Sinkhorn Algorithm for Lifted Assignment Problems

Yam Kushinsky, Haggai Maron, Nadav Dym, and Yaron Lipman

Weizmann Institute of Science

July 25, 2017

Abstract

Recently, Sinkhorn’s algorithm was applied for solving regularized linear programs emerging from
optimal transport very efficiently [1]. Sinkhorn’s algorithm is an efficient method of projecting a positive
matrix onto the polytope of doubly-stochastic matrices. It is based on alternating closed-form Bregman
projections on the larger polytopes of row-stochastic and column-stochastic matrices.

In this paper we generalize the Sinkhorn projection algorithm to higher dimensional polytopes origi-
nated from well-known lifted linear program relaxations of the Markov Random Field (MRF) energy
minimization problem and the Quadratic Assignment Problem (QAP). We derive a closed-form projec-
tion on one-sided local polytopes which can be seen as a high-dimensional, generalized version of the
row/column-stochastic polytopes. We then use these projections to devise a provably convergent algo-
rithm to solve regularized linear program relaxations of MRF and QAP. Furthermore, as the regularization
is decreased both the solution and the optimal energy value converge to that of the respective linear pro-
gram. The resulting algorithm is considerably more scalable than standard linear solvers and is able to
solve significantly larger linear programs.

1 Introduction

The Sinkhorn algorithm [2, 1] for optimal transport problems is a popular method which solves optimal
transport problems extremely efficiently, at the price of a minor modification of the energy to be minimized
which takes the form of an entropic regularization term. The Sinkhorn algorithm splits the regularized
optimal transport into two subproblems, each with a simple closed-form solution, and iteratively solves both
problems in an alternating fashion. This results in a provably convergent algorithm for optimal transport
problems that is significantly more scalable than generic LP solvers.

In this paper we propose Sinkhorn-type algorithms for linear relaxations of two popular pattern recog-
nition problems - The Markov Random Field (MRF) energy minimization problem and the Quadratic Assignment
Problem (QAP).

The MRF energy minimization problem is a central modeling paradigm in computer vision and pattern
recognition [3]. Given a graph with $n$ vertices, it is the problem of assigning a label, out of a possible set
of $m$ labels, to each of the vertices so that a quadratic energy is minimized. Typically, in computer vision
applications, the vertices are pixels of an image and the labels represent some semantic information (e.g.,
image segmentation where the labels are the classes of interest). The MRF minimization problem is known
to be NP-hard [4]. A popular method for approximately solving MRF is by a linear programming (LP)
relaxation. The relaxation is defined in a lifted (i.e., higher dimensional) variable space and minimizes a
linear energy over the local polytope. The definition of the local polytope is somewhat ambiguous in the
literature. The local polytope defined in [5, 4, 6, 7], which we refer to as the one-sided local polytope
is a superset of the two-sided local polytope as defined in [8, 9, 10]. Regardless of the definition of the local
polytope, the resulting LP relaxations are often too big to solve with generic (e.g., interior point) LP solvers
as they have $O(n^2 m^2)$ variables and constraints.

Our first observation is that the optimization of a regularized linear energy over the one-sided local
polytope has an easily computable closed-form solution. The time complexity of computing this closed-form
solution is linear in the size of the data. Based on this observation and the fact that the two-sided local
polytope is the intersection of two one-sided local polytopes, we propose an efficient, provably convergent
The **quadratic assignment problem** as introduced in Lawler [11] is the problem of finding a bijection between the $n$ vertices of two graphs minimizing a quadratic energy. Two well-known subproblems of the QAP are the traveling salesman problem and the Koopmans-Beckmann quadratic assignment problem. Approximately solving either one of these subproblems is known to be NP-hard in general [12]. The popular Johnson-Adams (JA) relaxation [13] for the QAP is an LP with $n^4$ variables and constraints. We represent the JA polytope as an intersection of four one-sided local polytopes, and obtain an efficient, provably convergent Sinkhorn-type algorithm for optimizing regularized linear energies over the JA polytope, by iteratively solving one-sided problems.

The Sinkhorn-type algorithm suggested in this paper: (i) Solves regularized large-scale LP relaxations of MRF and QAP; and (ii) the optimal energy and solution of these regularized problems converge to the optimal energy and solution of the respective LP relaxations of MRF and QAP as the regularization term is decreased to zero. We provide numerical experiments validating our algorithm on the standard QAP benchmark [14] achieving slightly inferior results to the best known lower-bounds for these problems. We note that these best lower-bounds were achieved with a plethora of different techniques. Furthermore, we present a comparison with state of the art MRF solver [4] on a set of random problems demonstrating favorable results.

## 2 Related work

**MRF energy minimization** The MRF energy minimization problem is a well-studied problem. Here, we focus on approaches for solving the local polytope LP relaxation and refer the reader to a survey [15] that contains a detailed comparison of LP relaxation methods as well as other methods for this problem. The local polytope LP relaxation is a very common method for approximately solving the MRF minimization problem. Many papers focus on the problem of optimizing the large LPs efficiently. Generally speaking, these algorithms have two components. The first component consists of splitting the original problem into a number of easier problems. Typically this splitting decomposes the graph to a disjoint union of simple sub-graphs (e.g., acyclic graphs) for which the optimization is easier. The second component is a strategy for solving the original problem by iteratively solving the easier subproblems. The authors of [16] and the celebrated TRW-S algorithm of [4] use a dual ascent algorithm which is monotone but does not necessarily converge to the global minimum of the LP. Subgradient methods [17] are guaranteed to converge, but their convergence is slow in practice. Augmented Lagrangian methods [18] require less iterations at the price of solving quadratic programs in each subproblem rather than linear problems. Strictly convex regularization ensures that dual ascent algorithms will converge [20] to the global minimum of the regularized objective.

The main difference between our algorithm and those mentioned is in the first component. Our splitting decomposes the LP into two subproblems, each of which have a closed form solution. In contrast, the splitting proposed in the algorithms above depend on the topology of the graphs and may include many more subproblems. Additionally the subproblems, while easier than the original problem, generally do not have a closed form solution. The second component in our algorithm resembles the strictly convex regularization strategy, although those methods regularize the non-smooth dual function while our algorithm regularizes the linear objective.

**Quadratic assignment problems** Convex relaxations are a common strategy for dealing with the hardness of the QAP. Small instances of the QAP ($n < 30$) can be solved using branch and bound algorithms which use convex relaxations to obtain lower bounds [23]. For larger problems the non-integer solution obtained from the relaxation is rounded to obtain a feasible (generally suboptimal) solution for the QAP. Examples include spectral relaxations [24] and quadratic programming relaxations over the set of doubly stochastic matrices [26]. Lifting methods, in which auxiliary variables that represent the quadratic terms are introduced, provide linear programming (LP) relaxations [13] or semi-definite programming relaxations [29] which are often more accurate, but solve convex problems with $n^4$ variables in contrast with the cheaper spectral and quadratic programming methods that solve problems with $n^2$ variables. As a result, lifting methods cannot be solved using generic convex programming solvers for $n > 20$. It is also possible to construct relaxations with $n^{2k}$, $k > 2$ variables to achieve even tighter relaxations [31] at an increased
computational price. The authors of [13, 34] suggest to deal with the computational complexity of the large JA linear program by using a greedy coordinate ascent algorithm to solve the dual LP. While the running time of this algorithm resembles the running time of our algorithm, it is not guaranteed to converge to the global minimum of the JA relaxations. The authors of [35] propose a specialized solver for a lifted SDP relaxation of QAP, and the authors of [36] propose a converging algorithm for the JA and SDP relaxations. However both algorithms can only handle quadratic assignment instances with up to 30 points. More on the QAP can be found in surveys such as [22].

Entropic regularization The successfulness of entropic regularization for optimal transport linear programs has motivated research aimed at extending this method to other optimization problems. In [37, 38, 39] it is shown that regularized quadratic energies over positive matrices with fixed marginal constraints can be solved efficiently by solving a sequence of regularized optimal transport problems. Cuturi et al. [40] compute Wasserstein barycenters using entropic regularization. Benamou et al. [41] also consider Wasserstein barycenters as well as several other problems for which entropic regularization can be applied. One of these problems is the multi-marginal optimal transport which is related to the JA linear program, although the latter is more complex as the marginals in the JA linear program are themselves variables constrained by certain marginal constraints.

3 Approach

Our goal is to construct efficient algorithms for solving the LP relaxations of the MRF and QAP problems. Our method is motivated by the successfulness of the highly scalable Sinkhorn algorithm [2, 1] in (approximately) solving optimal transport problems. We begin by reviewing the key ingredients of the Sinkhorn algorithm and then explain how we generalize it to higher order LP relaxations.

To solve optimal transport (OT) problems efficiently, it is suggested in [42, 40, 41] to add an entropic regularizer to the OT problem:

$$\min_{x \in DS} \langle \theta, x \rangle + \alpha \sum_{ij} x_{ij} \left( \log x_{ij} - 1 \right), \tag{1}$$

where $\alpha$ is some small positive number, $DS = DS(\mu, \nu) \subset \mathbb{R}_{\geq 0}^{n \times n}$ is the set of non-negative $n \times n$ matrices with specified positive marginals $\mu, \nu \in \mathbb{R}_{>0}^n$:

$$\sum_{j} x_{ij} = \mu_i, \quad \forall i \tag{2a}$$

$$\sum_{i} x_{ij} = \nu_j, \quad \forall j \tag{2b}$$

$$x_{ij} \geq 0 \tag{2c}$$

Adding the entropy to the energy has several benefits: First, it allows writing the energy as a Kullback-Leibler divergence w.r.t. some $z \in \mathbb{R}_{>0}^{n \times n}$,

$$\min_{x \in DS} KL(x|z), \tag{3}$$

where $KL(x|z) = \sum_{ij} x_{ij} \left( \log \frac{x_{ij}}{z_{ij}} - 1 \right)$ is the KL divergence. This turns (1) into an equivalent KL-projection problem. Secondly, it makes the energy strictly convex. Thirdly, since the entropy’s derivative explodes at the boundary of DS it serves as a barrier function which ensures that the inequality constraints (2c) are never active, resulting in significant simplification of the KKT equations for (2). Finally, due to this simplification, the KL-projection over the row-stochastic (RS($\mu$)) (defined by (2a)) and column-stochastic (CS($\nu$)) (defined by (2b)) matrices has a closed form solution:
Theorem 1. Given $z \in \mathbb{R}_{>0}^{n \times m}$, the minimizer of

$$\min_{x \in \text{RS}(\mu)} KL(x|z),$$

is realized by the equation

$$x^*_ij = \frac{z_{ij}}{\sum_s z_{is} \mu_i}, \quad (5)$$

that is the row normalized version of $z$. The projection onto CS is defined similarly as the column normalized version of $z$.

The theorem is proved by directly solving the KKT equations of (4) (see e.g. [11]). These observations are used to construct an efficient algorithm to approximate the solution of the regularized OT problem (1) by repeatedly solving KL-projections on RS($\mu$) and CS($\nu$). As proved in [42] this converges to the minimizer of (1).

Following [41], we note that the Sinkhorn algorithm is an instance of the Bregman iterative projection method that allows solving KL-projection problems over intersection of affine sets $C_1, C_2, \ldots, C_N$.

$$\begin{align*}
\min_{x \geq 0} & \quad KL(x|z) \\
\text{s.t.} & \quad x \in C_1 \cap C_2 \cap \cdots \cap C_N
\end{align*}$$

via alternate KL-projections on the sets $C_i$, that is

$$\begin{align*}
x_0 &= z \\
\n &= \arg\min_{x \in C_{\text{mod}(n-1,N)+1}} KL(x|x_{n-1}), \quad n \geq 1
\end{align*}$$

In [42] it is shown that this procedure is guaranteed to converge, under the conditions that: (i) the feasible set of (6), $C = \cap_i C_i$, has non-empty interior, i.e., it is strictly feasible; and (ii) the minimizer of (6a) over each $C_i$ is strictly positive. In fact, in the case of KL-divergence (in contrast to the general Bregman divergence dealt with in [42]), condition (ii) can be proved from (i) using the fact that the derivatives of the KL-divergence blow-up at the boundary of the set defined by $x \geq 0$. Lastly, condition (i) is satisfied in all the problems we discuss in this paper. For example, DS contains a feasible interior point $x = \frac{1}{n} \mu \nu^T > 0$. While Bregman iterations can always be used to solve problems of the form (6), the performance of the method greatly depends on the chosen splitting of the feasible convex set $C$ into convex subsets $C_i$, $i = 1, \ldots, N$. Generally speaking a good splitting will split $C$ into a small number of sets, where the KL-projection on every set is easy to compute. The success of the optimal transport solution can be attributed to the fact that the feasible set $C = DS(\mu, \nu)$ is split into only two sets $C_1 = \text{RS}(\mu)$, $C_2 = \text{CS}(\nu)$, and the projection onto each one of these sets has a closed form solution.

We will use Bregman iterations to approximate the solution of the linear program relaxations of the MRF and QAP problems. We split the feasible sets of these relaxations into two sets in the case of MRF, and four sets in the case of QAP, so that the projection on each one of these sets has a closed-form solution. We begin by describing the MRF and QAP problems and their linear relaxations:

The MRF energy minimization problem is the problem of minimizing a quadratic energy over the one-sided permutations

$$\min_{x \in \Pi_1} \sum_{ij} \theta_{ij} x_{ij} + \sum_{ijkl} \tau_{ijkl} x_{ij} x_{kl},$$

where the one-sided permutations $\Pi_1 \subset \mathbb{R}^{n \times m}$ are defined via

$$\begin{align*}
\sum_j x_{ij} &= 1, \quad \forall i \\
x_{ij} &\in \{0,1\}, \quad \forall i, j
\end{align*}$$

This formulation of the MRF problem is known as the "overcomplete representation of a discrete graphical model". One common and powerful approximation to the solution of (8) is achieved via an LP relaxation in
a lifted space. That is, (8) is relaxed by replacing quadratic terms \(x_{ij}x_{kl}\) with new auxiliary variables \(y_{ijkl}\) to obtain
\[
\min_{(x,y)\in C} \sum_{ij} \theta_{ij}x_{ij} + \sum_{ijkl} \tau_{ijkl} y_{ijkl},
\]
(10)
where \(C\) is the **two-sided local polytope** (TLP) which is a convex relaxation of the one-sided permutations set in the lifted \((x,y)\) space:
\[
\sum_{j} x_{ij} = 1, \quad \forall i \quad (11a)
\]
\[
\sum_{l} y_{ijkl} = x_{ij}, \quad \forall i,j,k \quad (11b)
\]
\[
\sum_{j} y_{ijkl} = x_{kl}, \quad \forall i,k,l \quad (11c)
\]
\[
x,y \geq 0 \quad (11d)
\]
where \(x \in \mathbb{R}^{n \times m}\), \(y \in \mathbb{R}^{n^2 \times m^2}\). It is indeed a relaxation to \(\Pi_1\) since every one-sided permutation \((x,y)\) satisfies \((x,y)\in TLP\) for \(y_{ijkl} = x_{ij}x_{kl}\). For notational convenience we let \(d = n \times m + n^2 \times m^2\) and denote \((x,y) \in \mathbb{R}^{d}\).

In this paper we are interested in the regularized version of (10),
\[
\min_{(x,y)\in C} \langle \theta, x \rangle + \langle \tau, y \rangle + \alpha \sum_{ij} x_{ij} (\log x_{ij} - 1) + \alpha \sum_{ijkl} y_{ijkl} (\log y_{ijkl} - 1),
\]
(12)
where we denote \(\langle \theta, x \rangle = \sum_{ij} \theta_{ij}x_{ij}\), and similarly \(\langle \tau, y \rangle\).

The QAP problem can be seen as a symmetric version of the MRF problem and deals with minimizing quadratic energies of the form (8) where \(m = n\) and the energy is minimized over the permutations \(\Pi \subset \mathbb{R}^{n \times n}\) rather than over \(\Pi_1\). A well-known LP relaxation to QAP is the Johnson-Adams (JA) relaxation \([13]\) defined with the same energy as (10) but constraining also \(x^T\) to be a one-sided permutation. That is, we denote \(y_{ijkl}^T = y_{jikl}\), and define the set \(TLP^T \subset \mathbb{R}^d\) by all the pairs \((x,y) \in \mathbb{R}^d\) satisfying \((x^T,y^T) \in TLP\). Using this notation, the JA relaxation amounts to minimizing the energy in (10) where \(C\) is taken to be the **Johnson-Adams polytope** \(JAP = TLP \cap TLP^T\). Similarly, the regularized JA relaxation amounts to minimizing the energy in (12) over \(C = JAP\).

Continuing to draw inspiration from the linear OT case we replace the energy of (12) with an equivalent (up to a multiplicative constant) KL divergence energy,
\[
KL(x|z) + KL(y|w),
\]
where
\[
z_{ij} = \exp \left( -\frac{\theta_{ij}}{\alpha} \right), \quad w_{ijkl} = \exp \left( -\frac{\tau_{ijkl}}{\alpha} \right)
\]
and use Bregman iterations to approximate the solutions of these equivalent KL-projection problems. The main building block in this algorithm is a novel formula for closed-form KL-projection on the **one-sided local polytope** (OLP) defined via
\[
\sum_{j} x_{ij} = 1, \quad \forall i \quad (14a)
\]
\[
\sum_{l} y_{ijkl} = x_{ij}, \quad \forall i,j,k \quad (14b)
\]
\[
x,y \geq 0 \quad (14c)
\]
The derivation of this formula will be presented in the next section. Provided with this closed-form projection we will write TLP as an intersection of two OLP sets, and consequently JAP as intersection of four OLP sets:

TLP can be represented by denoting \(y_{ijkl}^\circ = y_{klij}\) and defining OLP\(^\circ\) as the set of \((x,y)\) satisfying \((x,y^\circ)\in OLP\). Then,
\[
TLP = OLP \cap OLP^\circ.
\]
(15)
Consequently,

\[ \text{JAP} = \text{TLT} \cap \text{TLP} = \text{OLP} \cap \text{OLP}^{0} \cap \text{OLP}^{T} \cap \text{OLP}^{T_{0}}. \]

The Bregman iterations (7) are guaranteed to converge to the solution of the regularized LP (12), provided that the relaxations are strictly feasible. This is indeed the case; an example of a strictly feasible solution is \( x = \frac{1}{m} 11^{T}, y = \frac{1}{m} 11^{T} \). Furthermore, as the entropy coefficient \( \alpha \) tends to zero the optimal value of (12) converges to the optimal value of the non-regularized problem (10), and the minimizer of the regularized problem converges to the minimizer of (10) which has maximal entropy [41].

### 4 KL-Projections onto the one-sided local polytope

We now compute the closed-form solution for KL-projections over the one-sided local polytope (OLP) defined in (14). Namely for given \((z, w) \in \mathbb{R}_{>0}^{d}\) we seek to solve

\[
\min_{(x, y) \in \text{OLP}} KL(x \mid z) + KL(y \mid w),
\]

**Theorem 2.** Given \((z, w) \in \mathbb{R}_{>0}^{d}\), the minimizer of (17) is given by the equations:

\[
\begin{align*}
q_{ij} &= \exp \left( \sum_{k} \log(\sum_{s} w_{ijks}) + \log z_{ij} \right) / n + 1 \\
x_{ij} &= \frac{q_{ij}}{\sum_{j} q_{ij}} \\
y_{ijkl} &= x_{ij} \frac{w_{ijkl}}{\sum_{s} w_{ijks}}
\end{align*}
\]

**Proof.** The proof is based on two applications of Theorem 1. First, we will find the optimal \( y \) for any fixed \( x \). Indeed, fixing \( x \) decomposes (17) into \( n \times m \) independent problems, one for each pair of indices \( i, j \) in (14b). Each independent problem can be solved using the observation that the matrix \( u_{kl} = y_{ijkl} \) is in \( \text{RS}(\mu) \) where \( \mu \) is the constant vector \( \mu = x_{ij} 1 \), where \( 1 \) denotes the vector of ones. Thus using Theorem 1

\[
y_{ijkl} = u_{kl} = x_{ij} \frac{w_{ijkl}}{\sum_{s} w_{ijks}}.
\]

Now we can plug this back in (17) and end up with a problem in the variable \( x \) alone. Indeed,

\[
KL(x \mid z) + KL(y \mid w) = \sum_{ij} \left[ KL(x_{ij} \mid z_{ij}) + \sum_{kl} \frac{w_{ijkl}}{\sum_{s} w_{ijks}} KL(x_{ij} \mid \sum_{s} w_{ijks}) \right]
\]

\[
= (n + 1) \sum_{ij} KL(x_{ij} \mid \exp \left( \log z_{ij} + \sum_{k} \log(\sum_{s} w_{ijks}) \right) / n + 1),
\]

where in the second equality we used the following (readily verified) property of KL-divergence \( a_{k}, b_{k} > 0 \):

\[
\sum_{k} a_{k} KL(x \mid b_{k}) = \left( \sum_{k} a_{k} \right) KL \left( x \mid \exp \left( \frac{\sum_{k} a_{k} \log b_{k}}{\sum_{k} a_{k}} \right) \right).
\]

Finally, we are left with a single problem of the form (4) and applying Theorem 1 again proves (18).

**Incorporating zeros constrains** The LP relaxation (10) can be strengthened by noting that for one-sided permutation \( x \in \Pi_{1} \) there exists exactly one non-zero entry in each row and therefore \( x_{ij} x_{il} = 0 \) for all \( j \neq l \). In the lifted LP formulation this implies \( y_{ijjl} = 0 \) for all \( i \) and \( j \neq l \). For permutations \( x \in \Pi \) also the symmetric equations hold: \( y_{ijkj} = 0 \) for all \( j \) and \( i \neq k \). These constraints (which are sometimes called gangster constraints) are part of the standard JA relaxation, but to the best of our knowledge are usually omitted in the MRF-LP relaxation. These additional constraints can be incorporated seamlessly in our algorithm as we will now explain.
Denote multi-indices of \( y \) by \( \gamma \) and let \( \Gamma \) be the set of multi-indices \( \gamma \) for which the constraint \( y_\gamma = 0 \) is to be added. We eliminate the zero valued variables from the objective (10) and the constraints defining the polytope \( C \), and rewrite them as optimization problem in the variables \( x \) and \( (y_\gamma)_{\gamma \in \Gamma} \). We then add entropic regularization only with respect to these variables, and use the same Bregman iteration scheme described above for the reduced variables. The only modification needed to the algorithm is a minor modification to the formula (18), where \( w \) is replaced with \( \bar{w} \) which satisfies \( \bar{w}_\gamma = 0 \) if \( \gamma \in \Gamma \) and \( \bar{w}_\gamma = w_\gamma \) otherwise.

We note that also with respect to the reduced variables the strengthened relaxations are strictly feasible so that the alternating KL-projection algorithm converges. An example of a strictly feasible solution in TLP and JAP respectively being

\[
(x, y) = \frac{1}{|\Pi_1|} \sum_{x \in \Pi_1} (x, xx^T), \quad (x, y) = \frac{1}{|\Pi|} \sum_{x \in \Pi} (x, xx^T),
\]

where \( y = xx^T \) is defined via \( y_{ijkl} = x_{ij}x_{kl} \).

**Incorporating sparsity** The method of eliminating variables from the linear relaxation before adding entropic regularization can also be used to exploit the sparsity pattern which is common in MRF minimization problems. Where variables \( y_{ijkl} \) such that \((i, k)\) is not an edge in the graph, or such that the labeling \( i \rightarrow j \) or \( k \rightarrow l \) is prohibited, can be omitted from (10).

## 5 Results

**Evaluation** To check the effect of entropic regularization on the optimal energy value of the JA relaxation, we conducted the following experiment: We chose a low-dimensional problem (\( n = 15 \)) whose JA relaxation can be solved by standard conic solvers (we used MOSEK [43]). We then compared the objective value obtained by the JA minimizer found by MOSEK and the minimizer of the regularized JA relaxation (which we name Sinkhorn-JA) with varying values of \( \alpha \). The results are shown in Figure 1(a). Note that our objective is always higher than JA. This is because any minimizer of the regularized JA relaxation is a feasible point of the JA relaxation. It can be seen that when \( \alpha \sim 10^{-3} \) the difference between the regularized objective and the original JA objective is relatively small. We were not able to take \( \alpha \) smaller than \( 10^{-4} \) due to numerical problems. As explained in [11] this is a drawback of entropic regularization methods which are not able to achieve the precision of interior point methods which use self-concordant barriers. However the advantage of entropic regularization is the ability to achieve approximate solutions much faster than interior point methods. In Figure 1(b) we compared the running time of our method with the running time of MOSEK for the JA relaxation. We observe that MOSEK solves problems with \( n \sim 15 \) in 900 seconds (15 minutes), while our method solves such problems in a few seconds and solved problems with \( n \sim 100 \) in around 15 minutes. The figure shows timing results for our algorithm with two different choices of \( \alpha \): \( \alpha_0 = 0.003 \) and \( \alpha_1 = 0.01 \). The two points in Figure 1(a) show the energy obtained for these two choices of \( \alpha \) when \( n = 15 \).

**Quadratic assignment** We evaluate our algorithm’s performance for the JA relaxation using QAPLIB. QAPLIB [14] is an online library containing several data sets of quadratic assignment problems, and provides the best known lower and upper bounds for each problem. Many of the problems have been solved to optimality, in which case the lower bound and upper bound are equal.

In Figure 2 we compare the estimated objective we obtain from Sinkhorn-JA (\( \alpha = 0.001 \), and \( \alpha = 10^{-4} \) for the chr dataset), with the best known lower bound and upper bound in QAPLIB. As can be seen in the figure, in most problems our method gives similar (though slightly inferior) results to the optimal results in QAPLIB, which were achieved using a rather large collection of different algorithms that are typically far...
slower than our own. Our method does fail significantly on the esc dataset. This is because the unregularized JA relaxation fails for these problems as well. Note that in most cases the objective value of Sinkhorn-JA is lower than the true optimal objective value of the QAPs since it is an approximation of the convex JA relaxation. However since it is only an approximation, in cases where the lower bound provided by the JA relaxation is (very close to) the optimal value of the QAP as in the dataset chr, the objective of Sinkhorn-JA can be slightly higher than the optimal value of QAP. For algorithms which project the solution of the relaxation to obtain a feasible solution this should not be a problem, however this could be a limitation for algorithms such as branch-and-bound where convex relaxations are used to obtain lower bounds.

**MRF** We compared our algorithm (Sinkhorn-MRF) with the popular TRW-S algorithm [4] for minimization of MRF energy problems. For the comparison with TRW-S we solved 1,000 instances of MRF energy minimization problems. In each instance we used graphs with \( n = 1,000 \) vertices, and a label set with \( m = 2 \) labels. We obtained edges for the graphs by randomly assigning a point in \( \mathbb{R}^2 \) to each vertex, and then choosing as (undirected) edges the \( k \) neighbors nearest to each vertex. The unary term and pairwise cost were defined by randomly generating \( n \times m \) and \( m \times m \) matrices according to the uniform distribution on \([-1, 1]\). To obtain a feasible assignment \( \bar{x} \in \Pi_1 \) from the soft assignment \( x \) obtained from Sinkhorn-MRF we used the following standard rounding procedure (see [15]): For each row \( x_i \) whose maximal element \( x_{ij} \) is larger than \( \frac{1}{2} + \epsilon \) we set \( \bar{x}_{ij} = 1 \). This gives us an initial set of fixed variables \( \mathcal{F} \). We then fix the remaining variables one by one by optimizing the energy (8) with respect to one new variable and all previously fixed variables. The algorithm depends on the choice of \( \epsilon \), in practice we run the algorithm with several values of \( \epsilon \) and then choose the solution \( \bar{x} \in \Pi_1 \) with minimal energy. We note that a similar rounding procedure is used in TRW-S (adapted to handle the dual solutions obtained from their algorithm).

In all of the 1,000 instances the optimal value of the TLP relaxation as computed by both TRW-S and Sinkhorn-MRF (with \( \alpha = 2.5 \times 10^{-4} \)) was very similar. This indicates that although TRW-S is not guaranteed to achieve the optimal value of TLP, the algorithm did succeed in converging for these random instances. However the integer solution obtained by Sinkhorn-MRF consistently gave a modestly better upper bound to the MRF energy than the integer solution of TRW-S, as shown in Figure 3(a). We believe this is due to the fact that the rounding algorithm both methods use is more suited for primal solutions than for dual solutions.

In Figure 3(b) we show an example \((n = 100, m = 2)\) where we compare the TLP optimal value with...
TRW-S lower bound and Sinkhorn-MRF optimal values for decreasing $\alpha$; note, that for sufficiently small $\alpha$ Sinkhorn-MRF value converges to the TLP optimal value while TRW-S has a $\approx 1.5\%$ gap.

References

[1] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In Advances in Neural Information Processing Systems, pages 2292–2300, 2013.

[2] JJ Kosowsky and Alan L Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. Neural networks, 7(3):477–490, 1994.

[3] Stan Z Li. Markov random field modeling in image analysis. Springer Science & Business Media, 2009.

[4] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. IEEE transactions on pattern analysis and machine intelligence, 28(10):1568–1583, 2006.

[5] Martin J Wainwright, Tommi S Jaakkola, and Alan S Willsky. Map estimation via agreement on trees: message-passing and linear programming. IEEE transactions on information theory, 51(11):3697–3717, 2005.

[6] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF energy minimization and beyond via dual decomposition. IEEE transactions on pattern analysis and machine intelligence, 33(3):531–552, 2011.

[7] Qifeng Chen and Vladlen Koltun. Robust nonrigid registration by convex optimization. In Proceedings of the IEEE International Conference on Computer Vision, pages 2039–2047, 2015.

[8] Chandra Chekuri, Sanjeev Khanna, Joseph Seffi Naor, and Leonid Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms, pages 109–118. Society for Industrial and Applied Mathematics, 2001.

[9] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 1(1–2):1–305, 2008.

[10] Patrick Pletscher and Sharon Wulff. Lpqp for map: Putting lp solvers to better use. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), pages 783–790, 2012.

[11] Eugene L Lawler. The quadratic assignment problem. Management science, 9(4):586–599, 1963.

[12] Sartaj Sahni and Teofilo Gonzalez. P-complete approximation problems. Journal of the ACM (JACM), 23(3):555–565, 1976.

[13] Warren P Adams and Terri 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.

[14] Rainer E Burkard, Stefan E Karisch, and Franz Rendl. Qaplib—a quadratic assignment problem library. Journal of Global optimization, 10(4):391–403, 1997.

[15] Joerg Kappes, Bjorn Andres, Fred Hamprecht, Christoph Schnorr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard Kausler, Jan Lellmann, Nikos Komodakis, et al. A comparative study of modern inference techniques for discrete energy minimization problems. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1328–1335, 2013.

[16] Amir Globerson and Tommi S Jaakkola. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In Advances in neural information processing systems, pages 553–560, 2008.

[17] Jörg Hendrik Kappes, Bogdan Savchynskyy, and Christoph 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.
[18] Ofer Meshi and Amir Globerson. An alternating direction method for dual map lp relaxation. *Machine Learning and Knowledge Discovery in Databases*, pages 470–483, 2011.

[19] Pedro Aguiar, Eric P Xing, Mário Figueiredo, Noah A Smith, and André Martins. An augmented lagrangian approach to constrained map inference. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 169–176, 2011.

[20] Vladimir Jojic, Stephen Gould, and Daphne Koller. Accelerated dual decomposition for map inference. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 503–510, 2010.

[21] Bogdan Savchynskyy, Jörg Kappes, Stefan Schmidt, and Christoph Schnörr. A study of nesterov’s scheme for lagrangian decomposition and map labeling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1817–1823. IEEE, 2011.

[22] Bogdan Savchynskyy, Stefan Schmidt, Jörg Kappes, and Christoph Schnörr. Efficient mrf energy minimization via adaptive diminishing smoothing. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 746–755. AUAI Press, 2012.

[23] Eliane Maria Loiola, Nair Maria Maia de Abreu, Paulo Oswaldo Boaventura-Netto, Peter Hahn, and Tania Querido. A survey for the quadratic assignment problem. *European journal of operational research*, 176(2):657–690, 2007.

[24] Franz Rendl and Henry Wolkowicz. Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem. *Mathematical Programming*, 53(1):63–78, 1992.

[25] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, volume 2, pages 1482–1489. IEEE, 2005.

[26] Kurt M Austreicher and Nathan W Brixius. A new bound for the quadratic assignment problem based on convex quadratic programming. *Mathematical Programming*, 89(3):341–357, 2001.

[27] Mikhail Zaslavskiy, Francis Bach, and Jean-Philippe Vert. A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2227–2242, 2009.

[28] Fajwel Fogel, Rodolphe Jenatton, Francis Bach, and Alexandre d’Aspremont. Convex relaxations for permutation problems. In *Advances in Neural Information Processing Systems*, pages 1016–1024, 2013.

[29] Qing Zhao, Stefan E Karisch, Franz Rendl, and Henry Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *Journal of Combinatorial Optimization*, 2(1):71–109, 1998.

[30] Itay Kezurer, Shahar Z. Kovavsky, Ronen Basri, and Yaron Lipman. Tight relaxation of quadratic matching. *Comput. Graph. Forum*, 34(5):115–128, August 2015.

[31] Monique Laurent. A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0–1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003.

[32] Warren P Adams, Monique Guignard, Peter M Hahn, and William L Hightower. A level-2 reformulation–linearization technique bound for the quadratic assignment problem. *European Journal of Operational Research*, 180(3):983–996, 2007.

[33] Peter M Hahn, Yi-Rong Zhu, Monique Guignard, William L Hightower, and Matthew J Saltzman. A level-3 reformulation-linearization technique-based bound for the quadratic assignment problem. *INFORMS Journal on Computing*, 24(2):202–209, 2012.

[34] Stefan E Karisch, Eranda Cela, Jens Clausen, and Torben Espersen. A dual framework for lower bounds of the quadratic assignment problem based on linearization. *Computing*, 63(4):351–403, 1999.
[35] Franz Rendl and Renata Sotirov. Bounds for the quadratic assignment problem using the bundle method. *Mathematical programming*, 109(2):505–524, 2007.

[36] Samuel Burer and Dieter Vandenbussche. Solving lift-and-project relaxations of binary integer programs. *SIAM Journal on Optimization*, 16(3):726–750, 2006.

[37] Anand Rangarajan, Steven Gold, and Eric Mjolsness. A novel optimizing network architecture with applications. *Neural Computation*, 8(5):1041–1060, 1996.

[38] Anand Rangarajan, Alan L Yuille, Steven Gold, and Eric Mjolsness. A convergence proof for the softassign quadratic assignment algorithm. *Advances in neural information processing systems*, pages 620–626, 1997.

[39] Justin Solomon, Gabriel Peyré, Vladimir G. Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Trans. Graph.*, 35(4):72:1–72:13, July 2016.

[40] Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693, 2014.

[41] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[42] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[43] ED Andersen and KD Andersen. The mosek interior point optimization for linear programming: an implementation of the homogeneous algorithm. *High Performance Optimization*, pages 197–232.