

A Study of Lagrangean Decompositions and Dual Ascent Solvers for Graph Matching

Paul Swoboda¹, Carsten Rother², Hassan Abu Alhaija², Dagmar Kainmüller³ and Bogdan Savchynskyy²

¹IST Austria, ²TU Dresden, ³MPI-CBG

pswoboda@ist.ac.at

Abstract

We study the quadratic assignment problem, in computer vision also known as *graph matching*. Two leading solvers for this problem optimize the Lagrange decomposition duals with sub-gradient and dual ascent (also known as *message passing*) updates. We explore this direction further and propose several additional Lagrangean relaxations of the graph matching problem along with corresponding algorithms, which are all based on a common dual ascent framework. Our extensive empirical evaluation gives several theoretical insights and suggests a new state-of-the-art anytime solver for the considered problem. Our improvement over state-of-the-art is particularly visible on a new dataset with large-scale sparse problem instances containing more than 500 graph nodes each.

1. Introduction

In computer vision and beyond, the quadratic assignment problem, known also as *graph matching*, *feature correspondence* and *feature matching*, has attracted great interest. This problem is similar to Maximum-A-Posteriori (MAP) inference on a discrete pairwise graphical model, also called conditional random field (CRF) in the literature. It differs in an additional uniqueness constraint: Each label can be taken at most once. This uniqueness constraint makes it well-suited to attack e.g. tracking problems or shape matching. In both cases feature points or object parts have to be matched between multiple frames one-to-one. Unfortunately, the uniqueness constraint prevents naive application of efficient message passing solvers for MAP-inference to this problem. For this reason, many dedicated graph matching solvers were developed, see related work below.

On the other hand, efficient dual block-coordinate ascent (also known as *message passing*) algorithms like TRW-S [33] count among the most efficient solvers for MAP-inference in conditional random fields. Also, the graph matching problem, after possibly introducing many additional variables,

can be stated as a MAP-inference problem in a standard pairwise CRF. Such an approach already surpasses most state-of-the-art graph matching solvers.

Hence, it is desirable to devise specialized convergent message passing solvers exhibiting none of the drawbacks discussed above, i.e. (i) directly operating on a compact representation of the graph matching problem and (ii) using techniques from the MAP-inference community to gain computational efficiency.

To achieve this goal, we propose (i) several Lagrangean decompositions of the graph matching problem and (ii) novel efficient message passing solvers for these relaxations. We show their efficacy in an extensive empirical evaluation.

Related work The term *graph matching* refers to a number of different optimization problems in pattern recognition, see [18] for a review. We mean the special version known unambiguously as *quadratic assignment problem* (QAP) [36]. Recently, the graph matching was generalized to the *hyper-graph matching* problem (see [43] and references therein), which match between more than two graphs.

The quadratic assignment problem was first formulated in [13] back in 1957. Since a number of NP-complete problems such as traveling salesman, maximal clique, graph isomorphism and graph partitioning can be straightforwardly reduced to QAP, this problem is NP-hard itself. Its importance for numerous applications boosted its analysis a lot: The (already aged) overview [41] contains 362 references with over 150 works suggesting new algorithms and over 100 with new theoretical results related to this problem.

Nearly all possible solver paradigms were put to the test for QAP. These include, but are not limited to, convex relaxations based on Lagrangean decompositions [31, 52], linear [6, 22], convex quadratic [9] and semi-definite [46, 51, 63] relaxations, which can be used either directly to obtain approximate solutions or just to provide lower bounds. To tighten these bounds several cutting plane methods were proposed [11, 12]. On the other side, various primal heuristics, both (i) deterministic, such as local search [5, 44], graduated assignment methods [25], fixed point iterations [39], spectral

technique and its derivatives [17,38,53,62] and (ii) stochastic, like random walk [16] and Monte-Carlo sampling [37,49] were suggested to provide approximate solutions to the problem. Altogether these methods serve as building blocks for exact branch-and-bound [10,23,26] algorithms and other non-convex optimization methods [25,59,64]. The excellent surveys [15,41] contain further references.

As is usual for NP-hard problems, no single method can efficiently address all QAP instances. Different applications require different methods and we concentrate here on problem instances specific for computer vision. Traditionally within this community predominantly primal heuristics are used, since demand for low computational time usually dominates the need to obtain optimality guarantees. However, two recently proposed solvers [52,61] based on Lagrangean decomposition (also known as *dual decomposition* in computer vision) have shown superior results and surpassed numerous state-of-the-art primal heuristics.

The *dual decomposition solver* [52] represents the problem as a combination of MAP-inference for binary CRFs, the linear assignment problem and a number of small-sized QAPs over few variables; Lagrangean multipliers connecting these subproblems are updated with the sub-gradient method. Although the solver demonstrates superior results on computer vision datasets, we suspect that its efficiency can be further improved by switching to a different update method, such as bundle [30,32] or block-coordinate ascent [57]. This suspicion is based on comparison of such solvers for MAP-inference in CRFs [29] and similar observation related to other combinatorial optimization problems (see e.g. [45]).

Hungarian Belief Propagation (HBP) [61] considers a combination of a multilabel CRF and a linear assignment as subproblems; Lagrange multipliers are updated by a block-coordinate ascent (message passing) algorithm and the obtained lower bounds are employed inside a branch-and-bound solver. It is known [34], however, that efficiency of message passing significantly depends on the schedule of sending messages. Specifically, efficiency of dual ascent algorithms depends on selecting directions for the ascent (blocks of variables to optimize over) and the order in which these ascent operations are performed. Arguably, the underlying multilabel CRF subproblem is crucial and the message passing must deal with it efficiently. However, HBP [61] uses a message passing schedule similarly as in the MPLP algorithm [24], which was shown [29,34] to be significantly slower than the schedule of SRMP (TRW-S) [34].

Contribution We study several Lagrangean decompositions of the graph matching problem. Some of these are known, e.g. the one used in the HBP algorithm [61] and the one corresponding to the local polytope relaxation of the pairwise CRF representation of graph matching. The others have not been published so far, to our knowledge. For all these decompositions we provide efficient message passing

(dual ascent) algorithms based on a recent message passing framework [50]. In the case of the local polytope relaxation our algorithm coincides with the SRMP method [34], a higher-order generalization of the famous TRW-S algorithm [33].

Our experimental evaluation suggests a new state-of-the-art method for the graph matching problem, which outperforms both the dual decomposition [52] and the HBP [61] solvers. We propose tighter convex relaxations for all our methods. Also, we significantly improve performance of the HBP algorithm by changing its message passing schedule.

Proofs are given in the appendix. Code and datasets can be obtained at https://github.com/pawelwoboda/LP_MP.

Notation. Undirected graphs are denoted by $G = (V, E)$, where V is a finite *node set* and $E \subseteq \binom{V}{2}$ is the *edge set*. The set of neighboring nodes of $v \in V$ w.r.t. graph G is denoted by $\mathcal{N}_G(v) := \{u : uv \in E\}$. The convex hull of a set $X \subset \mathbb{R}^n$ is denoted by $\text{conv}(X)$.

2. CRFs and Graph Matching

First, we introduce conditional random fields and state the graph matching problem as one with additional uniqueness constraints. Second, we consider an inverse formulation of the graph matching problem, which, after being coupled with the original formulation, often leads to faster algorithms.

Conditional random fields (CRF). Let $G = (V, E)$ be an undirected graph. With each node $u \in V$ we associate a variable x_u taking its values in a finite *set of labels* $X_u \subseteq \{(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)\}$. Hence, each label corresponds to a unit vector. Notation X_A denotes the Cartesian product $\prod_{u \in A \subseteq V} X_u$. A vector $x \in X_V$ with coordinates $(x_u)_{u \in V}$ is called a *labeling*. Likewise, we use the notation $x_A \in X_A$ (a special case being $x_{uv} \in X_{uv} \equiv X_u \times X_v$) to indicate part of a labeling. Functions $\theta_u : X_u \rightarrow \mathbb{R}$, $u \in V$, and $\theta_{uv} : X_{uv} \rightarrow \mathbb{R}$, $uv \in E$, are *potentials*, which define a local quality of labels and label pairs.

The *energy minimization* or *MAP-inference* problem for CRFs is

$$\min_{x \in X_V} \sum_{u \in V} \theta_u(x_u) + \sum_{uv \in E} \theta_{uv}(x_{uv}). \quad (1)$$

The objective in (1) is called *energy* of the CRF.

A great number of applied problems can be efficiently cast in the format (1), see e.g. [29,54]. This defines its importance for computer vision, machine learning and a number of other branches of science [54]. While problem (1) is NP-hard in general, many exact and approximate solvers were proposed [29].

Graph Matching. Although the format of Problem (1) allows us to express many practically important optimization tasks efficiently, some applications require the resulting labelings x to satisfy additional constraints. In particular, for the graph matching problem no label may be taken twice.

Let a *common universe* \mathcal{L} of labels be given such that $X_u \subseteq \mathcal{L} \forall u \in V$. We require each label $s \in \mathcal{L}$ to be taken at most once, i.e. $|\{u \in V : x_u = s\}| \leq 1$. In other words, we seek an injective mapping $(x_u)_{u \in V} : V \rightarrow \mathcal{L}$. This problem can be stated as

$$\min_{x \in X_V} \sum_{u \in V} \theta_u(x_u) + \sum_{uv \in E} \theta_{uv}(x_{uv}) \quad \text{s.t. } x_u \neq x_v \forall u \neq v. \quad (2)$$

Graph matching is NP-hard, since it is equivalent to MAP-inference for CRFs (1) in the trivial case, when nodes of the graph contain mutually non-intersecting sets of labels.

Inverse Graph Matching A special case arises if the universe \mathcal{L} of labels to be matched has the same size as the set of nodes of the graph $|\mathcal{L}| = |V|$. Then every injective mapping $(x_u)_{u \in V} : V \rightarrow \mathcal{L}$ must also be a bijection. Hence, every feasible labeling $x \in X_V$ corresponds to a permutation of V . The graph matching problem (2) can in this case also be approached in terms of the inverse permutation. To this end let the *inverse graph* $G' = (V', E')$ be given by $V' = \mathcal{L}$; the *inverse label set* $X'_s = \{v \in V : s \in X_v\}$ is associated with each node $s \in V'$; respectively $X'_\mathcal{L} = \prod_{s \in \mathcal{L}} X'_s$ is the set of *inverse labelings* and X'_{st} denotes $X'_s \times X'_t$; the set of edges of the inverse graph is defined as $E' = \{st \in V' \times V' : \exists x_{st} \in X'_{st} \text{ s.t. } x_s x_t \in E\}$. The *inverse costs* θ' for $s, s' \in V', x_s \in X'_s$ read:

$$\theta'_s(x_s) = \theta_{x_s}(s), \quad \theta'_{st}(x_{st}) = \begin{cases} \theta_{x_{st}}(s, t), & x_{st} \in E \\ 0, & \text{otherwise.} \end{cases}$$

Consider the resulting *inverse graph matching* problem

$$\min_{x \in X'_\mathcal{L}} \sum_{s \in \mathcal{L}} \theta'_s(x_s) + \sum_{st \in E'} \theta'_{st}(x_{st}) \quad \text{s.t. } x_s \neq x_t \forall s \neq t. \quad (3)$$

Labeling $x \in X_V$ and inverse labeling $y \in X'_\mathcal{L}$ correspond to each other iff $x_u = s \in V \Leftrightarrow y_s = u \in V$.

Note that when the edge set E is sparse, the inverse edge set E' may be not. In such a case, computational complexity of the inverse problem is higher than of the original one.

3. Lagrangean Decompositions

Since the graph matching problem (2) is NP-hard, it is common to consider convex relaxations. Below, we present three Lagrangean decomposition based relaxations of the problem. These can be applied to the original graph matching problem (2), to the inverse one (3) and to a combination of both. Since all these relaxations are based on the famous local polytope relaxation [47, 55] of the MAP-inference for CRFs (1), we give a short overview of this relaxation first.

Local Polytope for CRFs. The MAP-inference problem (1) can be represented as an integer linear program (ILP) [35] using an *overcomplete representation* [54] by grouping potentials corresponding to each node and edge into separate vectors. That is, $\theta_w(x_w)$, $w \in V \cup E$ stands for a vector with coordinates $(\theta_w(x_w))_{x_w \in X_w}$. The real-valued vectors μ_w have the same dimensionality as θ_w and stand for the "relaxed" version of x_w . The corresponding linear programming (LP) relaxation reads:

$$\begin{aligned} \min \quad & \sum_{w \in V \cup E} \langle \theta_w, \mu_w \rangle \\ \text{s.t.} \quad & \sum_{x_w \in X_w} \mu_w(x_w) = 1, \mu_w(s) \geq 0, w \in V \cup E, s \in X_w, \\ & \sum_{x_v \in X_v} \mu_{uv}(x_{uv}) = \mu_u(x_u), uv \in E, u \in uv, x_u \in X_u. \end{aligned} \quad (4)$$

Constraints of (4) define the *local polytope* L_G . Note that adding integrality constraints $\mu_w \in \{0, 1\}^{|X_w|}$ makes the problem (4) equivalent to its combinatorial formulation (1).

Integer Relaxed Pairwise Separable Linear Programs (IRPS-LP) Below we describe a general problem format studied in [50], which generalizes the local polytope relaxation (4). Importantly, the same format fits also the Lagrangean decompositions of the graph matching problem, which we consider below. This makes it possible to consider all these relaxations at once from a general viewpoint.

Let a *factor graph* $\mathcal{G} = (\mathbb{F}, \mathbb{E})$ consist of nodes $\mathbb{F} = \{1, \dots, k\}$, called *factors* and edges \mathbb{E} , called *factor-edges*. Let $X_i \subset \{0, 1\}^{\dim(X_i)}$, $i \in \mathbb{F}$, be sets of binary vectors and (ii) $A_{(i,j)} \subset \{0, 1\}^{\dim(X_i) \times K_{ij}}$, $ij \in \mathbb{E}$, $K_{ij} \in \mathbb{N}$ be matrices with binary entries, which map binary vectors from X_i into binary vectors from $\{0, 1\}^{K_{ij}}$, i.e., $A_{(i,j)} : X_i \rightarrow \{0, 1\}^{K_{ij}}$. The IRPS-LP is a class of problems, which factorize according to \mathcal{G} .

$$\min_{\mu \in \Lambda_{\mathcal{G}}} \sum_{i \in \mathbb{F}} \langle \theta_i, \mu_i \rangle \quad (5)$$

$$\Lambda_{\mathcal{G}} := \left\{ (\mu_1 \quad \dots \quad \mu_k) : \begin{array}{ll} \mu_i \in \text{conv}(X_i) & i \in \mathbb{F} \\ A_{(i,j)} \mu_i = A_{(j,i)} \mu_j & \forall ij \in \mathbb{E} \end{array} \right\}.$$

Constraints $A_{(i,j)} \mu_i = A_{(j,i)} \mu_j$ are associated with each factor-edge and are called *coupling constraints*. When representing the local polytope relaxation (4) as (5) we assume $\mathbb{F} = V \cup E$ and $\mathbb{E} = \{\{u, uv\}, \{v, uv\} : uv \in E\}$. The convex hull of X_w is fully defined by the first line of constraints in (4), since X_w constitutes a set of unit binary vectors. The second line of constraints in (4) defines the coupling constraints.

We use variable names μ for (in general) non-binary vectors $\mu_i \in \text{conv}(X_i)$ and x for binary ones $x_i \in X_i$, $i \in \mathbb{F}$.

3.1. Graph Matching Problem Relaxations.

Below, we describe three relaxations of the graph matching (2) problem, which fit the IRPS-LP (5) format. The first

one results in a standard local polytope relaxation (4) of a specially constructed CRF, the second one utilizes additional coupling constraints on top of (4), while the third approach uses a network flow subproblem. Additionally, we use the inverse formulation (3) and build two additional IRPS-LPs.

(R1) Graph Matching as CRF. To build a CRF equivalent to the graph matching we start with the underlying CRF as in (2) and express the uniqueness constraints in the edge factors. To this end we (i) extend the edge set E with new edges connecting any two nodes having at least one common label, i.e. $\hat{E} := E \cup \{uv \in \binom{V}{2} : X_u \cap X_v \neq \emptyset\}$; (ii) assign edge potentials $\theta_{uv} \equiv 0$ to all new edges $\hat{E} \setminus E$; (iii) for all $uv \in \hat{E}$ we assign $\theta_{uv}(x, x) := \infty \forall x \in X_u \cap X_v$. Any solution of the resulting CRF (1) with cost $< \infty$ is an assignment. The relaxation in terms of an IRPS-LP is the local polytope (4).

This approach results in general in a quadratic number of additional edge potentials, which may become intractable as the size of the graph matching problem grows.

(R2) Relaxation with Label Factors. For each label $s \in \mathcal{L}$ we introduce an additional *label factor*, which keeps track of nodes which assign label s . The label set of this factor $X_s := \{u \in V : s \in X_u\} \cup \{\#\}$ consists of those nodes $u \in V$ which can be assigned label s and an additional dummy node $\#$ representing non-assignment of label s . Label $\#$ is necessary, as not every label needs to be taken. The set of factors becomes $\mathbb{F} = V \cup E \cup \mathcal{L}$, with the coupling constraint set $\mathbb{E} = \{\{u, uv\}, \{v, uv\} : uv \in E\} \cup \{\{u, l\} : u \in V, l \in X_u\}$. The resulting IRPS-LP formulation reads

$$\begin{aligned} \min \quad & \sum_{w \in V \cup E} \langle \theta_w, \mu_w \rangle + \sum_{s \in \mathcal{L}} \langle \tilde{\theta}_s, \tilde{\mu}_s \rangle \quad (\text{R2}) \\ & \mu \in \mathbf{L}_G \\ & \tilde{\mu}_s \in \text{conv}(X_s), \quad s \in \mathcal{L} \\ & \mu_u(s) = \tilde{\mu}_s(u), \quad s \in X_u. \end{aligned}$$

Here we introduced additional potentials $\tilde{\theta}_s$ for the label factor. Initially, we set $\tilde{\theta}_s \equiv 0$.

(R3) Relaxation with a Network Flow Factor. If one ignores the edge potentials θ_{uv} in (2), the problem can be equivalently reformulated as bipartite matching [7]:

$$\begin{aligned} \min_{\mu \in \mathcal{M}} \quad & \sum_{u \in V} \langle \theta_u, \mu_u \rangle, \quad \text{where} \quad (6) \\ \mathcal{M} = \quad & \left\{ (\mu_u)_{u \in V} \geq 0 : \begin{aligned} & \sum_{s \in X_u} \mu_u(s) = 1, u \in V \\ & \sum_{u \in V, s \in X_u} \mu_u(s) \leq 1, s \in \mathcal{L} \end{aligned} \right\} \end{aligned}$$

Here we substituted the uniqueness constraints with the linear inequalities $\sum_{u \in V, s \in X_u} \mu_u(s) \leq 1$, which is equivalent for $\mu_u \in \{0, 1\}^{|X_u|}$. It is known that \mathcal{M} is the convex hull of all binary vectors satisfying the conditions of \mathcal{M} [7], i.e. $\text{conv}(\mathcal{M} \cap \{0, 1\}^{\dim(\mathcal{M})}) = \mathcal{M}$. Therefore \mathcal{M} fits into the

IRPS-LP framework. Crucially for an efficient implementation, (6) can be efficiently solved by minimum cost flow solvers [7].

Below we treat (6) as a separate factor \mathcal{M} and link it with (4) to obtain an IRPS-LP. Its factor graph is defined by $\mathbb{F} = V \cup E \cup \{\mathcal{M}\}$ and $\mathbb{E} = \{\{u, uv\}, \{v, uv\} : uv \in E\} \cup \{\{u, \mathcal{M}\} : u \in V\}$. The resulting IRPS-LP formulation is

$$\begin{aligned} \min \quad & \sum_{w \in V \cup E} \langle \theta_w, \mu_w \rangle + \sum_{u \in V} \langle \tilde{\theta}_u, \tilde{\mu}_u \rangle \quad (\text{R3}) \\ & \mu \in \mathbf{L}_G, \quad \tilde{\mu} \in \mathcal{M}, \\ & \tilde{\mu}_u(s) = \mu_u(s), \quad u \in V, s \in X_u. \end{aligned}$$

Initially, we set $\tilde{\theta} \equiv 0$.

Representation (6) for the uniqueness constraints has been already used e.g., in [20]. However their optimization technique lacks both convergence guarantees and monotonicity of a lower bound, which our methods possess. The work [61] considered the Lagrange dual of (R3) as a relaxation the graph matching problem. Their relaxation is equivalent to (R3), but their algorithm differs from ours. We refer to Section 5 for a discussion of the differences and to Section 6 for an experimental comparison.

(R4-R5) Coupling Original Graph Matching (2) and its Inverse (3). In the special case when $|\mathcal{L}| = |V|$ we may solve the inverse graph matching problem (3) instead of the original one (2). Another alternative is to solve both problems simultaneously and couple them together by requiring that the labeling of (2) is the inverse permutation for the labeling from (3). Such an approach doubles the problem size, yet it may result in a smaller number of iterations required to obtain convergence. This approach works both for relaxations (R2) and (R3).

The resulting coupled IRPS-LP for (R2) reads

$$\begin{aligned} \min_{\mu, \mu'} \quad & \sum_{w \in V \cup E} \langle \theta_w, \mu_w \rangle + \sum_{w \in V' \cup E'} \langle \theta'_w, \mu'_w \rangle \quad (\text{R4}) \\ & \mu \in \mathbf{L}_G, \quad \mu' \in \mathbf{L}_{G'} \\ & \forall u \in V, u' \in X_u : \mu_u(u') = \mu'_{u'}(u). \end{aligned}$$

Here the role of label factors in (R2) has been taken over by the node factors of the inverse graph matching (3). We distribute the costs equally among θ and θ' initially.

Another coupled IRPS-LP, corresponding to (R3) reads

$$\begin{aligned} \min_{\mu, \mu', \tilde{\mu}} \quad & \sum_{w \in V \cup E} \langle \theta_w, \mu_w \rangle + \sum_{w \in V' \cup E'} \langle \theta'_w, \mu'_w \rangle + \sum_{u \in V} \langle \tilde{\theta}_u, \tilde{\mu}_u \rangle \quad (\text{R5}) \\ & \mu \in \mathbf{L}_G, \quad \mu' \in \mathbf{L}_{G'}, \quad \tilde{\mu} \in \mathcal{M} \\ & \forall u \in V, u' \in X_u : \mu_u(u') = \tilde{\mu}_u(u'), \quad \mu'_{u'}(u) = \tilde{\mu}_u(u') \end{aligned}$$

Here the network flow factor \mathcal{M} controls consistency of the original μ and inverse labelings μ' . Initially, we set $\tilde{\theta} \equiv 0$ and distribute costs in θ and θ' equally.

The optimal values obtained by relaxations (R1) – (R5) may deliver differing bounds to (2), as characterized below.

Proposition 1. (R2) = (R3) and (R4) = (R5). Relaxation (R1) is weaker than (R2) and (R3).

4. General Algorithm

In this section we define a general algorithm for IRPS-LP problems (5), which is applicable to the decompositions (R1)–(R5) of the graph matching problem considered in Section 3.1. Our algorithm is a simplified version of the algorithm [50], where we fixed several parameters to the values common to the relaxations (R1)–(R5).

Instead of optimizing IRPS-LP (5) directly, we consider its Lagrangean dual w.r.t. the coupling constraints $A_{(i,j)}\mu_i = A_{(j,i)}\mu_j$. The IRPS-LP problem (5) can be shortly written as $\min_{\mu} \{\langle \theta, \mu \rangle, \text{ s.t. } A\mu = 0, \mu \in P\}$, where μ stands for $(\mu_i)_{i=1}^k$, $A\mu = 0$ represents all coupling constraints $A_{(i,j)}\mu_i - A_{(j,i)}\mu_j = 0$ and P denotes a polytope encapsulating the rest of constraints. By dualizing $A\mu = 0$ with a vector of Lagrange multipliers Δ one obtains the Lagrange function $\langle \theta, \mu \rangle - \langle \Delta, A\mu \rangle = \langle \theta - A^\top \Delta, \mu \rangle$. After introducing $\theta^\Delta := \theta - A^\top \Delta$ the dual objective reads $D(\Delta) = \min_{\mu} \{\langle \theta^\Delta, \mu \rangle, \text{ s.t. } \mu \in P\}$. It is well-known [14] that $D(\Delta) \leq \langle \theta, \mu \rangle$ for any feasible μ and the dual problem consists in maximizing $D(\Delta)$ over Δ . Going from θ to θ^Δ is called an *equivalent transformation* or *reparametrization* in the literature or *message passing* in the CRF-community.

Now we apply the above considerations to the general IRPS-LP problem (5). Specifically, let $i, j \in \mathbb{F}$ be two neighboring factors in the factor-graph \mathbb{G} . Then for any μ_i and μ_j satisfying the coupling constraint for edge $ij \in \mathbb{E}$

$$\begin{aligned} & \langle \theta_i, \mu_i \rangle + \langle \theta_j, \mu_j \rangle \\ &= \langle \theta_i, \mu_i \rangle + \langle \theta_j, \mu_j \rangle + \underbrace{\langle \Delta_{(i,j)}, A_{(i,j)}\mu_i - A_{(j,i)}\mu_j \rangle}_{=0} \\ &= \langle \theta_i + A_{(i,j)}^\top \Delta_{(i,j)}, \mu_i \rangle + \langle \theta_j - A_{(j,i)}^\top \Delta_{(i,j)}, \mu_j \rangle. \end{aligned}$$

The values and sign of the Lagrange multipliers $\Delta_{(i,j)}$ define how much cost is "sent" from j to i or the other way around. When we consider a subset $J \subseteq \mathcal{N}_{\mathbb{G}}(i)$ of the neighboring factors for i , the resulting equivalent transformation reads:

$$\theta_i \rightarrow \theta_i + \sum_{j \in J} A_{(i,j)}^\top \Delta_{(i,j)} \text{ and } \theta_j \rightarrow \theta_j - A_{(j,i)}^\top \Delta_{(i,j)}. \quad (7)$$

Below we define a subclass of messages $\Delta_{(i,j)}$ in the setting (7) which improve the dual.

Definition 1. Messages $\Delta_{(i,j)}$, $j \in J$, are called *admissible*, if there exists $x_i^* \in \operatorname{argmin}_{x_i \in X_i} \langle \theta_i, \mu_i \rangle \cap \operatorname{argmin}_{x_i \in X_i} \langle \theta_i^\Delta, \mu_i \rangle$ and additionally

$$\Delta_{(i,j)}(s) \begin{cases} \geq 0, & \nu(s) = 1 \\ \leq 0, & \nu(s) = 0 \end{cases}, \text{ where } \nu := A_{(i,j)} x_i^*. \quad (8)$$

We denote the set of admissible vectors by $AD(\theta_i, x_i^*, J)$.

Lemma 1 ([50]). *Admissible messages do not decrease the dual value, i.e., $\Delta \in AD(\theta_i^\phi, x_i^*, J)$ implies $D(0) \leq D(\Delta)$.*

Example 1. Let us apply Definition 1 to the local polytope relaxation (4) of CRFs. Let ij correspond to $\{u, uv\}$, where $u \in \mathbb{V}$ is some node and $uv \in \mathbb{E}$ is any of its incident edges and $J = \{j\}$. Then x_i^* corresponds to a locally optimal label $x_u^* \in \operatorname{argmin}_{s \in X_u} \theta_u(s)$ and $\nu(s) = \llbracket s = x_u^* \rrbracket$. Therefore we may assign $\Delta_{u,uv}(s)$ to any value from $[0, \theta_u(x_u^*) - \theta_u(s)]$. This assures that (8) is fulfilled and x_u^* remains a locally optimal label after reparametrization even if there are multiple optima in X_u .

Sending Messages. Procedure 1 represents an elementary step of our optimization algorithm. It consists of sending messages from a node i to a subset of its neighbors J .

Procedure 1: Send messages from $i \in \mathbb{F}$ to $J \subseteq \mathcal{N}_{\mathbb{G}}(i)$

- 1 **Optimize factor:** $x_i^* \in \operatorname{argmin}_{x_i \in X_i} \langle \theta_i, \mu_i \rangle$
- 2 **Choose** $\delta \in \mathbb{R}^{d_i}$ s.t. $\delta(s) \begin{cases} \geq 0, & x_i^*(s) = 1 \\ \leq 0, & x_i^*(s) = 0 \end{cases}$
- 3 **Maximize admissible messages to J :**

$$\Delta_{(i,J)} := (\Delta_{(i,j)})_{j \in J} \in \operatorname{argmax}_{\hat{\Delta} \in D(\theta_i^\phi, x_i^*, J)} \langle \delta, \theta_i^{\hat{\Delta}} \rangle \quad (9)$$

- 4 **Update** θ^i and θ_j , $j \in J$, according to (7)
-

Procedure 1 first computes an optimal labeling for the factor i in line 1, then computes message updates in (9) and finally updates the costs θ in line 4. The costs δ in line 2 are chosen as ± 1 , except when $i = \mathcal{M}$ is the network flow factor for (R3) and (R5). In this case, we choose $\delta(u, x_u) = \begin{cases} 0, & x_u = x_u^* \\ 1 - |X_u|, & x_u \neq x_u^* \end{cases}$.

Computation (9) provides a maximally possible admissible message from i to $\{J\}$. Essentially, it makes the cost vector of the factor i as uniform as possible. So, in the setting of Example 1 $\Delta_{u,uv}(s)$ becomes equal to $\theta_u(x_u^*) - \theta_u(s)$ and therefore $\theta_i^\Delta(s) = \theta_u(x_u^*)$ for all $s \in X_u$. Since the result of (9) is an admissible message, Procedure 1 never decreases the dual objective, as follows from Lemma 1.

Dual Ascent Algorithm. Let the notation $\{j_1, \dots, j_n\}_<$ stand for an ordered set such that $j_k < j_{k+1}$, $k = 1, \dots, n$. Algorithm 2 below goes over some of the factors $i \in \mathbb{F}$ in a pre-specified order and calls Procedure 1 to send or receive messages to/from some of the neighbors.

Algorithm 2 works as follows: We choose an ordered subset of factors $\{i_1, \dots, i_k\}_<$. For each factor $i \in \mathbb{F}$ we

Algorithm 2: Dual Ascent for IRPS-LP

```

1 Input:  $I = \{i_1, \dots, i_k\}_< \subseteq \mathbb{F}$ ,  $(J_r(i) \subseteq \mathcal{N}_{\mathbb{G}}(i))_{i \in I}$ ,
    $(J_s(i) \subseteq \mathcal{N}_{\mathbb{G}}(i))_{i \in I}$ 
2 for  $iter = 1, \dots, \mathbf{do}$ 
3   for  $i = i_1, \dots, i_k$  do
4     Receive messages:
5     for  $j \in J_r(i)$  do
6       | Call Algorithm 1 with input  $(j, \{i\})$ .
7     end
8     Send messages:
9     Call Algorithm 1 with input  $(i, J_s(i))$ .
10  end
11  Reverse the order of  $i_1, \dots, i_k$  and exchange  $J_r \leftrightarrow J_s$ 
12 end

```

select a neighborhood $J_r(i) \subseteq \mathcal{N}_{\mathbb{G}}(i)$ of factors from which to receive messages and a neighborhood $J_s(i)$ to which messages are sent by Procedure 1. We run Algorithm 2 on $\{i_1, \dots, i_k\}_<$ (forward direction) and $\{i_k, \dots, i_1\}_<$ (backward direction) alternately until some stopping condition is met. Since Algorithm 2 reparametrizes the problem by Procedure 1 only and the latter is guaranteed to not decrease the dual, so is Algorithm 2. We refer to [50] for further theoretical properties of Algorithm 2.

5. Graph matching algorithms.

For each of the relaxations (R1)-(R5) of the graph matching problem we detail parameters of Algorithm 2 used in our experiments: we define the sets I , $J_r(i)$, $J_s(i)$.

Algorithm Names. We use the following shortcuts for specializations of Algorithm 2 to the relaxations (R1)-(R5): **GM** corresponds to (R1), **AMP** to (R2), **AMCF** to (R3). To obtain the relaxations (R1-R3) we use either the original graph, as in (2), or an inverse one, as in (3). These options are denoted by suffixes **-O** and **-I** respectively. Additionally, the two coupled relaxations (R4) and (R5), are addressed by algorithms **AMP-C** and **AMCF-C** respectively. All in all, we have eight algorithms **GM-O**, **GM-I**, **AMP-O**, **AMP-I**, **AMP-C**, **AMCF-O**, **AMCF-I** and **AMCF-C**.

The sets I , $J_r(i)$ and $J_s(i)$ are defined in Table 5. For algorithms with the suffix **-I** the values are the same as for those with **-O**, but corresponding to the inverse graph.

We assume the order of graph nodes $\mathbb{V} := \{u_1, \dots, u_n\}_<$ and labels $\mathcal{L} := \{s_1, \dots, s_{|\mathcal{L}|}\}_<$ to be given a priori. We define $u_n < \mathcal{M} < s_1$ for the matching factor \mathcal{M} and $u < uv < v$ for the edge factors $uv \in E$. Similarly, we define $s < ss' < s'$ for all edge factors $ss' \in E'$ in the inverse graph. We extend the resulting partial order to a total one, e.g., by topological sort. For $i \in \mathbb{F}$ we define $\mathcal{N}_{\mathbb{G}}(i)_< :=$

Algorithm	Ordered set I	$J_r(i)$	$J_s(i)$
GM-O	$\{u_1, \dots, u_n\}_<$	$\mathcal{N}_{\mathbb{G}}(i)_<$	$\mathcal{N}_{\mathbb{G}}(i)_>$
AMP-O	$\{u_1, \dots, u_n, s_1, \dots, s_{ \mathcal{L} }\}_<$	$\mathcal{N}_{\mathbb{G}}(i)_<$	$\mathcal{N}_{\mathbb{G}}(i)_>$
AMCF-O	$\{u_1, \dots, u_n, \mathcal{M}\}_<$	$\mathcal{N}_{\mathbb{G}}(i)_< \cap E$	$\mathcal{N}_{\mathbb{G}}(i)_>$
AMP-C	$\{u_1, \dots, u_n, s_1, \dots, s_{ \mathcal{L} }\}_<$	$\mathcal{N}_{\mathbb{G}}(i)_<$	$\mathcal{N}_{\mathbb{G}}(i)_>$
AMCF-C	$\{u_1, \dots, u_n, \mathcal{M}, l_1, \dots, l_{ \mathcal{L} }\}_<$	$\mathcal{N}_{\mathbb{G}}(i)_< \cap E$	$\mathcal{N}_{\mathbb{G}}(i)_>$

Table 1. Input sets for specializations of Algorithm 2. For algorithms with the suffix **-I** the sets are the same as for those with **-O**, but correspond to the inverse graph.

$\{j \in \mathcal{N}_{\mathbb{G}}(i) : j < i\}$ and $\mathcal{N}_{\mathbb{G}}(i)_> := \mathcal{N}_{\mathbb{G}}(i) \setminus J_r(i)$ as the sets of preceding and subsequent factors.

Sending a message by some factor automatically implies receiving this message by another, coupled factor. Therefore, there is no need to go over all factors in Algorithm 2. In particular, edge-factors are coupled to node-factors only, therefore processing all node factors in Algorithm 2 automatically means updating all edge-factors as well. In the processing order and selection of the sets $J_r(i)$ and $J_s(i)$ we follow the most efficient MAP-solvers TRWS [33] and SRMP [34] (the latter is a generalization of TRWS to higher order models and has a slightly different implementation for pairwise CRFs (1)). In the special case when all nodes contain disjoint subsets of labels the graph matching problem (2) turns into MAP-inference in CRFs (1). Then all our algorithms **GM**, **AMP** and **AMCF** reduce to SRMP [34].

It is worth mentioning that for CRFs there exist algorithms, such as MPLP [24], which go over edge-factors only and in this way implicitly process also node-factors. As empirically shown in SRMP [34], MPLP is usually slower than SRMP. In Section 6 we show that our methods also favorably compare to the recently proposed HBP [61], which is similar to **AMCF-O**, but uses an MPLP-like processing schedule.

Optimization Subproblems of Procedure 1. For each call of Procedure 1 one must find the best factor element in line 1 and compute the best messages by solving (9). The first subproblem is solved by explicitly scanning all elements of the factor for node-, edge- and label-factors. For optimizing over \mathcal{M} , we use a min-cost-flow solver. Solving (9) for all choices of factors and neighborhoods is possible through closed-form solutions or calling a minimum cost flow solver and is described in the appendix.

Primal Rounding. Algorithm 2 only provides lower bounds to the original problem (2). To obtain a primal solution one may ignore the edge potentials θ_{uv} and solve the resulting reparametrized bipartite matching problem (6) with a minimum cost flow solver, as done in [61]. Empirically we found that it is better to interleave rounding and message passing, similarly as in TRWS [33] and SRMP [34]. Assume we have already computed a primal integer solution x_v^* for all $v < u$ and we want to compute x_u^* . To this end, between

lines 4 and 5 of Algorithm 2 for $i = u$ we assign

$$x_u^* \in \operatorname{argmin}_{x_u: x_u \neq x_v^* \forall v < u} \theta_u(x_u) + \sum_{v < u: uv \in E} \theta_{uv}(x_u, x_v^*). \quad (10)$$

Time complexity If $X_u = \mathcal{L} \forall u \in \mathcal{V}$, time complexity per iteration is $O(|\mathcal{L}||\mathcal{V}| + |\mathcal{L}|^2|E|)$ for **GM**. For **AMP** we must add $|\mathcal{L}|^3$ and for **AMCF** the time to solve (6) (possible in $O(\mathcal{L}^3)$). Details and speedups are in the appendix.

Higher Order Extensions. Our approach is straightforwardly extendable to the graph matching problem with higher order factors, a special case being third order: Let $\mathcal{T} \subseteq \binom{\mathcal{V}}{3}$ be a subset of triplets of nodes and $\theta_{uvw} : X_{uvw} \rightarrow \mathbb{R}$ be corresponding triplet potentials. The corresponding third order graph matching problem reads

$$\begin{aligned} \min_{x \in \mathcal{X}_{\mathcal{V}}} \sum_{u \in \mathcal{V}} \theta_u(x_u) + \sum_{uv \in E} \theta_{uv}(x_{uv}) + \sum_{uvw \in \mathcal{T}} \theta_{uvw}(x_{uvw}) \\ \text{s.t. } x_u \neq x_v \forall u \neq v. \end{aligned} \quad (11)$$

The associated IRPS-LP can be constructed by including additional factors for all triplets in an analogous fashion as in (4), see e.g. [56] for the corresponding relaxation.

For relaxations (R1) – (R5) we use third order factors to enforce cycle inequalities, which we add in a cutting plane approach as in [48]. For this we set $\theta_{uvw} \equiv 0$ at the beginning. By this construction (11) is equivalent to (2), however the corresponding IRPS-LP are not: Triplet potentials make the relaxation tighter.

6. Experiments

Algorithms. We compare against the two Lagrangean decomposition based solvers [52, 61] described in Section 1.

- The dual decomposition solver **DD** [52]. We use local subproblems containing 4 nodes. Note that the comparison in [61] was made with subproblems of size 3, hence **DD**’s relaxation was weaker there.
- “Hungarian belief propagation” **HBP** [61]. In [61] a branch and bound solver is used on top of the dual ascent solver. For a fair comparison our reimplementation uses only the dual ascent component. As for **AMP** and **AMCF**, we append to **HBP** the suffixes **-O** and **-C** to denote the relaxations we let **HBP** run on.

According to [52, 61], these two algorithms outperformed competitors [16, 19, 20, 25, 27, 38, 39, 42, 46, 51, 58, 64] at the time of their publication, hence we do not compare against the latter ones.

We set a hard threshold of 1000 iterations for each algorithm, exiting earlier when the primal/dual gap vanishes or no dual progress was observed anymore. We compute primal solutions every 5-th iteration in our algorithms. For **GM**, **AMP**, **AMCF** and **HBP** we use the tightening extension discussed in Section 5 to improve the dual lower bound. We tighten our relaxation whenever no dual progress occurs.

dataset	#I	#V	#L	C
house	105	30	30	dense
hotel	105	30	30	dense
car	30	19-49	19-49	dense
motor	20	15-52	15-52	dense
graph flow	6	60-126	2-126	sparse
worms	30	≤ 600	20-60	sparse

Table 2. Dataset description. #I denotes number of instances, #V the number of nodes $|\mathcal{V}|$, #L the number $|X_u|$ of labels a node $u \in \mathcal{V}$ can be matched to and C the connectivity of the graph.

Datasets. We have compared on six datasets:

- **house** [3] and **hotel** [2] with costs as in [52]. The task is to find matching feature points between images capturing an object from different viewpoints.
- **car** and **motor**, both used in [40], contain pairs of cars and motorbikes with keypoints to be matched. The images are taken from the VOC PASCAL 2007 challenge [21]. Costs are computed from features as in [40].
- The **graph flow** dataset [1] comes from a tracking problem with large displacements [8]. Keypoints in frames of RGB-D images obtained by a Kinect camera [60] are matched. The depth information provided by the Kinect camera is taken into account when computing the potentials θ .
- The **worms** dataset [4] from bioimaging [28]. The goal is to annotate nuclei of *C. elegans*, a famous model organism used in developmental biology, by finding the corresponding nuclei in a precomputed model. The instances of **worms** are, to our knowledge, the largest graph matching datasets ever investigated in the literature.

A summary of dataset characteristics can be found in Table 2. Previous computational studies concentrated on small-scale problems having up to 60 nodes and labels. We have included the **worms** dataset with up to 500 nodes and $|\mathcal{L}| = 1500$ labels.

Results. Fig. 2 shows performance of the algorithms on all 6 considered datasets. Among all variants **-O**, **-I**, **-C** corresponding respectively to the original, inverse and coupled formulations we plotted only the best one. As expected, for dense graphs (datasets **house**, **hotel**, **car**, **motor**) the variant **-C** with coupling provided most robust convergence, being similar to the best of **-O** either **-I** and therefore is presented on the plots. For sparse graphs, the inverse representation becomes too expensive, as the inverse edge set E' may be dense even though E is sparse in (3). Therefore we stick to the original problem **-O**.

- **hotel** and **house** are easy datasets, and many instances needed < 5 iterations for convergence. **AMP**, **AMCF** and **DD** were able to solve all instances to optimality within few seconds or even faster. However, **DD** is the fastest method for this data.

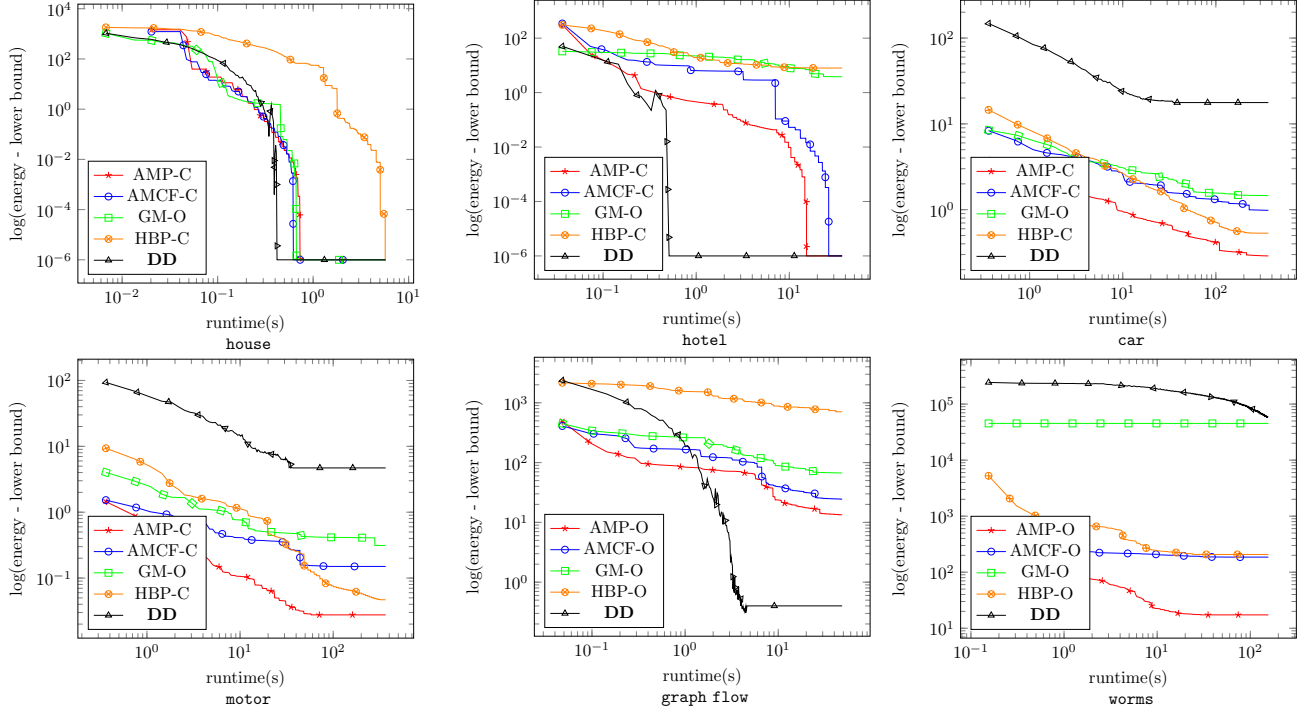


Figure 1. Plots showing convergence over time for house, hotel, car, motor, graph flow and worms datasets. Values denote $\log(\text{upper bound} - \text{lower bound})$. Values are averaged over all instances of the dataset. The x-axis and y-axis are logarithmic.

- `car` and `motor` were already harder and the 1000 iteration limit did not allow to ascertain optimality for all instances. **AMP** significantly outperforms its competitors, **DD** is significantly slower than the rest, whereas other algorithms show comparable results.
- on `worms` **AMP** significantly outperforms its competitors, **AMCF** and **HBP** converge to similar duality gap, although **AMCF** does it one-two orders of magnitude faster, **GM** and **DD** return results which are hardly competitive.
- `graph flow` is the only dataset, where **DD** clearly overcomes all competitors, followed by **AMP**. We attribute it to **DD**'s tighter relaxation, (its "local" subproblems contain 4 variables, whereas our subproblems have at most 3 variables after tightening.)

Insights and Conclusions

- **AMP** shows overall best performance for both small dense and large sparse datasets. It is the best anytime solver: it has the best performance in the first iterations. This is beneficial (i) if the run-time is limited or (ii) in branch-and-bound procedures, where a good early progress helps to efficiently eliminate non-optimal branches fast.
- Although **AMP**, **AMCF** and **HBP** address equivalent relaxations (having the same maximal dual value) their convergence speed is different. **AMCF** and **HBP** are generally slower than **AMP**, which we attribute to the sub-optimal redistribution of the costs by the min-cost-flow

factors $\{\mathcal{M}\}$ when maximizing messages in (9).

- **DD**'s relatively good performance is probably due to the large subproblems used by this method. First, this decreases the number of dual variables, which accelerates bound convergence; second, this makes the relaxation tighter, which decreases the duality gap. We attribute slow convergence of **DD** to the subgradient method.
- Summarizing, larger subproblems are profitable for the sub-gradient method, but not for message passing.
- **AMCF** outperforms **HBP** due to better message scheduling.
- We attribute the inferior performance of **GM** mostly to the weakest relaxation it optimizes. Even under this condition, due to a good message scheduling and fast message passing it outperforms **DD** and **HBP** on several datasets. A detailed evaluation of all instances is in the appendix.

7. Acknowledgments

The authors would like to thank Vladimir Kolmogorov for helpful discussions. The work is partially funded by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 616160 and under the European Unions Horizon 2020 research and innovation programme (grant agreement No 647769). The last author was supported by the DFG Grant "ERBI" SA 2640/1-1.

References

- [1] Graph flow dataset. <http://cvlab-dresden.de/research/iamge-matching/graphflow/>. 7
- [2] Hotel dataset. <http://vasc.ri.cmu.edu/idb/html/motion/hotel/>. 7
- [3] House dataset. <http://vasc.ri.cmu.edu/idb/html/motion/house/>. 7
- [4] Worms dataset. <https://datarep.app.ist.ac.at/57/>. 7
- [5] K. Adamczewski and Y. S. K. M. Lee. Discrete tabu search for graph matching. In *International Conference on Computer Vision*, 2015. 1
- [6] W. P. Adams and T. A. Johnson. Improved linear programming-based lower bounds for the quadratic assignment problem. *DIMACS series in discrete mathematics and theoretical computer science*, 16:43–75, 1994. 1
- [7] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1 edition, Feb. 1993. 4, 11, 13
- [8] H. A. Alhajja, A. Sellent, D. Kondermann, and C. Rother. Graphflow - 6d large displacement scene flow via graph matching. In J. Gall, P. V. Gehler, and B. Leibe, editors, *GCPR*, volume 9358 of *Lecture Notes in Computer Science*, pages 285–296. Springer, 2015. 7
- [9] K. M. Anstreicher and N. W. Bixius. A new bound for the quadratic assignment problem based on convex quadratic programming. *Mathematical Programming*, 89(3):341–357, 2001. 1
- [10] K. M. Anstreicher and N. W. Bixius. Solving quadratic assignment problems using convex quadratic programming relaxations. *Optimization Methods and Software*, 16(1-4):49–68, 2001. 2
- [11] M. Bazaraa and H. Sherali. New approaches for solving the quadratic assignment problem. In *Operations Research Verfahren*, volume 32, pages 29–46, 1979. 1
- [12] M. S. Bazaraa and H. D. Sherali. On the use of exact and heuristic cutting plane methods for the quadratic assignment problem. *Journal of the Operational Research Society*, 33(11):991–1003, 1982. 1
- [13] M. Beckman and T. Koopmans. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957. 1
- [14] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 5
- [15] R. E. Burkard, E. Cela, P. M. Pardalos, and L. S. Pitsoulis. The quadratic assignment problem. In *Handbook of combinatorial optimization*, pages 1713–1809. Springer US, 1998. 2
- [16] M. Cho, J. Lee, and K. M. Lee. Reweighted random walks for graph matching. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *ECCV (5)*, volume 6315 of *Lecture Notes in Computer Science*, pages 492–505. Springer, 2010. 2, 7
- [17] M. Cho, J. Sun, O. Duchenne, and J. Ponce. Finding matches in a haystack: A max-pooling strategy for graph matching in the presence of outliers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2083–2090, 2014. 2
- [18] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *IJPRAI*, 18(3):265–298, 2004. 1
- [19] T. Cour, P. Srinivasan, and J. Shi. Balanced graph matching. In *NIPS*. MIT Press, Cambridge, MA, 2007. 7
- [20] J. Duchi, D. Tarlow, G. Elidan, and D. Koller. Using combinatorial optimization within max-product belief propagation. In *NIPS*, pages 369–376. MIT Press, 2006. 4, 7
- [21] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. 7
- [22] A. Frieze and J. Yadegar. On the quadratic assignment problem. *Discrete applied mathematics*, 5(1):89–98, 1983. 1
- [23] J. W. Gavett and N. V. Plyter. The optimal assignment of facilities to locations by branch and bound. *Operations Research*, 14(2):210–232, 1966. 2
- [24] A. Globerson and T. S. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *NIPS*, pages 553–560, 2007. 2, 6
- [25] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(4):377–388, 1996. 1, 2, 7
- [26] P. Hahn, T. Grant, and N. Hall. A branch-and-bound algorithm for the quadratic assignment problem based on the hungarian method. *European Journal of Operational Research*, 108(3):629–640, 1998. 2
- [27] B. Jiang, J. Tang, C. Ding, and B. Luo. A local sparse model for matching problem. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI’15, pages 3790–3796. AAAI Press, 2015. 7
- [28] D. Kainmueller, F. Jug, C. Rother, and G. Myers. Active graph matching for automatic joint segmentation and annotation of *C. elegans*. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 81–88. Springer, 2014. 7
- [29] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, T. Kröger, J. Lellmann, N. Komodakis, B. Savchynskyy, and C. Rother. A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision*, 115(2):155–184, 2015. 2
- [30] J. H. Kappes, B. Savchynskyy, and C. Schnörr. A bundle approach to efficient MAP-inference by Lagrangian relaxation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1688–1695. IEEE, 2012. 2
- [31] S. E. Karisch and F. Rendl. Lower bounds for the quadratic assignment problem via triangle decompositions. *Mathematical Programming*, 71(2):137–151, 1995. 1
- [32] K. C. Kiwiel. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical programming*, 46(1-3):105–122, 1990. 2
- [33] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006. 1, 2, 6
- [34] V. Kolmogorov. A new look at reweighted message passing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(5):919–930, 2015. 2, 6

- [35] B. Korte, J. Vygen, B. Korte, and J. Vygen. *Combinatorial optimization*, volume 2. Springer, 2012. 3
- [36] E. L. Lawler et al. The quadratic assignment problem. *Management Science*, 9(4):586–599, 1963. 1
- [37] J. Lee, M. Cho, and K. M. Lee. A graph matching algorithm using data-driven markov chain monte carlo sampling. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2816–2819. IEEE, 2010. 2
- [38] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV*, pages 1482–1489. IEEE Computer Society, 2005. 2, 7
- [39] M. Leordeanu, M. Hebert, and R. Sukthankar. An integer projected fixed point method for graph matching and MAP inference. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *NIPS*, pages 1114–1122. Curran Associates, Inc., 2009. 1, 7
- [40] M. Leordeanu, R. Sukthankar, and M. Hebert. Unsupervised learning for graph matching. *International Journal of Computer Vision*, 96(1):28–45, 2012. 7
- [41] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-netto, P. Hahn, and T. Querido. An analytical survey for the quadratic assignment problem. *EUROPEAN JOURNAL OF OPERATIONAL RESEARCH*, pages 657–690, 2007. 1, 2
- [42] J. Maciel and J. Costeira. A global solution to sparse correspondence problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):187–199, 2003. 7
- [43] Q. Nguyen, A. Gautier, and M. Hein. A flexible tensor block coordinate ascent scheme for hypergraph matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5270–5278, 2015. 1
- [44] P. M. Pardalos, K. A. Murthy, and T. P. Harrison. A computational comparison of local search heuristics for solving quadratic assignment problems. *Informatica*, 4(1-2):172–187, 1993. 1
- [45] C. Ribeiro and M. Minoux. Solving hard constrained shortest path problems by Lagrangean relaxation and branch-and-bound algorithms. *Methods of Operations Research*, 53:303–316, 1986. 2
- [46] C. Schellewald and C. Schnörr. Probabilistic subgraph matching based on convex relaxation. In *EMMCVPR*, volume 3757 of *Lecture Notes in Computer Science*, pages 171–186. Springer, 2005. 1, 7
- [47] M. I. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika*, 4(113-130):1, 1976. 3
- [48] D. Sontag, D. K. Choe, and Y. Li. Efficiently searching for frustrated cycles in map inference. In *UAI*, pages 795–804. AUAI Press, 2012. 7
- [49] Y. Suh, M. Cho, and K. M. Lee. Graph matching via sequential monte carlo. In *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III*, pages 624–637, 2012. 2
- [50] P. Swoboda, J. Kuske, and B. Savchynskyy. A dual ascent framework for Lagrangean decomposition of combinatorial problems. In *CVPR*, 2017. 2, 3, 5, 6
- [51] P. H. S. Torr. Solving Markov random fields using semi definite programming. In *In: AISTATS*, 2003. 1, 7
- [52] L. Torresani, V. Kolmogorov, and C. Rother. A dual decomposition approach to feature correspondence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(2):259–271, 2013. 1, 2, 7
- [53] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(5):695–703, 1988. 2
- [54] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008. 2, 3
- [55] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007. 3
- [56] T. Werner. Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(8):1474–1488, 2010. 7
- [57] S. J. Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015. 2
- [58] J. Yarkony, C. C. Fowlkes, and A. T. Ihler. Covering trees and lower-bounds on quadratic assignment. In *CVPR*, pages 887–894. IEEE Computer Society, 2010. 7
- [59] M. Zaslavskiy, F. Bach, and J.-P. Vert. A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2227–2242, 2009. 2
- [60] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19(2):4–10, Apr. 2012. 7
- [61] Z. Zhang, Q. Shi, J. McAuley, W. Wei, Y. Zhang, and A. van den Hengel. Pairwise matching through max-weight bipartite belief propagation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2, 4, 6, 7
- [62] G. Zhao, B. Luo, J. Tang, and J. Ma. *Using Eigen-Decomposition Method for Weighted Graph Matching*, pages 1283–1294. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 2
- [63] Q. Zhao, S. E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *Journal of Combinatorial Optimization*, 2(1):71–109, 1998. 1
- [64] F. Zhou and F. D. la Torre. Factorized graph matching. In *CVPR*, pages 127–134. IEEE Computer Society, 2012. 2, 7