

Complexity of Discrete Energy Minimization Problems

Anonymous ECCV submission

Paper ID 1

Abstract. Discrete energy minimization is widely-used in computer vision and machine learning for problems such as MAP inference for graphical models. The problem, in general, is notoriously intractable, and finding the global optimal solution is known to be NP-hard. However, is it possible to approximate this problem with a reasonable ratio bound on the solution quality in polynomial time? We show in this paper that the answer is no. Specifically, we show that general energy minimization and planar energy minimization with three or more labels are exp-APX-complete. This finding rules out the existence of any approximation algorithm with a sub-exponential approximation ratio in the input size for these two problems, including constant factor approximations. Moreover, we collect and review the computational complexity of several subclass problems and arrange them on a complexity scale consisting of three major complexity classes – PO, APX, and exp-APX, corresponding to problems that are solvable, approximable, and inapproximable in polynomial time. Problems in the first two complexity classes can serve as alternative tractable formulations to the inapproximable ones. This paper can help vision researchers to select an appropriate model for an application or guide them in designing new algorithms.

Keywords: Energy minimization, complexity, NP-hard, APX, exp-APX, NPO, WCSP, min-sum, MAP MRF, planar graph

1 Introduction

Discrete energy minimization, also known as min-sum labeling [67] or weighted (valued) constraint satisfaction (WCSP)¹ [24], is a popular model for many problems in computer vision, machine learning, bioinformatics, and natural language processing. In particular, the problem arises in maximum a posteriori (MAP) inference for Markov (conditional) random fields (MRFs/CRFs) [42]. In the most frequently used pairwise case, the *discrete energy minimization problem* (simply “energy minimization” hereafter) is defined as

$$\min_{x \in \mathcal{L}^{\mathcal{V}}} \sum_{u \in \mathcal{V}} f_u(x_u) + \sum_{(u,v) \in \mathcal{E}} f_{uv}(x_u, x_v), \quad (1)$$

where x_u is the label for node u in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. When the variables x_u are binary (Boolean): $\mathcal{L} = \mathbb{B} = \{0, 1\}$, the problem can be written as a quadratic polynomial in x [10] and is known as quadratic pseudo-Boolean optimization (QPBO) [10].

¹ WCSP is a more general problem, considering a bounded plus operation. It is itself a special case of valued CSP, where the objective takes values in a more general valuation set.

In computer vision practice, energy minimization has found its place in semantic segmentation [50], pose estimation [69], scene understanding [56], depth estimation [43], optical flow estimation [68], image in-painting [58], and image denoising [7]. For example, tree-structured models have been used to estimate pictorial structures such as body skeletons or facial landmarks [69], multi-label Potts models have been used to enforce a smoothing prior for semantic segmentation [50], and general pairwise models have been used for optimal flow estimation [68]. However, it may not be fully understood that the energy minimization formulations used to model these vision problems have greatly varied degrees of tractability or *computational complexity*. For the three examples above, the first allows efficient exact inference, the second admits a constant factor approximation, and the third has no quality guarantee on the approximation of the optimum.

The study of complexity of energy minimization is a broad field. Energy minimization problems are often intractable in practice except for special cases. While many researchers analyze the time complexity of their algorithms (e.g., using big O notation), it is beneficial to delve deeper to address the difficulty of the underlying problem. The two most commonly known complexity classes are P (polynomial time) and NP (nondeterministic polynomial time, which includes problems whose solutions can be verified in polynomial time). However, these two complexity classes are only defined for *decision* problems. The analogous complexity classes for *optimization* problems are PO (P optimization) and NPO (NP optimization, which includes problems whose solution feasibility can be verified in polynomial time). Optimization problems form a superset of decision problems, since any decision problem can be cast as an optimization over the set {yes, no}, i.e., $P \subseteq PO$ and $NP \subseteq NPO$. The NP-hardness of an optimization problem means it is at least as hard as (under Turing reduction) the hardest decision problem in the class NP. If a problem is NP-hard, then it is not in PO assuming $P \neq NP$.

Although optimal solutions for problems in NPO, but not in PO, are intractable, it is sometimes possible to guarantee that a good solution (i.e., one that is worse than the optimal by no more than a given factor) can be found in polynomial time. These problems can therefore be further classified into class APX (constant factor approximation) and class exp-APX (inapproximable) with increasing complexity (Figure 1). We can arrange energy minimization problems on this more detailed complexity scale, originally established in [4], to provide vision researchers a new viewpoint for complexity classification with a focus on NP-hard optimization problems.

In this paper, we make three core contributions, as explained in the next three paragraphs. First, we prove the inapproximability result of QPBO and general energy minimization. Second, we show that the same inapproximability result holds when restricting to planar graphs with three or more labels. In the proof, we propose a novel micro-graph structure-based reduction that can be used for algorithmic design as well. Finally, we present a unified framework and an overview of vision-related special cases where the energy minimization problem can be solved in polynomial time or approximated with a constant, logarithmic, or polynomial factor.

Binary and multi-label case (Section 3). It is known that QPBO (2-label case) and the general energy minimization problem (multi-label case) are NP-hard [11] because they generalize such classical NP-hard optimization problems on graphs as vertex pack-

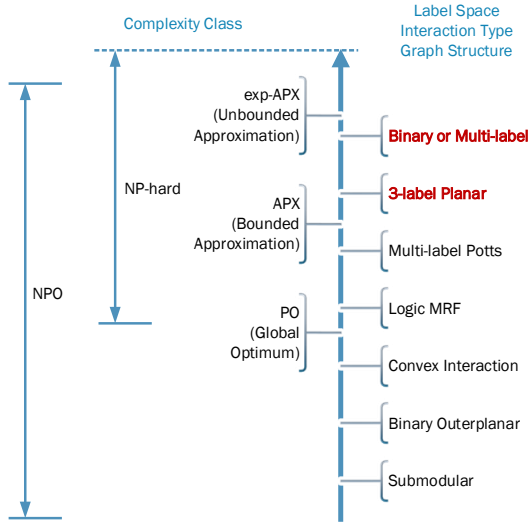


Fig. 1: Discrete energy minimization problems aligned on a complexity axis. Red, boldface indicates new results proven in this paper. Note that problems are not ranked within a complexity class. Some problems discussed in this paper are removed for the simplicity of the figure.

ing (maximum independent set) and the minimum and maximum cut problems [26]. In this paper, we show a much stronger conclusion. *We prove that QPBO as well as general energy minimization are complete (being the hardest problems) in the class exp-APX.* This implies that there can be no method in polynomial time to find an approximation within a constant factor of the optimal, and in fact, the only possible factor in polynomial time is exponential in the input size. In practice, this means that a solution may be essentially arbitrarily bad.

Planar three or more label case (Section 4). Planar graphs form the underlying graph structure for many computer vision and image processing tasks. It is known that efficient exact algorithms exist for some special cases of planar 2-label energy minimization problems [54]. In this paper, we show that for the case of three or more labels, planar energy minimization is exp-APX-complete, which means these problems are as hard as general energy minimization. The complexity of the general planar 2-label case remains an open question.

Subclass problems (Section 5). Special cases for some energy minimization algorithms relevant to computer vision are known to be tractable. However, detailed complexity analysis of these algorithms is patchy and spread across numerous papers. In Section 5, we classify the complexity of these subclass problems and illustrate some of their connections. Such an analysis can help computer vision researchers become acquainted with existing complexity results relevant to energy minimization used in computer vision and can aid in selecting an appropriate model for an application or in designing new algorithms.

1.1 Related Work

Much of the work on complexity in computer vision has focused on experimental or empirical comparison of inference methods, including influential studies on choosing the best optimization techniques for specific classes of energy minimization problems [60, 25] and the PASCAL Probabilistic Inference Challenge, which focused on the more general context of inference in graphical models [1]. In contrast, our work focuses on theoretical computational complexity, rather than experimental analysis.

On the theoretical side, the NP-hardness of certain energy minimization problems is well studied. It has been shown that 2-label energy minimization is, in general, NP-hard, but it can be in PO if it is submodular [29] or outerplanar [54]. For multi-label problems, the NP-hardness was proven by reduction from the NP-hard multi-way cut problem [12]. These results, however, say nothing about the complexity of *approximating* the global optimum for the intractable cases. The complexity involving approximation has been studied for classical combinatorial problems, such as MAX-CUT and MAX-2SAT, which are known to be APX-complete [45]. QPBO generalizes such problems and is therefore APX-hard.

Energy minimization is often used to solve MAP inference for undirected graphical models. In contrast to scarce results for energy minimization and undirected graphical models, researchers have more extensively studied the computational complexity of approximating the MAP solution for *Bayesian networks*, also known as *directed graphical models* [41]. Abdelbar and Hedetniemi first proved the NP-hardness for approximating the MAP assignment of directed graphical models in the value of probability, i.e., finding x such that

$$\frac{p(x^*)}{p(x)} \leq r(n) \quad (2)$$

with a constant or polynomial ratio $r(n) \geq 1$ is NP-hard, showing that this problem is poly-APX-hard [2]. The probability approximation ratio is closest to the method used in our work, but other approximation measures have also been studied. Kwisthout showed the NP-hardness for approximating MAPs with the measure of additive value- [39], structure- [40], and rank-approximation [41]. He also investigated the hardness of expectation approximation of MAP, and found that no randomized algorithm can expectation-approximate MAP in polynomial time with a bounded margin of error unless $\text{NP} \subseteq \text{BPP}$, an assumption that is highly unlikely to be true [41].

Unfortunately, the complexity results for directed models do not directly transfer to undirected models and vice versa. In directed and undirected models, the graph represents different conditional independence relations, thus the underlying family of probability distributions encoded by these two models is distinct. However, one can ask similar questions on the hardness of undirected models in terms of various approximation measures. In this work, we answer two questions, “How hard is it to approximate the MAP inference in the ratio of log probability (energy) and the ratio of probability?” The complexity of structure-, rank-, and expectation-approximation remain open questions for energy minimization.

2 Definitions and Notation

There are at least two different sets of definitions of what is considered an NP optimization problem [44, 4]. Here, we follow the notation of Ausiello et al [4] and restate the definitions needed for us to state and prove our theorems in Sections 3 and 4 with our explanation of their relevance to our proofs.

Definition 2.1 (Optimization Problem, [4] Def. 1.16). An *optimization problem* \mathcal{P} is characterized by a quadruple $(\mathcal{I}, \mathcal{S}, m, \text{goal})$ where

1. \mathcal{I} is the set of instances of \mathcal{P} ;
2. \mathcal{S} is a function that associates to any input instance $x \in \mathcal{I}$ the set of *feasible solutions* of x ;
3. and m is the *measure* function, defined for pairs (x, y) such that $x \in \mathcal{I}$ and $y \in \mathcal{S}(x)$. For every such pair (x, y) , $m(x, y)$ provides a positive integer.
4. $\text{goal} \in \{\text{MIN}, \text{MAX}\}$.

Notice the assumption that the cost is positive, and, in particular, it cannot be zero.

Analogous to NP and P for decision problems, we have NPO and PO for optimization problems:

Definition 2.2 (Class NPO, [4] Def 1.17). An optimization problem $\mathcal{P} = (\mathcal{I}, \mathcal{S}, m, \text{goal})$ belongs to the class of NP minimization (NPO) problems if the following hold:

1. The set of instances \mathcal{I} is recognizable in polynomial time.
2. There exists a polynomial q such that given an instance $x \in \mathcal{I}$, for any $y \in \mathcal{S}(x)$, $|y| < q(x)$ and, besides, for any y such that $|y| < q(x)$, it is decidable in polynomial time whether $y \in \mathcal{S}(x)$.
3. The measure function m is computable in polynomial time.

Definition 2.3 (Class PO, [4] Def 1.18). An optimization problem \mathcal{P} belongs to the class of PO if it is in NPO and there exists a polynomial-time computable algorithm that, for any instance $x \in \mathcal{I}$, returns an optimal solution $y \in \mathcal{S}^*(x)$, together with its value $m^*(x)$.

For intractable algorithms, it may be acceptable to seek an approximate solution that is sufficiently close to optimal.

Definition 2.4 (Approximation Algorithm, [4] Def. 3.1). Given an optimization problem $\mathcal{P} = (\mathcal{I}, \mathcal{S}, m, \text{goal})$ an algorithm \mathcal{A} is an *approximation algorithm* for \mathcal{P} if, for any given instance $x \in \mathcal{I}$, it returns an *approximate solution*, that is a feasible solution $\mathcal{A}(x) \in \mathcal{S}(x)$.

Definition 2.5 (Performance Ratio, [4], Def. 3.6). Given an optimization problem \mathcal{P} , for any instance x of \mathcal{P} and for any feasible solution $y \in \mathcal{S}(x)$, the *performance ratio*, *approximation ratio* or *approximation factor* of y with respect to x is defined as

$$R(x, y) = \max \left\{ \frac{m(x, y)}{m^*(x)}, \frac{m^*(x)}{m(x, y)} \right\}, \quad (3)$$

where $m^*(x)$ is the measure of the optimal solution for the instance x .

Since $m^*(x)$ is a positive integer, the performance ratio is well-defined. It is a rational number in $[1, \infty)$. Notice that from this definition, it follows that if finding a feasible solution, e.g. $y \in \mathcal{S}(x)$, is an NP-hard decision problem, then there exists no polynomial-time approximation algorithm for \mathcal{P} , irrespective of the kind of performance evaluation that one could possibly mean.

Definition 2.6 ($r(n)$ -approximation, [4], Def. 8.1). Given an optimization problem \mathcal{P} in NPO, an approximation algorithm \mathcal{A} for \mathcal{P} , and a function $r: \mathbb{N} \rightarrow (1, \infty)$, we say that \mathcal{A} is an $r(n)$ -approximate algorithm for \mathcal{P} if, for any instance x of \mathcal{P} such that $\mathcal{S}(x) \neq \emptyset$, the performance ratio of the feasible solution $\mathcal{A}(x)$ with respect to x verifies the following inequality:

$$R(x, \mathcal{A}(x)) \leq r(|x|). \quad (4)$$

Definition 2.7 (F -APX, [4], Def. 8.2). Given a class of functions F , F -APX is the class of all NPO problems \mathcal{P} such that, for some function $r \in F$, there exists a polynomial-time $r(n)$ -approximate algorithm for \mathcal{P} .

The class of constant functions for F yields the complexity class APX. Together with logarithmic, polynomial, and exponential functions applied in Definition 2.7, the following *complexity axis* is established:

$$\text{PO} \subseteq \text{APX} \subseteq \text{log-APX} \subseteq \text{poly-APX} \subseteq \text{exp-APX} \subseteq \text{NPO}.$$

Since the measure m needs to be computable in polynomial time for NPO problems, the largest measure and thus the largest performance ratio is an exponential function. But exp-APX does not equal to NPO (assuming $P \neq NP$) because NPO contains problems whose feasible solutions cannot be found in polynomial time. For an energy minimization problem, any label assignment is a feasible solution, implying that all energy minimization problems are in exp-APX.

The standard approach for proofs in complexity theory is to perform a reduction from a known NP-complete problem. Unfortunately, the most common polynomial-time reductions ignore the quality of the solution in the approximated case. For example, it is shown that any energy minimization problem can be reduced to a factor 2 approximable Potts model [47], however the reduction is not approximation preserving and is unable to show the hardness of general energy minimization in terms of approximation. Therefore, it is necessary to use an approximation preserving (AP) reduction to classify NPO problems that are not in PO, for which only the approximation algorithms are tractable. AP-preserving reductions preserve the approximation ratio in a linear fashion, and thus preserve the membership in these complexity classes. Formally,

Definition 2.8 (AP-reduction, [4] Def. 8.3). Let \mathcal{P}_1 and \mathcal{P}_2 be two problems in NPO. \mathcal{P}_1 is said to be AP-reducible to \mathcal{P}_2 , in symbols $\mathcal{P}_1 \leq_{\text{AP}} \mathcal{P}_2$, if two functions π and σ and a positive constant α exist such that:

1. For any instance $x \in \mathcal{I}_1$ and for any rational $r > 1$, $\pi(x, r) \in \mathcal{I}_2$.
2. For any instance $x \in \mathcal{I}_1$ and for any rational $r > 1$, if $S_1(x) \neq \emptyset$ then $S_2(\pi(x, r)) \neq \emptyset$.
3. For any instance $x \in \mathcal{I}_1$, for any rational $r > 1$ and for any $y \in S_2(\pi(x, r))$, $\sigma(x, y, r) \in S_1(x)$.

4. π and σ are computable by algorithms whose running time is polynomial for any fixed rational r .
5. For any instance $x \in \mathcal{I}_1$, for any rational $r > 1$, and for any $y \in S_2(\pi(x, r))$,

$$R_2(\pi(x, r), y) \leq r \quad \text{implies} \quad (5)$$

$$R_1(x, \sigma(x, y, r)) \leq 1 + \alpha(r - 1). \quad (6)$$

AP-reduction is the formal definition of the term ‘as hard as’ used in this paper unless otherwise specified. It defines a partial order among optimization problems. With respect to this relationship, we can formally define the subclass containing the hardest problems in a complexity class:

Definition 2.9 (\mathcal{C} -hard and \mathcal{C} -complete, [4] Def. 8.5). Given a class \mathcal{C} of NPO problems, a problem \mathcal{P} is \mathcal{C} -hard if, for any $\mathcal{P}' \in \mathcal{C}$, $\mathcal{P}' \leq_{\text{AP}} \mathcal{P}$. A \mathcal{C} -hard problem is \mathcal{C} -complete if it belongs to \mathcal{C} .

Intuitively, a complexity class \mathcal{C} specifies the upper bound on the hardness of the problems within, \mathcal{C} -hard specifies the lower bound, and \mathcal{C} -complete exactly specifies the hardness.

3 Inapproximability for the General Case

In this section, we show that QPBO and general energy minimization are inapproximable by proving they are exp-APX-complete. As previously mentioned, it is already known that these problems are NP-hard [11], but it was previously unknown whether useful approximation guarantees were possible in the general case. We start our proof with the 2-label case QPBO, then we prove the more general case.

The formal statement of QPBO as an optimization problem is as follows:

Problem 1. QPBO

INSTANCE: A pseudo-Boolean function $f: \mathbb{B}^{\mathcal{V}} \rightarrow \mathbb{N}$:

$$f(x) = \sum_{u \in \mathcal{V}} f_u(x_u) + \sum_{u, v \in \mathcal{V}} f_{uv}(x_u, x_v), \quad (7)$$

given by the collection of unary terms f_u and pairwise terms f_{uv} .

SOLUTION: Assignment of variables $x \in \mathbb{B}^{\mathcal{V}}$.

MEASURE: $\min f(x) > 0$.

Theorem 3.1. QPBO is exp-APX-complete.

Proof Sketch. (Full proof in Appendix A).

1. We observe that W3SAT-triv is known to be exp-APX-complete [4]. W3SAT-triv is a 3-SAT problem with weights on the variables and an artificial, trivial solution.
2. Each 3-clause in the conjunctive normal form can be represented as a polynomial consisting of three binary variables. Together with representing the weights with the unary terms, we arrive at a 3-label energy minimization problem.

3. We use the method of [23] to transform the 3-label problem into a 2-label one, with polynomially many additional variables, which is an instance of QPBO.
 4. The transformation defines an injective forward mapping π and then we define a surjective inverse mapping σ to set up the two-way correspondence between W3SAT-triv and the transformed QPBO.
 5. We prove that such mappings build an AP-reduction from W3SAT-triv to QPBO, i.e. $\text{W3SAT-triv} \leq_{\text{AP}} \text{QPBO}$. This proves that QPBO is exp-APX-hard.
 6. We observe that all energy minimization problems are in exp-APX and thereby concluding that QPBO is exp-APX-complete.
- This inapproximability result can be generalized to more than two labels.

Corollary 3.2. k -label energy minimization is exp-APX-complete for $k \geq 2$.

Proof Sketch. (Full proof in Appendix A). This theorem is proved by showing QPBO $\leq_{\text{AP}} k$ -label energy minimization for $k \geq 2$.

We show in Corollary A.3 the inapproximability in energy (log probability) transfer to probability in Equation (2) as well.

Taken together, this theorem and its corollaries form a very strong inapproximability result for general energy minimization. They imply not only NP-hardness, but also that there is no algorithm that can approximate general energy minimization with two or more labels with an approximation ratio better than some exponential function in the input size. In other words, any approximation algorithm of the general energy minimization problem can perform arbitrarily badly, and it would be pointless to try to prove a bound on the approximation ratio for existing approximation algorithms for the general case. While this conclusion is disappointing, these results serve as a clarification of grounds and guidance for model selection and algorithm design. Instead of counting on an oracle that solves the energy minimization problem, researchers should put efforts into selecting the proper formulation, trading off expressiveness for tractability.

4 Inapproximability for the Planar Case

Efficient algorithms for energy minimization have been found for special cases of 2-label planar graphs. Examples include planar 2-label problems without unary terms and outerplanar 2-label problems (i.e., the graph structure remains planar after connecting to a common node) [54]. Grid structures over image pixels naturally give rise to planar graphs in computer vision. Given their frequency of use in this domain, it is natural to consider the complexity of more general cases involving planar graphs. Figure 2 visualizes the current state of knowledge of the complexity of energy minimization

Planar 2-label Special Cases PO	Planar 2-label The General Case Unknown	Planar 3 and More Labels The General Case exp-APX-complete (This Paper)
--	--	--

Fig. 2: Complexity for planar energy minimization problems. The “general case” implies no restrictions on the pairwise interaction type. This paper shows that the third category of problems is not efficiently approximable.

problems on planar graphs. In this section, we prove that for the case of planar graphs with three or more labels, energy minimization is exp-APX-complete. This result is important because it significantly reduces the space of potentially efficient algorithms on planar graphs. The approximability of the general case for planar 2-label problems is an open question.

Theorem 4.1. Planar 3-label energy minimization is exp-APX-complete.

Proof Sketch. (Full proof in Appendix A).

1. We construct elementary gadgets to reduce any 3-label energy minimization problem to a planar one with polynomially many auxiliary nodes.
2. The above reduction defines an injective forward mapping and we define a surjective inverse mapping σ to set up the two-way correspondence between the original problem and the reduced planar one.
3. We prove that such mappings build an AP-reduction, i.e., 3-label energy minimization \leq_{AP} planar 3-label energy minimization.
4. Since 3-label energy minimization is exp-APX-complete (Corollary 3.2) and all energy minimization problems are in exp-APX, we thereby conclude that planar 3-label energy minimization is exp-APX-complete.

Corollary 4.2. Planar k -label energy minimization is exp-APX-complete, for $k \geq 3$.

Proof Sketch. (Full proof in Appendix A). This theorem is proved by showing planar 3-label energy minimization \leq_{AP} planar k -label energy minimization, for $k \geq 3$.

These theorems show that the restricted case of planar graphs with 3 or more labels is as hard as general case for energy minimization problems with the same inapproximable implications discussed in Section 3.

The most novel and useful aspect of the proof of Theorem 4.1 is the planar reduction in Step 1. The reduction creates an equivalent planar representation to any non-planar 3-label graph. That is, the graphs share the same optimal value. The reduction applies elementary constructions or “gadgets” to uncross two intersecting edges. This process is repeated until all intersecting edges are uncrossed. Similar elementary constructions were used to study the complexity of the linear programming formulation of energy minimization problems [48, 47]. Our novel gadgets have three key properties *at the same time*: 1) they are able to uncross intersecting edges, 2) they work on non-relaxed problems, i.e., all indicator variables (or pseudomarginals to be formal) are integral; and 3) they can be applied repeatedly to build an AP-reduction.

The two gadgets used in our reduction are illustrated in Figure 3. A 3-label node can be encoded as a collection of 3 indicator variables with a one-hot constraint. In the figure, a solid colored circle denotes a 3-label node, and a solid colored rectangle denotes the equivalent node expressed with indicator variables (white circles). For example, in Figure 3, $a = 1$ corresponds to the blue node taking the first label value. The pairwise potentials (edges on the left part of the figures) can be viewed as edge costs between the indicator variables (black lines on the right), e.g., $f_{uv}(3, 2)$ is placed onto the edge between indicator c and e and is counted into the overall measure if and only if $c = e = 1$. In our gadgets, drawn edges represent zero cost while omitted edges

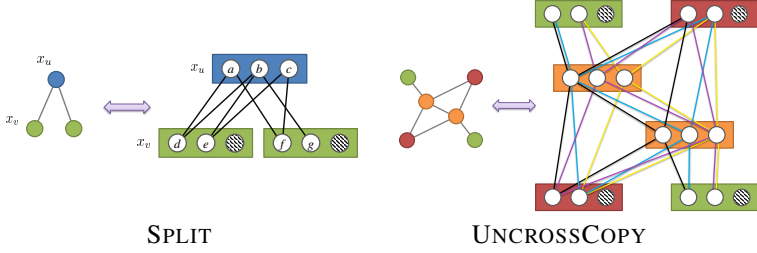


Fig. 3: Gadgets to represent a 3-label variable as two 2-label variables (SPLIT) and to copy the values of two diagonal pairs of 2-label variables without edge crossing (UNCROSSCOPY).

represent positive infinity². While the set of feasible solutions remains the same, the gadget encourages certain labeling relationships, which, if not satisfied, cause the overall measure to be infinity. Therefore, the encouraged relationships must be satisfied by any optimal solution. The two gadgets serve different purposes:

SPLIT A 3-label node (blue) is split into two 2-label nodes (green). The shaded circle represents a label with a positive infinite unary cost and thus creates a simulated 2-label node. The encouraged relationships are

- $a = 1 \leftrightarrow d = 1$ and $f = 1$.
- $b = 1 \leftrightarrow g = 1$.
- $c = 1 \leftrightarrow e = 1$ and $f = 1$.

Thus (d, f) encodes a , (d, g) and (e, g) both encode b and (e, f) encodes c .

UNCROSSCOPY The values of two 2-label nodes are encouraged to be the same as their diagonal counterparts respectively (red to red, green to green) without crossing with each other. The orange nodes are intermediate nodes that pass on the values. All types of lines represent the same edge cost, which is 0. The color differences visualize the verification for each of the 4 possible states of two 2-label nodes. For example, the cyan lines verify the case where the top-left (green) node takes the values $(1, 0)$ and the top-right (red) node takes the value $(0, 1)$. It is clear that the encouraged solution is for the bottom-left (red) node to take the value $(0, 1)$, and the bottom-right (green) node to take the value $(1, 0)$.

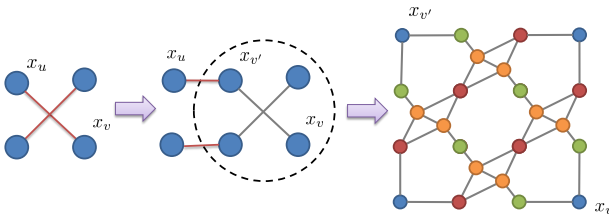


Fig. 4: Planar reduction for 3-label problems

² A very large number will also serve the same purpose, e.g., take the sum of the absolute value of all energy terms and add 1. Therefore, we are not expanding the set of allowed energy terms to include ∞ .

These two gadgets can be used to uncross the intersecting edges of two pairs of 3-label nodes (Figure 4, left). For a crossing edge (x_u, x_v) , first a new 3-label node x'_v is introduced preserving the same arbitrary interaction (red line) as before (Figure 4, middle). Then, the crossing nodes (shown within the dotted circle) are uncrossed by applying SPLIT and UNCROSSCOPY four times (Figure 4, right). Without loss of generality, we can assume that no more than two edges intersect at a common point except at their endpoints. This process can be applied repeatedly at each edge crossing until there are no edge crossings left in the graph [48].

5 Complexity of Subclass Problems

In this section, we classify some of the special cases of energy minimization according to our complexity axis (Figure 1). This classification can be viewed as a reinterpretation of existing results from the literature into a unified framework.

5.1 Class PO (Global Optimum)

Polynomial time solvability may be achieved by considering two principal restrictions: those restricting the *structure* of the problem, i.e., the graph G , and those restricting the type of allowed interactions, i.e., functions f_{uv} .

Structure Restrictions. When G is a chain, energy minimization reduces to finding a shortest path in the trellis graph, which can be solved using a classical dynamic programming method known, e.g., as the Viterbi algorithm [19]. The same dynamic programming (DP) principle applies to graphs of bounded tree-width. Fixing all variables in a separator set decouples the problem into independent optimization problems. For treewidth 1, the separators are just individual vertices and the problem is solved by a variant of DP [46, 53]. For higher treewidth, the respective optimization procedure is known as junction tree decomposition [42]. A loop is a simple example of a treewidth 2 problem. However, for a treewidth k , the time complexity is exponential in k [42]. Finally, when G is an outer-planar graph, the problem can be solved by the method of [54], which reduces it to a planar Ising model, for which efficient algorithms exist [59].

Interaction Restrictions. Submodularity is a restriction closely related to problems solvable by minimum cut. A quadratic pseudo-Boolean function f is *submodular* iff its quadratic terms are non-positive. It is then known to be equivalent with finding a minimum cut in a corresponding network [20]. Another way to state this condition for QPBO is $\forall (u, v) \in \mathcal{E}, f_{uv}(0, 1) + f_{uv}(1, 0) \geq f_{uv}(0, 0) + f_{uv}(1, 1)$. However, submodularity is more general. It extends to higher-order and multi-label problems. Submodularity is considered a discrete analog of convexity. Just as convex functions are relatively easy to optimize, general submodular function minimization can be solved in strongly polynomial time [55]. Kolmogorov and Zabini introduced submodularity in computer vision and showed that binary 2nd order and 3rd order submodular problems can be always reduced to minimum cut, which is much more efficient than general submodular function minimization [33]. Živný et al. and Ramalingam et al. give more results on functions reducible to minimum cut [66, 49]. For QPBO on an unrestricted

graph structure, the following *dichotomy* result has been proven by Cohen et al. [15]: either the problem is submodular and thus in PO or it is NP-hard (i.e., submodular problems are the only ones that are tractable in this case).

For multi-label problems Ishikawa proposed a reduction to minimum cut for problems with convex interactions, i.e., where $f_{uv}(x_u, x_v) = g_{uv}(x_u - x_v)$ and g_{uv} is convex and symmetric [22]. It is worth noting that when the unary terms are convex as well, the problem can be solved even more efficiently [21, 30]. The same reduction [22] remains correct for a more general class of submodular multi-label problems. In modern terminology, component-wise minimum $x \wedge y$ and component-wise maximum $x \vee y$ of complete labelings x, y for all nodes are introduced. These operations depend on the *order of labels* and, in turn, define a lattice on the set of labelings. The function f is called *submodular on the lattice* if $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$ for all x, y [63]. In the pairwise case, the condition can be simplified to the form of submodularity common in computer vision [49]:

$$f_{uv}(i, j + 1) + f_{uv}(i + 1, j) \geq f_{uv}(i, j) + f_{uv}(i + 1, j + 1). \quad (8)$$

In particular, it is easy to see that a convex f_{uv} satisfies it [22]. Kolmogorov [31] and Arora et al. [3] proposed maxflow-like algorithms for higher order submodular energy minimization. Schlesinger proposed an algorithm to find a reordering in which the problem is submodular if one exists [52]. However, unlike in the binary case, solvable multi-label problems are more diverse. A variety of problems are generalizations of submodularity and are in PO, including symmetric tournament pair, submodularity on arbitrary trees, submodularity on arbitrary lattices, skew bisubmodularity, and bisubmodularity on arbitrary domains (see references in [62]). Thapper and Živný [61] and Kolmogorov [32] characterized these tractable classes and proved a similar dichotomy result: **a of** problem of unrestricted structure is either solvable by LP-relaxation (and thus in PO) or it is NP-hard. It appears that LP relaxation is the most powerful and general solving technique [70].

Mixed Restrictions. In comparison, results with mixed structure and interaction restrictions are rare. One example is a planar Ising model without unary terms [59]. Since there is a restriction on structure (planarity) and unary terms, it does not fall into any of the classes described above. Another example is the restriction to supermodular functions on a bipartite graph, solvable by [52] and by LP relaxation, but not falling under the characterization [62] because of the graph restriction.

Algorithmic Applications. The aforementioned tractable formulations in PO can be used to solve or approximate harder problems. Such trees, cycles and planar problems are used in dual decomposition methods [34, 35, 8]. Binary submodular problems are used for finding an optimized crossover of two-candidate multi-label solutions. An example of this technique, the expansion move algorithm, achieves a constant approximation ratio for the Potts model [12]. Extended dynamic programming can be used to solve restricted segmentation problems [17] and as move-making subroutine [65]. LP relaxation also provides approximation guarantees for many problems [27, 14, 36, 5], placing them in the APX class.

5.2 Class APX (Bounded Approximation)

Problems that have bounded approximation in polynomial time usually have certain restriction on the interaction type. The Potts model may be the simplest and most common way to enforce the smoothness of the labeling. Each pairwise interaction depends on whether the neighboring labellings are the same, i.e. $f_{uv}(x_u, x_v) = c_{uv}\delta(x_u, x_v)$. Boykov et al. showed a reduction to this problem from the NP-hard multiway cut [12], also known to be APX-complete [4, 16]. They also proved that their constructed alpha-expansion algorithm is guaranteed to produce a factor two approximation to the global optimum. These results prove that the Potts model is in APX but not in PO. However, their reduction from multiway cut is not an AP-reduction, as it violates the third condition of AP-reducibility. Therefore, it is still an open problem whether the Potts model is APX-complete. Boykov et al. also showed that their algorithm can approximate the more general problem of metric labeling [12]. The energy is called *metric* if, for an arbitrary, finite label space \mathcal{L} , the pairwise interaction satisfies a) $f_{uv}(\alpha, \beta) = 0$, b) $f_{uv}(\alpha, \beta) = f_{uv}(\beta, \alpha) \geq 0$, and c) $f_{uv}(\alpha, \beta) \leq f_{uv}(\beta, \gamma) + f_{uv}(\gamma, \alpha)$, for any labels $\alpha, \beta, \gamma \in \mathcal{L}$ and any $uv \in \mathcal{E}$. Although their approximation algorithm has a bound on the performance ratio, the bound depends on the ratio of some pairwise terms, a number that can grow exponentially large. For metric labeling with k labels, Kleinberg et al. proposed an $O(\log k \log \log k)$ -approximation algorithm. This bound was further improved to $O(\log k)$ by Chekuri et al. [13], making metric labeling a problem in log-APX³.

We have seen that a problem with convex pairwise interactions is in PO. An interesting variant is its truncated counterpart, i.e., $f_{uv}(x_u, x_v) = w_{uv} \min\{d(x_u - x_v), M\}$, where w_{uv} is a non-negative weight, d is a convex symmetric function to define the distance between two labels and M is the truncating constant [64]. This problem is NP-hard [64], but Kumar et al. [38] have proposed an algorithm that yields bounded approximations with a factor of $2 + \sqrt{2}$ for linear distance functions and a factor of $O(\sqrt{M})$ for quadratic distance functions⁴. This bound is analyzed for more general distance functions by Kumar [37].

Another APX problem with implicit restrictions on the interaction type is logic MRF [6]. It is a powerful higher order model able to encode arbitrary logical relations of Boolean variables. It has energy function $f(x) = \sum_i^n w_i C_i$, where each C_i is a disjunctive clause involving a subset of Boolean variables x , and $C_i = 1$ if it is satisfied and 0 otherwise. Each clause C_i is assigned a non-negative weight w_i . The goal is to find an assignment of x to maximize $f(x)$. As disjunctive clauses can be converted into polynomials, this is essentially a pseudo-Boolean optimization problem. However, this is a special case of general 2-label energy minimization, as its polynomial basis spans a subspace of the basis of the latter. Bach et al. [6] proved that logic MRF is in APX by showing that it is a special case of MAX-SAT with non-negative weights.

³ Observe that $k = O(\text{poly}(|x|))$, where $|x|$ is the input instance size (see Corollary A.2).

⁴ In these truncated convex problems, the ratio bound is defined for the pairwise part of the energy (1). The approximation ratio in accordance to our definition is obtained assuming the unary terms are non-negative.

6 Discussion

The algorithmic implications of our inapproximability have been discussed above. Here, we focus on the discussion of practical implications. An approximation guarantee is most needed when inference is used as a subroutine of a larger task. In structural learning, it is acceptable to have a constant factor approximation for the inference subroutine when efficient exact algorithms are not available. Finley and Joachims proved that this constant factor approximation guarantee yields a multiplicative bound on the learning objective, providing a relative guarantee for the quality of the learned parameters [18]. An optimality guarantee is important, as the inference subroutine is called multiple times, and even a single poor approximation, which returns a not-so-bad worst violater, will lead to the early termination of structural learning algorithm.

However, despite having no approximation ratio guarantee, algorithms such as the extended roof duality algorithm for QPBO [51] are still widely used. This gap between theory and application applies not only to our results but to all other complexity results as well. We list several key reasons for the potential lack of correspondence between theoretical complexity guarantees and practical performance.

Complexity results address the worst case scenario. Our inapproximability result guarantees that for any polynomial time algorithm, there exists an input instance for which the algorithm will produce a very poor approximation. However, applications often do not encounter the worst case. Such is the case with the simplex algorithm, whose worst case complexity is exponential, yet it is widely used in practice.

Objective function is not the final evaluation criterion. In many image processing tasks, the final evaluation criterion is the number of pixels correctly labeled. The relation between the energy value and the accuracy is implicit. In many cases, a local optimum is good enough to produce a high labeling accuracy and a visually appealing labeling.

Other forms of optimality guarantee or indicator exist. Approximation measures in the distance of solutions or in the expectation of the objective value are likely to be prohibitive for energy minimization, as they are for Bayesian networks [39, 40, 41]. On the other hand, a family of energy minimization algorithms has the property of being *persistent* or *partial optimal*, meaning a subset of nodes have consistent labeling with the global optimal one [9, 10]. Rather than being an optimality guarantee, persistency is an optimality indicator. In the worst case, the set of persistent labeling could be empty, yet the percentage of persistent labeling over the all the nodes give us a notion of the algorithm’s performance on this particular input instance. Persistency is also useful in reducing the size of the search space [28, 57].

7 Conclusion

In this paper, we have shown inapproximability results for energy minimization in the general case and planar 3-label case. In addition, we present a unified overview of the complexity of existing energy minimization problems by arranging them in a fine-grained complexity scale. These altogether set up a new viewpoint for interpreting and classifying the complexity of optimization problems for the computer vision community. In the future, it will be interesting to consider the open questions of the complexity of structure-, rank-, and expectation-approximation for energy minimization.

References

- [1] The Probabilistic Inference Challenge (2011), <http://www.cs.huji.ac.il/project/PASCAL/>
- [2] Abdelbar, A., Hedetniemi, S.: Approximating MAPs for belief networks is NP-hard and other theorems. *Artificial Intelligence* 102(1), 21–38 (6 1998)
- [3] Arora, C., Banerjee, S., Kalra, P., Maheshwari, S.N.: Generic cuts: An efficient algorithm for optimal inference in higher order MRF-MAP. In: *ECCV*. pp. 17–30 (2012)
- [4] Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., Protasi, M.: Complexity and approximation: Combinatorial optimization problems and their approximability properties. Springer (1999)
- [5] Bach, S.H., Huang, B., Getoor, L.: Unifying local consistency and max sat relaxations for scalable inference with rounding guarantees. In: Lebanon, G., Vishwanathan, S.V.N. (eds.) *AISTATS. JMLR Proceedings*, vol. 38. JMLR.org (2015), <http://dblp.uni-trier.de/db/conf/aistats/aistats2015.html#BachHG15>
- [6] Bach, S.H., Huang, B., Getoor, L.: Unifying local consistency and max sat relaxations for scalable inference with rounding guarantees. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*. pp. 46–55 (2015)
- [7] Barbu, A.: Learning real-time mrf inference for image denoising. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. pp. 1574–1581. IEEE (2009)
- [8] Batra, D., Gallagher, A.C., Parikh, D., Chen, T.: Beyond trees: Mrf inference via outer-planar decomposition. In: *CVPR*. pp. 2496–2503. IEEE Computer Society (2010), <http://dblp.uni-trier.de/db/conf/cvpr/cvpr2010.html#BatraGPC10>
- [9] Boros, E., Hammer, P.L., Sun, X.: Network flows and minimization of quadratic pseudo-Boolean functions. *Tech. Rep. RRR 17-1991, RUTCOR* (May 1991)
- [10] Boros, E., Hammer, P.: Pseudo-Boolean optimization. *Tech. rep., RUTCOR* (October 2001)
- [11] Boros, E., Hammer, P.: Pseudo-Boolean optimization. *Discrete Applied Mathematics* 1-3(123), 155–225 (2002)
- [12] Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1222–1239 (November 2001), <http://www.csd.uwo.ca/~yuri/Abstracts/pami01-abs.shtml>
- [13] Chekuri, C., Khanna, S., Naor, J., Zosin, L.: A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM Journal on Discrete Mathematics* 18(3), 608–625 (2005)
- [14] Chekuri, C., Khanna, S., Naor, J., Zosin, L.: Approximation algorithms for the metric labeling problem via a new linear programming formulation. In: *In Symposium on Discrete Algorithms*. pp. 109–118 (2001)
- [15] Cohen, D., Cooper, M., Jeavons, P.: Principles and Practice of Constraint Programming – CP 2004: 10th International Conference, CP 2004, Toronto, Canada, September 27 -October 1, 2004. *Proceedings, chap. A Complete Characterization of Complexity for Boolean Constraint Optimization Problems*, pp. 212–226. Springer Berlin Heidelberg, Berlin, Heidelberg (2004), http://dx.doi.org/10.1007/978-3-540-30201-8_18
- [16] Dahlhaus, E., Johnson, D.S., Papadimitriou, C.H., Seymour, P.D., Yannakakis, M.: The complexity of multiterminal cuts. *SIAM Journal on Computing* 23(4), 864–894 (1994)
- [17] Felzenszwalb, P.F., Veksler, O.: Tiered scene labeling with dynamic programming. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. pp. 3097–3104 (June 2010)

- [18] Finley, T., Joachims, T.: Training structural svms when exact inference is intractable. In: Proceedings of the 25th international conference on Machine learning. pp. 304–311. ACM (2008)
- [19] Forney Jr, G.D.: The viterbi algorithm. Proceedings of the IEEE 61(3), 268–278 (1973)
- [20] Hammer, P.L.: Some network flow problems solved with pseudo-Boolean programming. Operation Research 13, 388–399 (1965)
- [21] Hochbaum, D.S.: An efficient algorithm for image segmentation, markov random fields and related problems. J. ACM 48(4), 686–701 (Jul 2001), <http://doi.acm.org/10.1145/502090.502093>
- [22] Ishikawa, H.: Exact optimization for Markov random fields with convex priors. PAMI 25(10), 1333–1336 (2003)
- [23] Ishikawa, H.: Transformation of general binary mrf minimization to the first-order case. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(6), 1234–1249 (2011)
- [24] Jeavons, P., Krokhin, A., Živný, S., et al.: The complexity of valued constraint satisfaction. Bulletin of EATCS 2(113) (2014)
- [25] Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Kröger, T., Lellmann, J., et al.: A comparative study of modern inference techniques for structured discrete energy minimization problems. International Journal of Computer Vision 115, 155–184 (2015)
- [26] Karp, R.M.: Reducibility among combinatorial problems. In: Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. pp. 85–103 (1972), <http://www.cs.berkeley.edu/~luca/cs172/karp.pdf>
- [27] Kleinberg, J., Tardos, E.: Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. J. ACM 49(5), 616–639 (Sep 2002), <http://doi.acm.org/10.1145/585265.585268>
- [28] Kohli, P., Shekhovtsov, A., Rother, C., Kolmogorov, V., Torr, P.: On partial optimality in multi-label mrfs. In: Proceedings of the 25th international conference on Machine learning. pp. 480–487. ACM (2008)
- [29] Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? PAMI 26(2), 147–159 (February 2004)
- [30] Kolmogorov, V.: Primal-dual algorithm for convex Markov random fields. Tech. Rep. MSR-TR-2005-117, Microsoft Research Cambridge (2005)
- [31] Kolmogorov, V.: Minimizing a sum of submodular functions. Discrete Applied Mathematics 160(15), 2246 – 2258 (2012), <http://www.sciencedirect.com/science/article/pii/S0166218X12002314>
- [32] Kolmogorov, V.: The power of linear programming for finite-valued CSPs: A constructive characterization. In: Fomin, F., Freivalds, R., Kwiatkowska, M., Peleg, D. (eds.) Automata, Languages, and Programming, Lecture Notes in Computer Science, vol. 7965, pp. 625–636. Springer Berlin Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-39206-1_53
- [33] Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? Pattern Analysis and Machine Intelligence, IEEE Transactions on 26(2), 147–159 (2004)
- [34] Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: ICCV. pp. 1–8 (2007)
- [35] Komodakis, N., Paragios, N.: Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles. In: ECCV. pp. 806–820 (2008)
- [36] Komodakis, N., Tziritas, G.: Approximate labeling via graph cuts based on linear programming. PAMI 29(8), 1436–1453 (2007)
- [37] Kumar, M.P.: Rounding-based moves for metric labeling. In: Advances in Neural Information Processing Systems. pp. 109–117 (2014)

- [38] Kumar, M.P., Veksler, O., Torr, P.H.: Improved moves for truncated convex models. *Journal of Machine Learning Research* 12, 31–67 (Feb 2011)
- [39] Kwisthout, J.: Most probable explanations in Bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning* 52(9), 1452 – 1469 (2011), <http://www.sciencedirect.com/science/article/pii/S0888613X11001095>, handling Incomplete and Fuzzy Information in Data Analysis and Decision Processes
- [40] Kwisthout, J.: Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 12th European Conference, ECSQARU 2013, Utrecht, The Netherlands, July 8-10, 2013. Proceedings, chap. Structure Approximation of Most Probable Explanations in Bayesian Networks, pp. 340–351. Springer Berlin Heidelberg, Berlin, Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-39091-3_29
- [41] Kwisthout, J.: Tree-width and the computational complexity of map approximations in bayesian networks. *Journal of Artificial Intelligence Research* pp. 699–720 (2015)
- [42] Lauritzen, S.L.: Graphical Models. No. 17 in Oxford Statistical Science Series, Oxford Science Publications (1998)
- [43] Liu, B., Gould, S., Koller, D.: Single image depth estimation from predicted semantic labels. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 1253–1260. IEEE (2010)
- [44] Orponen, P., Mannila, H.: On approximation preserving reductions: complete problems and robust measures (1987)
- [45] Papadimitriou, C.H., Yannakakis, M.: Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences* 43(3), 425 – 440 (1991), <http://www.sciencedirect.com/science/article/pii/002200009190023X>
- [46] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc. (1988)
- [47] Prusa, D., Werner, T.: How hard is the lp relaxation of the potts min-sum labeling problem. In: Energy Minimization Methods in Computer Vision and Pattern Recognition. pp. 57–70. Springer (2015)
- [48] Prusa, D., Werner, T.: Universality of the local marginal polytope. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 37(4), 898–904 (2015)
- [49] Ramalingam, S., Kohli, P., Alahari, K., Torr, P.H.: Exact inference in multi-label crfs with higher order cliques. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8. IEEE (2008)
- [50] Ren, X., Bo, L., Fox, D.: Rgb-(d) scene labeling: Features and algorithms. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 2759–2766. IEEE (2012)
- [51] Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.: Optimizing binary mrfs via extended roof duality. In: Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on. pp. 1–8. IEEE (2007)
- [52] Schlesinger, D.: Exact solution of permuted submodular minsum problems. In: Yuille, A.L., Zhu, S.C., Cremers, D., Wang, Y. (eds.) EMMCVPR. Lecture Notes in Computer Science, vol. 4679, pp. 28–38. Springer (2007), <http://dblp.uni-trier.de/db/conf/emmcvpr/emmcvpr2007.html#Schlesinger07>
- [53] Schlesinger, M.I., Hlaváč, V.: Ten lectures on statistical and structural pattern recognition, Computational Imaging and Vision, vol. 24. Kluwer Academic Publishers, Dordrecht, The Netherlands (2002)
- [54] Schraudolph, N.: Polynomial-time exact inference in NP-hard binary MRFs via reweighted perfect matching. In: Teh, Y.W., Titterton, D.M. (eds.) AISTATS. JMLR Proceedings, vol. 9, pp. 717–724. JMLR.org (2010), <http://dblp.uni-trier.de/db/journals/jmlr/jmlrp9.html#Schraudolph10>

- [55] Schrijver, A.: A combinatorial algorithm minimizing submodular functions in strongly polynomial time. *Journal of Combinatorial Theory Series B*(80), 346–355 (2000), <http://homepages.cwi.nl/~lex/files/minsubm6.ps>
- [56] Schwing, A.G., Urtasun, R.: Efficient exact inference for 3d indoor scene understanding. In: *Computer Vision–ECCV 2012*, pp. 299–313. Springer (2012)
- [57] Shekhovtsov, A., Swoboda, P., Savchynskyy, B.: Maximum persistency via iterative relaxed inference with graphical models. In: *CVPR* (2015)
- [58] Shekhovtsov, A., Kohli, P., Rother, C.: Curvature prior for MRF-based segmentation and shape inpainting. In: Axel Pinz, Thomas Pock, H.B.F.L. (ed.) *DAGM/OAGM*. pp. 41–51 (August 2012)
- [59] Shih, W.K., Wu, S., Kuo, Y.S.: Unifying maximum cut and minimum cut of a planar graph. *IEEE Transactions on Computers* 39(5), 694–697 (May 1990)
- [60] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *PAMI* 30(6), 1068–1080 (2008)
- [61] Thapper, J., Živný, S.: The power of linear programming for valued CSPs. In: *Symposium on Foundations of Computer Science (FOCS)*. pp. 669–678. IEEE (2012), <http://www.cs.ox.ac.uk/Stanislaw.Zivny/homepage/publications/tz12focs-preprint.pdf>
- [62] Thapper, J., Živný, S.: The complexity of finite-valued CSPs. In: *Symposium on the Theory of Computing (STOC)*. pp. 695–704. ACM (2013), <http://www.cs.ox.ac.uk/Stanislaw.Zivny/homepage/publications/tz13stoc-preprint.pdf>
- [63] Topkis, D.M.: Minimizing a submodular function on a lattice. *Operations Research* 26(2), 305–321 (1978), <http://www.jstor.org/stable/169636>
- [64] Veksler, O.: Graph cut based optimization for mrfs with truncated convex priors. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. pp. 1–8. IEEE (2007)
- [65] Vineet, V., Warrell, J., Torr, P.H.S.: A tiered move-making algorithm for general pairwise MRFs. In: *CVPR*. pp. 1632–1639 (2012)
- [66] Živný, S., Cohen, D.A., Jeavons, P.G.: The expressive power of binary submodular functions. *Discrete Applied Mathematics* 157(15), 3347 – 3358 (2009), <http://www.sciencedirect.com/science/article/pii/S0166218X09002819>
- [67] Werner, T.: A linear programming approach to max-sum problem: A review. *PAMI* 29(7), 1165–1179 (July 2007)
- [68] Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(9), 1744–1757 (2012)
- [69] Yang, Y., Ramanan, D.: Articulated pose estimation with flexible mixtures-of-parts. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. pp. 1385–1392. IEEE (2011)
- [70] Živný, S., Werner, T., Průša, D.a.: *The Power of LP Relaxation for MAP Inference*, pp. 19–42. The MIT Press, Cambridge, USA (December 2014)

A Formal Proofs

A.1 General Case

Theorem 3.1. QPBO is exp-APX-complete.

Proof. We reduce from the following problem.

Problem 2 ([4], Section 8.3.2). **W3SAT-triv**

INSTANCE: Boolean CNF formula F with variables x_1, \dots, x_n and each clause assuming exactly 3 variables; non-negative weights w_1, \dots, w_n .

SOLUTION: Truth assignment τ to the variables that satisfies F or the trivial all-true assignment.

MEASURE: $\min(1, \sum_{i=1}^n w_i \tau(x_i))$.

W3SAT-triv is known to be exp-APX-complete [4]. We use an AP-reduction from W3SAT-triv to prove the same completeness result for QPBO. The optimal value of W3SAT-triv is bounded above by $M := \sum_i w_i$ because the all-true assignment is feasible. The length of M is linear in the length of instance L . The objective weight is represented in QPBO as unary terms $f_i(x_i) = w_i x_i$. For every Boolean clause $C(x_i, x_j, x_k) \in F$ we construct a triple-wise term

$$\delta_{ijk}(x_i, x_j, x_k) = M(1 - C(x_i, x_j, x_k)). \quad (9)$$

This term takes large value M iff C is not satisfied and 0 otherwise. Further, the Boolean clause $C(x_i, x_j, x_k)$ can be represented as a cubic polynomial and so is

$$\delta_{ijk} = M(ax_i x_j x_k + \sum_J b_J \prod_{l \in J} x_l), \quad (10)$$

where $J \subset \{i, j, k\}$, $|J| \leq 2$ and each coefficient a and b_J is either +1 or -1. We now apply the quadratization techniques [23] to δ_{ijk} . Observe the following identities for auxiliary variable x_w with $w > n$:

$$-x_i x_j x_k = \min_{x_w \in \mathbb{B}} -x_w(x_i + x_j + x_k - 2) \quad (11)$$

if $a = -1$ and as $x_i x_j x_k =$

$$\min_{x_w \in \mathbb{B}} ((x_w - 1)(x_i + x_j + x_k - 1) + (x_i x_j + x_i x_k + x_j x_k)) \quad (12)$$

if $a = +1$. In either case, substituting the cubic term $a x_i x_j x_k$ with the expression inside the min operator, we can have a uniformed quadratic form for each triple-wise term

$$\psi_{ijk} := M \sum_{J_w} b_{J_w} \prod_{l \in J_w} x_l, \quad (13)$$

where $J_w \subset \{i, j, k, w\}$, $|J_w| \leq 2$ and $\prod_{i \in J_w} x_i$ is set to 1 if J_w is empty. In both cases, the quadratic form takes the same optimal values as its cubic counterpart given the optimal assignment, i.e.

$$\min_{x_i, x_j, x_k, x_w} \psi_{ijk} = \min_{x_i, x_j, x_k} \delta_{ijk}, \quad (14)$$

but the transformation expands the original range from $\{-1, 0\}$ to $\{-1, 0, 1, 2\}$ and from $\{0, 1\}$ to $\{0, 1, 3\}$ respectively. Therefore, the costs of the constructed instance of QPBO is bounded in the absolute value by $3M$ and the number of added variables is exactly the number of clauses in F . Clearly, this construction can be computed in polynomial time. *Note that when approximation is used, this transformation is no longer exact ($\psi_{ijk} \neq \delta_{ijk}$) as the optimality of the auxiliary variable x_w cannot be guaranteed. However, it can be verified that under all possible assignments (ignoring the min operator) in either case, $\psi_{ijk} \geq 0$, which is the key for the reduction to be an approximation preserving (AP) one.*

The construction above defines a mapping π from any instance of W3SAT-triv ($p_1 \in I_1$) to an instance of QPBO ($p_2 \in I_2$). The mapping σ from feasible solutions of p_2 ($x \in S_2(\pi(p_1))$) to that of p_1 is defined as follows: if $f(x) \geq M$, then let the mapped solution $\sigma(p_1, x)$ be the all true assignment, otherwise let the mapped solution $\sigma(p_1, x)$ to be $x_i, i \in \{1, \dots, n\}$. We need to show that (π, σ) together with a constant $\alpha \leq 1$ is an AP-reduction. For simplicity we omit the irrelevant number r required in the definition 2.8, and let m_1, m_2, m_1^* and m_2^* to be short for $m_1(p_1, \sigma(p_1, x)), m_2(\pi(p_1)), m_1^*(p_1)$ and $m_2^*(\pi(p_1))$ respectively.

First, note that $\sigma(p_1, x)$ is always feasible to W3SAT-triv: either it satisfies F or $f(x) \geq M$ and therefore $\sigma(p_1, x)$ is the all-true assignment. In the first case, since every quadratic term is non-negative, we have

$$m_1 = \sum_{i=1}^n x_i w_i \quad (15)$$

$$\leq \sum_{i=1}^n x_i w_i + \sum_{C_{ijk} \in F} \psi_{ijk}(x_i, x_j, x_k) = m_2. \quad (16)$$

In the second case, by construction

$$m_1 = M \leq f(x) = m_2. \quad (17)$$

Therefore, no matter which case $m_1 \leq m_2$.

Now for the optimal solution, if F is satisfiable, then by construction $m_1^* = m_2^*$. For any instances $p_1 \in I_1$ and for any rational $r > 1$, and for any $x \in S_2(\pi(p_1))$, if

$$R_2(\pi(p_1), x) = \frac{m_2}{m_2^*} \leq r, \quad (18)$$

then

$$m_1 \leq m_2 \leq r m_2^* = r m_1^* \quad (19)$$

$$R_1(p_1, \sigma(p_1, x)) = \frac{m_1}{m_1^*} \leq r \quad (20)$$

If F is not satisfiable, $m_1^* = M \leq m_2^*$ and $m_2 \geq m_2^* \geq M$, thus for any instances $p_1 \in I_1$ and for any rational $r > 1$, and for any $x \in S_2(\pi(p_1))$,

$$R_1(p_1, \sigma(p_1, x)) = \frac{m_1}{m_1^*} = \frac{M}{M} = 1 \leq r \quad (21)$$

Therefore $(\pi, \sigma, 1)$ is an AP-reduction. Since W3SAT-triv is exp-APX-complete and QPBO is in exp-APX, we prove that QPBO is exp-APX-complete. \square

Proof. (by contradiction). Assume that energy minimization is not exp-APX-complete but belongs to some simpler complexity class on the complexity axis. As QPBO is energy minimization with two labels, all of its instances are instances of general energy minimization (1). Therefore, general energy minimization could be employed for the two-label cases, contradicting the result of Theorem 3.1. In other words, general energy minimization can only be a harder problem than QPBO. \square

Corollary A.1. Probability ratio approximation is poly-APX-hard: it is NP-hard to approximate MAP in the value of probability (2) with any polynomial ratio $r(n)$.

Corollary A.2. There is no algorithm that can approximate k -label energy minimization with an approximation ratio better or equal than some polynomials in k .

Corollary A.3. It is NP-hard to approximate MAP in the value of probability (2) with any exponential ratio $\exp(r(n))$, where r is polynomial.

Proof. Recall that the probability $p(x)$ is given by the exponential map of the energy: $p(x) = \exp(-f(x))$. Assume for contradiction that there is a polynomial time algorithm \mathcal{A} that finds solution x and a polynomial $r(n) \geq 0$ for $n > 0$ such that

$$\frac{p(x^*)}{p(x)} \leq e^{r(n)} \quad (22)$$

for all instances of the problem. Taking the logarithm,

$$-f(x^*) + f(x) \leq r(n). \quad (23)$$

or,

$$f(x) \leq r(n) + f(x^*). \quad (24)$$

Divide by $f(x^*)$, which, by definition of NPO is positive, we obtain

$$\frac{f(x)}{f(x^*)} \leq 1 + \frac{1}{f(x^*)} r(n) \leq 1 + r(n). \quad (25)$$

where we have used that $f(x^*)$ is integer and positive and hence it is greater equal 1. Inequality (25) provides a polynomial ratio approximation for energy minimization. Since the latter is poly-APX-hard, this contradicts existence of the polynomial algorithm \mathcal{A} , unless $P = NP$. \square

A.2 Planar Case

Theorem 4.1. Planar 3-label energy minimization is exp-APX-complete.

Proof. Draft! The AP-reducibility can be easily verified after observing these facts The reverse mapping is multiple to one mapping (for generalization it is one to one). The

optimal measure are the same for ALL corresponding feasible solutions. $m_1 \preceq m_2$ all the time! thus $R_1 \preceq R_2 \preceq r$, $\alpha = 1$ And this same idea goes to all our theorems for generalizing from k labels to more labels.

M does not need to be infinity, just large enough is OK

□

Corollary 4.2. Planar k -label energy minimization is exp-APX-complete, for $k \geq 3$.

Proof.

□