Higher Order Maximum Persistency and Comparison Theorems

Alexander Shekhovtsov[1]

Graz University of Technology

shekhovtsov@icg.tugraz.at

Abstract

We address combinatorial problems that can be formulated as minimization of a partially separable function of discrete variables (energy minimization in graphical models, weighted constraint satisfaction, pseudo-Boolean optimization, 0-1 polynomial programming). For polyhedral relaxations of such problems it is generally not true that variables integer in the relaxed solution will retain the same values in the optimal discrete solution. Those which do are called persistent. Such persistent variables define a part of a globally optimal solution. Once identified, they can be excluded from the problem, reducing its size.

To any polyhedral relaxation we associate a sufficient condition proving persistency of a subset of variables. We set up a specially constructed linear program which determines the set of persistent variables maximal with respect to the relaxation. The condition improves as the relaxation is tightened and possesses all its invariances. The proposed framework explains a variety of existing methods originating from different areas of research and based on different principles. A theoretical comparison is established that relates these methods to the standard linear relaxation and proves that the proposed technique identifies same or larger set of persistent variables.

Keywords: Persistency, partial optimality, LP relaxation, discrete optimization, WCSP, graphical models

Contents

1. Introduction

Optimization models in the general form of minimizing a partially separable function of discrete variables, known as energy minimization, weighted/valued constraint satisfaction or max-sum labeling, proved useful in many areas. The function has the form $E_f(x) = \sum_{c \in \mathcal{E}} f_c(x_c)$. In computer vision and machine learning such models are largely motivated by maximum a posteriori inference in graphical models [71] used to model a variety of structured statistical recognition problems. In case variables take only two values 0 or 1, the problem is known as *pseudo-Boolean optimization* or 0-1 polynomial programming. Problems where terms (summands) involve at most two 0-1 variables at a time are called *quadratic*. We consider the general case, where terms may couple more that two variables at a time (higher order) and variables can take more then two values (multilabel).

One major trend for performing inference in graphical models is represented by graph cut methods. The basic capability is essentially to solve a binary pairwise submodular problem, *e.g.*, image segmentation [18], by reduction to a minimum cut / maximum flow problem. For the latter, many efficient algorithms exist and their running time is experimentally near linear for typical vision problems [9]. This basic method was extended to submodular multilabel problems [21, 51], to general multilabel problems by solving for an optimized crossover between two candidate solutions at a time [10], to higher-order 0-1 models reducible to a graph cut [34, 16] and to combinations of higher order and multilabel [41, 13].

Another technique that can be considered nowadays as a basic graph cut method is the roof dual relaxation [6] known in computer vision as quadratic pseudo-Boolean optimization (QPBO) [32]. It allows to find a partial optimal solution to a non-submodular binary problem and reduces to finding a minimum cut in a specially constructed network [7]. It can be interpreted [28] as solving a submodular relaxation of the initial problem. This basic method is again extended to multilabel problems by solving crossover problems [42] and to general higher order 0-1 problems by reduction (quadratization) techniques expressing the function as a quadratic function with auxiliary variables [22, 15, 5].

Another direction of extending graph cuts to higher order models relies on minimization of more general submodular functions. Several efficient max-flow based algorithms have been proposed [4, 29] for minimization of a sum of submodular functions (SoS). A natural extension of QPBO is represented by submodular and bisubmodular relaxations [28, 24].

Arguably, linear programming (LP) is a much more costly tool than computing a minimum cut. Yet, it provides theoretical insight to many methods [36, 37] and there has been solvers developed that can address (sometimes approximately) large scale problems. Dual decomposition methods [53, 35] or dual block-descent methods, in particular TRW-S [26], are competitive with graph cut based methods in terms of speed and quality. There are extensions of these specialized LP methods to higher order models [33, 31]. Smoothing [49] and proximal [47] methods are scalable and offer a theoretically guaranteed convergence speed. Cutting plane approaches [60, 73] are used to tighten the relaxation adaptively to the problem.

One drawback of relaxation based methods is that the final discrete solution is obtained by so-called rounding schemes and often appears inferior to solutions by graph cut methods as they stay feasible to the discrete space. Even in the case when many of the relaxed variables take integer values in the optimal relaxed solution, a fundamental problem remains that they may not take the same integer values in the optimal discrete solution. Therefore, unless the relaxation is tight, a local rounding technique cannot provide any guarantees for general models. The situation is dramatically different when we consider quadratic pseudo-Boolean functions. There, all variables that are integer in the relaxation correspond to at least one globally optimal discrete solution [44, 20]. This property of the relaxation is called *persistency*. For general 0-1 polynomial problems persistency was studied by [43, 2]. In their terminology persistency is associated with relaxations and is a property of the relaxed solution as a whole. In this work we call any

partial assignment of a subset of discrete variables *persistent* if it can be provably extended to a globally optimal solution based on the properties of the relaxation or any other sufficient condition. Success of relaxation based exact methods such as [50] on computer vision and machine learning problems suggests that often a large part of the relaxed solution is integral. In this case we are interested in determining the largest subset of such variables that is persistent.

Related Work The present work was to a large extent inspired by simple local sufficient conditions proposed in bioinformatics under the name *dead end elimination* (DEE) [14, 17]. In computer vision, several methods were proposed that identify a persistent assignment directly in the multi-label setting. These are methods by Kovtun [39, 40] and Swoboda et al. [63, 64]. Method [64] is applicable with a general polyhedral relaxation and it maximizes the subset of peristent variables for their sufficient condition (discussed in §4.5).

In the case of 0-1 variables there are several different techniques. Adams et al. [2] proposed a sufficient condition on dual multipliers to prove persistency of the integral part of the relaxation. Quadratization techniques by Ishikawa [22], Fix et al. [15], Boros and Gruber [5] introduce auxiliary variables in order to reduce the function to a quadratic form and infer persistency from the QPBO method. Lu and Williams [43] generalized the roof duality approach to higher order by using a higher order linear relaxation. Kolmogorov [28] generalized both QPBO and the construction by Lu and Williams [43] by proposing discrete submodular and bisubmodular relaxations. He argues that the key property of QPBO that needs to be generalized is the existence of a totally half integral optimal solution to the relaxation, *i.e.*, with values in $\{0, \frac{1}{2}, 1\}$. He characterized all totally half-integral relaxations as bisubmodular relaxations. Finding a good (bi)submodular relaxation appears to be a challenging problem. To our knowledge it was only resolved for the special case of mincut-reducible relaxations [24, 62], and even in this case it requires solving a series of linear programs. Even though the relaxed problem itself can be efficiently optimized (in particular when it is a sum of submodular functions), having a sound persistency result at a comparable computation cost is an open problem. No theoretical comparison seems to be possible between (bi)submodular relaxations and quadratization techniques [28].

Kohli et al. [25] reduce multilabel pairwise prob-

lems to 0-1 quadratic and Windheuser et al. [75] reduce multilabel higher order problems to submodular relaxations of [28].

Contribution In this work we settle the persistency capabilities achievable with a general polyhedral relaxation. The previously known results are in a certain sense unique, relying on a specific sufficient condition or on a specific type of the relaxation. We show that persistency guarantees are not that rare. To any polyhedral relaxation we associate clear sufficient conditions for persistency. We propose a polynomial time method to determine the largest strongly persistent subset of variables according to the sufficient condition. The method sets up a linear program connected to the given relaxation polytope and maximizes the number of strongly persistent variables. In comparison to QPBO-based or submodularity-based techniques, we employ a more costly optimization tool, but gain the following advantages:

- the new sufficient condition generalizes a wide variety of existing methods that span across different fields of research and apply different techniques;
- it is possible to pose formally and solve (under certain restrictions) the problem of determining the largest subset of persistent variables;
- the maximum w.r.t. to the proposed general sufficient condition is guaranteed to be at least as good as any of the individual methods or their combinations;
- the method is invariant to the permutation of labels and reparametrization of the problem as long as the relaxation is invariant;
- persistent assignments form a hierarchy when tightening the relaxation.

The author's previous work [54, 55] considered only pairwise models and the standard LP relaxation. This paper generalizes to higher order and arbitrary polyhedral relaxations, gives more complete proofs of some properties and establishes comparisons with a novel multilabel method [64] and higher-order 0-1 methods [28, 22, 15, 2].

Outline In §2 we propose a general approach to persistency with a general polyhedral relaxation. This includes the proposed linear program formulation of maximum persistency and general properties of the problem. In §3 we consider standard LP relaxations and specialize the construction for this case, many properties are simplified. In §4 we propose a theoretical comparison between the proposed frame-

work and other approaches. In §5 we validate our theoretical findings experimentally and compare performance on small random problems. In §6 there is a conclusion and discussion and §A contains proofs.

1.1. Notation

When generalizing to higher order models many statements and proofs simplify if we use a properly defined notion of the empty sum, the empty product and the empty Cartesian product. They are respectively: $\sum_{i \in \varnothing} x_i = 0$, $\prod_{i \in \varnothing} x_i = 1$ and $\prod_{i \in \varnothing} X_i = \{\varnothing\}$. The inclusion $\subset$ is non-strict and $\subsetneq$ is strict. LHS refers to the left hand side of an equation. $[\![\cdot]\!]$ denotes the Iverson bracket. argmin is the set of minimizers. Sets $\mathbb{R}, \mathbb{R}_+$ denote real and non-negative real numbers and $\mathbb{B} = \{0,1\}$ is the Boolean domain. A composition of functions is denoted as $(f \circ g)(x) = f(g(x))$. Finally, a *polytope* is assumed to be convex but may be unbounded and a *polyhedron* means the same as a polytope.

1.2. Energy Minimization

A hypergraph $(\mathcal{V}, \mathcal{E})$ is given by the set of nodes $\mathcal{V}$ and the set of hyperedges $\mathcal{E} \subset 2^{\mathcal{V}}$. We assume that $\mathcal{V}$ is totally ordered and each hyperedge $\mathrm{C} \in \mathcal{E}$ is identified with the tuple of elements of C ordered w.r.t. the total order of $\mathcal{V}$. We will further assume that $\varnothing \in \mathcal{E}$ and $(\forall s \in \mathcal{V})$ $\{s\} \in \mathcal{E}$. Let $\mathcal{X}_s$ be a finite set of *labels* associated to a node $s \in \mathcal{V}$. For a subset of nodes $\mathrm{C} \subset \mathcal{V}$ the set $\mathcal{X}_{\mathrm{C}}$ denotes the Cartesian product $\prod_{s \in \mathrm{C}} \mathcal{X}_s$ in the order defined on $\mathcal{V}$ and $\mathcal{X} = \mathcal{X}_{\mathcal{V}}$. The assignment of labels to all nodes $x: \mathcal{V} \to \mathcal{X}$ is called a *labeling*. Let x_{C} denote the restriction of x to $\mathrm{C} \subset \mathcal{V}$ (thus x_s is just a single coordinate) and $x_{\varnothing} = \varnothing$. Let us define the following functions (terms):

$$(\forall \mathrm{C} \in \mathcal{E}) \quad f_{\mathrm{C}}: \mathcal{X}_{\mathrm{C}} \to \mathbb{R}. \quad \text{(general hyperedge term)}$$

The special cases read

$$\begin{aligned} f_{\varnothing}: \{\varnothing\} &\to \mathbb{R} & \text{(constant term)}, \\ f_s: \mathcal{X}_s &\to \mathbb{R} & \text{(unary / 0 order term)}, \\ f_{\{s,t\}}: \mathcal{X}_{\{s,t\}} &\to \mathbb{R} & \text{(pairwise / 1st order term)} \end{aligned}$$

and so on. The constant term $f_{\varnothing}$ is nothing but a single number. The *energy function* $E_f: \mathcal{X} \to \mathbb{R}$ is defined by

$$E_f(x) = \sum_{\mathrm{C} \in \mathcal{E}} f_{\mathrm{C}}(x_{\mathrm{C}}). \tag{2}$$

It is a partially separable function of discrete variables x. In this paper we will use a graphical notation of the energy explained in Figure 1.
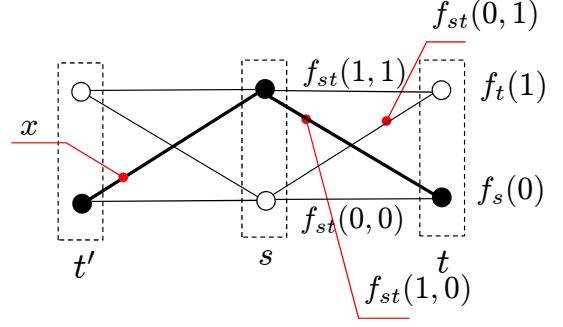

Figure 1: Graphical notation by Shlezinger [59]. Variables $x_s, x_t, x_{t'}$ are depicted as boxes, their possible states as circles and states of pairs of variables as lines. In the first order model (pairwise) the energy of a labeling x is the sum of the selected unary and pairwise costs.

The general energy minimization problem is NP-hard to approximate[2]. On the other hand, there are tractable subclasses. Works by Thapper and Živný [66, 67] and Kolmogorov [30] characterized all *languages* of energy functions with terms from a fixed finite set and unrestricted structure. They showed that there are no tractable languages other than those that can be solved by the basic LP relaxation (defined in §3), which proves that the relaxation is a universal and powerful technique.

1.3. General Polyhedral Relaxation

In this section we embed the energy minimization problem into the Euclidean space. A labeling x is represented as a 0-1 vector in order to linearize the energy and write it as scalar product of this vector with the cost vector f consisting of all components $f_{\mathrm{C}}(x'_{\mathrm{C}})$ for $\mathrm{C} \in \mathcal{E}$, $x'_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}}$. According to these components let us define the following set of *indices* $\mathcal{I} = \{(\mathrm{C}, x'_{\mathrm{C}}) \mid \mathrm{C} \in \mathcal{E}, \ x'_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}}\}$. The *embedding* $\delta: \mathcal{X} \to \mathbb{R}^{\mathcal{I}}: x \to \delta(x)$ is defined by its components

$$(\forall \mathrm{C} \in \mathcal{E}, \ x'_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}}) \quad \delta(x)_{\mathrm{C}}(x'_{\mathrm{C}}) = [\![x_{\mathrm{C}} = x'_{\mathrm{C}}]\!]. \tag{3}$$

The special cases read

$$\delta(x)_{\varnothing} = 1, \tag{4a}$$

$$\delta(x)_s(x'_s) = [\![x_s = x'_s]\!], \tag{4b}$$

$$\delta(x)_{\{s,t\}}(x'_{\{s,t\}}) = [\![x_s = x'_s]\!][\![x_t = x'_t]\!] \tag{4c}$$

and so on. Let $\langle \cdot, \cdot \rangle$ denote the scalar product in $\mathbb{R}^{\mathcal{I}}$. We can write the energy using the embedding δ as a linear function:

$$E_f(x) = \sum_{\mathrm{C} \in \mathcal{E}} \sum_{x'_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}}} f_{\mathrm{C}}(x'_{\mathrm{C}}) \delta(x)_{\mathrm{C}}(x'_{\mathrm{C}}) = \langle f, \delta(x) \rangle. \tag{5}$$

---

[2]*e.g.*, inapproximability of the traveling salesman problem [45].

The embedding δ is illustrated in Figure 2. The energy minimization can be expressed as:

$$\min_{x \in \mathcal{X}} \langle f, \delta(x) \rangle = \min_{\mu \in \delta(\mathcal{X})} \langle f, \mu \rangle = \min_{\mu \in \mathcal{M}} \langle f, \mu \rangle, \quad (6)$$

where $\delta(\mathcal{X})$ is the image of the set of labelings, *i.e.*, the set of corresponding points in $\mathbb{R}^{\mathcal{I}}$ and $\mathcal{M} = \operatorname{conv} \delta(\mathcal{X})$ is their convex hull, called *marginal polytope* [71]. The second equality follows from the fact that a convex combination of solutions is also a solution. Polytope $\mathcal{M}$ has in general exponentially many facets. A *relaxation* of the problem is obtained by replacing $\mathcal{M}$ with an outer approximation $\Lambda \supset \mathcal{M}$:

$$\min_{\mu \in \Lambda} \langle f, \mu \rangle. \quad (7)$$

A vector $\mu \in \Lambda \subset \mathbb{R}^{\mathcal{I}}$ will be called a *relaxed labeling*. We will consider polyhedral relaxations of the following general form:

$$\Lambda = \{ \mu \in \mathbb{R}^{\mathcal{I}} \mid A\mu \geqslant 0; \ \mu_{\varnothing} = 1; \ \mu \geqslant 0 \}, \quad (8)$$

where we assume that $A \in \mathbb{R}^{m \times |\mathcal{I}|}$ is such that Λ is bounded and $\mathcal{M} \subset \Lambda$. Since $\mathcal{M}$ is non-empty, it follows that Λ is non-empty. By these assumptions, relaxation (8) is a feasible and bounded linear program. Note that general inhomogenous equality and inequality constraints can be represented in this form by utilizing the component $\mu_{\varnothing}$. The dual problem to (7) and the conical hull of Λ are expressed conveniently as follows. Recall that for a convex set $\Lambda \subset \mathbb{R}^{\mathcal{I}}$ its conical hull is the set:

$$\operatorname{coni}(\Lambda) = \{ \alpha\mu \mid \mu \in \Lambda, \alpha \geqslant 0 \}. \quad (9)$$

Lemma 1.1. The conical hull of a relaxation polytope Λ (in the form (8), non-empty and bounded) is obtained by dropping the constraint $\mu_{\varnothing} = 1$:

$$\operatorname{coni}(\Lambda) = \{ \mu \in \mathbb{R}^{\mathcal{I}} \mid A\mu \geqslant 0; \mu \geqslant 0 \}. \quad (10)$$

Proof on p. 23.

The linear program (7) and its dual are expressed as

$$
\begin{array}{lll}
\min \langle f, \mu \rangle = & \max \psi\,, \\
A\mu \geqslant 0 & \varphi \in \mathbb{R}^m_+ \\
\mu_{\varnothing} = 1 & \psi \in \mathbb{R} & \text{(LP)} \\
\mu \geqslant 0 & f - A^{\mathsf{T}}\varphi - e_{\varnothing}\psi \geqslant 0
\end{array}
$$

where vector $e_{\varnothing} \in \mathbb{R}^{\mathcal{I}}$ is the basis vector for the component $\varnothing$ and the equality between the primal and the dual formulations holds because the primal problem is feasible and bounded. Let us introduce the notation $f^{\varphi} := f - A^{\mathsf{T}}\varphi$. Later on, when we consider
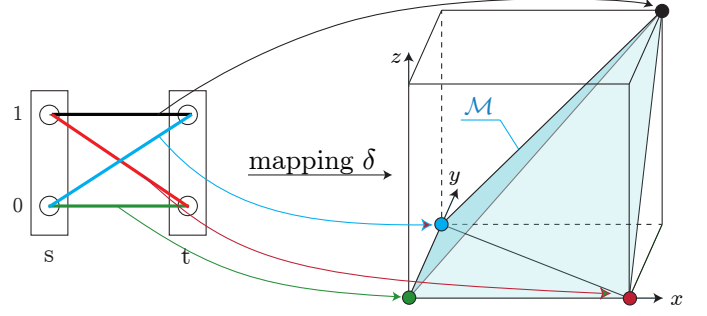

Figure 2: Mapping δ embeds discrete labelings as points in the space $\mathbb{R}^{\mathcal{I}}$. Left: 2 variables with 2 states, lines of different colors show possible assignments. Right: to each labeling x there correspond a point $\delta(x) \in \mathbb{R}^{\mathcal{I}}$. Axis x, y, z in the figure correspond respectively to $\delta_s(1)$, $\delta_t(1)$ and $\delta_{st}(1,1)$. In this representation the energy function is a linear functional. The minimization domain can be extended equivalently from the set of points $\delta(\mathcal{X})$ to their convex hull, the marginal polytope $\mathcal{M}$.

equality constraints of the form $A\mu = 0$, the vector f^{φ} will obtain the meaning of an equivalent problem and for now it is just an abbreviation.

2. Maximum Persistency

A *partial assignment* $y_{\mathcal{A}} \in \mathcal{X}_{\mathcal{A}}$, where $\mathcal{A} \subset \mathcal{V}$, is called *weakly persistent* if there exists an optimal solution x such that $x_{\mathcal{A}} = y_{\mathcal{A}}$. In other words, $y_{\mathcal{A}}$ can be extended to a global solution. Partial assignment $y_{\mathcal{A}}$ is called *strongly persistent* if $x_{\mathcal{A}} = y_{\mathcal{A}}$ holds for all optimal solutions x.

It may seem that there are no practical reasons to distinguish strongly and weakly persistent partial assignments as long as they allow to simplify the problem. However, it will become clear later that they have different theoretical properties leading to polynomially solvable versus NP-hard maximum persistency problems. It turns out that strong persistency is more tractable, whereas proofs are generally easier to obtain in the weak form and most results in the literature deliver weak persistency.

In the case of quadratic pseudo-Boolean functions the roof dual relaxation [6] is persistent: for any relaxed solution its integral part defines a partial assignment $y_{\mathcal{A}}$ which is optimal to the discrete problem. Moreover, for any labeling x, not necessarily optimal, replacing part of x on $\mathcal{A}$ with $y_{\mathcal{A}}$, the *overwrite* operation, denoted in [6] by $x[\mathcal{A}{\leftarrow}y]$, has the following *autarky* property:

$$(\forall x \in \mathcal{X}) \ E_f(x[\mathcal{A}{\leftarrow}y]) \leqslant E_f(x), \quad (11)$$

illustrated in Figure 3. We will generalize this property to the multilabel setting.
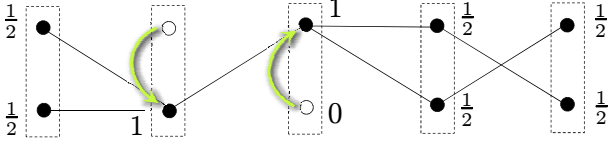
5

Figure 3: Quadratic pseudo-Boolean case (roof dual). There exist half-integral optimal solution to the relaxation (indicated by numbers). Its integer part (the assignment indicated by 1's) is persistent, *i.e.*, optimal to the original discrete optimization problem. The arrows show how an arbitrary labeling can be changed in order to improve the energy: switch to the optimal assignment for integer nodes and keep the assignment of the remaining nodes.
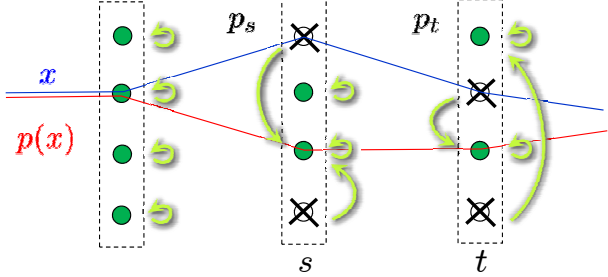

Figure 4: Improving mapping is a generalization of autarky. To every variable s a there is an associated mapping $p_s \colon \mathcal{X}_s \to \mathcal{X}_s$, shown by arrows. For any labeling x its image $p(x)$ has the same or better energy: $E_f(p(x)) \leqslant E_f(x)$. If label i is not in the range $p_s(\mathcal{X}_s)$ then it can be eliminated as shown by crosses.

2.1. Improving Mapping

The overwrite operation discussed above can be represented by a discrete mapping $p \colon \mathcal{X} \to \mathcal{X} \colon x \mapsto x[\mathcal{A}{\leftarrow}y]$. The following generalization of autarky to an arbitrary mapping is proposed.

Definition 2.1. A mapping $p \colon \mathcal{X} \to \mathcal{X}$ is called *(weakly) improving* for f if

$$(\forall x \in \mathcal{X}) \quad E_f(p(x)) \leqslant E_f(x), \qquad (12)$$

and *strictly improving* if

$$(p(x) \neq x) \implies E_f(p(x)) < E_f(x). \qquad (13)$$

The idea of the improving mapping is illustrated in Figure 4. It easily follows from the definition that if p is improving then there exists an optimal solution $x \in p(\mathcal{X})$ and if p is strictly improving then all optimal solutions are contained in $p(\mathcal{X})$. In this way an improving mapping reduces the search space from $\mathcal{X}$ to $p(\mathcal{X})$.

We will consider *node-wise* mappings, of the form $p(x)_s = p_s(x_s)$, where $(\forall s \in \mathcal{V}) \, p_s \colon \mathcal{X}_s \to \mathcal{X}_s$. Furthermore, we restrict ourselves to idempotent mappings, *i.e.*, satisfying $p \circ p = p$. This restriction is without loss of generality. Indeed, for an improving node-wise mapping p its compositional power
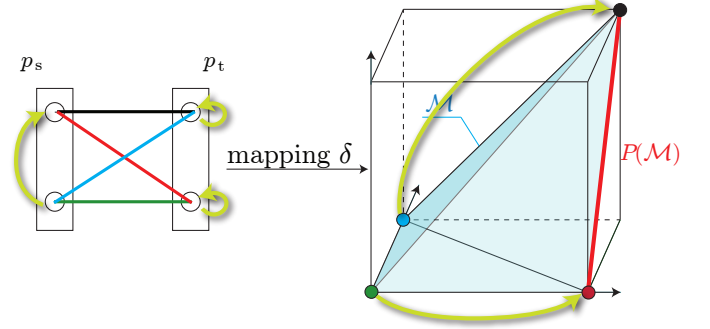
Figure 5: Embedding of a discrete mapping in $\mathbb{R}^{\mathcal{I}}$ (continues the example in Figure 2). Left: discrete node-wise mapping $p \colon \mathcal{X} \to \mathcal{X}$ is shown by arrows, it sends the green labeling to red and the blue one to black. Right: there is a corresponding linear map $P \colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ with this action on labelings embedded as vertices. It is an oblique projection which maps polytope $\mathcal{M}$ onto the red facet $P(\mathcal{M})$.

p^k will be idempotent for some k (*e.g.*, for $k = (\max_s |\mathcal{X}_s|)!$, which turns all cycles in the map to identity) and provides equally good or better reduction with $p^k(\mathcal{X}) \subseteq p(\mathcal{X})$. Idempotent maps have two following properties. Let X be a set and $p \colon X \to X$ idempotent.

- If $p(x) \neq x$ then no $y \in X$ is mapped to x;
- For $Y = p(X)$ the restriction of p to Y is the identity map $x \mapsto x$ and there holds $p(X) = \{x \in X \mid p(x) = x\}$;

It follows that knowing an improving mapping p, we can eliminate labels (s,i) for which $p_s(i) \neq i$ and there will remain at least one global minimizer of E_f.

Given a mapping p, the verification of the improving property (12) is NP-hard since already in the quadratic pseudo-Boolean case the verification of autarky property (11) is NP-hard [8]. A tractable sufficient condition will be constructed by embedding the mapping into the space $\mathbb{R}^{\mathcal{I}}$ and applying the relaxation there.

2.2. Relaxed Improving Mapping

Definition 2.2. A *linear extension* of $p \colon \mathcal{X} \to \mathcal{X}$ is a linear mapping $P \colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ that satisfies

$$(\forall x \in \mathcal{X}) \quad \delta(p(x)) = P\delta(x). \qquad (14)$$

See Figure 5 for illustration. Avoiding the discussion of uniqueness[3], we will only use the following linear extension for a node-wise mapping $p \colon \mathcal{X} \to \mathcal{X}$, which will be denoted $[p]$. The linear extension

---

[3]When a linear extension exists, its restriction to the affine hull of $\delta(\mathcal{X})$ is unique.

$P = [p]$ is defined by

$$(P\mu)_{\mathrm{C}}(x_{\mathrm{C}}) = \sum_{x'_{\mathrm{C}}} P_{\mathrm{C}, x_{\mathrm{C}}, x'_{\mathrm{C}}} \mu_{\mathrm{C}}(x'_{\mathrm{C}}) \qquad (15)$$

with coefficients

$$P_{\mathrm{C}, x_{\mathrm{C}}, x'_{\mathrm{C}}} := \prod_{s \in \mathrm{C}} [\![p_s(x'_s) = x_s]\!] = [\![p_{\mathrm{C}}(x'_{\mathrm{C}}) = x_{\mathrm{C}}]\!]. \qquad (16)$$

These coefficients should be understood as a "matrix" representation of P. To verify that (14) holds true we simply substitute an integer labeling $\delta(x)$ and expand the components as

$$\begin{aligned}(P\delta(x))_{\mathrm{C}}(x'_{\mathrm{C}}) &= \sum_{x''_{\mathrm{C}}} \prod_{s \in \mathrm{C}} [\![p_s(x''_s) = x'_s]\!] [\![x_{\mathrm{C}} = x''_{\mathrm{C}}]\!] \\ &= \delta(p(x))_{\mathrm{C}}(x'_{\mathrm{C}}).\end{aligned} \qquad (17)$$

Using the linear extension P of p we can write

$$E_f(p(x)) = \langle f, \delta(p(x)) \rangle = \langle f, P\delta(x) \rangle. \qquad (18)$$

This allows to express the condition of improving mapping (12) as

$$(\forall x \in \mathcal{X}) \quad \langle f, P\delta(x) \rangle \leqslant \langle f, \delta(x) \rangle \qquad (19)$$

or equivalently, fully in the embedding, as

$$(\forall \mu \in \delta(\mathcal{X})) \quad \langle f, P\mu \rangle \leqslant \langle f, \mu \rangle. \qquad (20)$$

Taking convex combinations in (20), we obtain an equivalent condition

$$(\forall \mu \in \mathcal{M}) \quad \langle f, P\mu \rangle \leqslant \langle f, \mu \rangle. \qquad (21)$$

Thus we have linearized the inequalities necessary for an improving mapping. However, the marginal polytope $\mathcal{M}$ is not tractable. We introduce a sufficient condition by requiring that the same inequality (21) is satisfied over a larger (tractable) polytope $\Lambda \supset \mathcal{M}$.

Definition 2.3. A linear mapping $P \colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ is (weak) Λ-*improving* for f if

$$(\forall \mu \in \Lambda) \quad \langle f, P\mu \rangle \leqslant \langle f, \mu \rangle; \qquad (22)$$

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

$$(\forall \mu \in \Lambda, \ P\mu \neq \mu) \quad \langle f, P\mu \rangle < \langle f, \mu \rangle. \qquad (23)$$

Statement 2.4. Let $P \colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ be a linear extension of $p \colon \mathcal{X} \to \mathcal{X}$ and Λ a relaxation polytope. If P is Λ-improving for f then p is improving for f.

Proof. A relaxed-improving mapping P satisfies inequality (22) over a superset Λ of $\mathcal{M}$, therefore condition (21) is satisfied, which by the definition of extension is equivalent to (12). $\square$

The set of mappings for which (22) (resp. (23)) is satisfied will be denoted $\mathbb{W}_f$ (resp. $\mathbb{S}_f$). For conve-

nience, we will use the term *relaxed improving* when the relaxation is clear from the context.

Naturally, a strict relaxed improving map is relaxed improving, *i.e.*, $\mathbb{S}_f \subset \mathbb{W}_f$. This is so because for all $\mu \in \Lambda$ such that $P\mu = \mu$ the inequality (22) is trivially satisfied.

Next we show that the verification of $P \in \mathbb{W}_f$ (resp. $P \in \mathbb{S}_f$) for a given P can be solved (decided) in polynomial time. The definition (22) of $P \in \mathbb{W}_f$ is equivalent to the expression

$$\min_{\mu \in \Lambda} \langle (I - P^{\mathsf{T}}) f, \mu \rangle \geqslant 0. \qquad (24)$$

The optimization problem in (24) will be therefore called the *verification LP*. As a linear program over a tractable polytope Λ, it can be solved in polynomial time and hence the decision problem $P \in \mathbb{W}_f$ is solvable in polynomial time.

In order to show that the verification of $[p] \in \mathbb{S}_f$ can also be decided in polynomial time we introduce the following equivalent reformulation.

Statement 2.5. Let $\mathcal{O} = \operatorname*{argmin}_{\mu \in \Lambda} \langle f, (I - P)\mu \rangle$. There holds $P \in \mathbb{S}_f$ iff

$$\mathcal{O} = P(\Lambda). \qquad (25)$$

Proof on p. 23.

The statement says that a strictly relaxed improving mapping must not change the set of all optimal solutions to the verification LP. This can be further expressed in components of the mapping and of the support set $\mathcal{O}$:

Statement 2.6. Let $\mathcal{O}_{\mathrm{C}} = \{ x_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}} \mid (\exists \mu \in \mathcal{O}) \ \mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0 \}$. There holds $[p] \in \mathbb{S}_f$ iff

$$(\forall \mathrm{C} \in \mathcal{E}) \ \mathcal{O}_{\mathrm{C}} = p_{\mathrm{C}}(\mathcal{X}_{\mathrm{C}}). \qquad (26)$$

Proof on p. 23.

Now, in order to solve the verification of $p \in \mathbb{S}_f$ in polynomial time we can solve the verification LP in (24), obtain C-support sets of its optimal solutions $\mathcal{O}_{\mathrm{C}}$ and check condition (26).

2.3. Properties

We next give necessary conditions for p in order that $[p] \in \mathbb{W}_f$ or $[p] \in \mathbb{S}_f$. They help to narrow down the set of maps to be considered. A relaxed improving map must preserve optimality of solutions to the relaxation and consequently their support set (again in components).

7

Lemma 2.7 (Necessary conditions I). Let $p\colon \mathcal{X} \to \mathcal{X}$ be node-wise and $P = [p]$. Let $\mathcal{O} = \operatorname{argmin}_{\mu\in\Lambda}\langle f, \mu\rangle$ and $\mathcal{O}_{\mathrm{C}} = \{x_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}} \mid (\exists \mu \in \mathcal{O})\ \mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0\}$. Then

  (i) For $p \in \mathbb{W}_f$ there holds

$$P(\mathcal{O}) \subset \mathcal{O}; \tag{27a}$$

$$(\forall \mathrm{C} \in \mathcal{E})\ \ p_{\mathrm{C}}(\mathcal{O}_{\mathrm{C}}) \subset \mathcal{O}_{\mathrm{C}}; \tag{27b}$$

  (ii) For $p \in \mathbb{S}_f$ there holds

$$P(\mathcal{O}) = \mathcal{O}; \tag{28a}$$

$$(\forall \mathrm{C} \in \mathcal{E})\ \ p_{\mathrm{C}}(\mathcal{O}_{\mathrm{C}}) = \mathcal{O}_{\mathrm{C}}; \tag{28b}$$

Proof on p. 23.

Next, we reformulate problems $[p] \in \mathbb{W}_f$ and $[p] \in \mathbb{S}_f$ dually, *i.e.*, not with quantifier $(\forall x \in \Lambda)$ as in (22) but with existence quantifiers. This will become important in the formulation of the maximum persistency problem where we optimize over p subject to the constraints $[p] \in \mathbb{W}_f$ (resp. $[p] \in \mathbb{S}_f$). Recall that the set $\mathbb{W}_f$ is defined for the relaxation polytope $\Lambda = \{\mu \in \mathbb{R}^{\mathcal{I}} \mid A\mu \geqslant 0;\ \mu_{\varnothing} = 1;\ \mu \geqslant 0\}$, where $A \in \mathbb{R}^{m\times|\mathcal{I}|}$.

Theorem 2.8 (Dual representation of $\mathbb{W}_f$). Set $\mathbb{W}_f$ can be expressed as

$$\{P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}} \mid (\exists \varphi \in \mathbb{R}_+^m)\ f - A^\mathsf{T}\varphi - P^\mathsf{T}f \geqslant 0\}. \tag{29}$$

Proof. Denote $g = (I - P^\mathsf{T})f$. Condition (24), equivalent to (22), can be stated yet equivalently for the conical hull of Λ:

$$\inf_{\mu\in\operatorname{coni}(\Lambda)} \langle g, \mu\rangle \geqslant 0. \tag{30}$$

This is because for any $\mu \in \Lambda$ and any $\alpha \geqslant 0$ vector $\alpha\mu$ will satisfy RHS of (22) as well. Using the expression for the conical hull of Λ in (10), we can write the minimization problem in (30) and its dual as

$$
\begin{array}{ll}
\inf\langle g, \mu\rangle & \max 0\,. \\
A\mu \geqslant 0 & \varphi \in \mathbb{R}_+^m \\
\mu \geqslant 0 & g - A^\mathsf{T}\varphi \geqslant 0
\end{array} \tag{31}
$$

Inequality (30) holds iff the primal problem is bounded, and it is bounded iff the dual is feasible, which is the case iff $(\exists \varphi \in \mathbb{R}_+^m)\ (f - A^\mathsf{T}\varphi) - P^\mathsf{T}f \geqslant 0$. $\square$

The set $\mathbb{S}_f$ is defined via a more complicated quantifier $(\forall \mu \in \Lambda, P\mu \neq \mu)$. Fortunately, the following dual reformulation holds for node-wise maps:

Theorem 2.9 (Dual representation of $\mathbb{S}_f$). Let $p\colon \mathcal{X} \to \mathcal{X}$ be node-wise. Then: (i) there exists $\varepsilon > 0$

such that $[p] \in \mathbb{S}_f$ iff

$$(\exists \varphi \in \mathbb{R}_+^m)\ \ f - A^\mathsf{T}\varphi - [p]^\mathsf{T}f \geqslant \varepsilon h, \tag{32}$$

where h is a function such that $h \geqslant 0$ and $h_{\mathrm{C}}(x_{\mathrm{C}}) = 0$ iff $p_{\mathrm{C}}(x_{\mathrm{C}}) = x_{\mathrm{C}}$; and (ii) for rational inputs (including h) the value of ε in (i) is a rational number of polynomial bit length. Proof on p. 24.

The constraint $[p] \in \mathbb{S}_f$ can thus be reduced to nearly the same representation as (29), with an addition of an εh slack term. By construction, this term is zero iff $[p]\mu = \mu$. In practice, taking a larger value of ε always results in a sufficient condition for $\mathbb{S}_f$ and hence does not break correctness. In theory, we want a very small ε but not so small that it would break polynomiality of the reformulation, which is ensured by part (ii). Note, while the set $\mathbb{S}_f$ in the space of all maps $\mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ was convex but not closed (as seen from definition (23)), the theorem encloses the discrete maps of our interest, $\{[p] \mid p\colon \mathcal{X} \to \mathcal{X}$ node-wise$\}$ in a closed (convex) polytope.

Finally we give a necessary condition for $\mathbb{W}_f$. The theorem has a primal and a dual counterpart. The primal counterpart states that when solving the verification LP, because its objective $(I - P^\mathsf{T})f$ is in the null space of P, the constrains of the problem can be projected onto the same subspace providing a simplification. The dual counterpart states that there always exist dual multipliers such that the improving property holds component-wise for reparametrized costs. This is useful in proofs, providing an alternative reformulation of local inequalities (29).

Theorem 2.10 (Necessary conditions II). Let $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ be idempotent, $P(\Lambda) \subset \Lambda$ and $P \in \mathbb{W}_f$. Then

$$\inf_{\substack{A(I-P)\mu\geqslant 0 \\ \mu\geqslant 0}} \langle (I - P)^\mathsf{T}f, \mu\rangle = 0; \tag{33a}$$

$$(\exists \varphi \in \mathbb{R}_+^m)\ \ (I - P^\mathsf{T})(f - A^\mathsf{T}\varphi) \geqslant 0. \tag{33b}$$

Proof on p. 24.

These conditions become necessary and sufficient for standard relaxations as discussed in §3.1. The constraint $A(I - P)\mu \geqslant 0$ in (33a) replaces the constraint $A\mu \geqslant 0$ in (31) and simplifies the problem.

2.4. Maximum Relaxed Improving Mapping

We showed in §2.2 that weak/strict relaxed-improving property can be verified in polynomial time and have described sets $\mathbb{W}_f$, $\mathbb{S}_f$. Any relaxed-improving map, with the exception of the identity,

order	labels	maps	(MAX-SI)	(MAX-WI)
pairwise	2	all	P	P
higher-order	2	all	P	NP-hard
any	any	$\mathcal{P}^{1,y}$ or $\mathcal{P}^{2,y}$	P	P
any	any	$\mathcal{P}^1$	P	NP-hard
pairwise	$\geqslant 4$	$\mathcal{P}^2$	NP-hard	NP-hard

Table 1: Polynomiality of finding the maximum strict/weak relaxed-improving mapping for a general relaxation.

eliminates some labels as non-optimal. Recall that the label (s, i) is eliminated by node-wise mapping p if $p_s(i) \neq i$. We formulate the following *maximum persistency* problem:

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

i.e. we directly maximize the number of eliminated labels. The strict variant, with constraint $[p] \in \mathbb{S}_f$, will be denoted MAX-SI.

The problem may look difficult to solve. Indeed, it optimizes over discrete maps and involves a general polyhedral relaxation in the specification of constraints. Nevertheless, if we place some additional restrictions on the set of maps, it turns out to be solvable in polynomial time in a number of cases summarized in Table 1. One of them is the pseudo-Boolean case, where there are only 3 possible idempotent maps for every node: $(0,1) \mapsto (1,1)$, $(0,1) \mapsto (0,0)$ and $(0,1) \mapsto (0,1)$. Problem (MAX-SI) turns out to be solvable in this case. For multilabel problems, node-wise mappings are more diverse. Motivated by the goal to include/generalize existing multilabel methods, the following sets of maps are introduced:

all-to-one maps. The set $\mathcal{P}^{1,y}$ of maps p of the form $p \colon x \mapsto x[\mathcal{A} \leftarrow y]$ for all $\mathcal{A} \subset \mathcal{V}$ and fixed $y \in \mathcal{X}$. This class is a straightforward generalization of the overwrite operation in the autarky (11). A mapping $p \in \mathcal{P}^{1,y}$ is illustrated in Figure 10(a). There are only two possible choices for every node s. The mapping p_s either contracts $\mathcal{X}_s$ to a single label $\{y_s\}$ or retains $\mathcal{X}_s$ unchanged. This class allows to explain one-against-all method of Kovtun [39] and the central part of the method of Swoboda et al. [64] as discussed in §4.4, §4.5.

all-to-one-unknown maps. Set $\mathcal{P}^1 = \bigcup_{y \in \mathcal{X}} \mathcal{P}^{1,y}$. A mapping $p \in \mathcal{P}^1$ has the same form as above, $p \colon x \mapsto x[\mathcal{A} \leftarrow y]$, however the labeling y is not fixed now but a part of the specification of the mapping, see Figure 11. In every node there are $|\mathcal{X}_s| + 1$ choices for p_s: send all labels to a single one (which may be chosen) or
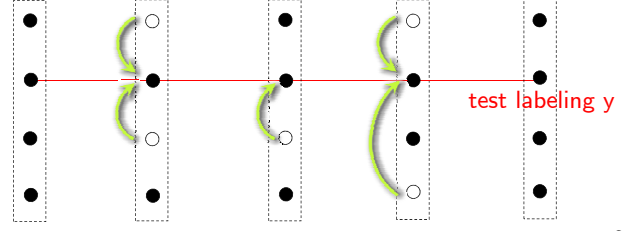

Figure 6: Example of a map in the subset-to-one class $\mathcal{P}^{2,y}$. Labeling y is fixed while a map p can select a subset of labels in every node s that are sent to y_s. Nodes without an outgoing arrow are mapped to themselves.

change nothing. It is easy to see that in the case of two labels, $\mathcal{P}^1$ contains all idempotent node-wise maps. As will be shown later the (MAX-SI) problem over this class decomposes into sufficient conditions to determine y from the integral part of the solution to the relaxation and the (MAX-SI) problem over $\mathcal{P}^{1,y}$.

subset-to-one maps. The set of maps $\mathcal{P}^{2,y}$ is defined as follows. Let $V = \{(s,i) \mid s \in \mathcal{V}, i \in \mathcal{X}_s\}$ – the set of labels in all nodes. Let $\zeta \in \{0,1\}^V$. Mapping $p_\zeta \in \mathcal{P}^{2,y}$ in every node either preserves the label x_s or overwrites it with y_s:

$$p_\zeta(x)_s = \begin{cases} y_s & \text{if } \zeta_{s,x_s} = 0, \\ x_s & \text{if } \zeta_{s,x_s} = 1. \end{cases} \qquad (34)$$

Vector $(\zeta_{s,i} \mid i \in \mathcal{X}_s)$ serves as the indicator of the subset of labels in node s that stay immovable while all other labels are mapped to y_s, see Figure 6. In a node s there are $2^{|\mathcal{X}_s|-1}$ choices for p_s. Clearly, this class generalizes $\mathcal{P}^{1,y}$.

The main result of this paper is that both (MAX-WI) and (MAX-SI) problems are tractable for the class $\mathcal{P}^{2,y}$. Other tractability results in Table 1 are obtained as corollaries. Intractability results are shown to hold for the basic LP relaxation in §3.1.

2.5. Formulation for Subset-to-one Maps

In the following three subsections we gradually show that (MAX-WI) problem over $\mathcal{P}^{2,y}$ class can be written as a mixed integer linear program in which integrality constraints can be relaxed without loss of tightness and thus we obtain an equivalent LP formulation.

Using the dual representation of the constraint $[p] \in \mathbb{W}_f$ (29) and the form of the mapping (34),

9

monomial	new variable	(S1)
$\prod_{s\in D} z_s$, $D \subset C$	$\rightarrow \zeta_D \in \mathbb{R}$, $\zeta_\varnothing = 1$	
multilinear polynomial	linearization	(S2)
$g(z) = \sum_{D\subset C} \alpha_D \prod_{s\in D} z_s$	$\rightarrow G(\zeta) = \sum_{D\subset C} \alpha_D \zeta_D$	
linearization properties		(S3)
$g_1(z) + g_2(z)$	$\rightarrow G_1(\zeta) + G_2(\zeta)$	
$\alpha g(z)$	$\rightarrow \alpha G(\zeta)$	
$\forall z \in \{0,1\}^C$ $g(z) = 0$	$\Leftrightarrow G(\zeta) \equiv 0$	
identity inequality for $B \subset C$	new constraint	(S4)
$\prod_{s\in B} z_s \prod_{s\in C\setminus B}(1 - z_s) \geq 0$	$\rightarrow \sum_{D\subset C\setminus B}(-1)^{|D|}\zeta_{D\cup B} \geq 0$	
any other identity inequality	identity inequality	(S5)
$\forall z \in \{0,1\}^C$ $g(z) \geq 0$	$\Leftrightarrow (\forall \zeta \in \mathcal{Z}_C)\ G(\zeta) \geq 0$	

Table 2: Summary of correspondences in the relaxation approach [58] within a hyperedge C.

the problem (MAX-WI) becomes

$$\min \sum_{s\in\mathcal{V}} \sum_{i\in\mathcal{X}_s} \zeta_{s,i} \tag{35a}$$

$$(\zeta_{s,i} \in \{0,1\} \mid s \in \mathcal{V},\ i \in \mathcal{X}_s);\ \varphi \in \mathbb{R}_+^m;$$

$$f - A^\mathsf{T}\varphi - [p_\zeta]^\mathsf{T} f \geq 0. \tag{35b}$$

Constraints (35b) involve a complicating expression $[p_\zeta]$. Let us express coefficients P_{c,x_c,x'_c} (16) of the linear extension $P = [p_\zeta]$. Substituting mapping p_ζ (34) they are expressed as polynomials in ζ:

$$P_{C,x_C,x'_C} = \prod_{s\in C} [\![p_\zeta(x')_s = x_s]\!] \tag{36}$$

$$= \prod_{s\in C} \Big([\![x'_s = x_s]\!] \zeta_{s,x'_s} + [\![y_s = x_s]\!](1 - \zeta_{s,x'_s}) \Big).$$

It appears that after expanding $[p_\zeta]$ using (36) the constraint that we need to represent (35b) will involve products of binary variables $\prod_{s\in D} \zeta_{s,x'_s}$ for all $C \in \mathcal{E}$, $D \subset C$ and $x'_D \in \mathcal{X}_D$. To reach the ILP formulation we are going to replace each such product with a substitute variable ζ_{D,x'_D}. This is achieved with the help of the relaxation of Sherali and Adams [58].

2.6. Relaxation of Sherali and Adams

The relaxation of Sherali and Adams [58] is applicable to polynomial programs with binary variables $z \in \{0,1\}^\mathcal{V}$. The relaxation of order d performs a simultaneous lifting for all subsets of variables $C \subset \mathcal{V}$ with $|C| = d$. Let us focus on a single hyperedge C chosen for generality from the set of hy-

peredges $\mathcal{E} \subset 2^\mathcal{V}$. The construction and its properties (within hyperedge C) are summarized in Table 2. For every product $\prod_{s\in D} z_s$, $D \subset C$, a new variable ζ_D is introduced (S1). A pseudo-Boolean function $g\colon \{0,1\}^C \to \mathbb{R}$ is linearized by writing it as a multilinear polynomial and replacing each monomial $\prod_{s\in D} z_s$ with the new variable ζ_D, (S2). From this definition we have linearity properties (S3), in particular:

Lemma 2.11 (Identity Equality (S3)). Let $G(\zeta)$ be the linearization of $g(z)$. Then $g(z) = 0$ for all $z \in \{0,1\}^C$ iff $G(\zeta) = 0$ for all $\zeta \in \mathbb{R}^{2^C}$. Proof on p. 25.

Next, constraints on new variables are added which correspond to identity inequalities $\prod_{s\in B} z_s \prod_{s\in C\setminus B}(1 - z_s) \geq 0$ for each $B \subset C$. Clearly this inequality holds for all $z \in \{0,1\}^C$. By expanding this expression one obtains its equivalent multilinear polynomial $g(z) = \sum_{D\subset C\setminus B}(-1)^{|D|}\prod_{s\in D\cup B} z_s \geq 0$. Constraints (S4) ensure that the linearization of this expression is nonnegative. The set of all such constraints defines the polytope

$$\mathcal{Z}_C = \{\zeta \in \mathbb{R}^{2^C} \mid \zeta_\varnothing = 1,\ (\forall B \subset C) \sum_{D\subset C\setminus B}(-1)^{|D|}\zeta_D \geq 0\}. \tag{37}$$

In fact, polytope $\mathcal{Z}_C$ is the convex hull of all binary vectors ζ corresponding to configurations z:

Lemma 2.12 (Convex hull). Polytope $\mathcal{Z}_C$ equals the convex hull

$$\mathrm{conv}\Big\{\zeta(z) \in \mathbb{R}^{2^C} \mid \zeta(z)_D = \prod_{s\in D} z_s,\ \forall D \subset C, \forall z \in \{0,1\}^C\Big\}. \tag{38}$$

Proof on p. 26.

From the convex hull representation there naturally follows an equivalence of identity inequalities before and after linearization:

Lemma 2.13 (Identity inequality (S5)). Let $G(\zeta)$ be the linearization of $g(z)$. Then $g(z) \geq 0$ for all $z \in \{0,1\}^C$ iff $G(\zeta) \geq 0$ for all $\zeta \in \mathcal{Z}_C$. Proof on p. 25.

In particular, for $\zeta \in \mathcal{Z}_C$ there holds $0 \leq \zeta_D \leq 1$ for $D \subset C$, a relation which is rather difficult to prove directly form (37). Finally, for our construction the next two results are necessary.

Theorem 2.14 (Lemma 2 of [58]). If $\zeta \in \mathcal{Z}_C$ and unary components ζ_s are integer (*i.e.*, equal to some $z_s \in \{0,1\}$) for all $s \in C$, then there holds $\zeta_D = \prod_{s\in D} z_s$

for all $D \subset C$.

Lemma 2.15 (Product)**.** For $\zeta \in \mathcal{Z}_C$ there holds $\zeta^2 \in \mathcal{Z}_C$, where the product $\zeta^2 = \zeta\zeta$ is component-wise. Proof on p. 26.

When applying the linearization to all hyperedges simultaneously, a variable $\zeta_{A \cap B}$ is introduced only once for (overlapping) hyperedges $A, B \in \mathcal{E}$. All local properties described above continue to hold for each hyperedge $C \in \mathcal{E}$ individually but of course they need not hold for the whole set $\mathcal{V}$.

2.7. Solution via Linear Program Formulation

Let us return to the reformulation (35) of (MAX-WI). It is clear that by opening brackets in (36), the coefficients P_{C,x_C,x'_C} can be expressed as

$$P_{C,x_C,x'_C} = \sum_{D \subset C} c_{C,D}(x_C, x'_D) \prod_{s \in D} \zeta_{s,x'_s}, \qquad (39)$$

where $c_{C,D}(x_C, x'_D)$ are appropriate constants not depending on ζ (detailed in §A.3). Because for $x'_s = y_s$ there holds $p_\zeta(y)_s = y_s$ irrespectively of ζ_{s,y_s} (label y_s is always mapped to itself) we may assume that $\zeta_{s,y_s} = 0$ as well as all products involving it.

The relaxation of Sherali and Adams is applied as follows. Let us denote $\tilde{\mathcal{X}}_s = \mathcal{X}_s \setminus \{y_s\}$ and respectively $\tilde{\mathcal{X}}_C = \prod_{s \in C} \tilde{\mathcal{X}}_s$. We substitute new variables $\zeta_{D,x'_D} \in \mathbb{R}$ in place of products $\prod_{s \in D} \zeta_{s,x'_s}$ in (39). For zero products, *i.e.*, for $x'_D \notin \tilde{\mathcal{X}}_D$, we let $\zeta_{D,x'_D} = 0$. From now on, let ζ denote the vector of relaxed variables

$$\zeta = (\zeta_{D,x_D} \in \mathbb{R} \mid \forall C \in \mathcal{E}, \ D \subset C, x_D \in \mathcal{X}_D). \qquad (40)$$

New variables ζ must satisfy the following constraints, defining a polytope $\mathcal{Z}$:

$$\zeta_\varnothing = 1, \qquad (41a)$$

$$(\forall C \in \mathcal{E}, \ \forall x'_C \in \mathcal{X}_C \setminus \tilde{\mathcal{X}}_C, \ \forall D \subset C) \ \ \zeta_{D,x'_D} = 0, \qquad (41b)$$

$$(\forall C \in \mathcal{E}, \ \forall x'_C \in \tilde{\mathcal{X}}_C, \ \forall D \subset C) \sum_{D \subset C \setminus B} (-1)^{|D|} \zeta_{D,x'_D} \geqslant 0. \qquad (41c)$$

Polytope $\mathcal{Z}$ is the intersection of polytopes $\mathcal{Z}_C$ (37) lifted to the space of all variables over $C \in \mathcal{E}$ and $x'_C \in \tilde{\mathcal{X}}_C$. Let P_ζ denote the extension-linearization of p_ζ (34), according to (39) and (S2) defined by:

$$(P_\zeta)_{C,x_C,x'_C} = \sum_{D \subset C} c_{C,D}(x_C, x'_D) \zeta_{D,x'_D}. \qquad (42)$$

For our purpose it is necessary that the linearized map P_ζ preserves the relaxation polytope Λ: $P_\zeta(\Lambda) \subset$

Λ. This constraint expresses as $(\forall \mu \in \Lambda)$

$$(P_\zeta \mu)_\varnothing = 1; \qquad (43a)$$

$$P_\zeta \mu \geqslant 0; \qquad (43b)$$

$$A P_\zeta \mu \geqslant 0. \qquad (43c)$$

We trivially have $(P_\zeta)_\varnothing = \zeta_\varnothing = 1$. It is also easy to show that $(P_\zeta)_{C,x_C,x'_C} \geqslant 0$ for $\zeta \in \mathcal{Z}$: before linearization, coefficients P_{C,x_C,x'_C} in the expression (36) are clearly non-negative and by property (S5) it is guaranteed that $(P_\zeta)_{C,x_C,x'_C} \geqslant 0$ holds on $\mathcal{Z}$. Then for $\mu \geqslant 0$ there holds $P_\zeta \mu \geqslant 0$. Interestingly, the converse is also true (but this result is not necessary in the subsequent construction):

Theorem 2.16. Inequalities (41c) in the definition of polytope $\mathcal{Z}$ can be equivalently replaced with $P_\zeta \geqslant 0$. Proof on p. 26.

There remains constraint (43c). In the case of standard local relaxations (to be defined later) constraint (43c) holds automatically and needs not be enforced. To account for general relaxations, we include constraint (43c) explicitly by representing it similarly to Theorem 2.8 in the dual form as:

$$(\exists \Phi \in \mathbb{R}_+^{m \times |\mathcal{I}|}) \ \ A P_\zeta - \Phi^\mathsf{T} A \geqslant 0. \qquad (44)$$

We arrive at the following relaxation of (MAX-WI) as a linear program:

$$\min_{\zeta,\varphi,\Phi} \sum_{s,i} \zeta_{s,i} \qquad (L1)$$

$$(I - P_\zeta^\mathsf{T}) f - A^\mathsf{T} \varphi \geqslant 0; \ \ \varphi \geqslant 0; \qquad (45a)$$

$$A P_\zeta - \Phi^\mathsf{T} A \geqslant 0; \ \ \Phi \geqslant 0; \qquad (45b)$$

$$\zeta \in \mathcal{Z}. \qquad (45c)$$

Constraint (45a) ensures that mapping P_ζ is relaxed-improving, constraints (45b) that it preserves the polytope: $P_\zeta(\Lambda) \subset \Lambda$ and constraint (45c) ensures that for each $C \in \mathcal{E}$ relaxed variables $(\zeta_{D,x_D} \mid D \subset C)$ stay in the local convex hull for C.

We claim that this relaxation is tight. As shown below, rounding down all components of ζ in a feasible solution maintains feasibility (with possibly different values of φ, Φ) and can only improve the objective. The rounding is performed by constructing the composite mapping $P_\zeta P_\zeta$. If P_ζ is relaxed-improving then so is $(P_\zeta)^2$ provided that it satisfies all feasibility constraints. The auxiliary lemma below establish this feasibility: it verifies that $P_\zeta^2 = P_{\zeta^2}$. Starting from a non-integer ζ and building a feasible sequence by taking $\zeta \mapsto \zeta^2$ we get each next point closer and closer to the integer limit.

11

Lemma 2.17. For $\zeta \in \mathcal{Z}$ there holds $P_\zeta^2 = P_{\zeta^2}$. Proof on p. 27.

Theorem 2.18. In a solution (ζ, φ, Φ) to (L1) vector ζ is integer.

Proof. Because ζ is feasible to (L1), the mapping P_ζ is Λ-improving for f. Note, at this point, unless ζ is integer it is not guaranteed that $P_\zeta(\mathcal{M}) \subset \mathcal{M}$ and we cannot draw any partial optimality from it, neither P_ζ is guaranteed to be idempotent. By constraints (45b), (45c), there holds $P_\zeta(\Lambda) \subset \Lambda$. Therefore

$$(\forall \mu \in \Lambda) \quad \langle f, P_\zeta P_\zeta \mu \rangle \leqslant \langle f, P_\zeta \mu \rangle \leqslant \langle f, \mu \rangle. \quad (46)$$

It follows that $P_\zeta^2 = P_\zeta P_\zeta$ is Λ-improving. Since $P_\zeta(\Lambda) \subset \Lambda$, it is also $P_\zeta^2(\Lambda) \subset P_\zeta(\Lambda) \subset \Lambda$.

By Lemma 2.17, there holds $P_\zeta^2 = P_{\zeta^2}$ and by Lemma 2.15 $\zeta^2 \in \mathcal{Z}$. By induction, there holds $P_\zeta^{2^n} = P_{\zeta^{2^n}}$, $P_\zeta^{2^n}(\Lambda) \subset \Lambda$ and $P_\zeta^{2^n}$ is Λ-improving. Let

$$\zeta^*_{C,x_C} = \lim_{n \to \infty} \zeta^{2^n}_{C,x_C} = [\![\zeta_{C,x_C} = 1]\!]. \quad (47)$$

Since P_{ζ^*} is Λ-improving and $\zeta^* \in \mathcal{Z}$, it is feasible to (L1). Assume for contradiction that there exist (s', i') such that $0 < \zeta_{s'i'} < 1$. From (47) we have $\zeta^*_{si} \leqslant \zeta_{si}$ for all s, i and $\zeta^*_{s'i'} < \zeta_{s'i'}$. It follows that ζ^* achieves a strictly better objective value, which contradicts the optimality of ζ. If all unary components $\zeta_{s,i}$ are integer then by Theorem 2.14 ζ is integer. $\square$

Corollary 2.19. The optimal solution to (L1) is unique.

Proof. Assume for contradiction that ζ_1, ζ_2 are two distinct integer solutions to (L1). Since (L1) is a linear program, their combination $\zeta = (\zeta_1 + \zeta_1)/2$ is an optimal solution (values of φ, Φ are omitted for clarity). But if $\zeta_1 \neq \zeta_2$ then ζ is not integer, which contradicts Theorem 2.18. $\square$

Clearly, for an integer vector $\zeta \in \mathcal{Z}$ the linearization P_ζ coincides with the extension of the discrete mapping p_ζ (34). It follows that the unique optimal solution to (L1) is the solution to (MAX-WI).

2.8. Perturbation for Strong Persistency

Problem (MAX-SI) over $\mathcal{P}^{2,y}$ can be reduced to (MAX-WI) with a perturbed cost vector $\tilde{f}$ as follows. It is sufficient to show that dual representation (32) of constraint $[p] \in \mathbb{S}_f$ can be reduced to that of $[p] \in \mathbb{W}_{\tilde{f}}$. For $p \in \mathcal{P}^{2,y}$ we can choose components of vector h

in the dual representation (32) of $\mathbb{S}_f$ as

$$h_C(x_C) = g_C(p_C(x_C)) - g_C(x_C), \quad (48a)$$

$$g_C(x_C) = \sum_{s \in C} [\![x_s = y_s]\!]. \quad (48b)$$

Clearly, $p_C(x_C) = x_C$ iff $h_C(x_C) = 0$ and for $p_C(x_C) \neq x_C$ there holds $1 \leqslant h_C(x_C) \leqslant |C|$. With such a vector h the dual representation of $\mathbb{S}_f$ can be written as

$$(f + \varepsilon g) - [p]^\mathsf{T}(f + \varepsilon g) - A^\mathsf{T}\varphi \geqslant 0, \quad (49)$$

i.e., the same constraint as (29) must hold but for an ε-perturbed cost vector

$$\tilde{f} := f + \varepsilon g. \quad (50)$$

Since the solution ζ to the perturbed problem is integer and unique it is the optimal solution to (MAX-SI).

2.9. Two-Phase Method

Let us consider the class $\mathcal{P}^{1,y}$. Formulation (L1) can be adopted by incorporating additional constraints on ζ (making variables ζ_{s,x_s} equal for all x_s). The proof of Theorem 2.18 is based on the fact that for a feasible ζ also ζ^2 is feasible. Clearly, this property is not destroyed by any equality constraints between components of ζ. Therefore Theorem 2.18 continues to hold and thus both (MAX-SI) and (MAX-WI) problems over $\mathcal{P}^{1,y}$ are tractable.

For class $\mathcal{P}^1$ the problem (MAX-SI) can be solved as proposed by Algorithm 1. It first solves the LP-relaxation in order to determine the *test labeling* y and then solves the (MAX-SI) problem for fixed y using perturbed (L1) for class $\mathcal{P}^{1,y}$.

Algorithm 1: Two Phase Method

1 $\mu \in \operatorname{argmin}_{\mu \in \Lambda}\langle f, \mu \rangle$;   /* solve (LP) */
2 For all s if there exists $i \in \mathcal{X}_s$ such that $\mu_s(i) = 1$ then set $y_s = i$, otherwise set y_s arbitrarily;
3 For strong persistency apply perturbation (50);
4 Solve the problem (L1) for the class of maps $\mathcal{P}^{2,y}$ or $\mathcal{P}^{1,y}$;

Theorem 2.20. Algorithm 1 solves (MAX-SI) over $\mathcal{P}^1$.

Proof. The necessary conditions of Lemma 2.7 for the optimal solution of LP-relaxation require that a strictly-improving mapping does not change optimal relaxed solutions. From the component-wise condition (28b) follows that when μ_s if fractional for some s then p_s (assuming $p \in \mathcal{P}^1$) must be identity. When

μ_s is integer, the only possible value of y_s qualifying necessary conditions must correspond to $\mu_s(y_s) = 1$. Applying perturbation in step 3 and optimizing over $\mathcal{P}^{1,y}$ in step 4 we obtain the optimal solution. $\square$

As a general heuristic, we can apply the same two-phase method, optimizing in step 4 over $\mathcal{P}^{1,y}$ or $\mathcal{P}^{2,y}$ with or without perturbation. The persistent assignment found by the heuristic is guaranteed to be at least as large as the solution of (MAX-SI) over $\mathcal{P}^1$.

3. Local LP Relaxations

In this section we consider a special case of *local* (or standard) LP relaxations in energy minimization [59, 58, 11, 38, 70], see also the survey by Werner [72]. In our notation local relaxations are described by the polytope Λ of the form

$$\Lambda = \{\mu \in \mathbb{R}^{\mathcal{I}} \mid A\mu = 0;\ \mu_\varnothing = 1;\ \mu \geqslant 0\}. \quad (51)$$

Recall that in the embedding δ, different components of a relaxed labeling μ, *e.g.*, μ_C and μ_D for $D \subsetneq C$ represent overlapping subsets of variables. In order that they represent all discrete labelings consistently they must satisfy marginalization constraints of the form

$$(\forall x_D \in \mathcal{X}_D) \sum_{x_{C \backslash D}} \mu_C(x_C) = \mu_D(x_D). \quad (52)$$

Werner [74] considers a family of LP relaxations generated by enforcing constraint (52) for some pairs of subsets $C \in \mathcal{E}, D \subsetneq C$. The set of such pairs is called the *coupling structure* [74]. For $C, D \subset \mathcal{V}$ we define coupling relation $D \mathrel{\dot\subsetneq} C$ of order d: let $D \mathrel{\dot\subsetneq} C$ iff

$$D \subsetneq C,\ C, D \in \mathcal{E},\ |D| \leqslant d. \quad (53)$$

Subsequently, we will consider two possibilities: to include only first order constraints or all of them. Zero order constraints (52) define just normalization:

$$\sum_{x_C} \mu_C(x_C) = \mu_\varnothing.$$

Together with non-negativity they guarantee boundedness (which was assumed in the general case §1.3). The first order constraints (52) add marginalization constraints of the form

$$(\forall s \in C,\ \forall x_s \in \mathcal{X}_s) \sum_{x_{C \backslash \{s\}}} \mu_C(x_C) = \mu_s(x_s). \quad (54)$$

And so on. By specifying larger d, we introduce more coupling between relaxed variables.

Note that any relaxation in the form (51) is local, *i.e.*, tied to the hypergraph. We cannot add more facets (inequalities) without increasing the number of variables and the variables are defined by the fixed embedding δ. Tightening the relaxation is thus only possible by enlarging the hypergraph (adding zero interactions in [72]), which results in an exponential increase in the number of relaxed variables. An example of a non-local relaxation is the cutting plane method [60], which progressively adds facet-defining inequalities coupling many variables at a time. While general results of §2 are applicable, the local representation would not be tractable.

The primal and dual LP relaxation problems for coupling $\mathrel{\dot\subsetneq}$ are expressed as follows:

$$(\forall D \mathrel{\dot\subsetneq} C) \quad
\begin{aligned}
\min\langle f, \mu\rangle &= \max \psi \\
\sum_{x_{C \backslash D}} \mu_C(x_C) = \mu_D(x_D), &\quad \varphi_{D,C}(x_D) \in \mathbb{R}, \\
\mu_\varnothing = 1, &\quad \psi \in \mathbb{R},
\end{aligned}$$

$$(\forall C \in \mathcal{E},\ \forall x_C)\ \mu_C(x_C) \geqslant 0, \quad f_C^\varphi(x_C) \geqslant \begin{cases} 0, & C \neq \varnothing, \\ \psi, & C = \varnothing. \end{cases}$$

Matrix A corresponds to primal equality constraints. Vector $f^\varphi = f - A^{\mathsf{T}}\varphi$ is an *equivalent transformation* [59] or *reparametrization* [70] of f. Its components are expressed as

$$f_C^\varphi(x_C) = f_C(x_C) - \sum_{D \mathrel{\dot\subsetneq} C} \varphi_{D,C}(x_D) + \sum_{H \mathrel{\dot\supsetneq} C} \varphi_{C,H}(x_C). \quad (55)$$

In particular, components $f_\varnothing^\varphi$ and $f_s^\varphi(x_s)$ are expressed as

$$f_\varnothing^\varphi = f_\varnothing + \sum_{C \mathrel{\dot\supsetneq} \varnothing} \varphi_{\varnothing,C}; \quad (56a)$$

$$f_s^\varphi(x_s) = f_s(x_s) - \varphi_{\varnothing,s} + \sum_{H \mathrel{\dot\supsetneq} \{s\}} \varphi_{s,H}(x_s). \quad (56b)$$

For any $\varphi \in \mathbb{R}^m$ there holds

$$(\forall \mu \in \Lambda)\ \langle f^\varphi, \mu\rangle = \langle f, \mu\rangle - \langle \varphi, A\mu\rangle = \langle f, \mu\rangle. \quad (57)$$

Since $\Lambda \supset \mathcal{M} \supset \delta(\mathcal{X})$, it follows that $E_f(x) = E_{f^\varphi}(x)$ for all $x \in \mathcal{X}$. Hence f^φ is indeed equivalent to f in defining the energy function.

The dual problem can be equivalently written as

$$\max\{f_\varnothing^\varphi \mid (\forall C \neq \varnothing)\ f_C^\varphi(x_C) \geqslant 0,\ \varphi \in \mathbb{R}^m\}, \quad (58)$$

we therefore can speak of the dual solution as just φ.

Complementary slackness Complementary slackness for (LP) reads that a feasible primal-dual pair (μ, φ) is optimal iff $(\forall C \in \mathcal{E} \backslash \{\varnothing\},\ \forall x_C)$

$$\mu_C(x_C) > 0 \implies f_C^\varphi(x_C) = 0. \quad (59)$$

Because a feasible dual solution satisfies $f_C^\varphi \geqslant 0$, the condition on the RHS of (59) implies that assignment x_C is *locally minimal* for f_C^φ: $x_C \in \arg\min f_C^\varphi(\cdot)$.

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

$$\mu_C(x_C) > 0 \iff f_C^\varphi(x_C) = 0. \tag{60a}$$

Clearly, a strictly complementary pair is complementary and thus it is optimal. Such a pair always exists and can be found by interior point algorithms (see *e.g.*, [68]). It is known that μ is a relative interior point of the primal optimal facet and φ is a relative interior point of the dual optimal facet.

Arc Consistency The following conditions, known as *arc consistency* (AC, *e.g.*, [74]), are satisfied for strictly complementary pairs:

- If $f_C^\varphi(x_C) = 0$ then $(\forall D \subsetneq C)\ f_D^\varphi(x_D) = 0$.
- If $f_D^\varphi(x_D) = 0$ then $(\forall C \supsetneq D)\ (\exists x_C' \in \mathcal{X}_C \mid x_D' = x_D)$ $f_C^\varphi(x_C') = 0$.

These conditions say that the set of local minimizers must be consistent over overlapping hyperedges. Arc consistency is a necessary but, in general, not sufficient condition for strict complementarity.

BLP The relaxation with marginalization constraints of order 1 is known as Basic LP relaxation (BLP) [74]. Note, if we do not enforce marginalization constraints of at least order 1 there may occur integer feasible solutions to the relaxation which are not consistent, *i.e.*, do not correspond to a global assignment. Out of all local relaxations BLP is the least constrained useful one. It is remarkable that it is tight for all tractable languages [66, 67, 30]. However, for certain purposes BLP is not sufficient, as can be illustrated with pseudo-Boolean functions. Suppose we would like to express a pseudo-Boolean function of 3 variables as a cubic polynomial. We know it can be expressed in this form, however, such a desired equivalent transformation of the problem appears to be not equivalent for BLP and hence not equivalent for the maximum persistency problem. Another difficulty is that fixing a variable to its optimal value is not the same as eliminating this variable. Example in Figure 7 illustrates that eliminating a persistent variable tightens the relaxation. It follows that under BLP relaxation we won't be able to compare theoretically neither to quadratization techniques (as they perform general equivalent transformations) nor to generalized roof duality [23], which incrementally eliminates persistent variables.
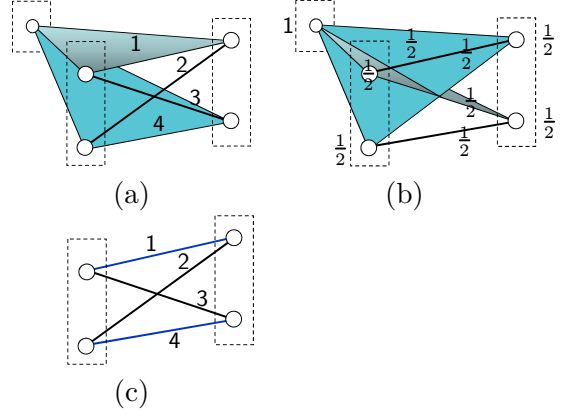
Figure 7: An example when fixing a variable in BLP relaxation is not equivalent to eliminating it. (a) Energy in 3 variables (the leftmost variable has only one possible assignment). Costs 2 and 3 are assigned to the pairwise term (solid lines) and 1 and 4 to triplewise (faces), other costs are zero. (b) An optimal solution to BLP relaxation (of cost 0). BLP relaxation does not enforce marginalization between the triple and the pair. (c) The energy after elimination of the dummy variable. Now BLP relaxation is tight and can determine the optimal labeling of cost 1.

FLP The relaxation with all marginalization constraints present will be refereed to as Full local LP relaxation (FLP). For every hyperedge all its subsets are assumed to be contained in $\mathcal{E}$ and all constraints of the form (52) with d equal to the order of the problem are included. In case of pairwise model, individual nodes are the only proper subsets of edges and hence BLP and FLP are the same. In the pseudo-Boolean case, FLP matches the relaxation of Sherali and Adams [58] as discussed in §A.9.

3.1. Maximum Persistency with Local Relaxations

In this section we summarize how the general construction and formulation of (L1) simplifies for local relaxations. First, the constraint $P_\zeta(\Lambda) \subset \Lambda$ holds automatically and needs not be enforced. It is shown in two steps: first we consider the linear extension $[p]$ of any node-wise mapping p and then the linearized mapping P_ζ, $\zeta \in \mathcal{Z}$.

Lemma 3.1. Node-wise mapping $[p]$ preserves the local polytope Λ. Proof on p. 27.

Lemma 3.2. Mapping P_ζ for $\zeta \in \mathcal{Z}$ preserves the local polytope: $P_\zeta(\Lambda) \subset \Lambda$. Proof on p. 27.

In short, P_ζ satisfies all the equality constraints satisfied by $[p]$ and has all components non-negative for $\zeta \in \mathcal{Z}$. As a consequence of Lemma 3.2, the constraint of polytope preservation (45b) in the maximum persistency problem (L1) can be dropped. We can write

14

(MAX-WI) as

$$\min_{\zeta,\varphi} \sum_{s,i} \zeta_{s,i} \qquad (61)$$

$$(I - P_\zeta^\mathsf{T})f - A^\mathsf{T}\varphi \geqslant 0; \ \varphi \in \mathbb{R}^m;$$

$$\zeta \in \mathcal{Z}.$$

Further properties of improving mappings for local relaxations are as follows. Conditions that are necessary for $[p] \in \mathbb{W}_f$ in the general case (Theorem 2.10) become necessary and sufficient for local relaxations and can be now summarized together with the dual representation Theorem 2.8:

Theorem 3.3 (Characterizations)**.** For a local relaxation Λ all of the following are equivalent:

   (a) $P \in \mathbb{W}_f$;
   (b) $(\exists \varphi \in \mathbb{R}^m) \ f^\varphi - P^\mathsf{T}f \geqslant 0$;
   (c) $(\exists \varphi \in \mathbb{R}^m) \ (I - P^\mathsf{T})f^\varphi \geqslant 0$;
   (d) $\inf\{\langle f - P^\mathsf{T}f, \mu\rangle \mid \mu \in \mathbb{R}_+^\mathcal{I}, \ A(I - P)\mu = 0\} = 0$.

Proof on p. 28.

We have transitions from a global property $(\forall \mu \in \Lambda)$ $\langle(I - P^\mathsf{T})f, \mu\rangle \geqslant 0$ (a) to component-wise local inequalities (b) and (c). Inequalities (c) offer an equivalent reparametrization f^φ in which mapping p improves every component independently:

$$(\forall \text{C} \subset \mathcal{E}, \ \forall x_\text{C} \in \mathcal{X}_\text{C}) \ f_\text{C}^\varphi(p_\text{C}(x_\text{C})) \leqslant f_\text{C}^\varphi(x_\text{C}). \qquad (62)$$

This is a fairly simple condition similar in spirit to the idea of equivalent transformations by Shlezinger [59] (find an equivalent f^φ such that the global minimum may be recovered from independent component-wise minima). Condition (d) is a primal reformulation which has fewer equality constraints than the verification LP and hence is simpler.

Some properties expressed for all hyperedges $\text{C} \in \mathcal{E}$ can be simplified if we assume at least the BLP relaxation. In Statement 2.6 it is sufficient that only unary components satisfy $(\forall s \in \mathcal{V}) \ \mathcal{O}_s = p_s(\mathcal{X}_s)$. For other components the constraint is implied by marginalization. For the same reason, in the perturbation (50) it is sufficient to have only unary components $f_s(y_s)$ increased by ε for all s and leave higher-order terms intact.

Lastly, there are following NP-hardness results with BLP relaxation:

Theorem 3.4. Problem (MAX-WI) over the $\mathcal{P}^1$ class of maps and the BLP relaxation is solvable in polynomial time for the quadratic pseudo-Boolean case and otherwise (when the problem is multilabel or higher order) it is NP-hard. Proof on p. 28.

Theorem 3.5. Problem (MAX-SI) with 4 or more labels over the class of maps $\mathcal{P}^2 = \bigcup_{y \in \mathcal{X}} \mathcal{P}^{2,y}$ and BLP relaxation is NP-hard. Proof on p. 28.

We see that the difference between weak and strong persistency leads to different complexity classes for the maximum persistency problem. The question of complexity of (MAX-SI) with 3 labels is not resolved.

4. Comparison Theorems

This section is devoted to theoretical comparison between different persistency techniques. The firs result is the following:

Theorem 4.1. Let $\Lambda \subset \Lambda'$ and P be a Λ'-improving mapping for f. Then P is Λ-improving for f.

Proof. The claim follows from Definition 2.2 and nesting of polytopes $\Lambda \subset \Lambda'$. $\qquad\square$

We therefore have a natural hierarchy: if we can identify some variables as persistent by the proposed sufficient condition with relaxation Λ' then for any tighter relaxation $\Lambda \subset \Lambda'$ we are guaranteed to find at least the same persistent variables. Other nesting results under different reformulations of the problem are obtained in §4.6, §A.7.

Table 3 gives an overview of the obtained comparisons to other methods. The first comparison column establishes that all listed methods correspond to a relaxed-improving mapping under standard relaxations (recall that in the pairwise case FLP = BLP). For cases when Algorithm 1 is optimal, as indicated in Table 1, it is guaranteed to find the same or larger set of persistent labels than any other method. This fills the second comparison column in Table 3. In the remainder of this section we give a more detailed overview of different methods and comparison results.

4.1. DEE

We will consider Goldstein's *simple* DEE [17] (which is stronger than original DEE by Desmet et al. [14]) in the pairwise setting. For every node s this method considers its neighbors in the graph, $\mathcal{N}(s)$, and for a pair of labels $\alpha, \beta \in \mathcal{X}_s$ verifies the condi-

		Corresponds to a BLP/FLP-improving mapping	Dominated by Algorithm 1
pairwise multilabel	Simple DEE [17]	✓	-
	MQPBO [25]	✓	-
	Kovtun's one-against-all [39]	✓	✓
	Kovtun's iterative [40]	✓	-
	Swoboda et al. [64]**	✓	✓
higher order pseudo-Boolean	Roof dual / QPBO [44, 20, 6, 32]	✓	=*
	Reductions: HOCR [22], Fix et al. [15]	FLP	✓*
	Bisubmodular relaxations [27]***	BLP	✓*
	Generalized Roof Dualilty [23]	FLP	✓*
	Persistency by Adams et al. [2]	FLP	✓*

Table 3: Summary of theoretical comparisons. *Result holds for strong persistency variants and resp. strict version of Algorithm 1. **[64] is higher order but the comparison proof is for the pairwise case. ***Result holds for sum of bisubmodular functions over the same hypergraph as the BLP relaxation.
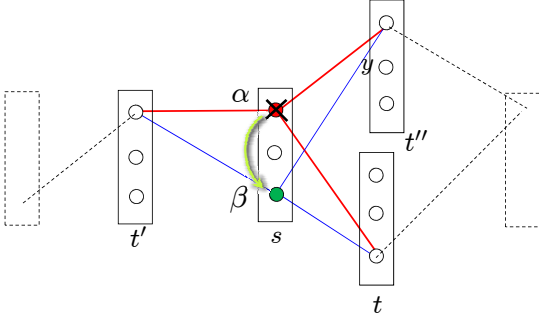

Figure 8: Improving mapping corresponding to an individual DEE condition. The full DEE method iterates over all nodes and labels and composes the found improving maps.

tion

$$(\forall x \in \mathcal{X}_{\mathcal{N}(s)}) \qquad (63)$$
$$f_s(\alpha) - f_s(\beta) + \sum_{t \in \mathcal{N}(s)} [f_{\{s,t\}}(\alpha, x_t) - f_{\{s,t\}}(\beta, x_t)] \geqslant 0,$$

illustrated in Figure 8. If the condition is satisfied it means that a (weakly) improving switch from α to β exists for an arbitrary labeling x. In this case (s, α) can be eliminated while preserving at least one optimal assignment.

It is trivial to construct an improving mapping for this case. We let $p_s(\alpha) = \beta$, $p_s(i) = i$ for $i \neq \alpha$; and $p_t(i) = i$ for all $t \neq s$. The non-zero terms of the problem $g = (I - P^\mathsf{T})f$ form a tree with root node s and other nodes $t \in \mathcal{N}(s)$ being leaves. It is known that in this case the FLP relaxation is tight and therefore p is FLP-improving. Similarly, the strict inequality in (63) implies $[p] \in \mathbb{S}_f$.

4.2. QPBO

Let $\mathcal{X}_s = \mathbb{B}$. The *weak persistency* theorem [44, 20] can be formulated as follows. Let $\mu \in$ $\mathrm{argmin}_{\mu \in \Lambda} \langle f, \mu \rangle$. Let $\mathcal{O}_s = \{i \in \mathbb{B} \mid \mu_s(i) > 0\}$. Then

$$(\exists x \in \mathrm{argmin}_x E_f(x)) \ (\forall s \in \mathcal{V}) \ x_s \in \mathcal{O}_s. \qquad (64)$$

In the case $|\mathcal{O}_s| = 1$ vector μ_s is necessarily integer and the theorem states that there is an optimal solution x to the discrete problem which is consistent with the integer part of the relaxed solution μ. The largest weakly persistent assignment is obtained in the case μ is the solution with the maximum number of integer components.

The *strong persistency* theorem [44, 20] can be formulated as follows.

Theorem 4.2 ([44, 20]). Let (μ, φ) be a strictly complementary primal-dual pair. Let $\mathcal{O}_s$ be defied as above. Then

$$(\forall x \in \mathrm{argmin}_x E_f(x)) \ (\forall s \in \mathcal{V}) \ x_s \in \mathcal{O}_s. \qquad (65)$$

The difference to (64) is in the quantifier $\forall$ *vs.* $\exists$. Note, a strictly complementary solution has the minimum number of integer components.

Theorem 4.3 ([55]). Weak (resp. strong) persistency by QPBO corresponds to an FLP-improving (resp. strict FLP-impriving) mapping.

The mapping is defined by $p_s(i) = 0$ if $O_s = \{0\}$, $p_s(i) = 1$ if $O_s = \{1\}$ and $p_s(i) = i$ otherwise. The idea of the proof is to show that the dual optimal solution φ provides the reparametrization in which the mapping improves every component independently, *i.e.*, satisfies the inequalities of the characterization Theorem 3.3(c).

It follows from the theorem that solution by Algorithm 1 with perturbation coincides with the strong QPBO persistency.

4.3. MQPBO

The MQPBO method [25] extends partial optimality properties of QPBO to multilabel problems via the reduction of the problem to 0-1 variables. The reduction, known as "K to 2" transform [51] ($K = |\mathcal{X}_s|$), depends on the linear ordering of labels in $\mathcal{X}_s$. The method outputs two labelings $x^{\min}$ and $x^{\max}$ with the guarantee that there exists an optimal labeling x that satisfies $x_s \in [x_s^{\min}, x_s^{\max}]$. The corresponding improving mapping has the form $p \colon x \mapsto (x \vee x^{\min}) \wedge x^{\max}$, where $\vee$ and $\wedge$ are component-wise minimum and maximum, resp. in a given ordering of labels. The mapping is illustrated in Figure 9. Because the reduction is component-wise
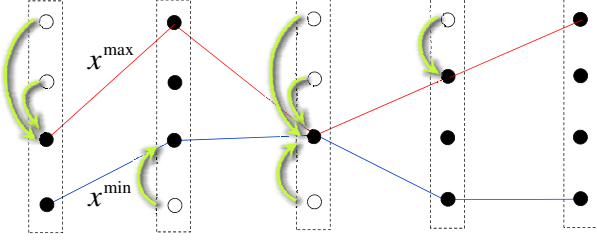
Figure 9: Improving mapping found by the MQPBO method: all labels above $x^{\max}$ are mapped to $x^{\max}$ and all lables below $x^{\min}$ to $x^{\min}$. Labelings $x^{\max}$ and $x^{\min}$ are determined by the method from the underlying roof dual relaxation.

and component-wise inequalities hold for QPBO it follows that the component-wise conditions of Theorem 3.3(c) hold for p (proof in [55]). For $(x^{\min}, x^{\max})$ obtained from weak (resp. strong) persistency by QPBO there holds $[p] \in \mathbb{W}_f$ (resp. $[p] \in \mathbb{S}_f$). Since the class of mappings of the form $x \mapsto (x \vee x^{\min}) \wedge x^{\max}$ is not among the cases for which Algorithm 1 is optimal, the question of tractability of (MAX-SI) for this class remains open.

4.4. Auxiliary Submodular Problems by Kovtun

There were several methods proposed [39, 40] which differ in detail. All methods construct an auxiliary submodular energy E_g. A minimizer y of E_g has the property that $E_g(x \vee y) \leqslant E_g(x)$, implied by submodularity. It follows that mapping $p \colon x \mapsto x \vee y$ is improving for g. Figure 10 illustrates such mappings found by two of the methods in [40]. In case (a) the test labeling y must be the highest (the maximum) in the selected order of label. It follows that the mapping p is essentially of the form $x \mapsto x[\mathcal{A} \leftarrow y]$, *i.e.*, from the class $\mathcal{P}^{1,y}$. The construction of the auxiliary function ensures that improvement in f is at least as big as improvement in g and so p is improving for f.

Theorem 4.4 ([55])**.** Persistency by any method [39, 40] corresponds to an FLP-improving mapping.

Since in the case (a) the mapping is in the class $\mathcal{P}^1$, we know that the strict version of the method is dominated by Algorithm 1. In the case (b), the class of maps is a subset of maps considered in MQPBO and tractability of (MAX-SI) problem is also open.

Computationally, methods of Kovtun [40] have an advantage as they rely on the minimization of a pairwise submodular function. In the case of the Potts model, the method [39] for all "flat" test labelings $y_s = \alpha$ for $\alpha = 1 \ldots K$, $(K = |\mathcal{X}_s|)$, can be efficiently performed using $\log(K)$ maximum flow computations [19]. It is very practical in some vision
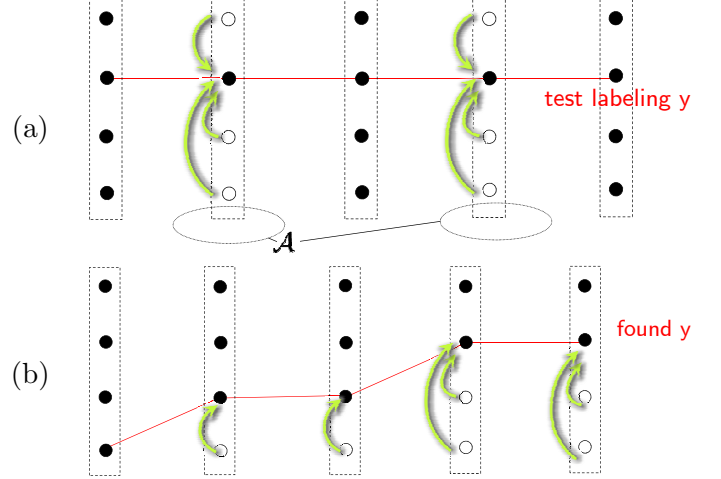
Figure 10: Improving mappings in Kovtun's methods. (a) One-against-all method for fixed test labeling y. The method determines a subset $\mathcal{A}$ of vertices for which the optimal labeling is y. (b) Iterative method [40] in which labeling y is found incrementally and with respect to a predefined ordering of labels.
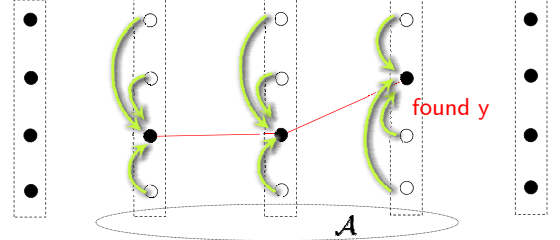
Figure 11: Improving mapping in the method of Swoboda et al. [64]. The method finds the labeling y and a subset $\mathcal{A}$. Outside $\mathcal{A}$ the mapping is identity.

problems (*e.g.* results [39, 3, 19]), where unary costs are determining. At the same time experiments on difficult random problems in §5 reveal very poor performance of this method.

4.5. Iterative Pruning by Swoboda et al.

The iterative Pruning method was first proposed [63] for the Potts model and then extended to general pairwise and higher order energies [64]. The method can be interpreted as finding an improving mapping in the class $\mathcal{P}^1$ (Figure 11).

Theorem 4.5. Persistency by method [64] in the pairwise multilabel case corresponds to an FLP-improving mapping. Proof on p. 28.

In fact the optimal value of y is determined in [64] by the initial relaxation, similarly to how it is determined in Algorithm 1. Therefore, Algorithm 1 without perturbation identifies the same or better weak persistency. Algorithm 1 with perturbation identifies the same or larger set $\mathcal{A}_{\text{strong}}$ as theoretically guaranteed in [64].

4.6. Quadratization Techniques

We now turn to the higher order pseudo-Boolean case. There is a number of different reductions proposed [22, 15, 5] which represent the initial function of 0-1 variables as a minimum of a quadratic function over auxiliary 0-1 variables. Persistency is obtained then by applying the QPBO method to the reduced problem. Since QPBO solves the FLP relaxation, our goal is to compare local relaxations as well as relaxed-improving maps before and after the reduction. Fortunately, full reductions [22, 15] are defined by chaining certain elementary reductions applied to separate cliques or groups of cliques. We define a sufficient set of atomic reductions with the following property: the maximum persistent subset by an FLP-improving mapping for the reduced problem is not larger than that one for the initial problem. Chaining these atomic reductions we obtain the following comparisons.

Theorem 4.6. Persistency by Higher Order Clique Reduction (HOCR) of Ishikawa [22] corresponds to an FLP-improving mapping. Proof on p. 31.

Theorem 4.7. Persistency by method of Fix et al. [15] corresponds to an FLP-improving mapping. Proof on p. 32.

Ishikawa [22] proposed a family of elementary reductions (called γ-flipping) and posed the problem of what sequence of reductions gives in a certain sense the best overall reduction. This is a difficult combinatorial problem. While we do not address it directly, it follows that FLP maximum persistency by Algorithm 1 dominates persistencies that can be obtained by any reduction from the family and hence also the best one.

4.7. Bisubmodular Relaxations

Submodular/bisubmodular relaxations were introduced by Kolmogorov [27] as a natural generalization of the roof duality approach to higher order pseudo-Boolean functions. Kolmogorov showed that all totally half-integral relaxations are bisubmodular relaxations and vice versa. Similar to roof duality, (bi)submodular relaxations have a global persistency property. However, to a given function many different (bi)submodular relaxations can be constructed. There are two challenges in this approach. One is that the class of all (bi)submodular relaxations is very large and it is not tractable to parametrize it. The other challenge is to answer the question of which relaxation provides the largest persistent assignment.

Kahl and Strandmark [23] build upon graph-cut reducible submodular relaxations. They propose that the relaxation which corresponds to the best lower bound on the energy is the optimal one. Their algorithm solves a series of linear programs to build the tightest graph-cut reducible submodular relaxation. However, not all submodular relaxations are graph-cut reducible (it is a hard problem to determine which ones actually are [69] with the exception of cubic functions). Moreover, it is not clear whether the relaxation that gives the best lower bound is also the best one w.r.t. the size of the persistent assignment.

We consider a more general case when the relaxation is a sum of bisubmodular functions (SoB) over the same hypergraph as f. This class includes all graph-cut reducible submodular relaxations. Exploiting the property that SoB function can be minimized exactly by BLP relaxation [66] and properties [27], we obtain the following theorem.

Theorem 4.8. Persistency by SoB relaxation [27] corresponds to a BLP-improving mapping. Proof on p. 34.

The work of Lu and Williams [43] is a special case of SoB relaxation, it follows that their result corresponds to a BLP-improving mapping as well.

4.8. Generalized Roof Duality

As discussed above, the method of Kahl and Strandmark [23] finds persistencies by SoB relaxation and all such relaxations are dominated by BLP-improving maps. There is however a catch. The method reduces the problem progressively by finding in each iteration a BLP-improving map. While Lemma 2.7 guarantees that fixing variables to their persistent values does not tighten the BLP-relaxation, eliminating persistent variables actually does (as explained in Figure 7). It follows that their method is not in general dominated by a single BLP-improving map. On the other hand, we can easily claim domination by a single FLP-improving map (as it is stronger than BLP and is not tightened by the elimination of persistent variables), which is also confirmed experimentally.

While avoiding the difficult question [27, 23] of how to find the best SoS or SoB relaxation, we give an answer to how to find same or larger strong persistent assignment. In the case of 3rd order energies

(quartic terms), Kolmogorov [27] gives an example where there is a tight bisubmodular relaxation but no tight submodular relaxation. It follows that in this case Algorithm 1 can determine strictly larger strong persistent assignment than [23]. We give experimental confirmation of larger persistent set for both cubic and quartic problems. The proposed two-phase algorithm is seen more computationally attractive than the series of LPs of Kahl and Strandmark [23].

Windheuser et al. [75] extended generalized roof duality [27, 23] to multilabel case. For pairwise models they showed equivalence with MQPBO. For higher order models the approach can be seen as a combination of K to 2 transform [51] and application of submodular relaxation [27, 23]. As we have given comparison with (bi)submodular relaxations and K to 2 transform is component-wise, it should follow that the sufficient condition of [75] corresponds to an FLP-improving mapping.

4.9. Persistency in 0-1 Polynomial Programming by Adams et al.

Adams et al. [2] proved a persistency result for the 0-1 polynomial programming problem. The result is based on the relaxation of Sherali and Adams [58], which can be identified with the FLP relaxation. They proposed a sufficient condition on the dual multipliers in the relaxation which provides a persistency guarantee. The sufficient condition is a linear feasibility program that can be verified for a given partial assignment, similar in spirit to our verification LP. No method to find a persistent partial assignment except for the case when the integer part of the optimal relaxed solution turns out to be persistent is proposed. We show the following.

Theorem 4.9. Persistency by the sufficient condition of Adams et al. [2, Lemma 3.2] corresponds to an FLP-improving mapping. Proof on p. 36.

In fact their sufficient conditions splits the problem into two overlapping parts: one part, where an optimal assignment is unknown (call it the inner problem) and the second part, containing all the assigned variables and the coupled unassigned variables (call it outer problem). It can be seen that the sufficient condition guarantees that any choice of unassigned variables together with the assigned ones delivers an optimal solution to the outer problem. Thus what happens in the inner problem can be efficiently ignored. Any feasible solution of the inner problem is optimal to the outer one.

5. Experiments

We propose two families of experiments: for pairwise multilabel energies and higher-order pseudo-Boolean energies. We first discuss linear program (L1) for FLP relaxation in these two cases.

5.1. Details of L1 Program

Explicit form of (L1) for the pairwise multilabel case is given in [55]. It is expressed with variables ξ which are related to ζ by $\xi_{s,i} = 1 - \zeta_{s,i}$; $\xi_{st,ij} = 1 - \zeta_{s,i} - \zeta_{t,j} + \zeta_{st,ij}$, so that $\xi_{st,ij}$ is the linearization of $\xi_{s,i}\xi_{t,j} = (1 - \zeta_{s,i})(1 - \zeta_{t,j})$. This parametrization is more convenient in the pairwise case. Its drawback is a more complex representation of P_ζ^2 (which is however not needed except for the proof).

Let us consider now the pseudo-Boolean case. Without loss of generality we assume that $y = 0$ (otherwise variable values can be flipped). Let f be given in the form of a multilinear polynomial: $f_C(x_C) = \eta_C \prod_{s \in C} x_s$. We let ζ_C denote $\zeta_{C,1_C}$. The expression $P_\zeta f$ simplifies as

$$(P_\zeta f)_C(x'_C) = \sum_{x_C \in \mathcal{X}_C} \sum_{D \subset C} c_{C,D}(x_C, x'_C) \zeta_{D, x'_D} f_C(x_C) \quad (66)$$

$$= \sum_{D \subset C} \prod_{s \in D} (\llbracket x'_s {=} 1 \rrbracket - \llbracket y_s {=} 1 \rrbracket) \prod_{s \in C \setminus D} \llbracket y_s {=} 1 \rrbracket \zeta_{D, x'_D} f_C(1_C)$$

$$= \llbracket x'_C {=} 1_C \rrbracket \zeta_C \eta_C.$$

Components of $(I - P_\zeta^\mathsf{T}) f$ simplify as

$$f_C(x'_C) - \llbracket x'_C {=} 1_C \rrbracket \zeta_C \eta_C = \llbracket x'_C {=} 1_C \rrbracket (1 - \zeta_C) \eta_C.$$

The set $\tilde{\mathcal{X}}_s = \mathcal{X}_s \setminus \{y_s\}$ equals simply $\{1\}$ and thus polytope $\mathcal{Z}$ simplifies as

$$\zeta_\varnothing = 1, \quad (67)$$

$$(\forall C \in \mathcal{E}, \ \forall D \subset C) \quad \sum_{D \subset C \setminus B} (-1)^{|D|} \zeta_D \geqslant 0.$$

Problem (L1) can be written as

$$\min_{\zeta, \varphi} \sum_{s \in \mathcal{V}} \zeta_s \quad (68)$$

$$(\forall C \in \mathcal{E}) \qquad \zeta_C \eta_C - (A^\mathsf{T} \varphi)_C (1_C) \geqslant 0;$$
$$(\forall C \in \mathcal{E})(\forall x'_C \neq 1_C) \qquad -(A^\mathsf{T} \varphi)_C (x'_C) \geqslant 0;$$
$$-(A^\mathsf{T} \varphi)_\varnothing \geqslant 0;$$
$$\zeta \in \mathcal{Z}.$$

Implementation in matlab is available at `http://www.icg.tugraz.at/Members/shekhovtsov/persistency` for research purposes.

We evaluated all methods on small random problems. The purpose of the experiments is to validate the theory and to verify whether the improvement obtained by the new method is not negligible. For all methods, including ours, for each instance we numerically verified that:

- map p constructed by the method is relaxed improving w.r.t. FLP relaxation (by solving the verification LP (24)).
- the persistency guarantee is correct (by solving exactly the initial and the reduced problems).

We measure *solution completeness* as $\frac{n_{\text{elim}}}{|\mathcal{V}|(K-1)}100\%$, where K is the number of labels in every node ($|\mathcal{X}_s| = K$) and n_{elim} is the total number of labels ($s \in \mathcal{V}, i \in \mathcal{X}_s$) eliminated by the method as non-optimal.

Pairwise Multilabel Models We report results on random problems with Potts interactions and full interactions. Both types have unary weights $f_s(i) \sim U[0, 100]$ (uniformly distributed). Full random energies have pairwise terms $f_{st}(i,j) \sim U[0, 100]$ and Potts energies have $f_{st}(i,j) = -\gamma_{st}(i)[\![i{=}j]\!]$, where $\gamma_{st}(i) \sim U[0, 50]$. All costs are integer to allow for exact verification of correctness. Only instances with non-zero integrality gap w.r.t. FLP relaxation are considered (non-FLP-tight). The results are shown in Figure 12, while Table 4 gives details of the methods.

Higher Order Binary Models The proposed evaluation of higher order 0-1 models is based on the `submodular` library [61, 24]. The library interfaces quadratization techniques **HOCR** [22] and **Fix *et al.*** [15] and implements three variants of generalized roof duality [24], **GRD***. Figure 13 shows evaluation on random polynomials of degrees 3 and 4, sampled by the library. In the first series, we reproduce results [24] with similar parameters but smaller problems (*e.g.*, $n = 100$ variables and $T = 30$ multilinear terms *vs.* $n = 1000$ and $T = 300$ in [24]). The results for baseline methods are consistent with [24]. It turned out however that most of the instances are FLP-tight. Our method, as well as [64], reduces in this case to solving the FLP relaxation and gives the trivial 100% persistency result. In the second series we increased the complexity by adding more terms as well as selecting only non-FLP-tight instances. The proposed approach determines a significantly larger persistent assignment.

In Figure 14, we generated grid problems of degree 3 with hyperedges $\{(i,j), (i+1,j), (i,j+1)\}$ and of degree 4 with hyperedges $\{(i,j), (i+1,j), (i,j+1), (i+1,j+1)\}$ at every grid location $(i,j) \in \{1, \ldots N-1\}^2$. For each such hyperedge C we sampled the term f_C as a random posiform ($2^{|C|}$ uniformly distributed numbers, one per configuration, as opposed to sampling coefficients of multilinear polynomials in Figure 13). This results in somewhat more difficult problems to solve as there is no bias from an unsymmetrical treatment. We further selected only non-FLP-tight instances. It turns out that for the class of the problems of degree 4 none of the baseline methods identified more than $1 - 2\%$ of the optimal solution, in contrast to the proposed method.

Running Time Figure 14(b) gives a rough idea of running times when using CPLEX to solve linear programs. The running time for L1-FLP and Swoboda *et al.* includes only the time to solve linear programs and excludes all data preparation in matlab. The method of Swoboda *et al.* is the slowest one because it needs to solve several LP relaxations in the inner loop (but see [64, 65] for applicability with suboptimal solvers and incremental computation). The proposed two-phase methods (L1-FLP) solves two linear programs. Somewhat unexpectedly, the initialization phase (FLP) takes more than a half of the total time. The optimal version of GRD performs similarly to the proposed method but determines less variables. GRD-heur is much faster while the result is comparable to GRD. It can be concluded that for practical applicability of the proposed method a feasible but possibly only approximately maximal solution should be found.

6. Conclusions

Techniques for partial optimality avoid the NP-hardness of the energy minimization problem by exploiting different sufficient conditions by which a part of optimal solution can be found. We proposed a new sufficient condition corresponding to a given polyhedral relaxation and verifiable in polynomial time. The condition generalizes the mechanism of improving mapping which is present in many works (although often in a hidden form) and allows to explain them from this perspective. We can explain variety of methods originating in different fields and relate these methods to linear relaxations. In particular, it follows that all covered methods cannot be used to tighten FLP relaxation. Applying them as a preprocessing in solving the FLP relaxation may only
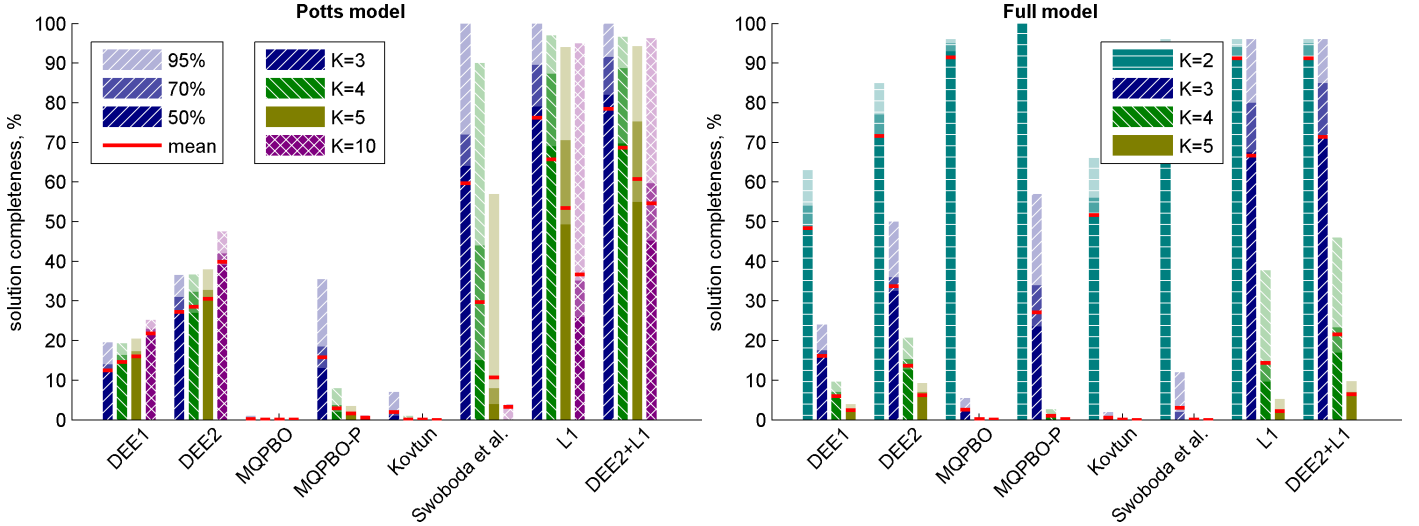
Figure 12: Evaluation of pairwise multilabel methods. Problems size is 10x10, 4-connected. Bars of different shades indicate the portion of the sample under the given solution completeness value (statistics over 100 instances).

DEE1	Goldstein's Simple DEE [17]: 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: $f_s(\alpha_s) - f_s(\beta_s) + f_t(\alpha_t) - f_t(\beta_t) + f_{st}(\alpha_{st}) - f_{st}(\beta_{st}) + \sum_{t' \in \mathcal{N}(s) \setminus \{t\}} \min_{x_{t'}} [f_{st'}(\alpha_s, x_{t'}) - f_{st'}(\beta_s, x_{t'})] + \sum_{t' \in \mathcal{N}(t) \setminus \{s\}} \min_{x_{t'}} [f_{tt'}(\alpha_t, x_{t'}) - f_{tt'}(\beta_t, x_{t'})] \geqslant 0$.
MQPBO(-P)	The method of Kohli $et\ al.$ [25]. The problem reduced to $\{0,1\}$ variables is solved by QPBO(-P) [48], where "-P" is the variant with probing [8]. 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 [40]. 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. Reimplementation.
Swoboda $et\ al.$	Iterative Pruning method of Swoboda et al. [64] using CPLEX [1] for each iteration. Reimplementation.
L1	The proposed method in Algorithm 1 for class P^2 without perturbation, both phases solved with CPLEX [1].
DEE2+L1	Sequential application of DEE2 and L1. Note, DEE2 uses a condition on pairs which is not covered by the proposed sufficient condition under pairwise BLP relaxation.

Table 4: List of tested methods for pairwise multilabel evaluation.

speed it up but cannot change the set of optimal relaxed solutions. We formally posed and studied the problem of determining the largest set of persistent variables subject to the general sufficient condition. It appeared that there are reasonably large classes of this problem (restricted by the set of allowed mappings) which can be solved in polynomial time. While the proposed solution might not be the most efficient, its generality allows to subsume multiple problem reformulations, reductions, equivalent transformations and choices that other existing techniques depend on. In bisubmodular relaxations, this is the choice of a bisubmodular lower bound function, in method [40] the choice of the order of labels and the test labeling, in methods [22, 15] choice of the sequence of the re-

ductions and flips. While optimizing these methods w.r.t. to all such choices does not seem tractable, it is tractable to find a persistent assignment (by the proposed method) which is at least as good as if these choices were optimized over.

In the experimental evaluation we verified that our theoretical comparisons hold true, $i.e.$ that all evaluated methods (except DEE2 for which we do not claim anything) have output FLP-improving maps in all test cases. Our linear program (L1) had always integer optimal solution[4] ζ. The persistent assignment found by our method with FLP-relaxation was larger per instance and significantly larger on average.

---

[4]With exception of few cases when CPLEX experienced a numerical error.
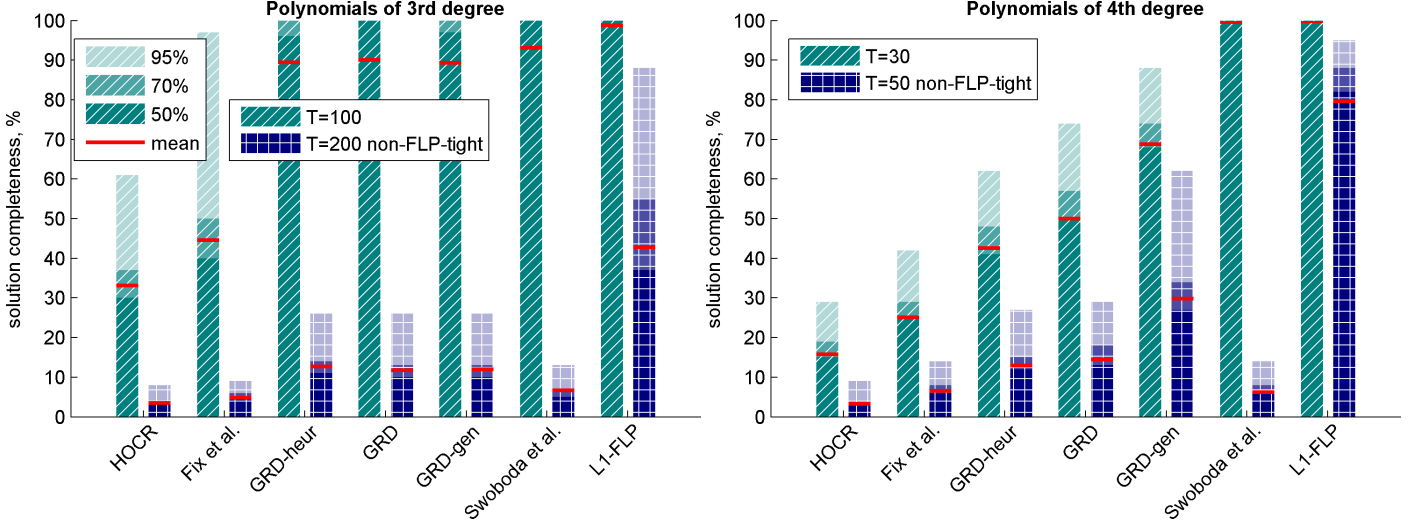
Figure 13: Evaluation of higher-order binary methods on random polynomials generated as [24] for $n = 100$ variables. Plots with $(d = 3, T = 100)$ and $(d = 4, T = 30)$ reproduce results reported in [24]. In is seen that most of the instances are FLP-tight and thus solved exactly by L1. We increase complexity by evaluating $(d = 3, T = 200)$ and $(d = 4, T = 50)$ and selecting only non FLP-tight instances.
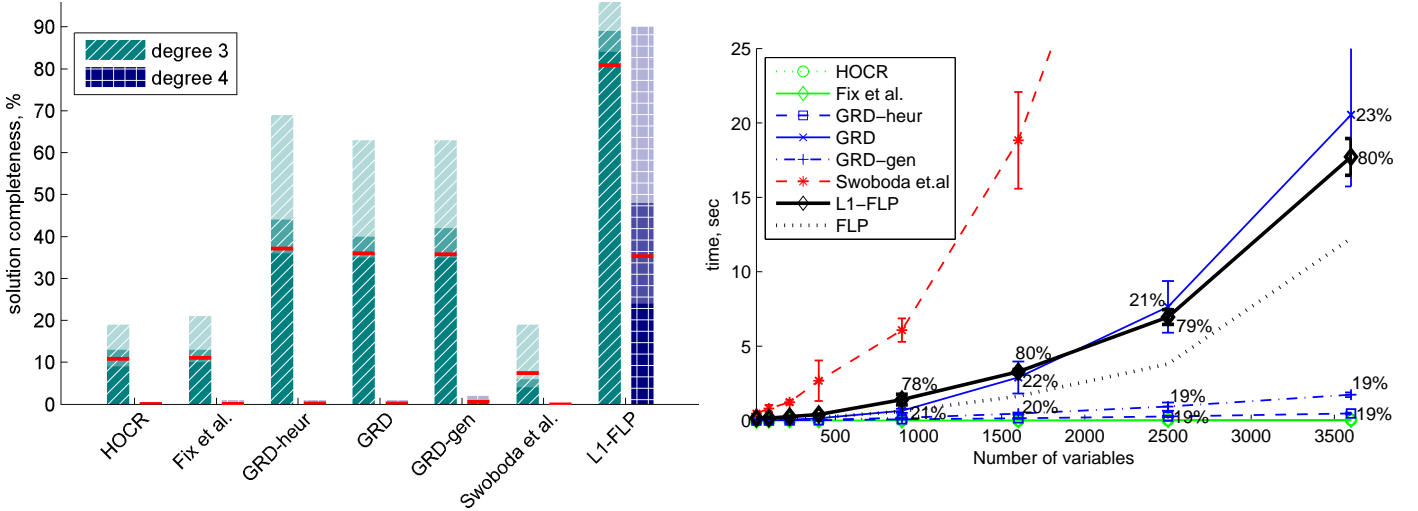
Figure 14: Evaluation of higher-order methods on non-FLP-tight problems on grid. Left: Problems of degree 3 and 4 on grid size 10×10. Note, degree 4 appears difficult for existing methods. Right: Running time for problems of degree 3 and varied size up to 60×60. Percentage indicates solution completeness for selected points.

6.1. Discussion

Iterative Application Do we get more persistencies if the algorithm is run iteratively?

If we consider FLP relaxation in the cases when maximality is guaranteed, a subsequent application of the method cannot give an improvement (it would contradict maximality). Maximality is achieved in pseudo-Boolean or multi-label class $\mathcal{P}^1$ under strong persistency. It is also achieved if we keep the test labeling y fixed and consider the class $\mathcal{P}^{2,y}$ (for both weak and strong persistency). In the other cases it would be possible to improve by iterating the method. Because for BLP relaxation excluding persistent vari-

ables may lead to a tighter relaxation (see Figure 7), the method can be iterated similarly to generalized roof duality, but the result is still dominated by L1-FLP. In the multi-label case we can iterate while varying the test labeling y, however this is computationally expensive and does not seem practical.

Efficiency The present work focused on theoretical aspects. Practical applicability of the method requires some further research and development of efficient specialized methods that use approximate solutions of the relaxation as [64] or a windowing technique [56] or alike. Method of Swoboda et al. [64] performs not the best in Figure 14 and is also the

22

slowest when implemented with CPLEX. However, it can be made optimal for class $\mathcal{P}^1$ as proposed in [65] and fast in practice using scalable dual solvers. Our most recent work in this direction [57] proposes an algorithm of this type for the pairwise multilabel case and $\mathcal{P}^{2,y}$ class of maps. It can be viewed as an alternative (combinatorial w.r.t. the mapping) algorithm for the problem (L1) and achieves the necessary efficiency.

Open Questions It was shown that strict persistency leads to a tractable problem for a larger set of maps. It guarantees not to remove ambiguous solutions. By increasing ε in the perturbation method one gets a potentially stronger guarantee w.r.t. the uncertainty of the data, which may be explored. The general approach holds for an arbitrary bounded polytope, allowing one to incorporate also global linear constraints. This suggests a generalization to linearly constrained discrete optimization problems or mixed integer linear programs. Another interesting direction is how to combine the proposed persistency method with cutting plane techniques. Finally, among the questions that remained open is polynomiality of (MAX-SI) problem for the following cases: (i) For all node-wise maps in a problem with 3 labels. (ii) For maps of the form $x \mapsto (x \vee x^{\min}) \wedge x^{\max}$ with $x^{\min}$ and $x^{\max}$ being free variables, which is relevant when labels have a natural linear ordering. The optimal solution to this case would improve over the iterative method of Kovtun [40], MQPBO and the method [75].

A. Proofs

Lemma 1.1. The conical hull of a relaxation polytope Λ (in the form (8), non-empty and bounded) is obtained by dropping the constraint $\mu_\varnothing = 1$:

$$\operatorname{coni}(\Lambda) = \{\mu \in \mathbb{R}^{\mathcal{I}} \mid A\mu \geqslant 0; \mu \geqslant 0\}. \tag{10}$$

Proof. We will show that the defining set (9) is contained in (10) and vice versa. Let $\mu \in \Lambda$, let $\alpha \geqslant 0$. Then $A\alpha\mu = \alpha A\mu \geqslant 0$ and $\alpha\mu \geqslant 0$. Therefore $\alpha\mu$ is contained in the set (10). Now let μ belong to the set (10). If $\mu_\varnothing > 0$, we can select $\alpha = \mu_\varnothing$ and vector $\mu' = \mu/\mu_\varnothing \in \Lambda$ and conclude that $\mu = \alpha\mu'$ belongs to (9). Let $\mu_\varnothing = 0$. Assume for contradiction that $\mu \neq 0$. Set Λ is non-empty by assumption, let $\mu' \in \Lambda$. Then for any $\alpha \geqslant 0$ there holds $\mu' + \alpha\mu \in \Lambda$. But $\|\mu' + \alpha\mu\| \geqslant \alpha\|\mu\|$, which is unbounded and contradicts boundedness of Λ. Therefore $\mu = 0$, which

belongs to the set (9). $\qquad\square$

Statement 2.5. Let $\mathcal{O} = \operatorname*{argmin}_{\mu \in \Lambda}\langle f, (I - P)\mu\rangle$. There holds $P \in \mathbb{S}_f$ iff

$$\mathcal{O} = P(\Lambda). \tag{25}$$

Proof. By idempotency we have that $P(\Lambda) = \{\mu \in \Lambda, P\mu = \mu\}$. Let $\mu' \in P(\Lambda)$. Since $\langle(I - P^{\mathsf{T}})f, \mu'\rangle = 0$ it is clear that the value of verification LP, $\min_{\mu \in \Lambda}\langle(I - P^{\mathsf{T}})f, \mu\rangle$ is not positive.

Direction "$\Leftarrow$". Let (25) hold. For $\mu \in P(\Lambda) = \mathcal{O}$ condition (23) is trivially satisfied and we find that the value of verification LP is zero. For $\mu \in \Lambda\backslash P(\Lambda)$ from (25) follows that μ is not a minimizer, therefore the objective must be strictly larger than zero and the strict inequality in (23) is satisfied.

Direction "$\Rightarrow$". Let P satisfy (23). Then the value of verification LP is zero. Since any $\mu \in P(\Lambda)$ satisfies $P\mu = \mu$ and achieves zero objective we have $P(\Lambda) \subset \mathcal{O}$. Now let $\mu \notin P(\Lambda)$. In this case $P\mu \neq \mu$ and inequality (23) is strict. Therefore $\mu \notin \mathcal{O}$. $\qquad\square$

The condition (25) can be further expressed in the components of mapping p as follows.

Statement 2.6. Let $\mathcal{O}_{\mathrm{C}} = \{x_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}} \mid (\exists\mu \in \mathcal{O})\ \mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0\}$. There holds $[p] \in \mathbb{S}_f$ iff

$$(\forall \mathrm{C} \in \mathcal{E})\ \mathcal{O}_{\mathrm{C}} = p_{\mathrm{C}}(\mathcal{X}_{\mathrm{C}}). \tag{26}$$

Proof. Direction "$\Rightarrow$": We prove the negation of the implication. Assume $(\exists \mathrm{C}, \exists x_{\mathrm{C}} \in \mathcal{O}_{\mathrm{C}})\ p_{\mathrm{C}}(x_{\mathrm{C}}) \neq x_{\mathrm{C}}$. Then $\exists\mu \in \mathcal{O}$ such that $\mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0$. By idempotency, there is no x_{C}' such that $p_{\mathrm{C}}(x_{\mathrm{C}}') = x_{\mathrm{C}}$. By (15) $([p]\mu)_{\mathrm{C}}(x_{\mathrm{C}}) = 0 \neq \mu_{\mathrm{C}}(x_{\mathrm{C}})$ and therefore $[p]\mu \neq \mu$. From (25) follows that $[p] \notin \mathbb{S}_f$.

Direction "$\Leftarrow$": Let (26) hold. Let $\mu \in \mathcal{O}$. Then for all C, x_{C} such that $\mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0$ there holds $p_{\mathrm{C}}(x_{\mathrm{C}}) = x_{\mathrm{C}}$. It follows from (15) that $([p]\mu)_{\mathrm{C}} = \mu_{\mathrm{C}}$ and thus $[p]\mu = \mu$. Therefore $[p](\mathcal{O}) = \mathcal{O}$ and thus $\mathcal{O} \subset P(\Lambda)$ and the value of the verification LP is zero. Because $P(\Lambda) = \{\mu \in \Lambda \mid P\mu = \mu\}$ the value of the objective on $P(\Lambda)$ is zero and thus $P(\Lambda) \subset \mathcal{O}$. From (25) follows $[p] \in \mathbb{S}_f$. $\qquad\square$

A.1. Properties

Lemma 2.7 (Necessary conditions I)**.** Let $p \colon \mathcal{X} \to \mathcal{X}$ be node-wise and $P = [p]$. Let $\mathcal{O} = \operatorname{argmin}_{\mu \in \Lambda}\langle f, \mu\rangle$ and $\mathcal{O}_{\mathrm{C}} = \{x_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}} \mid (\exists\mu \in \mathcal{O})\ \mu_{\mathrm{C}}(x_{\mathrm{C}}) > 0\}$. Then

(i) For $p \in \mathbb{W}_f$ there holds

$$P(\mathcal{O}) \subset \mathcal{O}; \tag{27a}$$

$$(\forall \mathrm{C} \in \mathcal{E}) \quad p_\mathrm{C}(\mathcal{O}_\mathrm{C}) \subset \mathcal{O}_\mathrm{C}; \tag{27b}$$

(ii) For $p \in \mathbb{S}_f$ there holds

$$P(\mathcal{O}) = \mathcal{O}; \tag{28a}$$

$$(\forall \mathrm{C} \in \mathcal{E}) \quad p_\mathrm{C}(\mathcal{O}_\mathrm{C}) = \mathcal{O}_\mathrm{C}; \tag{28b}$$

Proof. **(i)** Assume $(\exists \mu \in \mathcal{O}) \ P\mu \in \Lambda \backslash \mathcal{O}$. Then $\langle f, P\mu \rangle > \langle f, \mu \rangle$, therefore $p \notin \mathbb{W}_f$. This proves equation (27a).

Assume for contradiction that (27b) does not hold, *i.e.* $(\exists \mathrm{C} \in \mathcal{E}, \ \exists x_\mathrm{C} \in \mathcal{O}_\mathrm{C}) \ x'_\mathrm{C} := p_\mathrm{C}(x_\mathrm{C}) \neq x_\mathrm{C}$. Because $x_\mathrm{C} \in \mathcal{O}_\mathrm{C}$ there exists $\mu \in \mathcal{O}$ such that $\mu_\mathrm{C}(x_\mathrm{C}) > 0$. From expression (16) it follows that $([p]\mu)_\mathrm{C}(x'_\mathrm{C}) > 0$ but $x'_\mathrm{C} \notin \mathcal{O}_\mathrm{C}$ and therefore $[p]\mu \notin \mathcal{O}$, which contradicts to (27a).

(ii) Assume $(\exists \mu \in \mathcal{O}) \ P\mu \neq \mu$. Then $\langle f, P\mu \rangle \geqslant \langle f, \mu \rangle$ and therefore $p \notin \mathbb{S}_f$. This proves equation (28a).

Assume for contradiction that (28b) does not hold, *i.e.* $(\exists \mathrm{C} \in \mathcal{V}, \ \exists x_\mathrm{C} \in \mathcal{O}_\mathrm{C}) \ p_\mathrm{C}(x_\mathrm{C}) \neq x_\mathrm{C}$. Because $x_\mathrm{C} \in \mathcal{O}_\mathrm{C}$ there exists $\mu \in \mathcal{O}$ such that $\mu_\mathrm{C}(x_\mathrm{C}) > 0$. Since $p_\mathrm{C}(x_\mathrm{C}) \neq x_\mathrm{C}$, from idempotency and (15) follows that $([p]\mu)_\mathrm{C}(x_\mathrm{C}) = 0 \neq \mu_\mathrm{C}(x_\mathrm{C})$ and thus $[p]\mu \neq \mu$, which contradicts (28a). $\qquad\square$

Theorem 2.9 (Dual representation of $\mathbb{S}_f$). Let $p \colon \mathcal{X} \to \mathcal{X}$ be node-wise. Then: (i) there exists $\varepsilon > 0$ such that $[p] \in \mathbb{S}_f$ iff

$$(\exists \varphi \in \mathbb{R}^m_+) \quad f - A^\mathsf{T}\varphi - [p]^\mathsf{T} f \geqslant \varepsilon h, \tag{32}$$

where h is a function such that $h \geqslant 0$ and $h_\mathrm{C}(x_\mathrm{C}) = 0$ iff $p_\mathrm{C}(x_\mathrm{C}) = x_\mathrm{C}$; and (ii) for rational inputs (including h) the value of ε in (i) is a rational number of polynomial bit length.

Proof. (i) The "if" part. Condition (32) implies a weaker condition

$$f^\varphi - [p]^\mathsf{T} f \geqslant 0, \tag{69}$$

i.e. it satisfies dual representation of $\mathbb{W}_f$ (29) and therefore p is relaxed-improving. It remains to prove strictness. The value of the verification LP in (24) is zero. The value of its dual problem

$$\max\{\psi \in \mathbb{R} \mid f^\varphi - [p]^\mathsf{T} f - e_\varnothing \psi; \ \varphi \in \mathbb{R}^m_+\} \tag{70}$$

is thus also zero. It follows that φ is optimal to (70). We need to show that for $\mu \in \mathcal{O}$ there holds $[p]\mu = \mu$. By multiplying (32) with $\mu_\mathrm{C}(x_\mathrm{C})$ and summing over

C and x_C we obtain

$$\langle (I - [p]^\mathsf{T}) f, \mu \rangle - \langle \varphi, A\mu \rangle \geqslant \varepsilon \langle h, \mu \rangle. \tag{71}$$

Because μ, φ are optimal primal and dual solutions, by complementary slackness $\langle \varphi, A\mu \rangle = 0$. Assume for contradiction that $[p]\mu \neq \mu$. Then $(\exists \mathrm{C} \in \mathcal{E}, \ \exists x_\mathrm{C} \in \mathcal{X}_\mathrm{C})$ $([p]\mu)_\mathrm{C}(x_\mathrm{C}) \neq \mu_\mathrm{C}(x_\mathrm{C})$. We consider now two cases

Case 1: if $p_\mathrm{C}(x_\mathrm{C}) \neq x_\mathrm{C}$, then by idempotency for all x'_C holds $p_\mathrm{C}(x'_\mathrm{C}) \neq x_\mathrm{C}$ and therefore from (15) we calculate that $([p]\mu)_\mathrm{C}(x_\mathrm{C}) = 0$. In this case from the assumption it must be $\mu_\mathrm{C}(x_\mathrm{C}) > 0$ and

$$\langle \mu, h \rangle \geqslant \mu_\mathrm{C}(x_\mathrm{C}) h_\mathrm{C}(x_\mathrm{C}) > 0. \tag{72}$$

Case 2: if $p_\mathrm{C}(x_\mathrm{C}) = x_\mathrm{C}$ then from (15) and the assumption follows $(\exists x'_\mathrm{C} \neq x_\mathrm{C})$ such that $p_\mathrm{C}(x'_\mathrm{C}) = x_\mathrm{C}$ and $\mu_\mathrm{C}(x'_\mathrm{C}) > 0$. In this case

$$\langle \mu, h \rangle \geqslant \mu_\mathrm{C}(x'_\mathrm{C}) h_\mathrm{C}(x'_\mathrm{C}) > 0. \tag{73}$$

In both cases 1 and 2 we have $\langle (I - [p]^\mathsf{T}) f, \mu \rangle > 0$, which contradicts optimality of μ.

We now prove the "only if" part of (i). Let $g = (I - P^\mathsf{T}) f$ and let $(\mu, (\varphi, \psi))$ be a primal-dual strictly complementary pair of solutions to

$$\min_{\mu \in \Lambda} \langle g, \mu \rangle. \tag{74}$$

Let $\mathcal{O}_\mathrm{C}$ be the C-support set of primal solutions: $\mathcal{O}_\mathrm{C} = \{x_\mathrm{C} \in \mathcal{X}_\mathrm{C} \mid \mu_\mathrm{C}(x_\mathrm{C}) > 0\}$. By Statement 2.6 and idempotency, there holds $p_\mathrm{C}(x_\mathrm{C}) = x_\mathrm{C}$ for $x_\mathrm{C} \in \mathcal{O}_\mathrm{C}$. By strict complementarity, for $x_\mathrm{C} \in \mathcal{O}_\mathrm{C}$ there holds $g^\varphi_\mathrm{C}(x_\mathrm{C}) = 0$ and for $x_\mathrm{C} \notin \mathcal{O}_\mathrm{C}$ there holds $g^\varphi_\mathrm{C}(x_\mathrm{C}) > 0$. We let

$$\varepsilon = \min_{\substack{\mathrm{C} \in \mathcal{E}, \\ x_\mathrm{C} \in \mathcal{X}_\mathrm{C} \backslash \mathcal{O}_\mathrm{C}}} \frac{g^\varphi_\mathrm{C}(x_\mathrm{C})}{h_\mathrm{C}(x_\mathrm{C})} > 0. \tag{75}$$

Since for $x_\mathrm{C} \in \mathcal{O}_\mathrm{C}$ $h_\mathrm{C}(x_\mathrm{C}) = 0$, we can bound now components of g^φ as follows

$$(\forall \mathrm{C} \in \mathcal{E}, \ \forall x_\mathrm{C}) \quad g^\varphi_\mathrm{C}(x_\mathrm{C}) \geqslant \varepsilon h_\mathrm{C}(x_\mathrm{C}). \tag{76}$$

Expanding components of g^φ as $g^\varphi_\mathrm{C}(x_\mathrm{C}) = f_\mathrm{C}(x_\mathrm{C}) - f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) - (A^\mathsf{T}\varphi)_\mathrm{C}(x_\mathrm{C})$, we obtain relations (32).

The statement of part (ii) of the theorem is proved as follows. The bit length of the rational dual solution φ is polynomially bounded as well as the bit length of rational numbers h. It follows that ε calculated by (75) is a rational number of polynomially bounded bit length. $\qquad\square$

Theorem 2.10 (Necessary conditions II). Let $P \colon \mathbb{R}^\mathcal{I} \to \mathbb{R}^\mathcal{I}$ be idempotent, $P(\Lambda) \subset \Lambda$ and $P \in \mathbb{W}_f$.

Then

$$\inf_{\substack{A(I-P)\mu\geqslant 0 \\ \mu\geqslant 0}} \langle (I-P)^\mathsf{T} f, \mu \rangle = 0; \tag{33a}$$

$$(\exists \varphi \in \mathbb{R}^m_+) \ \ (I - P^\mathsf{T})(f - A^\mathsf{T}\varphi) \geqslant 0. \tag{33b}$$

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

$$\inf_{\substack{A\mu\geqslant 0 \\ \mu\geqslant 0}} \langle f - P^\mathsf{T} f, \mu \rangle \overset{(b)}{\leqslant} \inf_{\substack{AP\mu\geqslant 0 \\ A(I-P)\mu\geqslant 0 \\ \mu\geqslant 0}} \langle f - P^\mathsf{T} f, \mu \rangle$$

$$\overset{(c)}{=} \inf_{\substack{A(I-P)\mu\geqslant 0 \\ \mu\geqslant 0}} \langle f - P^\mathsf{T} f, \mu \rangle \overset{(d)}{=} \sup_{\substack{\varphi\in\mathbb{R}^m_+ \\ (I-P^\mathsf{T})(f-A^\mathsf{T}\varphi)\geqslant 0}} 0 \, . \tag{77}$$

On the LHS we have problem (30). If $P \in \mathbb{W}_f$, this problem is bounded and the value of the problem is zero. Equalities (b), (c) essentially claims boundedness of the other two minimization problems in the chain.

Inequality (b) is verified as follows. Inequality $\leqslant$ holds because by summing two inequalities

$$AP\mu \geqslant 0, \tag{78a}$$
$$A(I - P)\mu \geqslant 0 \tag{78b}$$

we get $A\mu \geqslant 0$.

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 $\mu = \mu_1 + \mu_2$, where

$$\mu_1 = P\mu; \qquad \mu_2 = (I - P)\mu. \tag{79}$$

There holds

$$(I - P)\mu_1 = (I - P)P\mu = (P - P^2)\mu = 0, \\ P\mu_2 = P(I - P)\mu = 0, \tag{80}$$

i.e., $\mu_1 \in \text{null}(I - P)$ and $\mu_2 \in \text{null}(P)$. Let us chose μ_1' such that

$$\mu_1' \geqslant \mu_1 \qquad \text{and} \qquad A\mu_1' \geqslant 0. \tag{81}$$

For example, the relaxed labeling

$$\mu_1' = \gamma \frac{1}{|\mathcal{X}|} \sum_{x\in\mathcal{X}} \delta(x) \tag{82}$$

will satisfy these constraints for sufficiently large $\gamma > 0$. Indeed, all components of μ_1' are strictly positive, it belongs to $\mathcal{M}$ as a convex combination of integer labelings and therefore satisfies constraints of the relaxation $A\mu_1' \geqslant 0$ for any $\gamma > 0$. It remains to chose γ large enough so as to have $(\mu_1') \geqslant \mu_1$ satisfied.

Notice, that $(\mu_1')_\varnothing = \gamma$. Let $\mu_1'' = P\mu_1'$. Because

$P \geqslant 0$, we have

$$\mu_1'' = P\mu_1' \geqslant P\mu_1 = \mu_1. \tag{83}$$

Because $\mu_1' \in \text{coni}(\Lambda)$, there holds

$$AP\mu_1'' = APP\mu_1' = AP\mu_1' \geqslant 0.$$

By idempotency, $(I - P)\mu_1'' = (I - P)P\mu_1' = 0$.

Let $\mu^* = \mu_1'' + \mu_2$. It preserves the objective,

$$\langle f - P^\mathsf{T} f, \mu^* \rangle = \langle f, (I - P)(\mu_1'' + \mu_2) \rangle \tag{84}$$
$$= \langle f, (I - P)\mu_2 \rangle = \langle f, (I - P)\mu \rangle.$$

We also have that

$$\mu^* = \mu_1'' + \mu_2 \geqslant \mu_1 + \mu_2 = \mu \geqslant 0, \\ A(I - P)\mu^* = A(I - P)\mu_2 = A(I - P)\mu \geqslant 0, \tag{85} \\ AP\mu^* = AP\mu_1'' \geqslant 0.$$

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.

$\square$

A.2. Relaxation of Sherali and Adams

Lemma 2.11 (Identity Equality (S3))**.** Let $G(\zeta)$ be the linearization of $g(z)$. Then $g(z) = 0$ for all $z \in \{0, 1\}^\mathrm{C}$ iff $G(\zeta) = 0$ for all $\zeta \in \mathbb{R}^{2^\mathrm{C}}$.

Proof. The correspondence between g and G is through coefficients α:

$$g(z) = \sum_{\mathrm{D}\subset\mathrm{C}} \alpha_\mathrm{D} \prod_{s\in\mathrm{D}} z_s; \tag{86a}$$

$$G(\zeta) = \sum_{\mathrm{D}\subset\mathrm{C}} \alpha_\mathrm{D} \zeta_\mathrm{D}. \tag{86b}$$

We have $g \equiv 0$ iff all coefficients of the (unique) multilinear polynomial representation α are zero and it is the case iff $G \equiv 0$. $\square$

For subsequent proofs let us introduce the correspondence between binary variables z and their lifted representation ζ as the mapping $\zeta(z)$ from $\{0, 1\}^\mathrm{C}$ to $\mathbb{R}^{2^\mathrm{C}}$ with components

$$\zeta(z)_\mathrm{D} = \prod_{s\in\mathrm{D}} z_s. \tag{87}$$

Lemma 2.13 (Identity inequality (S5))**.** Let $G(\zeta)$ be the linearization of $g(z)$. Then $g(z) \geqslant 0$ for all $z \in \{0, 1\}^\mathrm{C}$ iff $G(\zeta) \geqslant 0$ for all $\zeta \in \mathcal{Z}_\mathrm{C}$.

Proof. ($\Leftarrow$) This part follows from the fact that $\zeta(z)$ satisfies all constraints of $\mathcal{Z}_\mathrm{C}$ and thus $g(z) = G(\zeta(z)) \geqslant 0$ for all $z \in \{0, 1\}^\mathrm{C}$.

($\Rightarrow$) Note that a special case when $g(z) = \prod_{s\in B} z_s \prod_{s\in A}(1 - z_s)$ and $A \cap B = \varnothing$, A, $B \subset C$ is proven in [58, Lemma 2]. Here is a different general proof.

Since g is non-negative, it can be represented as a posiform [6, Proposition 1]:

$$g(z) = \sum_{B\subset C} \alpha_B \prod_{s\in B} z_s \prod_{s\in C\setminus B}(1 - z_s), \qquad (88)$$

where $\alpha_B = g(\mathbb{1}_B) \geqslant 0$ and $\mathbb{1}_B \in \{0,1\}^C$ is the indicator of the set B defined as: $(\mathbb{1}_B)_s := [\![s \in B]\!]$. By linear combination property (S3), $G(\zeta)$ can be written as

$$G(\zeta) = \sum_{B\subset C} \alpha_B \sum_{D\subset C\setminus B} (-1)^{|D|} \zeta_{D\cup B}, \qquad (89)$$

which is a non-negative combination of non-negative summands, as ensured by constraints of $\mathcal{Z}_C$. $\qquad\square$

Lemma 2.12 (Convex hull). Polytope $\mathcal{Z}_C$ equals the convex hull

$$\mathrm{conv}\left\{ \zeta(z) \in \mathbb{R}^{2^C} \,\middle|\, \zeta(z)_D = \prod_{s\in D} z_s, \ \forall D \subset C, \forall z \in \{0,1\}^C \right\}. \qquad (38)$$

Proof. Let $\mathcal{H}$ denote the convex hull (38). Clearly, any vertex of $\mathcal{H}$ is in $\mathcal{Z}_C$ and therefore $\mathcal{H} \subset \mathcal{Z}_C$.

Let $G(\zeta) = \sum_{D\subset C} \alpha_D \zeta_D \geqslant 0$ be a facet-defining inequality of $\mathcal{H}$. Let us show it holds for all $\zeta \in \mathcal{Z}_C$. Let $g(z) = \sum_{D\subset C} \alpha_D \prod_{s\in D} z_s$, $z \in \{0,1\}^C$, a multilinear polynomial corresponding to G. For all vertices $\zeta(z)$ of $\mathcal{H}$ there holds $G(\zeta(z)) = g(z)$ and at the same time $G(\zeta) \geqslant 0$. It follows that $g(z) \geqslant 0$ for all $z \in \{0,1\}^C$ and, by Lemma 2.13, $G(\zeta) \geqslant 0$ for all $\zeta \in \mathcal{Z}_C$. We have proven that $\mathcal{Z}_C \subset \mathcal{H}$. $\qquad\square$

Lemma 2.15 (Product). For $\zeta \in \mathcal{Z}_C$ there holds $\zeta^2 \in \mathcal{Z}_C$, where the product $\zeta^2 = \zeta\zeta$ is component-wise.

Proof. Using the convex hull property we can represent $\zeta \in \mathcal{Z}_C$ as convex combination of vertices, *i.e.*, $\zeta_D = \sum_{k=1}^n \alpha_k \prod_{s\in D} z_s^k$, $D \subset C$, where $z^k \in \{0,1\}^C$ for $k = 1,\ldots,n$ and $\alpha_k \geqslant 0$ and $\sum_{k=1}^n \alpha_k = 1$. Then

$$\zeta_D^2 = \sum_{k_1=1}^n \sum_{k_2=1}^n \alpha_{k_1}\alpha_{k_2} \prod_{s\in D} z_s^{k_1} \prod_{s\in D} z_s^{k_2} \qquad (90)$$

$$= \sum_{k_1,\,k_2=1}^n \alpha_{k_1}\alpha_{k_2} \prod_{s\in D} z_s^{k_1,k_k}, \qquad (91)$$

where $z^{k_1,k_2} \in \{0,1\}^C$ is the coordinate-wise product of z^{k_1} and z^{k_2}. Note that $\alpha_{k_1}\alpha_{k_2} \geqslant 0$ and $\sum_{k_1,\,k_2=1}^n \alpha_{k_1}\alpha_{k_2} = 1$. Expression (90) proves that ζ^2 is representable as a convex combination of vertices

and thus belongs to $\mathcal{Z}_C$. One could similarly show that for $\zeta, \eta \in \mathcal{Z}_C$ their product $\zeta\eta$ also belongs to $\mathcal{Z}_C$. $\qquad\square$

A.3. L1 construction

In §2.7 we used representation of coefficients $P_{C,x_C,x_C'}$ of the linear extension $[p_\zeta]$ in the form of a polynomial (39). This representation is obtained as follows. Starting from definition (36), we express:

$$P_{C,x_C x_C'} = \prod_{s\in C}\left(([\![x_s' = x_s]\!] - [\![y_s = x_s]\!])\zeta_{s,x_s'} + [\![y_s = x_s]\!] \right)$$

$$= \sum_{D\subset C} c_{C,D}(x_C, x_D') \prod_{s\in D} \zeta_{s,x_s'}, \qquad (92)$$

where

$$c_D(x_s, x_s') = \begin{cases} [\![x_s' = x_s]\!] - [\![y_s = x_s]\!], & s \in D \\ [\![y_s = x_s]\!], & s \in C\setminus D \end{cases} \qquad (93)$$

$$c_{C,D}(x_C, x_D') = \prod_{s\in C} c_D(x_s, x_s').$$

Theorem 2.16. Inequalities (41c) in the definition of polytope $\mathcal{Z}$ can be equivalently replaced with $P_\zeta \geqslant 0$.

Proof. The fact that inequalities (41c) imply $P_\zeta \geqslant 0$, assuming equality constraints (41a)-(41b), was shown in §2.7. We show now that $P_\zeta \geqslant 0$ implies inequalities (41c).

The inequality $P_\zeta \geqslant 0$ for the linear mapping P_ζ means that all its matrix elements are non-negative, *i.e.*, the defining coefficients $P_{C,x_C,x_C'}$ for $C \in \mathcal{E}$, $x_C \in \mathcal{X}_c$, $x_C' \in \mathcal{X}_C$ are non-negative. Let us detail the constraint

$$P_{C,x_C,x_C'} \geqslant 0. \qquad (94)$$

Let

$$A = \{s \in C \,|\, x_s \neq x_s'\}, \quad B = \{s \in C \,|\, x_s \neq y_s\}. \qquad (95)$$

and let $\bar{A}$ denote the complement in C. From (93) we have

$$c_D(x_s, x_s') = \begin{cases} 1, & s \in \bar{A} \cap B \cap D, \\ -1, & s \in A \cap \bar{B} \cap D, \\ 0, & s \in \overline{(A\triangle B)} \cap D, \\ 1 & s \in \bar{B} \cap \bar{D}, \\ 0, & s \in B \cap \bar{D}. \end{cases} \qquad (96)$$

Coefficient $c_D(x_C, x_C')$ in (93), which is the product of (96) over $s \in C$, expresses as

$$c_D(x_C, x_C') = (-1)^{|A\cap\bar{B}\cap D|} 0^{|D\setminus(A\triangle B)|} 0^{|B\setminus D|}. \qquad (97)$$

26

It is non-zero only when $D\backslash(A\triangle B) = \varnothing$ and $B\backslash D = \varnothing$ or equivalently

$$B \subset D, \quad D \subset A\triangle B; \tag{98a}$$

$$\Leftrightarrow B \subset D, \quad A \subset \bar{B}, \quad D \subset A \cup B. \tag{98b}$$

Using sets A, B and coefficients (97) we obtain

$$P_{C,x_C,x_C'} = \sum_{\substack{D \\ B\subset D\subset A\cup B}} (-1)^{|D\backslash B|}\zeta_{D,x_D'} \geqslant 0 \tag{99}$$

which is equivalent to

$$\sum_{D\subset A} (-1)^{|D|}\zeta_{D\cup B,x_{D\cup B}'} \geqslant 0. \tag{100}$$

In order to obtain inequalities (41c) we need to show that for all $x_C' \in \tilde{\mathcal{X}}_C$, by varying $x_C \in \mathcal{X}_C$, the set B ranges over all subsets in C while at the same time A equals to $\bar{B}$. Since $x_C' \in \tilde{\mathcal{X}}_C$ we have $x_s' \neq y_s'$ for all $s \in C$. An arbitrary given set $\tilde{B}$ can be realized by the choice $x_s = x_s'$ if $s \in \tilde{B}$ and $x_s = y_s$ otherwise. At the same time for this choice of x there holds $A = \bar{B}$. $\quad\square$

Lemma 2.17. For $\zeta \in \mathcal{Z}$ there holds $P_\zeta^2 = P_{\zeta^2}$.

Proof. Let us calculate the expression of $(P^2)_{C,x_Cx_C''}$. It is equal to

$$\sum_{x_C'\in\mathcal{X}_C} P_{C,x_Cx_C'}P_{C,x_C'x_C''} = \sum_{D_1\subset C}\sum_{D_2\subset C} \tag{101a}$$

$$\sum_{x_C'\in\mathcal{X}_C}\prod_{s\in C} c_{D_1}(x_s,x_s')c_{D_2}(x_s',x_s'')\zeta_{D_1,x_{D_1}'}\zeta_{D_2,x_{D_2}''} \tag{101b}$$

The expression in line (101b) factors as

$$\Big(\sum_{x_{D_1}'\in\mathcal{X}_{D_1}}\prod_{s\in D_1} c_{D_1}(x_s,x_s')c_{D_2}(x_s',x_s'')\zeta_{D_1,x_{D_1}'}\Big) \tag{102a}$$

$$\Big(\prod_{s\in C\backslash D_1}\sum_{x_s'\in\mathcal{X}_s} c_{D_1}(x_s,x_s')c_{D_2}(x_s',x_s'')\Big)\Big(\zeta_{D_2,x_{D_2}''}\Big). \tag{102b}$$

The term of the second factor for $s \in C\backslash D_1$ equals

- Case $s \notin D_1$, $s \in D_2$:

$$\sum_{x_s'} [\![y_s{=}x_s]\!](-[\![y_s{=}x_s']\!] + [\![x_s''{=}x_s']\!]) = 0. \tag{103}$$

- Case $s \notin D_1$, $s \notin D_2$:

$$\sum_{x_s'} [\![y_s{=}x_s]\!][\![y_s{=}x_s']\!] = [\![y_s{=}x_s]\!]. \tag{104}$$

It follows from (103) that for $D_2 \not\subset D_1$ the factors vanishes and hence expression (101b) vanishes. In the first factor the coefficient $c_{D_1}(x_s,x_s')c_{D_2}(x_s',x_s'')$ expresses as:

- Case $s \in D_1$, $s \in D_2$:

$$(-[\![y_s{=}x_s]\!] + [\![x_s'{=}x_s]\!])(-[\![y_s{=}x_s']\!] + [\![x_s''{=}x_s']\!])$$
$$= (-[\![y_s{=}x_s]\!] + [\![x_s''{=}x_s]\!])[\![x_s'{=}x_s'']\!]. \tag{105}$$

- Case $s \in D_1$, $s \notin D_2$:

$$(-[\![y_s{=}x_s]\!] + [\![x_s'{=}x_s]\!])[\![y_s{=}x_s']\!]$$
$$= -[\![y_s{=}x_s{=}x_s']\!] + [\![x_s'{=}x_s{=}y_s]\!] = 0. \tag{106}$$

Therefore, if $D_1 \not\subset D_2$, for each value of x_{D_1}' the product $\prod_{s\in D_1} c_{D_1}(x_s,x_s')c_{D_2}(x_s',x_s'')$ vanishes and hence the sum (102a) and the expression (101b) vanish. It follows that we need to count expression (101b) only for the case $D_1 = D_2 =: D$. In this case, carrying the summation over x_{D_1}' in (102a) we obtain

$$\prod_{s\in D}(-[\![y_s{=}x_s]\!] + [\![x_s''{=}x_s]\!])\zeta_{D,x_D''}. \tag{107}$$

From the full expression (101) there remains

$$\sum_{D\subset C}\prod_{s\in D}(-[\![y_s{=}x_s]\!] + [\![x_s''{=}x_s]\!])\prod_{s\in C\backslash D}[\![y_s = x_s]\!]\zeta_{D,x_D''}\zeta_{D,x_D''}$$
$$= \sum_{D\subset C} c_{C,D}(x_C',x_C'')\zeta_{C,x_D''}^2 = (P_{\zeta^2})_{C,x_C,x_C''}. \tag{108}$$

The claim $(P_\zeta)^2 = P_{\zeta^2}$ is proven. $\quad\square$

A.4. Maximum Persistency for Local Relaxations

Lemma 3.1. Node-wise mapping $[p]$ preserves the local polytope Λ.

Proof. Clearly, $[p]$ preserves non-negativity. Let μ satisfy marginalization constraint (52) for some $C \in \mathcal{E}$, $D \subsetneq C$ and $x_D \in \mathcal{X}_D$. Then

$$\sum_{x_{C\backslash D}}([p]\mu)_C(x_C) = \sum_{x_{C\backslash D}}\sum_{x_C'\in\mathcal{X}_C} [\![p_C(x_C'){=}x_C]\!]\mu_C(x_C')$$

$$= \sum_{x_C'}\Big(\sum_{x_{C\backslash D}}[\![p_C(x_C'){=}x_C]\!]\Big)\mu_C(x_C') \tag{109a}$$

$$= \sum_{x_C'}[\![p_D(x_D'){=}x_D]\!]\mu_C(x_C') \tag{109b}$$

$$= \sum_{x_D'\in\mathcal{X}_D}[\![p_D(x_D'){=}x_D]\!]\mu_D(x_D') = ([p]\mu)_D(x_D). \tag{109c}$$

$$\square$$

Lemma 3.2. Mapping P_ζ for $\zeta \in \mathcal{Z}$ preserves the local polytope: $P_\zeta(\Lambda) \subset \Lambda$.

Proof. We need to prove that $(\forall \mu \in \Lambda)$

$$(P_\zeta\mu)_\varnothing = 1; \tag{110a}$$

$$P_\zeta\mu \geqslant 0; \tag{110b}$$

$$AP_\zeta\mu = 0. \tag{110c}$$

27

Constraints (110a), (110b) are satisfied for any relaxation polytope (same as (43)). Constraint (110c) is local: for each $C \subset \mathcal{E}$ it is given by equalities (52) for $(P\mu)_C$, and hence involves only $(\zeta_D \,|\, D \subset C)$. This constraint holds for all integer ζ by Lemma 3.1 and thus also for all $\zeta \in \mathcal{Z}$. $\qquad\square$

Theorem 3.3 (Characterizations). For a local relaxation Λ all of the following are equivalent:
  (a) $P \in \mathbb{W}_f$;
  (b) $(\exists \varphi \in \mathbb{R}^m)\ f^{\varphi} - P^{\mathsf{T}} f \geqslant 0$;
  (c) $(\exists \varphi \in \mathbb{R}^m)\ (I - P^{\mathsf{T}}) f^{\varphi} \geqslant 0$;
  (d) $\inf\{\langle f - P^{\mathsf{T}} f, \mu\rangle \,|\, \mu \in \mathbb{R}_+^{\mathcal{I}}, \ A(I - P)\mu = 0\} = 0$.

Proof. We already have (a) $\Leftrightarrow$ (b) by Theorem 2.8 and (a) $\Rightarrow$ (d) $\Leftrightarrow$ (c) by Theorem 2.10 (equivalence is the duality relation discussed in the proof). Let us prove (c) $\Rightarrow$ (a). Let $\mu \in \Lambda$. We multiply component inequalities (c) with non-negative numbers $\mu_C(x_C)$ and sum over x_C and C. We get inequality

$$\langle P^{\mathsf{T}} f, \mu\rangle - \langle P^{\mathsf{T}} A^{\mathsf{T}} \varphi, \mu\rangle \leqslant \langle f, \mu\rangle - \langle A^{\mathsf{T}} \varphi, \mu\rangle.$$

The sum $\langle A^{\mathsf{T}}\varphi, \mu\rangle = \langle \varphi, A\mu\rangle$ vanishes because $A\mu = 0$ and similarly $\langle P^{\mathsf{T}} A^{\mathsf{T}}\varphi, \mu\rangle = \langle \varphi, AP\mu\rangle = 0$ since by Lemma 3.1 $P\mu \in \Lambda$ and thus $AP\mu = 0$. The remaining inequality proves that $P \in \mathbb{W}_f$. $\qquad\square$

A.5. NP Hardness Results

Theorem 3.4. Problem (MAX-WI) over the $\mathcal{P}^1$ class of maps and the BLP relaxation is solvable in polynomial time for the quadratic pseudo-Boolean case and otherwise (when the problem is multilabel or higher order) it is NP-hard.

Proof. In the quadratic pseudo-Boolean case the solution is given by the optimal relaxed labeling with the maximum number of integer components. For a more special case of vertex packing problem it was proven polynomial by Picard and Queyranne [46], whose proof we extended in [55, Statement 5]. For the general quadratic pseudo-Boolean case, the solution can be found efficiently by analyzing connected components in the network flow model [7], [32, §2.3].

Next we prove that (MAX-WI) is NP-hard if either: there are more than two labels or the order of the problem is 2 (cubic terms) or higher. We use reduction from pairwise constraint satisfaction problem (CSP) which is NP-complete when variables can take 3 or more values or when constraints can couple 3 or more variables at a time (this case includes 3-SAT).

This problem can be represented as energy minimization with constraints $f_C \colon \mathcal{X}_C \to \{0, 1\}$. The CSP is satisfiable iff the minimum of the energy is zero. Let V be the value of the LP-relaxation. Then either it is larger than zero and in this case the CSP is not satisfiable or it equals zero. In the latter case, the CSP is satisfiable iff there exist an integer solution with cost zero, *i.e.*, the relaxation is tight. This is the case iff (MAX-WI) determines all variables as persistent. Thus if (MAX-WI) was in P we could solve CSP, which is a contradictions. $\qquad\square$

Theorem 3.5. Problem (MAX-SI) with 4 or more labels over the class of maps $\mathcal{P}^2 = \bigcup_{y \in \mathcal{X}} \mathcal{P}^{2,y}$ and BLP relaxation is NP-hard.

Proof. We again use a reduction from CSP with 3 labels, $\mathcal{X}_s = \{1, 2, 3\}$. Let the constraints of CSP be defined by $g_C \colon \mathcal{X}_C \to \{0, 1\}$. We construct an energy minimization problem with 4 labels as follows:

$$f_C(x_C) = \begin{cases} g_C(x_C), & x_C \in \{1, 2, 3\}^C; \\ \varepsilon & x_C = 4_C; \\ B, & \text{otherwise}, \end{cases} \qquad (111)$$

where $B > |\mathcal{E}|$ and $\varepsilon < 1/|\mathcal{E}|$. Let V be the value of the BLP relaxation. If it is larger than zero, then the CSP is not satisfiable. Otherwise, the relaxation is tight iff the CSP is satisfiable. It is clear that if an integer solution of zero cost y^* exists, it must take values in $\{1, 2, 3\}$. If it exists, then mapping $p_s \colon (1, 2, 3, 4) \mapsto (1, 2, 3, y_s^*)$ is strictly relaxed improving as it replaces components of cost $\varepsilon > 0$ with components of cost 0. Let q be the maximum strict relaxed improving mapping in class $\mathcal{P}^2$. Let $x_s = 4$ for all $s \in \mathcal{V}$. If $(\exists s \in \mathcal{V})\ q(x)_s = 4$, then the CSP is not satisfiable, as q is not larger than p. Otherwise $(\forall s \in \mathcal{V})\ q(x)_s \neq 4$ and $E_f(q(x)) < |\mathcal{E}|\varepsilon < 1$. It follows that $g(q(x)) = 0$ and hence $q(x)$ is a solution to CSP. We showed that CSP was reduced in polynomial time to (MAX-SI). $\qquad\square$

A.6. Method of Swoboda et al.

Theorem 4.5. Persistency by method [64] in the pairwise multilabel case corresponds to an FLP-improving mapping.

Proof. The method constructs a subset $\mathcal{A} \subset \mathcal{V}$, a labeling y on $\mathcal{A}$ and an auxiliary energy E_g defined

by:

$$(\forall s \in \mathcal{A}) \quad g_s = f_s, \qquad (112)$$

$$(\forall st \in \mathcal{E}, \ s \in \mathcal{A}, t \in \mathcal{A}) \quad g_{st} = f_{st},$$

$$(\forall st \in \mathcal{E}, \ s \in \mathcal{A}, t \notin \mathcal{A}, \ \forall ij)$$

$$g_{st}(i,j) = \begin{cases} \max_{j' \in \mathcal{X}_t} f_{st}(i,j'), & i = y_s, \\ \min_{j' \in \mathcal{X}_t} f_{st}(i,j'), & i \neq y_s, \end{cases} \qquad (113)$$

with remaining terms set to zero. It can be seen that energy E_g depends on the assignment of y only on the *boundary* $\partial \mathcal{A} = \{s \in \mathcal{A} \mid (\exists t \in \mathcal{V} \backslash \mathcal{A}) \ \{s,t\} \in \mathcal{E}\}$. 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 [64, Corollary 1] implies that $\delta(y) \in \operatorname{argmin}_{\mu \in \Lambda} \langle g, \mu \rangle$ (the relaxation is tight). We construct mapping p as

$$p_s(i) = \begin{cases} y_s & \text{if } s \in \mathcal{A}, \\ i & \text{if } s \notin \mathcal{A}, \end{cases} \qquad (114)$$

i.e., p replaces part of a labeling x on $\mathcal{A}$ with the labeling y. This mapping is illustrated in Figure 11. We claim that mapping p is relaxed-improving.

We first show that g is *auxiliary* for f, *i.e.*, that

$$(\forall \mu \in \Lambda) \ \langle (I-P)^\mathsf{T} f, \mu \rangle \geqslant \langle (I-P)^\mathsf{T} g, \mu \rangle. \quad (115)$$

We trivially have $f_s(i) - f_s(p_s(i)) = g_s(i) - g_s(p_s(i))$. We also have equality of pairwise terms

$$f_{st}(i,j) - f_{st}(p_s(i), p_t(j)) = g_{st}(i,j) - g_{st}(p_s(i), p_t(j))$$

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 $t \notin \mathcal{A}$, $i = y_s$.

It remains to verify the inequality for boundary pairs $s \in \mathcal{A}$, $t \notin \mathcal{A}$ in the case $i \neq y_s$. We have

$$\begin{aligned}
& f_{st}(i,j) - f_{st}(p_s(i), p_t(j)) \\
& \geqslant \min_{j' \in \mathcal{X}_t} \big(f_{st}(i,j') - f_{st}(p_s(i), p_t(j')) \big) \\
& \geqslant \min_{j'} f_{st}(i,j') - \max_{j'} f_{st}(y_s, p_t(j')) \\
& = g_{st}(i,j) - g_{st}(p_s(i), p_t(j)).
\end{aligned} \qquad (116)$$

Because component-wise inequalities hold it follows that (115) holds.

The second step is to show that p is relaxed-improving for g. By assumption, we have $\delta(y) \in \operatorname{argmin}_{\mu \in \Lambda} \langle g, \mu \rangle$. Given a labeling x, mapping p replaces part over $\mathcal{A}$ to the optimal labeling y. It follows that $(\forall \mu \in \Lambda) \ \langle g, P\mu \rangle = \langle g, \delta(y) \rangle \leqslant \langle g, \mu \rangle$. Combining this inequality with (115), we obtain that $[p] \in \mathbb{W}_f$. $\qquad\square$

A.7. Quadratization Techniques

In this section we define a sufficient set of atomic reductions to represent methods [22, 15]. We first define a rather general reduction.

Definition A.1. An *injective reduction* is a procedure that for a given energy minimization problem described by $\mathcal{V}, \mathcal{E}, \mathcal{X}, f$ specifies:

  • The reduced energy minimization problem described by $\mathcal{V}', \mathcal{E}', \mathcal{X}', f'$;
  • An injective mapping $\pi \colon \mathcal{X} \to \mathcal{X}'$;
  • A left inverse of π, mapping $\sigma \colon \operatorname{dom}(\sigma) \to \mathcal{X}$, where $\mathcal{X}' \supset \operatorname{dom}(\sigma) \supset \pi(\mathcal{X})$.

The energies are related by $E_{f'}(\pi(x)) = E_f(x)$ for all $x \in \mathcal{X}$.

Mapping π establishes a correspondence between labelings that preserves distinctness. Its left inverse always exists but may be non-unique (when $\pi(\mathcal{X}) \neq \mathcal{X}'$). For some labelings in $\mathcal{X}'$ there may be no meaningful correspondence in $\mathcal{X}$. For this reason the domain of σ is allowed to be specified. We introduce the following atomic injective reductions. First, the ones not changing the hypergraph:

  • `reparametrize`: Let $f' := f - A^\mathsf{T}\varphi$, where A is the matrix of a local relaxation. Mappings π and σ are identity.
  • `permute_labels`: Apply a permutation (a bijection) of labels: for each $s \in \mathcal{V}$, mapping $\pi_s \colon \mathcal{X}_s \to \mathcal{X}_s$ is a bijection, $\mathcal{X}'_s = \mathcal{X}_s$ and $\sigma_s = \pi_s^{-1}$. The reduced energy is $f'_\mathrm{C}(\pi(x_\mathrm{C})) = f_\mathrm{C}(x_\mathrm{C})$.
  • `add_labels`: Expand variable domains as $\mathcal{X}_s \subset \mathcal{X}'_s$ and extend the objective component-wise: let $f'_\mathrm{C}(x'_\mathrm{C}) = f_\mathrm{C}(x'_\mathrm{C})$ for $x'_\mathrm{C} \in \mathcal{X}_\mathrm{C}$ and arbitrary for $x'_\mathrm{C} \in \mathcal{X}'_\mathrm{C} \backslash \mathcal{X}_\mathrm{C}$. Mapping $\pi \colon \mathcal{X} \to \mathcal{X}'$ is the injection $x \mapsto x$. Mapping $\sigma \colon \mathcal{X} \to \mathcal{X}$ is the identity $(\operatorname{dom}(\sigma) = \mathcal{X})$.

Now, atomic reductions changing the hypergraph:

  • `remove_zero`: If the term $f_\mathrm{C}(x_\mathrm{C}) = 0$ for all $x_\mathrm{C} \in \mathcal{X}_\mathrm{C}$, exclude C from $\mathcal{E}$: let $\mathcal{E}' = \mathcal{E} \backslash \{\mathrm{C}\}$. Mapping π is the identity. This operation is the converse of adding zero interactions in [72].
  • `clique_aux`: For a given $\mathrm{C} \in \mathcal{E}$, introduce a new node ω, an auxiliary variable $y \in \mathcal{X}_\omega$ and represent the term f_C as

$$f_\mathrm{C}(x_\mathrm{C}) = \min_y f'_{\mathrm{C} \cup \omega}(x_\mathrm{C}, y). \qquad (117)$$

Define $\mathcal{V}' = \mathcal{V} \cup \{\omega\}$ and

$$\mathcal{E}' = \mathcal{E} \cup \{\mathrm{D} \cup \{\omega\} \mid \mathrm{D} \in \mathcal{E}, \ \mathrm{D} \subset \mathrm{C}\}. \qquad (118)$$

For hyperedges $\mathrm{D} \in \mathcal{E} \backslash \{\mathrm{C}\}$ the energy terms

are copied: $f'_{\mathrm{D}}(x_{\mathrm{D}}) = f_{\mathrm{D}}(x_{\mathrm{D}})$ and the remaining terms of f' (other than in $(\mathcal{E}\setminus\{\mathrm{C}\}) \cup \{\mathrm{C} \cup \{\omega\}\}$) are zero. The new labeling domain is $\mathcal{X}' = \mathcal{X} \times \mathcal{X}_\omega$. Mapping $\pi\colon \mathcal{X} \to \mathcal{X}'$ is defined as $\pi\colon x \mapsto (x, y(x))$ where

$$y(x) = y(x_{\mathrm{C}}) \in \operatorname*{argmin}_{y} f'_{\mathrm{C}\cup\{\omega\}}(x_{\mathrm{C}}, y). \quad (119)$$

Mapping $\sigma\colon \mathcal{X}' \to \mathcal{X}$ is the restriction $(x,y) \mapsto x$.

- `group_aux`: Let $\mathcal{H} \subset \mathcal{E}$ and intersection $\mathrm{C} = \bigcap_{\mathrm{H}\in\mathcal{H}} \mathrm{H} \in \mathcal{E}$. Introduce a new node ω and a variable $y \in \mathcal{X}_\omega$. Let function $y\colon \mathcal{X}_{\mathrm{C}} \to \mathcal{X}_\omega$ be such that $(\forall \mathrm{H} \in \mathcal{H})$

$$f_{\mathrm{H}}(x_{\mathrm{H}}) = f'_{\mathrm{H}\cup\{\omega\}}(x_{\mathrm{H}}, y(x_{\mathrm{C}})) = \min_{y\in\mathcal{X}_\omega} f'_{\mathrm{H}\cup\{\omega\}}(x_{\mathrm{H}}, y).$$

Define $\mathcal{V}' = \mathcal{V} \cup \{\omega\}$ and

$$\mathcal{E}' = \mathcal{E} \cup \bigcup_{\mathrm{H}\in\mathcal{H}} \{\mathrm{D} \cup \{\omega\} \mid \mathrm{D} \in \mathcal{E},\ \mathrm{D} \subset \mathrm{H}\}. \quad (120)$$

Define $\pi\colon \mathcal{X} \to \mathcal{X}'\colon x \mapsto (x, y(x_{\mathrm{C}}))$ and $\sigma\colon \mathcal{X}' \to \mathcal{X}\colon (x,y) \mapsto x$.

In order to relate relaxed-improving maps before and after the reduction we need a correspondence between them. However, not all maps $p\colon \mathcal{X}' \to \mathcal{X}'$ of the reduced problem make sense for the initial problem.

Definition A.2. A node-wise mapping $p'\colon \mathcal{X}' \to \mathcal{X}'$ is *admissible* for a reduction if $p'(\pi(\mathcal{X})) \subset \operatorname{dom}(\sigma)$. In this case $p = \sigma \circ p' \circ \pi$ is the *corresponding* mapping for p'.

All the above atomic reductions fulfill the requirements of Definition A.1 and, with the exception of `add_labels` reduction, all mappings $p'\colon \mathcal{X}' \to \mathcal{X}'$ are admissible.

Definition A.3. A reduction *has inclusion* of relaxed-improving maps for relaxations Λ and Λ' if for every admissible Λ'-improving mapping p' for f' its corresponding mapping p is Λ-improving for f.

Theorem A.4. Atomic reductions `reparametrize` and `permute_labels` have inclusion of relaxed-improving maps for local relaxations.

Proof. Let Λ be a local relaxation. Let p' be Λ-improving for f'. In case of `permute_labels`, marginalization constraints are invariant w.r.t. the order of labels. We trivially get that $\sigma \circ p' \circ \pi$ is Λ-improving for f. In case of `reparametrize`, there holds $\langle f - A^{\mathsf{T}}\varphi, \mu \rangle = \langle f, \mu \rangle$ for all $\mu \in \Lambda$ as well as for all $\mu \in [p'](\Lambda) \subset \Lambda$ and so $p = p'$ is Λ-improving for f. $\qquad\square$

To a hypergraph $(\mathcal{V}, \mathcal{E})$ let us associate an *induced* local relaxation $\Lambda_{\mathcal{E}}$ as follows. For $\mathrm{A}, \mathrm{B} \subset \mathcal{V}$ let $\mathrm{A} \mathrel{\dot{\subsetneq}} \mathrm{B}$ if $\mathrm{A}, \mathrm{B} \in \mathcal{E}$ and $\mathrm{A} \subsetneq \mathrm{B}$. Let $\Lambda_{\mathcal{E}}$ be the local relaxation for the coupling $\dot{\subsetneq}$.

Theorem A.5. Atomic reductions `remove_zero`, `add_labels` have inclusion of admissible relaxed-improving maps for local relaxations $\Lambda_{\mathcal{E}}$ and $\Lambda' = \Lambda_{\mathcal{E}'}$.

Proof (`remove_zero`). Because $\mathcal{E}'$ does not contain hyperedge C, all variables μ_{C}, as well as the associated marginalization constraints, are absent in Λ'. Polytope Λ' can be still represented in $\mathbb{R}^{\mathcal{I}}$ as having unconstrained variables $\mu'_{\mathrm{C}}(x_{\mathrm{C}})$. In this representation $\Lambda \subset \Lambda'$ and Theorem 4.1 applies. $\qquad\square$

Proof (`add_labels`). Let p' be relaxed-improving for f':

$$(\forall \mu' \in \Lambda')\ \langle f', [p']\mu' \rangle \leqslant \langle f', \mu' \rangle. \quad (121)$$

By the dual representation Theorem 3.3(b), there exists dual multipliers ψ for Λ' such that

$$(\forall \mathrm{C} \in \mathcal{E},\ \forall x'_{\mathrm{C}} \in \mathcal{X}'_{\mathrm{C}})\ f'_{\mathrm{C}}(p'(x'_{\mathrm{C}})) \leqslant f'^{\psi}_{\mathrm{C}}(x'_{\mathrm{C}}). \quad (122)$$

It must be that $p'_s(\mathcal{X}_s) \subset \mathcal{X}_s$, otherwise p' is not admissible. The corresponding mapping $p_s\colon \mathcal{X}_s \to \mathcal{X}_s$ is the restriction of p'_s to $\mathcal{X}_s$. Restricting (122) to $\mathcal{X}_s$ we obtain $(\forall \mathrm{C} \in \mathcal{E},\ \forall x_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}})$

$$f_{\mathrm{C}}(p(x_{\mathrm{C}})) \leqslant f_{\mathrm{C}}(x_{\mathrm{C}}) - ((A')^{\mathsf{T}}\psi)_{\mathrm{C}}(x_{\mathrm{C}}). \quad (123)$$

The RHS expression is of the form (55) and thus is a valid reparametrization of f in Λ. Therefore (123) satisfies component-wise inequalities and by Theorem 3.3(b), p is relaxed-improving for f. $\qquad\square$

Theorem A.6. Atomic reductions `clique_aux` and `group_aux` have inclusion of relaxed-improving maps for local relaxations $\Lambda_{\mathcal{E}}$ and $\Lambda' = \Lambda_{\mathcal{E}'}$.

Proof (`clique_aux`). First, let us show that a relaxed solution $\mu \in \Lambda$ can be mapped to a relaxed solution of the reduced problem $\mu' \in \Lambda'$. Let

$$(\forall \mathrm{D} \in \mathcal{E})\ (\forall x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}})\ \mu'_{\mathrm{D}}(x_{\mathrm{D}}) := \mu_{\mathrm{D}}(x_{\mathrm{D}}), \quad (124a)$$

$$(\forall \mathrm{D} \subset \mathrm{C},\ \mathrm{D} \in \mathcal{E})\ (\forall x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}}, \forall y \in \mathcal{X}_\omega) \quad (124b)$$

$$\mu'_{\mathrm{D}\cup\{\omega\}}(x_{\mathrm{D}}, y) := \sum_{x_{\mathrm{C}\setminus\mathrm{D}}} \mu_{\mathrm{C}}(x_{\mathrm{C}})[\![y(x_{\mathrm{C}}){=}y]\!].$$

In particular, $\mu'_{\mathrm{C}\cup\{\omega\}}(x_{\mathrm{C}}, y) = \mu_{\mathrm{C}}(x_{\mathrm{C}})[\![y(x_{\mathrm{C}}){=}y]\!]$ and $\mu'_{\{\omega\}}(y) = \sum_{x_{\mathrm{C}}} \mu_{\mathrm{C}}(x_{\mathrm{C}})[\![y(x_{\mathrm{C}}){=}y]\!]$. Clearly, solution μ' satisfies all those marginalization constraints that μ does. The additional marginalization constraints of

$\Lambda_{\mathcal{E}'}$ are given for $\text{B} \subset \text{C}$, $\text{A} \subsetneq \text{B}$, $\text{A}, \text{B} \in \mathcal{E}$ by couplings

$$\text{A}, \quad \text{B}, \quad \text{A} \cup \{\omega\} \subsetneq \text{B} \cup \{\omega\}. \tag{125}$$

Marginalization for B and $\text{B} \cup \{\omega\}$:

$$\sum_y \mu'_{\text{B} \cup \{\omega\}}(x_{\text{B}}, y) = \sum_y \sum_{x_{\text{C} \backslash \text{B}}} \mu_{\text{C}}(x_{\text{C}}) [\![y(x_{\text{C}}) = y]\!]$$

$$= \sum_{x_{\text{C} \backslash \text{B}}} \mu_{\text{C}}(x_{\text{C}}) = \mu_{\text{B}}(x_{\text{B}}) = \mu'_{\text{B}}(x_{\text{B}}). \tag{126}$$

Marginalization for A and $\text{B} \cup \{\omega\}$:

$$\sum_{x_{\text{B} \backslash \text{A}}, y} \mu'_{\text{B} \cup \{\omega\}}(x_{\text{B}}, y) = \sum_{x_{\text{B} \backslash \text{A}}, y} \sum_{x_{\text{C} \backslash \text{B}}} \mu_{\text{C}}(x_{\text{C}}) [\![y(x_{\text{C}}) = y]\!]$$

$$= \sum_{x_{\text{C} \backslash \text{A}}} \mu_{\text{C}}(x_{\text{C}}) = \mu_{\text{A}}(x_{\text{A}}) = \mu'_{\text{A}}(x_{\text{A}}). \tag{127}$$

Marginalization for $\text{A} \cup \{\omega\}$ and $\text{B} \cup \{\omega\}$:

$$\sum_{x_{\text{B} \backslash \text{A}}} \mu'_{\text{B} \cup \{\omega\}}(x_{\text{B}}, y) = \sum_{x_{\text{B} \backslash \text{A}}} \sum_{x_{\text{C} \backslash \text{B}}} \mu_{\text{C}}(x_{\text{C}}) [\![y(x_{\text{C}}) = y]\!]$$

$$= \sum_{x_{\text{C} \backslash \text{A}}} \mu_{\text{C}}(x_{\text{C}}) [\![y(x_{\text{C}}) = y]\!] = \mu'_{\text{A} \cup \{\omega\}}(x_{\text{A}}, y). \tag{128}$$

Therefore $\mu' \in \Lambda'$. Since we have the equality $f_{\text{C}}(x_{\text{C}}) = f'_{\text{C} \cup \{\omega\}}(x_{\text{C}}, y(x_{\text{C}}))$, there holds

$$\sum_{x_{\text{C}}} f_{\text{C}}(x_{\text{C}}) \mu_{\text{C}}(x_{\text{C}}) = \sum_{x_{\text{C}}} f'_{\text{C} \cup \{\omega\}}(x_{\text{C}}, y(x_{\text{C}})) \mu_{\text{C}}(x_{\text{C}})$$

$$= \sum_{x_{\text{C}}, y} f'_{\text{C} \cup \{\omega\}}(x_{\text{C}}, y) [\![y = y(x_{\text{C}})]\!] \mu_{\text{C}}(x_{\text{C}})$$

$$= \sum_{x_{\text{C}}, y} f'_{\text{C} \cup \{\omega\}}(x_{\text{C}}, y) \mu'_{\text{C} \cup \{\omega\}}(x_{\text{C}}, y). \tag{129}$$

For other components $\text{D} \in \mathcal{E} \backslash \{\text{C}\}$ we have

$$\sum_{x_{\text{D}}} f_{\text{D}}(x_{\text{D}}) \mu_{\text{D}}(x_{\text{D}}) = \sum_{x_{\text{D}}} f'_{\text{D}}(x_{\text{D}}) \mu'_{\text{D}}(x_{\text{D}}) \tag{130}$$

and for all $\text{D} \subsetneq \text{C}$, $\text{D} \in \mathcal{E}$ we have $f'_{\text{D} \cup \{\omega\}} = 0$. Let us denote the linear mapping $\mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}'}$ (124) by Π. It follows that for all $\mu \in \Lambda$ there holds $\Pi\mu \in \Lambda'$ and from (129),(130) that

$$\langle f, \mu \rangle = \langle f', \Pi\mu \rangle. \tag{131}$$

This proves that

$$\min_{\mu \in \Lambda} \langle f, \mu \rangle \geqslant \min_{\mu' \in \Lambda'} \langle f', \mu' \rangle, \tag{132}$$

i.e., relaxation Λ' is not tighter than Λ.

Let $p' \colon \mathcal{X}' \to \mathcal{X}'$ be node-wise Λ'-improving for f' and $p = \sigma \circ p' \circ \pi$, *i.e.*, $p_s = p'_s$ for $s \in \mathcal{V}$. Let $P' = [p']$, the extension to $\mathbb{R}^{\mathcal{I}'}$. Similarly to (129) and (130) we can express parts of the scalar product

$$\langle f, [p]\mu \rangle = \sum_{\text{D} \in \mathcal{E}} \sum_{x_{\text{D}} \in \mathcal{X}_{\text{D}}} f_{\text{D}}(p'_{\text{D}}(x_{\text{D}})) \mu_{\text{D}}(x_{\text{D}}) \tag{133}$$

for each $\text{D} \in \mathcal{E}$ as follows. For $\text{D} \in \mathcal{E} \backslash \{\text{C}\}$ we have

$$\sum_{x_{\text{D}}} f_{\text{D}}(p'_{\text{D}}(x_{\text{D}})) \mu_{\text{D}}(x_{\text{D}}) \tag{134}$$

$$= \sum_{x_{\text{D}}} \sum_{\tilde{x}_{\text{D}}} f_{\text{D}}(\tilde{x}_{\text{D}}) [\![\tilde{x}_{\text{D}} = p'(x_{\text{D}})]\!] \mu'_{\text{D}}(x_{\text{D}})$$

$$= \sum_{\tilde{x}_{\text{D}}} f'_{\text{D}}(\tilde{x}_{\text{D}}) (P'\mu')_{\text{D}}(\tilde{x}_{\text{D}}). \tag{135}$$

For C let $\text{C}' = \text{C} \cup \{\omega\}$, we have

$$\sum_{x_{\text{C}}} f_{\text{C}}(p'_{\text{C}}(x_{\text{C}})) \mu_{\text{C}}(x_{\text{C}}) \tag{136a}$$

$$= \sum_{x_{\text{C}}} \sum_{\tilde{x}_{\text{C}}} f_{\text{C}}(\tilde{x}_{\text{C}}) [\![\tilde{x}_{\text{C}} = p'(x_{\text{C}})]\!] \mu_{\text{C}}(x_{\text{C}})$$

$$= \sum_{x_{\text{C}}} \sum_{\tilde{x}_{\text{C}}} f'_{\text{C}'}(\tilde{x}_{\text{C}}, y(\tilde{x}_{\text{C}})) [\![\tilde{x}_{\text{C}} = p'(x_{\text{C}})]\!] \mu_{\text{C}}(x_{\text{C}}) \tag{136b}$$

$$\leqslant \sum_{x_{\text{C}}} \sum_{\tilde{x}_{\text{C}}} f'_{\text{C}'}(\tilde{x}_{\text{C}}, p'_{\omega}(y(x_{\text{C}}))) [\![\tilde{x}_{\text{C}} = p'(x_{\text{C}})]\!] \mu_{\text{C}}(x_{\text{C}})$$

$$= \sum_{x_{\text{C}}, y} \sum_{\tilde{x}_{\text{C}}, \tilde{y}} f'_{\text{C}'}(\tilde{x}_{\text{C}}, \tilde{y}) [\![\tilde{x}_{\text{C}} = p'(x_{\text{C}})]\!] [\![\tilde{y} = p'_{\omega}(y)]\!]$$
$$[\![y = y(x_{\text{C}})]\!] \mu_{\text{C}}(x_{\text{C}}) \tag{136c}$$

$$= \sum_{x_{\text{C}}, y} \sum_{\tilde{x}_{\text{C}}, \tilde{y}} f'_{\text{C}'}(\tilde{x}_{\text{C}}, \tilde{y}) P'_{\text{C}', (\tilde{x}_{\text{C}}, \tilde{y}), (x_{\text{C}}, y)} \mu'_{\text{C}'}(x_{\text{C}}, y) \tag{136d}$$

$$= \sum_{\tilde{x}_{\text{C}}, \tilde{y}} f'_{\text{C} \cup \{\omega\}}(\tilde{x}_{\text{C}}, \tilde{y}) (P'\mu')_{\text{C} \cup \{\omega\}}(\tilde{x}, \tilde{y}). \tag{136e}$$

The inequality is due to $y(\tilde{x}_{\text{C}}) \in \arg\min_y f'_{\text{C}'}(\tilde{x}_{\text{C}}, y)$. For all $\text{D} \subsetneq \text{C}$, $\text{D} \in \mathcal{E}$ there holds $f'_{\text{D} \cup \{\omega\}} = 0$. It follows that

$$\langle f, [p]\mu \rangle = \langle f', [p']\mu' \rangle \tag{137}$$
$$\leqslant \langle f', \mu' \rangle = \langle f', \Pi\mu \rangle = \langle f, \mu \rangle.$$

Therefore $[p] \in \mathbb{W}_f$. $\qquad\square$

Proof (`group_aux`). Because $y(x_{\text{C}})$ minimizes all group terms simultaneously, arguments of the proof for `clique_aux` apply to each $\text{H} \in \mathcal{H}$. It follows that the final relation (137) is satisfied. $\qquad\square$

Theorem 4.6. Persistency by Higher Order Clique Reduction (HOCR) of Ishikawa [22] corresponds to an FLP-improving mapping.

Proof. Reductions used in HOCR are of the following form:

$$x_1 x_2 \ldots x_n = \min_{y_1, y_2, \ldots y_k} g(x, y), \tag{138}$$

where function $g(x, y)$ is by design partially separable so that the reduction decreases the order of the problem. This reduction can be implemented as follows:

- Use `clique_aux` to introduce new variables

$y_1, \ldots y_k$, one at a time, *e.g.*, first let $x_1 x_2 \ldots x_n = \min_{y_1} f^1(x, y_1)$, where $f^1(x, y_1) = \min_{y_2, \ldots y_k} g(x, y)$, and so on.

- Use `reparametrize` to rewrite the term $g(x, y)$ that gets assigned to the hyperedge over all variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_k$ as a sum of terms over smaller subsets of variables as guaranteed by design of g.
- Use `remove_zero` to clean up the hypergraph so that all now zero higher order terms are excluded.

HOCR full reduction can be implemented by iterating the steps above. Starting from the FLP relaxation of the initial problem, the necessary `reparametrize` operations are feasible as we add all possible interactions during `clique_aux`. Each atomic reformulation has inclusion of relaxed-improving maps. It follows that the whole reformulation has inclusion of relaxed-improving maps. The persistency in the final reformulation is obtained with QPBO, which by Theorem 4.3 is an FLP-improving mapping. It follows that there is a corresponding FLP-improving mapping in the initial formulation. □

Theorem 4.7. Persistency by method of Fix et al. [15] corresponds to an FLP-improving mapping.

Proof. Method [15] uses reductions of HOCR and a new reduction for a group of cliques [15, Theorem 3.1]. Let us show that the group reduction has the form that can be implemented using `group_aux`, `reparametrize` and `remove_zero`. The optimal value of the auxiliary variable y in [15, Theorem 3.1] depends on the assignment of x_C only: it is given by $y(x_C) = 1 - \prod_{j \in C} x_j$. Since all hyperedges $H \in \mathcal{H}$ contain C, the reduced expression [15, equation (2)] equals

$$\sum_{H \in \mathcal{H}} \alpha_H \Big(y(x_C) \prod_{j \in C} x_j + (1 - y(x_C)) \prod_{j \in H \setminus C} x_j \Big), \quad (139)$$

where $\alpha_H > 0$, *i.e.*, it has the form required by `group_aux`. The actual simplification of the problem is achieved by applying `reparametrize` and `remove_zero` similarly to HOCR.

It follows that reduction [15] can be implemented using operations `clique_aux`, `group_aux`, `reparametrize` and `remove_zero`. □

A.8. Bisubmodular Relaxations

The section is organized as follows. We review definitions and the persistency property of bisubmodular functions. Then we define the sum of bisubmodular functions relaxation via injection of label space $\{0, 1\}$ into $\{0, \frac{1}{2}, 1\}$ (with operation `add_labels`). The comparison result is then achieved in two steps. We show that persistency for a bisubmodular function is a BLP-improving mapping on 3 labels ($\{0, \frac{1}{2}, 1\}$) admissible for `add_labels` and then apply the result that `add_labels` has inclusion of BLP-improving maps to transfer all persistencies by bisubmodular relaxations to the BLP-relaxation of the initial problem.

We follow the notation of [28]. A sum of bisubmodular functions (SoB) relaxation is constructed as follows. Let $\mathcal{X}_s = \{0, 1\}$ and

$$\mathcal{K}^{1/2} = \{0, \frac{1}{2}, 1\}^{\mathcal{V}}. \quad (140)$$

Define binary operations $\sqcap, \sqcup : \mathcal{K}^{1/2} \to \mathcal{K}^{1/2}$ component-wise according to:

$\sqcap$	0	$\frac{1}{2}$	1		$\sqcup$	0	$\frac{1}{2}$	1
0	0	$\frac{1}{2}$	$\frac{1}{2}$		0	0	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{2}$	0	$\frac{1}{2}$	1
1	$\frac{1}{2}$	$\frac{1}{2}$	1		1	$\frac{1}{2}$	1	1

$$(141)$$

Definition A.7. Function $f : \mathcal{K}^{1/2} \to \mathbb{R}$ is called *bisubmodular* if

$$(\forall x, y \in \mathcal{K}^{1/2}) \quad f(x \sqcap y) + f(x \sqcup y) \leqslant f(x) + f(y). \quad (142)$$

Definition A.8. Function $f' : \mathcal{K}^{1/2} \to \mathbb{R}$ is a *sum of bisubmodular functions relaxation* (SoB relaxation) for $f : \mathcal{X} \to \mathbb{R}$ if for every $C \in \mathcal{E}$ there holds:

- $f'_C : \mathcal{K}_C^{1/2} \to \mathbb{R}$ is bisubmodular;
- $(\forall e \in \mathcal{E}, \forall x_C \in \mathcal{X}_C) \quad f'_C(x) = f_C(x)$;

The next lemma reviews persistency according to [28, Proposition 12].

Lemma A.9. Let $x^* \in \mathcal{K}^{1/2}$ be a minimizer of f'. Define the following mappings

$$p'_s : \mathcal{K}_s^{1/2} \to \mathcal{K}_s^{1/2} : x_s \mapsto (x_s \sqcup x_s^*) \sqcup x_s^*; \quad (143a)$$

$$p_s : \mathcal{X}_s \to \mathcal{X}_s : x_s \mapsto \begin{cases} x_s^*, & \text{if } x_s^* \neq \frac{1}{2}; \\ x_s, & \text{otherwise.} \end{cases} \quad (143b)$$

There holds
- (a) Autarky: $(\forall x \in \mathcal{K}^{1/2}) \quad f'(p'(x)) \leqslant f'(x)$;
- (b) $p'(\mathcal{X}) \subset \mathcal{X}$ and p is the restriction of p' to $\mathcal{X}$.

Proof. Part (a) follows from bisubmodularity (142) using that x^* is a minimizer: for all $y \in \mathcal{K}^{1/2}$ there

holds

$$f'(y \sqcup x^*) \leqslant \big(f'(x^*) - f'(y \sqcap x^*)\big) + f'(y) \leqslant f'(y). \tag{144}$$

It follows that $(\forall x \in \mathcal{K}^{1/2})$

$$f'((x \sqcup x^*) \sqcup x^*) \leqslant f'(x \sqcup x^*) \leqslant f'(x). \tag{145}$$

Part (b) can be verified explicitly by calculating the tables for mappings p and p'. They are represented respectively as:

$p_s(x_s)$					$p'_s(x_s)$			
$x_s \backslash x_s^*$	0	$\frac{1}{2}$	1		$x_s \backslash x_s^*$	0	$\frac{1}{2}$	1
0	0	0	1		0	0	0	1
1	0	1	1		$\frac{1}{2}$	0	$\frac{1}{2}$	1
					1	0	1	1

(146)

It is seen that for non-fractional x_s the results match.
□

We now upgrade the autarky property by Lemma A.9(a) to the statement that p' is BLP-improvinf for f'. In order to show this, we construct a reparametrization in which the improving inequality holds component-wise. This reparametrization is given by an optimal dual point, provided that it preserves bisubmodularity of all components. The next Lemma shows such a dual solution exists.

Lemma A.10. Let $f \colon \mathcal{K}^{1/2} \to \mathbb{R}$ be a sum of bisubmodular functions. BLP for f admits an optimal dual solution φ such that each component

$$f_{\mathrm{C}}^{\varphi}(x_{\mathrm{C}}) = f_{\mathrm{C}}(x_{\mathrm{C}}) - (A^{\mathsf{T}}\varphi)_{\mathrm{C}}(x_{\mathrm{C}}) \tag{147}$$

is bisubmodular.

Proof. The plan of the proof is as follows:

- Reformulate bisubmodular function $f \colon \mathcal{K}^{1/2} \to \mathbb{R}$ as a function $g \colon \mathbb{B}^{2\mathcal{V}} \to \mathbb{R}$.
- Show there exist a symmetric optimal dual solution to this BLP. Such symmetric solution defines a reparametrization which preserves the sum of bisubmodular functions property for g and consequently for f.

The following reformulation of bisubmodular function $f \colon \mathcal{K}^{1/2} \to \mathbb{R}$ as a function $g \colon \mathbb{B}^{2\mathcal{V}} \to \mathbb{R}$ is according to [28]. Node s is represented by a pair of nodes (s, s'). Label $x_s \in \mathcal{K}^{1/2}$ is represented by a pair of labels (u_s, u'_s) as follows:

$$0 \to (1, 0); \quad 1 \to (0, 1); \quad \frac{1}{2} \to (0, 0). \tag{148}$$

To a hyperedge C there corresponds hyperedge $\mathrm{C} \cup \mathrm{C}'$,

where $\mathrm{C}' = \{s' \mid s \in \mathrm{C}\}$. In the hypergraph $(\mathcal{V} \cup \mathcal{V}', \mathcal{G})$ there are edges of the form $\mathrm{C} \cup \mathrm{C}'$, $\{s\}$ and $\{s'\}$ for $\mathrm{C} \in \mathcal{E}$ and $s \in \mathcal{V}$. We will denote components $g_{\mathrm{C} \cup \mathrm{C}'}$ simply by g_{C} for $|\mathrm{C}| \geqslant 1$. The energy g is defined by

$$(\forall u \in \mathbb{B}^{\mathrm{C} \times \mathrm{C}'}) \quad g_{\mathrm{C}}(u_{\mathrm{C}}, u_{\mathrm{C}'}) = f_{\mathrm{C}}\left(\frac{u_{\mathrm{C}} + \bar{u}_{\mathrm{C}'}}{2}\right), \tag{149}$$

where $\bar{u} := 1 - u$. By definition, g has the symmetry property:

$$(\forall u \in \mathbb{B}^{\mathrm{C} \times \mathrm{C}'}) \quad g_{\mathrm{C}}(u_{\mathrm{C}}, u_{\mathrm{C}'}) = g(\bar{u}_{\mathrm{C}'}, \bar{u}_{\mathrm{C}}). \tag{150}$$

Let $\mathcal{X}_s^- = \{(u_s, u_{s'}) \in \mathbb{B}^2 \mid (u_s, u_{s'}) \neq (1, 1)\}$ and $\mathcal{X}^- = \prod_{s \in \mathcal{V}} \mathcal{X}_s^-$. For $u, v \in \mathcal{X}^-$ operations $\sqcup$, $\sqcap$ are defined by

$$u \sqcap v = u \wedge v, \tag{151a}$$
$$u \sqcup v = \mathtt{reduce}(u \vee v), \tag{151b}$$

where $\mathtt{reduce}(w)$ is the labeling obtaining from w by changing labels $(w_s, w_{s'})$ from $(1, 1)$ to $(0, 0)$ for all $s \in \mathcal{V}$. It can be seen that these definitions are consistent with equations (141). Components g_{C} satisfy

$$(\forall u_{\mathrm{C}}, v_{\mathrm{C}} \in \mathcal{X}_{\mathrm{C}}^-) \quad g_{\mathrm{C}}(u_{\mathrm{C}} \sqcap v_{\mathrm{C}}) + g_{\mathrm{C}}(u_{\mathrm{C}} \sqcup v_{\mathrm{C}}) \\ \leqslant g_{\mathrm{C}}(u_{\mathrm{C}}) + g_{\mathrm{C}}(v_{\mathrm{C}}). \tag{152}$$

With this reformulation we proceed as follows. Let φ be dual optimal to g. According to the hypergraph $(\mathcal{V} \cup \mathcal{V}', \mathcal{G})$, there are only components $\varphi_{s, \mathrm{C} \cup \mathrm{C}'}(i)$, $\varphi_{s', \mathrm{C} \cup \mathrm{C}'}(i)$ for $\mathrm{C} \in \mathcal{E}$, $|\mathrm{C}| > 1$, $s \in \mathrm{C}$, and $i \in \mathbb{B}$ and components $\varphi_{\varnothing, \mathrm{C} \cup \mathrm{C}'}$ for $\mathrm{C} \in \mathcal{E}$. Similarly to function g, we can denote the index $\mathrm{C} \cup \mathrm{C}'$ as just C.

If $\varphi_{\varnothing, \mathrm{C}} \neq 0$ for some $\mathrm{C} \in \mathcal{E}$, we can apply the transformation $g_{\mathrm{C}} := g_{\mathrm{C}} - \varphi_{\varnothing, \mathrm{C}}$; $g_{\varnothing} := g_{\varnothing} + \varphi_{\varnothing, \mathrm{C}}$. Clearly this constant transformation does not change bisubmodularity of g_{C}. Without loss of generality let us assume now that $\varphi_{\varnothing, \mathrm{C}} = 0$ for all $\mathrm{C} \in \mathcal{E}$.

Dual feasibility of BLP relaxation for each component $\mathrm{C} \cup \mathrm{C}'$ where $\mathrm{C} \in \mathcal{E}$, $|\mathrm{C}| > 1$ reads $(\forall u_{\mathrm{C} \cup \mathrm{C}'})$

$$g_{\mathrm{C}}(u_{\mathrm{C}}, u_{\mathrm{C}'}) - \sum_{s \in \mathrm{C}} \big(\varphi_{s, \mathrm{C}}(u_s) + \varphi_{s', \mathrm{C}}(u_{s'})\big) \geqslant 0. \tag{153}$$

And for unary terms (*i.e.*, $\mathrm{C} = \{s\}$), it is $(\forall u_{\{s, s'\}})$

$$g_s(u_s, u_{s'}) + \sum_{\mathrm{D} \supsetneq \{s\}} \big(\varphi_{s, \mathrm{D}}(u_{s'}) + \varphi_{s', \mathrm{D}}(u_s)\big) \geqslant 0. \tag{154}$$

By symmetry of g, condition (153) is equivalent to $(\forall u_{\mathrm{C} \cup \mathrm{C}'})$

$$g_{\mathrm{C}}(\bar{u}_{\mathrm{C}'}, \bar{u}_{\mathrm{C}}) - \sum_{s \in \mathrm{C}} \big(\varphi_{s, \mathrm{C}}(u_s) + \varphi_{s', \mathrm{C}}(u_{s'})\big) \geqslant 0 \tag{155}$$

and, by flipping all bound variables $u_{\mathrm{C} \cup \mathrm{C}'}$, to $(\forall u_{\mathrm{C} \cup \mathrm{C}'})$

$$g_{\mathrm{C}}(u_{\mathrm{C}}, u_{\mathrm{C}'}) - \sum_{s \in \mathrm{C}} \big(\varphi_{s, \mathrm{C}}(\bar{u}_{s'}) + \varphi_{s', \mathrm{C}}(\bar{u}_s)\big) \geqslant 0. \tag{156}$$

Similarly, for unary terms there holds $(\forall u_{\{s,s'\}})$

$$g_s(u_s, u_{s'}) + \sum_{\mathrm{D} \rightleftharpoons \{s\}} \big(\varphi_{s,\mathrm{D}}(\bar{u}_{s'}) + \varphi_{s',\mathrm{D}}(\bar{u}_s)\big) \geqslant 0. \quad (157)$$

Therefore the dual point φ' with the following components is feasible:

$$\begin{aligned} \varphi'_{s,\mathrm{C}}(i) &:= \varphi_{s',\mathrm{C}}(\bar{i}), \\ \varphi'_{s',\mathrm{C}}(i) &:= \varphi_{s,\mathrm{C}}(\bar{i}) \end{aligned} \quad (158)$$

for $\mathrm{C} \in \mathcal{E}$, $|\mathrm{C}| > 1$, $s \in \mathrm{C}$, $i \in \mathbb{B}$ (components $\varphi'_{\varnothing,\mathrm{C}} = 0$ for $\mathrm{C} \in \mathcal{E}$ are omitted in this and subsequent steps). Solution φ' is clearly optimal to BLP of g since

$$g_{\varnothing}^{\varphi} = g_{\varnothing} = g_{\varnothing}^{\varphi'}.$$

Therefore the following symmetrized solution is optimal:

$$\begin{aligned} \hat{\varphi}_{s,\mathrm{C}}(u_s) &:= \frac{1}{2}\Big(\varphi_{s,\mathrm{C}}(u_s) + \varphi_{s',\mathrm{C}}(\bar{u}_s)\Big), \\ \hat{\varphi}_{s',\mathrm{C}}(u_{s'}) &:= \frac{1}{2}\Big(\varphi_{s',\mathrm{C}}(u_{s'}) + \varphi_{s,\mathrm{C}}(\bar{u}_{s'})\Big). \end{aligned} \quad (159)$$

Let us check it is bisubmodular. We map this dual solution back to BLP for f by reversing (149) as

$$\begin{aligned} \tilde{\varphi}_{s,\mathrm{C}}(0) &:= \hat{\varphi}_{s,\mathrm{C}}(1) + \hat{\varphi}_{s',\mathrm{C}}(0) = \varphi_{s,\mathrm{C}}(0) + \varphi_{s',\mathrm{C}}(1), \\ \tilde{\varphi}_{s,\mathrm{C}}(1) &:= \hat{\varphi}_{s,\mathrm{C}}(0) + \hat{\varphi}_{s',\mathrm{C}}(1) = \varphi_{s,\mathrm{C}}(1) + \varphi_{s',\mathrm{C}}(0), \\ \tilde{\varphi}_{s,\mathrm{C}}\Big(\frac{1}{2}\Big) &:= \hat{\varphi}_{s,\mathrm{C}}(0) + \hat{\varphi}_{s',\mathrm{C}}(0) \quad (160) \\ &= \frac{1}{2}\Big(\varphi_{s,\mathrm{C}}(0) + \varphi_{s,\mathrm{C}}(1)\Big) + \frac{1}{2}\Big(\varphi_{s',\mathrm{C}}(0) + \varphi_{s',\mathrm{C}}(1)\Big) \end{aligned}$$

It satisfies $\tilde{\varphi}_{s,\mathrm{C}}(\frac{1}{2}) = \frac{1}{2}(\tilde{\varphi}_{s,\mathrm{C}}(0) + \tilde{\varphi}_{s,\mathrm{C}}(1))$ and therefore both $\tilde{\varphi}_{s,\mathrm{C}}$ and $-\tilde{\varphi}_{s,\mathrm{C}}$ are bisubmodular.

It remains to show that $\tilde{\varphi}$ is optimal to BLP of f'. While it is well known that BLP relaxation is tight for SoB function $f' \colon \mathcal{K}^{1/2} \to \mathbb{R}$, $e.g.$ [66], it is not obvious that BLP relaxation for the reformulation $g \colon \mathbb{B}^{2\mathcal{V}} \to \mathbb{R}$ is tight as well. Let us show this is the case. It will follow then that $\tilde{\varphi}$, constructed from an optimal dual solution $\hat{\varphi}$ to BLP of g, is optimal to BLP of f'.

Statement A.11. The BLP relaxation for sum of bisubmodular functions $g \colon \mathbb{B}^{2\mathcal{V}} \to \mathbb{R}$ is tight.

Proof. The schema of the proof is similar to $e.g.$ [12, T.6.2] or [52] who considered sum of submodular functions. We construct a primal integer optimal solution from an arc-consistent optimal dual solution. In the construction we will need that the reparametrized problem g^{φ} is a sum of bisubmodular functions. Therefore we need an arc consistent symmetric optimal dual solution.

We start by taking a pair (μ, φ) that satisfies strict complementarity slackness for BLP of g. Since arc consistency is a necessary condition for strict complementarity, φ is arc consistent. As was shown before, φ' defined by equations (158) is dual optimal to BLP of g. It is arc consistent because φ was arc consistent. By taking a symmetrized dual solution $\hat{\varphi}$ defined by equations (159) we obtain an optimal symmetric arc consistent dual solution. By symmetry, it preserves component-wise bisubmodularity.

Let now $\varphi := \hat{\varphi}$. We construct an integer solution as:

$$u_s^* = \bigwedge\{i \in \mathbb{B} \mid g_s^{\varphi}(i) = 0\}. \quad (161)$$

In order to show that u^* is optimal we prove complementarity with φ. Let $\mathrm{C} \in \mathcal{E}$. By arc consistency, for every $s \in \mathrm{C}$ there exists $u_{\mathrm{C}}^{(s)}$ such that $u_{\mathrm{C}}^{(s)} = u_s^*$ and $g_{\mathrm{C}}^{\varphi}(u_{\mathrm{C}}^{(s)}) = 0$. It also follows that $(\forall t \in \mathrm{C} \mid t \neq s)$ $g_t^{\varphi}(u_t^{(s)}) = 0$ and hence $u_t^{(s)} \geqslant u_t^*$. It follows that $u_{\mathrm{C}}^* = \bigwedge_{s \in \mathrm{C}} u_{\mathrm{C}}^{(s)}$ and by component-wise bisubmodularity we have that

$$g_{\mathrm{C}}^{\varphi}\Big(\bigwedge_{s \in \mathrm{C}} u_{\mathrm{C}}^{(s)}\Big) = 0.$$

From feasibility and complementarity slackness follows u^* is optimal. Therefore BLP relaxation for g is tight. $\qquad\square$

It follows that $\tilde{\varphi}$ constructed in (160) not only feasible to BLP of f' but achieves the optimal dual objective $f_{\varnothing}^{\tilde{\varphi}} = g_{\varnothing}^{\varphi} = \min_{u \in \mathbb{B}^{2\mathcal{V}}} E_g(u)$. $\qquad\square$

Lemma A.12. Mapping p' defined in Lemma A.9 by an optimal solution $x^* \in \mathcal{K}^{1/2}$ is BLP-improving for f'.

Proof. Let φ provide a component-wise bisubmodular optimal reparametrization f'^{φ} for BLP of f' which exists by Lemma A.10. Since BLP for f' is tight, φ and $\delta(x^*)$ must satisfy complementary slackness. For every hyperedge $\mathrm{C} \in \mathcal{E}$, $\mathrm{C} \neq \varnothing$ by dual feasibility $f_{\mathrm{C}}'^{\varphi} \geqslant 0$ and by complementary slackness

$$f_{\mathrm{C}}'^{\varphi}(x_{\mathrm{C}}^*) = 0. \quad (162)$$

Therefore, labeling x_{C}^* is a minimizer of $f_{\mathrm{C}}'^{\varphi}$. Since $f_{\mathrm{C}}'^{\varphi}$ is bisubmodular it follows by the same argument as in equations (144), (145) that for each hyperedge C we have the component-wise autarky property:

$$(\forall x_{\mathrm{C}} \in \mathcal{K}_{\mathrm{C}}^{1/2}) \quad f_{\mathrm{C}}'^{\varphi}(p'_{\mathrm{C}}(x_{\mathrm{C}})) \leqslant f_{\mathrm{C}}'^{\varphi}(x_{\mathrm{C}}); \quad (163)$$

By the characterization Theorem 3.3(c), mapping p' is BLP-improving for f'. $\qquad\square$

Theorem 4.8. Persistency by SoB relaxation [27] corresponds to a BLP-improving mapping.

Proof. Let p' and p be mappings defined in Lemma A.9, corresponding to the persistency [27]. We need to show that mapping p is BLP-improving for f. By Lemma A.12, p' is BLP-improving for f'. The discrete relaxation f' is obtained from f with `add_labels` which has inclusion of admissible BLP-improving maps (Theorem A.5). By Lemma A.9(b), p' is admissible and thus Theorem A.5 applies. $\square$

A.9. Persistency in 0-1 Polynomial Programming by Adams et al.

The 0-1 polynomial programming problem is the following optimization problem:

$$\min \sum_{J \subset \mathcal{V}} c_J \prod_{j \in J} x_j, \qquad \text{(PP)}$$

where x is binary and coefficients $c_J \in \mathbb{R}$ are not necessarily all non-zero. The objective of (PP) is a multilinear polynomial expression, which can be written uniquely for any polynomial in 0-1 variables applying the identity $x_j^2 = x_j$. Adams et al. [2] considered a hierarchy of relaxations of Sherali and Adams [58] for this problem. A relaxation of level d can be constructed assuming that for all $|J| > d$ coefficients c_J are zero.

To match their result we need the following:

- A hypergraph $(\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ contains all subsets of $\mathcal{V}$ of cardinality up to d (in case $d = 2$ it is a fully connected graph):

$$\mathcal{E} = \{ J \subset \mathcal{V} \mid |J| \leqslant d \}$$

- Represent the problem (PP) as energy minimization with terms

$$f_J(x_J) = \begin{cases} c_J & \text{if } x_J = 1_J, \\ 0 & \text{otherwise.} \end{cases} \qquad (164)$$

- Consider the FLP relaxation.

The persistency result [2, Lemma 3.2] can be described as follows. Their lemma partitions the set of nodes as $\mathcal{V} = N^+ \cup N^- \cup N^f$. A sufficient condition is proposed implying that the partial assignment $x_{N^-} = 0$, $x_{N^+} = 1$ is globally optimal.

We interpret their result in the dual decomposition framework. The hypergraph $(\mathcal{V}, \mathcal{E})$ is split into two parts:

- Nodes $\mathcal{V}_1 = N^f$, hyperedges $\mathcal{E}_1 = \{ J \in \mathcal{E} \mid J \subset N^f \}$.
- Nodes $\mathcal{V}_2 = N^+ \cup N^- \cup \mathcal{B}$, hyperedges $\mathcal{E}_2 = \{ J \in \mathcal{E} \mid J \nsubseteq N^f \}$, where $\mathcal{B}$ is the following *boundary*

set:

$$\mathcal{B} = \{ v \in N^f \mid (\exists J \in \mathcal{E}) \; v \in J, \; J \nsubseteq N^f \}. \quad (165)$$

It can be seen that hyperedges $\mathcal{E}_1$, $\mathcal{E}_2$ form a partition of $\mathcal{E}$ whereas the sets of nodes $\mathcal{V}_1$, $\mathcal{V}_2$ overlap over $\mathcal{B}$. Accordingly to the two hypergraphs we define the decomposition:

$$f(x) = f^1(x_{\mathcal{V}_1}) + f^2(x_{\mathcal{V}_2}), \qquad (166)$$

where

$$f^1 \colon \mathcal{V}_1 \to \mathbb{R} \colon x \mapsto \sum_{J \in \mathcal{E}_1} c_J x_J, \qquad (167a)$$

$$f^2 \colon \mathcal{V}_2 \to \mathbb{R} \colon x \mapsto \sum_{J \in \mathcal{E}_2} c_J x_J. \qquad (167b)$$

If we found an optimal solution to f^1 and an optimal solution to f^2 and accidentally they were consistent over the overlap part $\mathcal{B}$ we would have obtained an optimal solution to f.

We will show that the conditions of Adams et al. [2] imply that an arbitrary solution $x_{\mathcal{V}_1}$ to f^1 defines an optimal solution x' to f^2 given by

$$x'_s = \begin{cases} x_s, & s \in \mathcal{B} \\ 0, & s \in N^- \\ 1, & s \in N^+. \end{cases} \qquad (168)$$

In other words, any extension of $x_{\mathcal{V}_1}$ to $\mathcal{V}_2$ defines an optimal solution to f^2. In fact, under these conditions, what happens inside the problem f^1 is irrelevant.

Recall that the relaxation [2] is obtained by introducing a relaxed variable $w_J \in [0, 1]$ in place of every product $\prod_{s \in J} x_J$. The constraints are represented by the following linear forms:

$$f_d(J_1, J_2) = \sum_{J' \subset J_2} (-1)^{|J'|} w_{J_1 \cup J'} \qquad (169)$$

for each $J_1 \cap J_2 = \varnothing$ and $|J_1 \cup J_2| = d$. The relaxation, denoted $\mathrm{LP}(d/n)$ is given by

$$\min \sum_{J \subset \mathcal{V}} c_J w_J \qquad (170)$$

$$(\forall J_1, J_2 \mid |J_1 \cup J_2| = d, J_1 \cap J_2 = \varnothing) \; f_d(J_1, J_2) \geqslant 0.$$

This relaxation can be matched to FLP by the relation

$$\mu_J(x_J) = f_d(J_1, J_2), \qquad (171)$$

where $J_1 \cap J_2 = \varnothing$, $J_1 \cup J_2 = J$, $d = |J|$, $x_{J_1} = 1$, $x_{J_2} = 0$. In particular,

$$\mu_J(1_J) = w_J, \qquad (172)$$

where 1_J is a $|J|$-vector of ones. For w feasible to

LP(d/n), the terms $f_d(J_1, J_2)$ are non-negative and satisfy marginalization constraints according to [2, equation 2.3]. It follows from these marginalization constraints that a feasible solution satisfies further inequality constraints:

$$(\forall p \leqslant d, \forall (J_1, J_2)_d) \ f_p(J_1, J_2) \geqslant 0.$$

This ensures that the corresponding solution μ is feasible to FLP and hence the relaxations are identical. The persistency lemma formulated in [2, Lemma 3.2] has the following form.

Lemma A.13 (Lemma 3.2 of Adams et al. [2])**.** Let $N^+ \cup N^- \cup N^f$ be a partition of $\mathcal{V}$. Let there exist dual multipliers satisfying certain sufficient condition (as defined in [2]). Then there exists an optimal solution to LP(d/n) having $w_s = 1$ for $s \in N^+$ and $w_s = 0$ for $s \in N^-$.

For an interested reader we remark that the sufficient conditions of their lemma ensure a dual feasible solution for f^2 which is complementary to the solution defined by N^+, N^- on $N^+ \cup N^-$ and is zero on all boundary constraints $J \subset \mathcal{B}$. This ensures complementarity with any feasible primal solution consistent with N^+, N^-. We are not going to prove this claim formally, but use the existing lemma together with the observation that their sufficient condition does not depend on the coefficients $\{c_J \mid J \subset N^f\}$.

Lemma A.14. Assume that the conditions of Lemma A.13 are satisfied. Let $w_s = 1$ for $s \in N^+$, $w_s = 0$ for $s \in N^-$ and let w be feasible to LP(d/n). Then w is optimal to LP$(d/|\mathcal{V}_2|)$ for f^2.

Proof. It can be seen from [2, equation 3.4d] that if the conditions of their Lemma are satisfied then they are also satisfied with coefficients $c_J = 0$ for all $J \subset N^f$. Since w is feasible, it is also feasible to their equation 3.5[d]. The objective of the latter is identically zero, therefore w is optimal to their equation 3.5[d]. Lemma A.13 proves that w is optimal to LP(d/n). Since we made f^1 zero, w is optimal to LP$(d/|\mathcal{V}_2|)$ for f^2. $\square$

With this refined result we can easily prove the relaxed-improving property. Define the mapping

$$p_s(x_s) = \begin{cases} 0, & s \in N^-, \\ 1, & s \in N^+, \\ x_s, & \text{otherwise.} \end{cases} \tag{173}$$

Theorem 4.9. Persistency by the sufficient condition of Adams et al. [2, Lemma 3.2] corresponds to an FLP-improving mapping.

Proof. We need to show that mapping p is FLP-improving. Let $\mu \in \Lambda$. Then $\mu' := [p]\mu \in \Lambda$ is optimal to f^2 by Lemma A.14. At the same time $(\forall J \in \mathcal{E}, J \cap N^f \neq \varnothing)$ we have $\mu'_J = \mu_J$ and hence $\langle f^1, \mu_J \rangle = \langle f^1, \mu'_J \rangle$. It follows that

$$\langle f, [p]\mu \rangle = \langle f^1, [p]\mu \rangle + \langle f^2, [p]\mu \rangle \leqslant \langle f, \mu \rangle. \tag{174}$$
$$\square$$

We obtained that FLP maximum persistency dominates the persistency result of Adams et al. [2]. In case of strong persistency, the former is found by Algorithm 1.

In the pairwise case, conditions [2] are always satisfied for the same persistency assignment as the roof dual. However, in the higher order case our conditions are less restrictive as can be seen from the dual representation (29): we require less inequalities than the complementarity slackness imposed on all solutions of the unassigned part.

References

References

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

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

[3] Alahari, K., Kohli, P., and Torr, P. H. S. (2008). Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In *CVPR*.

[4] Arora, C., Banerjee, S., Kalra, P., and Maheshwari, S. N. (2012). Generic cuts: An efficient algorithm for optimal inference in higher order MRF-MAP. In *ECCV*, pages 17–30.

[5] Boros, E. and Gruber, A. (2012). On quadratization of pseudo-Boolean functions. In *ISAIM*.

[6] Boros, E. and Hammer, P. (2002). Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 1-3(123):155–225.

[7] Boros, E., Hammer, P. L., and Sun, X. (1991). Network flows and minimization of quadratic pseudo-Boolean functions. Technical Report RRR 17-1991, RUTCOR.

[8] Boros, E., Hammer, P. L., and Tavares, G. (2006). Pre-processing of unconstrained quadratic binary optimization. Technical Report RRR 10-2006, RUTCOR.

[9] Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *PAMI*, volume 26, pages 1124–1137.

[10] Boykov, Y., Veksler, O., and Zabih, R. (1999). Fast approximate energy minimization via graph cuts. In *ICCV*, volume 1, pages 377–384.

[11] Chekuri, C., Khanna, S., Naor, J., and Zosin, L. (2001). Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *In Symposium on Discrete Algorithms*, pages 109–118.

[12] Cooper, M. C. (2008). Minimization of locally defined submodular functions by optimal soft arc consistency. *Constraints*, 13(4):437–458.

[13] Delong, A., Osokin, A., Isack, H. N., and Boykov, Y. (2012). Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 96(1):1–27.

[14] 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.

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

[16] Freedman, D. and Drineas, P. (2005). Energy minimization via graph cuts: Settling what is possible. In *CVPR*, pages 939–946.

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

[18] Greig, D., Porteous, B., and Seheult, A. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279.

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

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

[21] Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *PAMI*, 25(10):1333–1336.

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

[23] Kahl, F. and Strandmark, P. (2011). Generalized roof duality for pseudo-Boolean optimization. In *ICCV*, pages 255–262.

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

[25] Kohli, P., Shekhovtsov, A., Rother, C., Kolmogorov, V., and Torr, P. (2008). On partial optimality in multi-label MRFs. In McCallum, A. and Roweis, S., editors, *ICML*, pages 480–487. Omnipress.

[26] Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583.

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

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

[29] Kolmogorov, V. (2012b). Minimizing a sum of submodular functions. *Discrete Applied Mathematics*, 160(15):2246 – 2258.

[30] Kolmogorov, V. (2013). The power of linear programming for finite-valued CSPs: A constructive characterization. In Fomin, F., Freivalds, R., Kwiatkowska, M., and Peleg, D., editors, *Automata, Languages, and Programming*, volume 7965 of *Lecture Notes in Computer Science*, pages 625–636. Springer Berlin Heidelberg.

[31] Kolmogorov, V. (2014). A new look at reweighted message passing. *PAMI*, PP(99):1–1.

[32] Kolmogorov, V. and Rother, C. (2007). Minimizing non-submodular functions with graph cuts – a review. *PAMI*.

[33] Kolmogorov, V. and Schoenemann, T. (2012). Generalized sequential tree-reweighted message passing. *CoRR*, abs/1205.6352.

[34] Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159.

[35] Komodakis, N., Paragios, N., and Tziritas, G. (2011). MRF energy minimization and beyond via dual decomposition. *PAMI*, 33(3):531–552.

[36] Komodakis, N. and Tziritas, G. (2007). Approximate labeling via graph cuts based on linear programming. *PAMI*, 29(8):1436–1453.

[37] Komodakis, N., Tziritas, G., and Paragios, N. (2008). Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. *CVIU*, 112(1):14–29.

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

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

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

[41] Ladicky, L., Russell, C., Kohli, P., and Torr, P. (2010). Graph cut based inference with co-occurrence statistics. In *ECCV*.

[42] Lempitsky, V. S., Roth, S., and Rother, C. (2008). Fusionflow: Discrete-continuous optimization for optical flow estimation. In *CVPR*.

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

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

[45] Orponen, P. and Mannila, H. (1990). On approximation preserving reductions: Complete problems and robust measures (revised version). Technical report.

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

[47] Ravikumar, P. D., Agarwal, A., and Wainwright, M. J. (2010). Message-passing for graph-structured linear programs: Proximal methods and rounding schemes. *Journal of Machine Learning Research*, 11:1043–1080.

[48] Rother, C., Kolmogorov, V., Lempitsky, V., and Szummer, M. (2007). Optimizing binary MRFs via extended roof

duality. In *CVPR*.

[49] 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.

[50] Savchynskyy, B., Kappes, J. H., Swoboda, P., and Schnörr, C. (2013). Global MAP-optimality by shrinking the combinatorial search area with convex relaxation. In *NIPS*.

[51] Schlesinger, D. and Flach, B. (2006). Transforming an arbitrary minsum problem into a binary one. Technical Report TUD-FI06-01, Dresden University of Technology.

[52] Schlesinger, M. I. and Flach, B. (2000). Some solvable subclasses of structural recognition problems. In Svoboda, T., editor, *Czech Pattern Recognition Workshop 2000, invited paper*.

[53] Schlesinger, M. I. and Giginyak, V. V. (2007). Solution to structural recognition (max,+)-problems by their equivalent transformations. *Control Systems and Computers*, (1,2):3–15.

[54] Shekhovtsov, A. (2013). *Exact and Partial Energy Minimization in Computer Vision*. PhD Thesis, Czech Technical University in Prague.

[55] Shekhovtsov, A. (2014a). Maximum persistency in energy minimization. Technical report, Graz University of Technology.

[56] Shekhovtsov, A. (2014b). Maximum persistency in energy minimization. In *CVPR*, pages 1162–1169.

[57] Shekhovtsov, A., Swoboda, P., and Savchynskyy, B. (2015). Maximum persistency via iterative relaxed inference with graphical models. In *CVPR*, page 8. to appear.

[58] Sherali, H. D. and Adams, W. P. (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430.

[59] Shlezinger, M. (1976). Syntactic analysis of two-dimensional visual signals in the presence of noise. *Cybernetics and Systems Analysis*, 4:113–130. See review by Werner [72].

[60] Sontag, D., Choe, D. K., and Li, Y. (2012). Efficiently searching for frustrated cycles in MAP inference. In *Uncertainty in Artificial Intelligence (UAI)*, pages 795–804, Corvallis, Oregon. AUAI Press.

[61] Strandmark, P. Submodular library v.1.0.

[62] Strandmark, P. (2012). *Discrete optimization in early vision : model tractability versus fidelity*. PhD thesis, Lund. Diss. Lund : Lunds universitet, 2013.

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

[64] Swoboda, P., Savchynskyy, B., Kappes, J., and Schnrr, C. (2014a). Partial optimality by pruning for MAP-inference with general graphical models. In *CVPR*, pages 1170–1177.

[65] Swoboda, P., Shekhovtsov, A., Kappes, J. H., Schnörr, C., and Savchynskyy, B. (2014b). Partial optimality by pruning for MAP-inference with general graphical models. *CoRR*, abs/1410.6641.

[66] Thapper, J. and Živný, S. (2012). The power of linear programming for valued CSPs. In *Symposium on Foundations of Computer Science (FOCS)*, pages 669–678. IEEE.

[67] Thapper, J. and Živný, S. (2013). The complexity of finite-valued CSPs. In *Symposium on the Theory of Computing (STOC)*, pages 695–704. ACM.

[68] Vanderbei, R. J. (2001). *Linear Programming: Foundations and Extensions*. Department of operations and research and financial engineering, Princeton university.

[69] Živný, S., Cohen, D. A., and Jeavons, P. G. (2009). The expressive power of binary submodular functions. *Discrete Applied Mathematics*, 157(15):3347 – 3358.

[70] Wainwright, M., Jaakkola, T., and Willsky, A. (2003). Exact MAP estimates by (hyper)tree agreement. In S. Becker, S. T. and Obermayer, K., editors, *NIPS*, pages 809–816. MIT Press.

[71] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.

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

[73] Werner, T. (2008). High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). In *CVPR*.

[74] Werner, T. (2010). Revisiting the linear programming relaxation approach to gibbs energy minimization and weighted constraint satisfaction. *PAMI*, 32(8):1474–1488.

[75] Windheuser, T., Ishikawa, H., and Cremers, D. (2012). Generalized roof duality for multi-label optimization: Optimal lower bounds and persistency. In *ECCV*.