector constructed as in Def. 6.1. Let also $q \in \mathcal{P}^{2,y}, q \leq p$. Then $q \in \mathbb{S}_f$ iff $q \in \mathbb{S}_{\bar{g}}$. Proof in § A.

From Theorem 6.2 and Corollary 4.4 it follows that $q \in \mathbb{S}_f$ iff $q_u(\mathcal{O}_u^*) = \mathcal{O}_u^*$, where \mathcal{O}_u^* are the support sets of optimal solutions to the *reduced verification LP*,

$$\operatorname{argmin}_{\mu \in \Lambda} \langle \bar{g}, \mu \rangle. \quad (29)$$

Therefore it is valid for algorithms 1 and 2 to consider this reduced LP and prune all substitutions q that do not satisfy the property $q_v(\mathcal{O}_v^*) = \mathcal{O}_v^*$. The optimal relaxed solutions and their support sets can in general differ from those of the original verification LP, however for the purpose of the algorithm it is an equivalent replacement potentially affecting only the order in which substitutions are pruned.

The reduction has the following advantages:

- subsets of labels $\mathcal{X}_v \setminus \mathcal{Y}_v$ can be contracted to a single representative label y_v , because associated unary and pairwise costs are equal;
- It will allow (see § 7) to relax the requirements on approximate dual solvers needed to establish termination and correctness of the algorithm.
- It is useful for the speed up heuristics (§ 8). In particular, it is easier to find a labeling with a negative cost since we have decreased many edge costs. It will be shown that such a labeling allows for an early stopping of the dual solver and a pruning of substitution without loss of maximality.

7 PERSISTENCY WITH A FINITE NUMBER OF DUAL UPDATES

We assume that a suboptimal dual solver is iterative and can be represented by a procedure `dual_update`, which given a current dual point φ makes a step resulting in a new dual point and a guess of a primal integer solution x .

In this setting we propose Algorithm 4. In its inner loop, the algorithm calls `dual_update` (line 7) checks whether a speed-up shortcut is available (line 8) and verifies whether it can already

Procedure 3: dual correct (φ, \bar{g})

```

1 for  $uv \in \mathcal{E}$  do
2    $(\forall i \in \mathcal{X}_u) \varphi_{uv}(i) := \varphi_{uv}(i) + \min_{ij} \bar{g}_{uv}^\varphi(i, j);$ 
3    $(\forall i \in \mathcal{X}_u) \varphi_{uv}(i) := \varphi_{uv}(i) + \min_j \bar{g}_{uv}^\varphi(i, j);$ 
4    $(\forall j \in \mathcal{X}_v) \varphi_{vu}(j) := \varphi_{vu}(j) + \min_i \bar{g}_{uv}^\varphi(i, j);$ 
5  $(\forall u \in \mathcal{V}) \varphi_u := \varphi_u + \min_i g_u^\varphi(i);$  /* Normalize */
6 return  $\varphi;$ 

```

terminate (lines 11-13). If neither occurs in a certain number of iterations (stopping condition in line 14), the pruning based on the currently active labels is executed (line 15). After that the cost vector \bar{g} is rebuilt, but the dual solver continues from the last found dual point (warm start).

The speed-ups will be explained in the next section, they are not critical for the overall correctness. Now we focus on the new termination conditions (lines 11-13). A correction step (line 11) is introduced whose purpose is to move the slacks from pairwise terms to unary terms so that active labels become more decisive. This procedure is defined in Procedure 3. The correction is not intermixed with dual updates but serves as a proxy between the solver and the termination conditions. It has the following property.

Lemma 7.1. Output φ of Procedure 3 is feasible and satisfies

$$(\forall u \in \mathcal{V}) \min_{i \in \mathcal{X}_u} g_u^\varphi(i) = 0, \quad (30)$$

$$(\forall uv \in \mathcal{E}, ij \in \mathcal{X}_{uv}) \min_{i' \in \mathcal{X}_u} \bar{g}_{uv}^\varphi(i', j) = \min_{j' \in \mathcal{X}_v} \bar{g}_{uv}^\varphi(i, j') = 0. \quad (31)$$

Moreover, if the input φ is feasible, the lower bound f_\emptyset^φ does not decrease.

Proof. Line 2 of Procedure 3 moves a constant from an edge to node. This turns the minimum of terms $g_{uv}^\varphi(i, j)$ to zero. Lines 3 and 4 turn to zero the minimal pairwise value attached to each label, which provides (31). Line 5 provides (30). In case of feasibility of the initial φ , which implies $g^\varphi \geq 0$, all values of φ can only increase during steps 2-4 and hence the unary potentials g_u^φ remain non-negative. Therefore step 5 can not decrease the lower bound value f_\emptyset^φ . \square

According to Lemma 7.1 Procedure 3 can not worsen the lower bound attained by a dual solver. The following theorem guarantees that when no further pruning is possible, the corrected dual point constitutes an optimal solution, ensuring persistency.

Theorem 7.2. Let φ be a dual point for reduced problem \bar{g} satisfying (30)-(31). Then either

1. $g_\emptyset^\varphi = 0$, φ is dual optimal and $\delta(y)$ is primal optimal, or
2. $(\exists u \in \mathcal{V}) \mathcal{O}_u(\varphi) \cap \mathcal{Y}_u \neq \emptyset$.

Proof. Assume (b) does not hold: $(\forall u \in \mathcal{V}) \mathcal{O}_u(\varphi) \subseteq \mathcal{X}_u \setminus \mathcal{Y}_u$. Let us pick in each node u a label $z_u \in \mathcal{O}_u(\varphi)$. As ensured by (31), for each edge uv there is a label $j \in \mathcal{X}_v$ such that $\bar{g}_{uv}^\varphi(z_u, j) = 0$ and similarly, there exists $i \in \mathcal{X}_u$ such that $\bar{g}_{uv}^\varphi(i, z_v) = 0$. By partial submodularity of \bar{g} , we have

$$\bar{g}_{uv}^\varphi(z_u, z_v) + \bar{g}_{uv}^\varphi(i, j) \leq \bar{g}_{uv}^\varphi(z_u, j) + \bar{g}_{uv}^\varphi(i, z_v) = 0. \quad (32)$$

Therefore, $\bar{g}_{uv}^\varphi(z_u, z_v) \leq -\bar{g}_{uv}^\varphi(i, j) \leq 0$. Hence $\bar{g}_{uv}^\varphi(z_u, z_v) = 0$ and it is active. Therefore $\delta(z)$ and dual point φ satisfy complementarity slackness conditions and hence they are primal-dual optimal and $g_\emptyset^\varphi = E_{\bar{g}}(z) = 0 = E_{\bar{g}}(y)$. \square

Algorithm 4: Efficient Iterative Pruning

Input: Problem $f \in \mathbb{R}^{\mathcal{I}}$, test labeling $y \in \mathcal{X}$;
Output: Improving substitution $p \in \mathcal{P}^{2,y} \cap \mathbb{S}_f$;

- 1 $(\forall u \in \mathcal{V}) \mathcal{Y}_u := \mathcal{X}_u \setminus \{y_u\}$;
- 2 Set φ to the initial dual solution if available;
- 3 **while true**
- 4 Apply single node pruning ; /* speed-up */
- 5 Construct reduced verification LP \bar{g} from f and current sets \mathcal{Y}_u , according to Definition 6.1;
- 6 **repeat**
- 7 $(\varphi, x) := \text{dual_update}(\bar{g}, \varphi)$;
- 8 **if** $E_{\bar{g}}(x) < 0$ **then** /* speed-up */
- 9 Apply pruning cut with x ;
- 10 **goto** step 4 to rebuild \bar{g} ;
- /* Verification of Optimality */
- 11 $\varphi' := \text{dual_correct}(\bar{g}, \varphi)$;
- 12 $\mathcal{O}_u := \{i \mid \bar{g}'_u(i) = 0\}$;
- 13 **if** $(\forall u \in \mathcal{V}) \mathcal{O}_u \cap \mathcal{Y}_u = \emptyset$ **then return** p by (7);
- 14 **until** any stopping condition (e.g., iteration limit);
- 15 Prune: $(\forall u \in \mathcal{V}) \mathcal{Y}_u := \mathcal{Y}_u \setminus \mathcal{O}_u$;
- 16 **Procedure** $\text{dual_update}(\bar{g}, \varphi)$
- Input:** Cost vector \bar{g} , dual point φ ;
- Output:** New dual point φ , approximate primal integer solution x ;

Theorem 7.3 (Termination and Correctness of Algorithm 4). For any stopping condition in line 14, Algorithm 4 terminates in at most $\sum_v (|\mathcal{X}_v| - 1)$ outer iterations and returns $p \in \mathbb{S}_f$.

Proof. When the algorithm has not yet terminated some further pruning is guaranteed to be possible (compare conditions in lines 13 and 15). The iteration limit follows. When Algorithm 4 terminates, from Theorem 7.2 it follows that φ' is dual optimal and hence $\mathcal{O}_u(\varphi') \supseteq \mathcal{O}_u^*$. Therefore, $(\forall u \in \mathcal{V}) \mathcal{O}_u^* \cap \mathcal{Y}_u = \emptyset$, which is sufficient for p to be strictly Λ -improving according to Corollary 4.4. \square

In [40] we prove that a similar result holds for a TRW-S iteration without correction by arguing on complete chain subproblems instead of individual nodes. The correction might be needed in case the algorithm does not keep slacks on the nodes, e.g. for SRMP [22].

The stopping condition in line 14 of Algorithm 4 controls the aggressiveness of pruning. Performing fewer iterations may result only in the found p not being the maximum, but in any case it is guaranteed that the Algorithm 4 does not stall and identifies a *correct persistency*. In the case when the solver does have convergence and optimality guarantees, the time budget controls the degree of approximation to the maximum persistency.

8 SPEED-UPS

8.1 Inference Termination Without Loss of Maximality

Next, we propose several sufficient conditions to quickly prune some substitutions without worsening the final solution found by the algorithm. As follows from Definition 3.2, existence of a labeling x such that $\langle (I - [p])^\top f, \delta(x) \rangle \leq 0$ and $x \neq p(x)$ is sufficient to prove that substitution p is *not* strictly Λ -improving. Hence one could consider updating the current substitution p without waiting for an exact solution of the inference problem in line 4. The tricky part is to find labels that can be pruned without

loss of optimality of the algorithm. Lemma 8.1 below suggests to solve a simpler verification LP, $\min_{\mu \in \Lambda'} \langle \bar{g}, \mu \rangle$ over a subset Λ' of Λ . This does not guarantee to remove all non-improving substitutions (which implies one has to switch to Λ afterwards), but can be much more efficient than the optimization over Λ . After the lemma we provide two examples of such efficient procedures.

Lemma 8.1. Let $p \in \mathcal{P}^{2,y}$ and \bar{g} be defined by (27) (depends on p). Let $q \in \mathbb{S}_f \cap \mathcal{P}^{2,y}$, $q \leq p$, $Q = [q]$. Let $\Lambda' \subseteq \Lambda$, $Q(\Lambda') \subseteq \Lambda'$ and $\mathcal{O}^* = \text{argmin}_{\mu \in \Lambda'} \langle \bar{g}, \mu \rangle$. Then $(\forall v \in \mathcal{V}) q_v(\mathcal{O}_v^*) = \mathcal{O}_v^*$. Proof in § A.

Note, while Theorem 6.2 is necessary and sufficient for pruning, Lemma 8.1 is only sufficient.

8.2 Pruning of Negative Labelings

Assume we found an integer labeling x such that $E_{\bar{g}}(x) \leq 0$ and $p(x) \neq x$. Lemma 8.1 gives an answer, for which nodes v the label x_v can be pruned from the set \mathcal{Y}_v without loss of optimality. Define the following restriction of the polytope Λ :

$$\Lambda_x = \{\mu \in \Lambda \mid (\forall v \in \mathcal{V}) \mu(y_v) + \mu(x_v) = 1\} \subseteq \Lambda. \quad (33)$$

Polytope Λ_x corresponds to the restriction of Λ to the label set $\{y_v, x_v\}$ in each node $v \in \mathcal{V}$. According to Lemma 8.1 we need to solve the problem

$$\mathcal{O}^* := \text{argmin}_{\mu \in \Lambda_x} \langle \bar{g}, \mu \rangle \quad (34)$$

and exclude x_v from \mathcal{Y}_v if $x_v \in \mathcal{O}_v^*$. Due to the partial submodularity of \bar{g} the problem (34) is submodular and can be solved by min-cut/max-flow algorithms [23]. Because x was found to have non-positive energy, it is necessarily that for some nodes v there will hold $x_v \in \mathcal{O}_v^* \cap \mathcal{Y}_v$ and therefore some pruning will take place.

8.3 Single Node Pruning

Let us consider “a single node” polytope $\Lambda_{u,i} := \{\mu \in \Lambda \mid \mu_u(y_u) + \mu_u(i) = 1; (\forall v \neq u) \mu_v(y_v) = 1\}$. It is a special case of Λ_x when y and x differ in a single node u only and $x_u = i$. In this case problem (34) amounts to calculating $\bar{g}_u(x_u) + \sum_{v \in \mathcal{N}(u)} \bar{g}_{uv}(x_u, y_v)$. If this value is non-positive, x_u must be excluded from \mathcal{Y}_u . The single node pruning can be applied to all pairs (u, i) exhaustively, but it is more efficient to keep track of the nodes for which sets \mathcal{Y}_v have changed (either due to a negative labeling pruning, active labels pruning in line 15 or the single node pruning itself) and check their neighbors.

8.4 Efficient Message Passing

The main computational element in dual coordinate ascent solvers like TRWS or MPLP is *passing a message*, i.e., an update of the form $\min_{i \in \mathcal{X}_u} (f_{uv}(i, j) + a(i))$. In many practical cases the message passing for f can be computed in time linear in the number of labels [15, 28, 3]. This is the case when f_{uv} is a convex function of $i - j$ (e.g., $|i - j|$, $(i - j)^2$) or a minimum of few such functions (e.g. Potts model is $\min(1, |i - j|)$). However, in Algorithm 1 we need to solve the problem with the cost vector $g = (I - P^\top)f$, resp. \bar{g} (27) if we apply the reduction. It turns out that whenever there is a fast message passing method for f , the same holds for \bar{g} .

Our-CPLEX	Our Algorithm 1 (Iterative Relaxed Inference) using CPLEX [17].
Our-TRWS	Our Algorithm 4 using TRW-S [20]. Initial solution uses at most 1000 iterations (or the method has converged). All speedups.
[43]-CPLEX	Method [43] with CLPEX.
[43]-TRWS	Method [43] with TRW-S.
[44]-TRWS	Improved version of the above with optimal reparametrization [44].
ϵ -L1 [35]	Single LP formulation of the maximum strong persistency [35] solved with CPLEX.
Kovtun [24]	One-against-all method of Kovtun [25].
MQPBO	Multilabel QPBO [19].
MQPBO-10	MQPBO with 10 random permutations, accumulating persistency.

TABLE 1: List of Evaluated Methods

Theorem 8.2 (Fast message passing). Message passing for an edge term \bar{g}_{uv} (27) can be reduced to that for f_{uv} in time $O(|\mathcal{Y}_u| + |\mathcal{Y}_v|)$. Proof in § A.

This complexity is proportional to the size of the sets \mathcal{Y}_u . The more labels are pruned from sets \mathcal{Y}_u in the course of the algorithm, the less work is required.

Note, that contrary to limiting the number of iterations of a dual solver, described in § 7, the speedups presented in this section do *not* sacrifice the persistence maximality (6). In our experiments for some instances, Algorithm 4 finished before ever reaching step 15. In such cases the found substitution $p \in \mathbb{S}_f$ is the maximum.

9 EXPERIMENTAL EVALUATION¹

In the experiments we study how well we approximate the maximum persistency [35], Table 2; illustrate the contribution of different speedups, Table 4; give an overall performance comparison to a larger set of relevant methods, Table 3; and provide a more detailed direct comparison to the most relevant scalable method [43] using exact and approximate LP solvers, Table 5. As a measure of persistency we use the percentage of labels eliminated by the improving substitution p :

$$\frac{\sum_{v \in \mathcal{V}} |\mathcal{X}_v| p_v(\mathcal{X}_v)}{\sum_{v \in \mathcal{V}} (|\mathcal{X}_v| - 1)} = \frac{\sum_{v \in \mathcal{V}} |\mathcal{Y}_v|}{\sum_{v \in \mathcal{V}} (|\mathcal{X}_v| - 1)}. \quad (35)$$

9.1 Random Instances

Table 2 gives comparison to [43] and [35] on random instances generated as in [35] (small problems on 4-connected grid with uniformly distributed integer potentials for “full” model and of the Potts type for “Potts” model, all not LP-tight). It can be seen that our exact Algorithm 1 performs identically to the ϵ -L1 formulation [35]. Although it solves a series of LPs, as opposed to a single LP solved by ϵ -L1, it scales better to larger instances. Instances of size 20x20 in the ϵ -L1 formulation are already too difficult for CPLEX: it takes excessive time and sometimes returns a computational error. The performance of the dual Algorithm 4 confirms that we loose very little in terms of persistency but gain significantly in speed.

¹The implementation of our method is available at <http://cmp.felk.cvut.cz/~shekhovt/persistency>

9.2 Benchmark Problems

Table 3 summarizes average performance on the OpenGM MRF benchmark [18]. The datasets include previous benchmark instances from computer vision [45] and protein structure prediction [29, 51] as well as other models from the literature. Results per instance are given in § B.

9.3 Speedups

In this experiment we report how much speed improvement was achieved with each subsequent technique of § 8. The evaluation in Table 4 starts with a basic implementation (using only a warm start). The solver is allowed to run at most 50 iterations in the partial optimality phase until pruning is attempted. We expect that on most datasets the percentage of persistent labels improves when we apply the speedups (since they are without loss of maximality).

9.4 Discussion

Tables 2 and 5 demonstrate that Our-TRWS, which is using a suboptimal dual solver, closely approximates the maximum persistency [35]. Our method is also significantly faster and scales much better. The method [43] is the closest contender to ours in terms of algorithm design. Tables 2, 3 and 5 clearly show that our method determines a larger set of persistent variables. This holds true with exact (CPLEX) as well as approximate (TRWS) solvers. There are two reasons for that as discussed in § 4.3. First, we optimize over a larger set of substitutions than [43], *i.e.*, we identify per-label persistencies while [43] is limited to the whole-variable persistencies. Second, even in the case of the whole-variable persistencies the criterion in [43] is in general weaker than (18) and depends on the initial reparametrization of the problem. This later difference does not matter for Potts models [44], the examples Figs. 6 and 8, but does matter, *e.g.*, in Fig. 9. Although our method searches over a significantly larger space of possible substitutions, it needs fewer TRW-S iterations due to speedup techniques. Details on iteration counts can be found in § B. In the comparison of running time it should be taken into account that different methods are optimized to a different degree. Nevertheless, it is clear that the algorithmic speedups were crucial in making the proposed method much more practical than [43] and [35] while maintaining high persistency recall quality.

To provide more insights to the numbers reported, we illustrate in Figs. 1 and 6 to 9 some interesting cases. Fig. 6 shows “the hardest” instance of color-seg-n4 family. Identified persistencies allow to fix a single label in most of the pixels, but for some pixels more than one possible label remains. The *remainder* of the problem has the reduced search space $p(\mathcal{X})$, which can be passed to further solvers. The tsukuba image Fig. 1 is interesting because it has appeared in many previous works. The performance of graph-cut based persistency methods relies very much on strong unary costs, while the proposed method is more robust. Fig. 7 shows an easy example from object-seg, where LP relaxation is tight, the dual solver finds the optimal labeling y and our verification LP confirms that this solution is unique. In Fig. 8 we show a hard instance of mrf-stereo. Partial reason for its hardness is integer costs, leading to non-uniqueness of the optimal solution. In Fig. 9, photomontage/pano instance, we report 79% solution completeness, but most of these 79% correspond to trivial forbidden labels in the problem (very big unary costs). At the same time other methods perform even worse. This problem

Problem family	[43]-CPLEX	[43]-TRWS	ϵ -L1 [35]	Our-CPLEX	Our-TRWS
10x10 Potts-3	0.18s 58.46%	0.05s 58.38%	0.05s 72.27%	0.18s 72.27%	0.04s 72.21%
10x10 full-3	0.24s 2.64%	0.09s 1.22%	0.06s 62.90%	0.24s 62.90%	0.05s 62.57%
20x20 Potts-3	3.25s 73.95%	0.21s 68.49%	0.87s 87.38%	2.43s 87.38%	0.06s 87.38%
20x20 full-3	2.81s 0.83%	0.37s 0.83%	0.95s 72.66%	3.03s 72.66%	0.07s 72.31%
20x20 Potts-4	12.45s 23.62%	0.39s 18.43%	19.40s 74.28%	8.56s 74.28%	0.08s 73.63%
20x20 full-4	3.96s 0.01%	0.39s 0.01%	21.08s 6.58%	12.41s 6.58%	0.08s 6.58%

TABLE 2: Performance evaluation on random instances of [35]. For each problem family (size, type of potentials and number of labels) average performance over 100 samples is given. To allow for precise comparison all methods are initialized with the same test labeling y found by LP relaxation. Our-TRWS closely approximates Our-CPLEX, which matches ϵ -L1 [35], and scales much better.

Problem family	#I	#L	#V	MQPBO	MQPBO-10	Kovtun [24]	[43]-TRWS	Our-TRWS
mrf-stereo	3	16-60	> 100000	†	†	1.0s 0.23%	2.5h 13%	117s 73.56%
mrf-photomontage	2	5-7	≤ 514080	93s 22%	866s 16%	0.4s 15.94%	3.7h 16%	483s 41.98%
color-seg	3	3-4	≤ 424720	22s 11%	87s 16%	0.3s 98%	1.3h 99.88%	61.8s 99.95%
color-seg-n4	9	3-12	≤ 86400	22s 8%	398s 14%	0.2s 67%	321s 90%	4.9s 99.26%
ProteinFolding	21	≤ 483	≤ 1972	685s 2%	2705s 2%	0.02s 4.56%	48s 18%	9.2s 55.70%
object-seg	5	4-8	68160	3.2s 0.01%	†	0.1s 93.86%	138s 98.19%	2.2s 100%

TABLE 3: Performance on OpenGM benchmarks. Columns #I,#L,#V denote the number of instances, labels and variables respectively. For each method an average over all instances in a family is reported. † – result is not available (memory / implementation limitation).

Instance	Initialization (1000 it.)	Extra time for persistency				
		no speedups	+reduction	+node pruning	+labeling pruning	+fast msgs
Protein folding 1CKK	8.5s	268s (26.53%)	168s (26.53%)	2.0s (26.53%)	2.0s (26.53%)	2.0s (26.53%)
colorseg-n4 pfau-small	9.3s	439s (88.59%)	230s (93.41%)	85s (93.41%)	76s (93.41%)	19s (93.41%)

TABLE 4: Evaluation of speedups on selected examples: computational time drops, as from left to right we add techniques described in § 6. 1CKK: an example when the final time for persistency is only a fraction of the initialization time. pfau-small: an example when times for initialization and persistency are comparable; speedups also help to improve the persistency as they are based on exact criteria.

Instance	#L	#V	[43]-CPLEX	[43]-TRWS	Our-CPLEX	Our-TRWS
1CKK	≤ 445	38	2503s 0%	46s 0%	2758s 27%	8.5+2s 26.53%
1CM1	≤ 350	37	2388s 0%	51s 0%	4070s 34%	9+3.9s 29.97%
1SY9	≤ 425	37	1067s 0%	67s 0%	2629s 51%	11+4.2s 57.98%
2BBN	≤ 404	37	9777s 0%	5421s 0%	9677s 9%	16+4.3s 14.17%
PDB1B25	≤ 81	1972	325s 22%	120s 22%	1599s 84%	4.3+7.3s 87.84%
PDB1D2E	≤ 81	1328	483s 59%	83s 59%	154s 98%	1.6+1.8s 98.25%

TABLE 5: Comparison to [43] using exact and approximate LP solvers. Examples of hard ProteinFolding instances [29, 51]. For Our-TRWS the initialization + persistency time is given. An occasionally better persistency of Our-TRWS vs. Our-CPLEX is explained by different test labelings produced by the CPLEX and TRW-S solvers (unlike in Table 2). The results of ϵ -L1 [35] would be identical to Our-CPLEX, as has been proven and verified on random instances. Unfortunately the existing implementation of ϵ -L1 cannot actually run on these instances to compare the timing (it assumes the same number of labels in all variables, letting 81 labels in each variable of the smallest PDB1D2E problem, the implementation did not terminate in 6 hours).

has hard interaction constraints. It seems that hard constraints and ambiguous solutions pose difficulties to all methods including ours.

10 CONCLUSIONS AND OUTLOOK

We presented an approach to find persistencies for an exp-APX-complete problem employing only solvers for a convex relaxation. Using a suboptimal solver for the relaxed problem, we still correctly identify persistencies while the whole approach becomes scalable. Our method with an exact solver matches the maximum persistency [35] and with a suboptimal solver closely approximates it, outperforming state of the art persistency techniques [43, 19, 25]. The speedups we have developed allow to achieve this at a reasonable computational cost making the method much more practical than the works [35, 43] we build on. In fact, our approach takes an approximate solver, like TRW-S, and turns it into a method with partial optimality guarantees at a reasonable computational overhead.

We believe that many of the presented results can be extended to higher order graphical models and tighter relaxations. Practical applicability with other approximate solvers can be explored. A further research direction that seems promising is mixing different optimization strategies such as persistency and cutting plane methods.

ACKNOWLEDGEMENT

Alexander Shekhovtsov was supported by the Austrian Science Fund (FWF) under the START project BIVISION, No. Y729. Paul Swoboda and Bogdan Savchynskyy were supported by the German Research Foundation (DFG) within the program “Spatio-/Temporal Graphical Models and Applications in Image Analysis”, grant GRK 1653. Bogdan Savchynskyy was also supported by European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 647769).

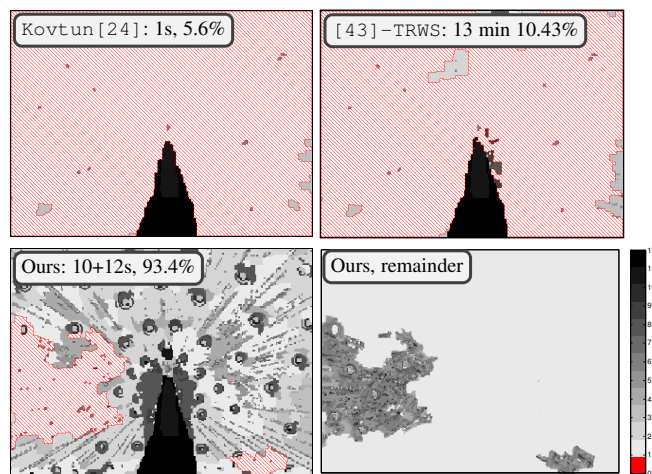


Fig. 6: Performance on a hard segmentation problem. The remainder of the problem visualizes $|p_u(X_u)|$ for all pixels. The result of the improved method [44]-TRWS is the same as [43]-TRWS, because this is a Potts model.

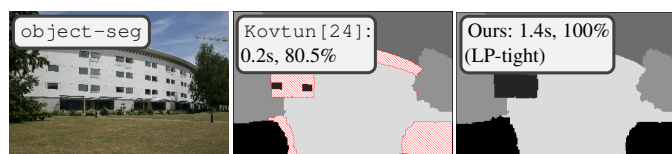


Fig. 7: Examples of an easy problem, *object-seg*. TRWS finds the optimal solution (zero integrality gap) and therefore our method as well as [43]-TRWS, [44]-TRWS report 100%.

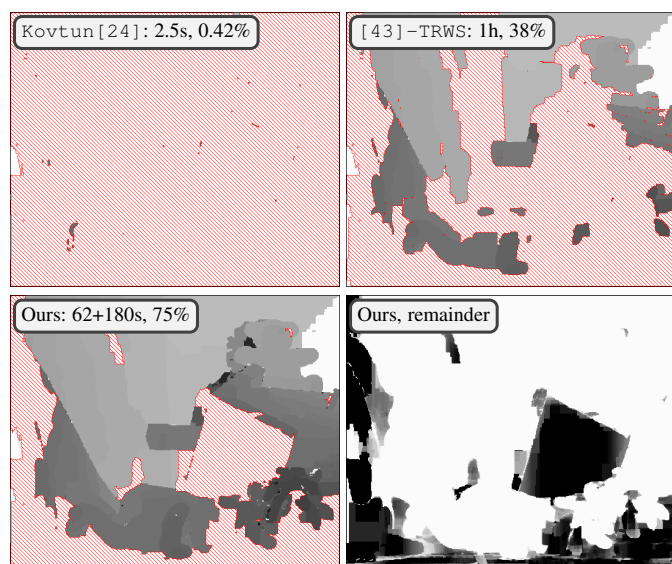


Fig. 8: Examples of a hard stereo problem, *ted-gm*. The result [44]-TRWS is the same as [43]-TRWS, because this is a Potts model too.

REFERENCES

[1] OpenGM benchmark. <http://hci.iwr.uni-heidelberg.de/opengm2/?l0=benchmark>.
 [2] Adams, W. P., Lassiter, J. B., and Sherali, H. D. (1998). Persistence in 0-1 polynomial programming. *Mathematics of Operations Research*, 23(2):359–389.
 [3] Aggarwal, A., Klawe, M. M., Moran, S., Shor, P., and Wilber, R. (1987). Geometric applications of a matrix-searching algorithm. *Algorithmica*, 2(1):195–208.

[4] Alahari, K., Kohli, P., and Torr, P. H. S. (2008). Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In *CVPR*, pages 1–8.
 [5] Boros, E. and Hammer, P. L. (2002). Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123(1):155 – 225.
 [6] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239.
 [7] Choi, J. and Rutenbar, R. A. (2012). Hardware implementation of MRF MAP inference on an FPGA platform. In *Field Programmable Logic*, pages 209–216.
 [8] Cooper, M., de Givry, S., Sanchez, M., Schiex, T., Zytnicki, M., and Werner, T. (2010). Soft arc consistency revisited. *Artificial Intelligence*, 174(78):449 – 478.
 [9] de Givry, S., Prestwich, S. D., and O’Sullivan, B. (2013). Dead-end elimination for weighted CSP. In *Constraint Programming*, volume 8124, pages 263–272.
 [10] Desmet, J., Maeyer, M. D., Hazes, B., and Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, 356:539–542.
 [11] Freuder, E. C. (1991). Eliminating interchangeable values in constraint satisfaction problems. In *National Conference on Artificial Intelligence*, pages 227–233.
 [12] Globerson, A. and Jaakkola, T. (2007). Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *NIPS*.
 [13] Goldstein, R. F. (1994). Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophysical Journal*, 66(5):1335–1340.
 [14] Gridchyn, I. and Kolmogorov, V. (2013). Potts model, parametric maxflow and k-submodular functions. In *ICCV*, pages 2320–2327.
 [15] Hirata, T. (1996). A unified linear-time algorithm for computing distance maps. *Information Processing Letters*, 58(3):129 – 133.
 [16] Hurkat, S., Choi, J., Nurvitadhi, E., Martnez, J. F., and Rutenbar, R. A. (2015). Fast hierarchical implementation of sequential tree-reweighted belief propagation for probabilistic inference. In *Field Programmable Logic*, pages 1–8.
 [17] ILOG, Inc. ILOG CPLEX: High-performance software for mathematical programming and optimization. See <http://www.ilog.com/products/cplex/>.
 [18] Kappes, J. H., Andres, B., Hamprecht, F. A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B. X., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., and Rother, C. (2015). A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision*, pages 1–30.
 [19] Kohli, P., Shekhovtsov, A., Rother, C., Kolmogorov, V., and Torr, P. (2008). On partial optimality in multi-label MRFs. In *ICML*, pages 480–487.
 [20] Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583.
 [21] Kolmogorov, V. (2012). Generalized roof duality and bisubmodular functions. *Discrete Applied Mathematics*, 160(4-5):416–426.
 [22] Kolmogorov, V. (2015). A new look at reweighted message passing. *PAMI*, 37(5):919–930.
 [23] Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159.
 [24] Kovtun, I. (2003). Partial optimal labeling search for a NP-hard subclass of (max, +) problems. In *DAGM-Symposium*, pages 402–409.
 [25] Kovtun, I. (2011). Sufficient condition for partial optimality for (max, +) labeling problems and its usage. *Control Systems and Computers*, 2:35–42. Special issue.
 [26] Lecoutre, C., Roussel, O., and Dehane, D. E. (2012). WCSP integration of soft neighborhood substitutability. In Milano, M., editor, *CP*, volume 7514 of *Lecture Notes in Computer Science*, pages 406–421.
 [27] Li, M., Shekhovtsov, A., and Huber, D. (2016). Complexity of discrete energy minimization problems. In *ECCV*, pages 834–852.
 [28] Meijster, A., Roerdink, J. B. T. M., and Hesselink, W. H. (2000). *A General Algorithm for Computing Distance Transforms in Linear*

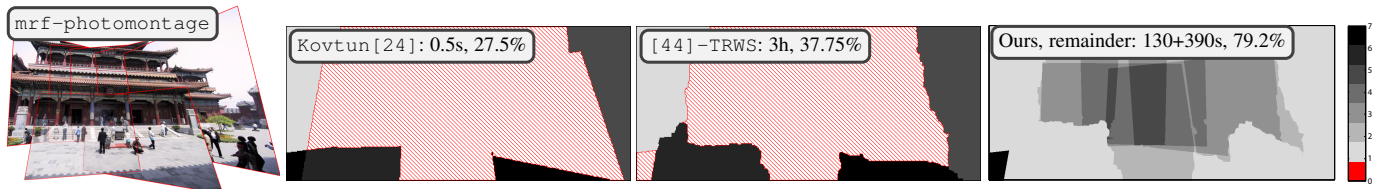


Fig. 9: Example of a very hard instance, *photomontage*. The number of views covering a pixel is usually smaller than the total number of views = number of labels. The 79% by our method mostly originate from the elimination of these redundant labels. The result of [43] - TRWS gives 27.55% (visually same as Kovtun [24]) in 3.9h. [44] - TRWS improves on this due to choosing the optimal reparametrization [44]. Note that our partial labeling (the part with one label remaining) is larger.

Time, pages 331–340.

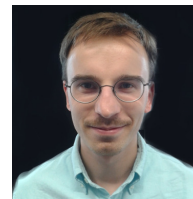
- [29] PIC2011. The probabilistic inference challenge (PIC2011). <http://www.cs.huji.ac.il/project/PASCAL/>.
- [30] Rother, C., Kolmogorov, V., Lempitsky, V., and Szummer, M. (2007). Optimizing binary MRFs via extended roof duality. In *CVPR*, pages 1–8.
- [31] Savchynskyy, B., Kappes, J. H., Schmidt, S., and Schnörr, C. (2011). A study of Nesterov’s scheme for Lagrangian decomposition and MAP labeling. In *CVPR*, pages 1817–1823.
- [32] Savchynskyy, B., Schmidt, S., Kappes, J., and Schnörr, C. (2012). Efficient MRF energy minimization via adaptive diminishing smoothing. *Uncertainty in Artificial Intelligence*, pages 746–755.
- [33] Schlesinger, M. I. and Antoniuk, K. V. (2011). Diffusion algorithms and structural recognition optimization problems. *Cybernetics and Sys. Anal.*, 47(2):175–192.
- [34] Shekhovtsov, A. (2013). *Exact and Partial Energy Minimization in Computer Vision*. PhD Thesis CTU–CMP–2013–24, CMP, Czech Technical University in Prague.
- [35] Shekhovtsov, A. (2014a). Maximum persistency in energy minimization. In *CVPR*, pages 1162–1169.
- [36] Shekhovtsov, A. (2014b). Maximum persistency in energy minimization. *CoRR*, abs/1404.3653.
- [37] Shekhovtsov, A. (2016). Higher order maximum persistency and comparison theorems. *CVIU*, 143(C):54–79.
- [38] Shekhovtsov, A., Reinbacher, C., Graber, G., and Pock, T. (2016). Solving dense image matching in real-time using discrete-continuous optimization. In *Computer Vision Winter Workshop*, pages 1–13.
- [39] Shekhovtsov, A., Swoboda, P., and Savchynskyy, B. (2015a). Maximum persistency via iterative relaxed inference with graphical models. In *CVPR*, pages 521–529.
- [40] Shekhovtsov, A., Swoboda, P., and Savchynskyy, B. (2015b). Maximum persistency via iterative relaxed inference with graphical models. *CoRR*, abs/1508.07902.
- [41] Shlezinger, M. (1976). Syntactic analysis of two-dimensional visual signals in the presence of noise. *Cybernetics and Systems Analysis*, 4:113–130. See review [49].
- [42] Swoboda, P., Savchynskyy, B., Kappes, J., and Schnörr, C. (2013). Partial optimality via iterative pruning for the Potts model. In *SSVM*, pages 477–488.
- [43] Swoboda, P., Savchynskyy, B., Kappes, J. H., and Schnörr, C. (2014). Partial optimality by pruning for MAP-inference with general graphical models. In *CVPR*, pages 1170–1177.
- [44] Swoboda, P., Shekhovtsov, A., Kappes, J. H., Schnörr, C., and Savchynskyy, B. (2016). Partial optimality by pruning for MAP-inference with general graphical models. *PAMI*, 38(7):1370–1382.
- [45] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *PAMI*, 30(6):1068–1080.
- [46] Vanderbei, R. J. (2001). *Linear Programming: Foundations and Extensions*. Department of operations and research and financial engineering, Princeton university.
- [47] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends*

Mach. Learn., 1(1-2):1–305.

- [48] Wang, C. and Zabih, R. (2016). Relaxation-based preprocessing techniques for Markov random field inference. In *CVPR*, pages 5830–5838.
- [49] Werner, T. (2007). A linear programming approach to max-sum problem: A review. *PAMI*, 29(7):1165–1179.
- [50] Windheuser, T., Ishikawa, H., and Cremers, D. (2012). *Generalized Roof Duality for Multi-Label Optimization: Optimal Lower Bounds and Persistence*, pages 400–413.
- [51] Yanover, C., Schueler-Furman, O., and Weiss, Y. (2008). Minimizing and learning energy functions for side-chain prediction. *Jour. of Comp. Biol.*, 15(7):899–911.
- [52] Zach, C. (2014). A principled approach for coarse-to-fine MAP inference. In *CVPR*, pages 1330–1337.



Alexander Shekhovtsov graduated in applied mathematics from the National Technical University of Ukraine “Kiev Polytechnic Institute”. From 2001 to 2005 worked in Kiev under supervision of Schlesinger, M.I. From 2005 to 2013 researcher and PhD student at Czech Technical University in Prague. In 2014 defended PhD thesis “exact and partial energy minimization in computer vision”, supervised by Hlaváč, V. Since 2014 postdoc at Graz University of Technology.



Paul Swoboda graduated and obtained a PhD in mathematics from Heidelberg University in 2011 and 2016 respectively. Since then, he is postdoc at IST Austria. His main interests are optimization problems for image analysis, in particular combinatorial methods.



Bogdan Savchynskyy graduated from the National Technical University of Ukraine “Kiev Polytechnic Institute” in 2002. In 2007 he defended his PhD Thesis at the National Academy of Sciences of Ukraine. During 2009-2014 he worked at Heidelberg University and since 2015 at Dresden University of Technology. His main interests are optimization problems for image and video analysis, in particular inference and learning with Markov random fields.