Solving the Uncapacitated Facility Location Problem Using Message Passing Algorithms

Nevena Lazic, Brendan J. Frey, Parham Aarabi
University of Toronto

Abstract

The Uncapacitated Facility Location Problem (UFLP) is one of the most widely studied discrete location problems, whose applications arise in a variety of settings. We tackle the UFLP using probabilistic inference in a graphical model - an approach that has received little attention in the past. We show that the fixed points of max-product linear programming (MPLP), a convexified version of the max-product algorithm, can be used to construct a solution with a 3-approximation guarantee for metric UFLP instances. In addition, we characterize some scenarios under which the MPLP solution is guaranteed to be globally optimal. We evaluate the performance of both max-sum and MPLP empirically on metric and non-metric problems, demonstrating the advantages of the 3-approximation construction and algorithm applicability to non-metric instances.

1 INTRODUCTION

The Uncapacitated Facility Location Problem (UFLP) is one of the most widely studied discrete location problems. Its applications arise in a variety of settings, including distribution system design (Klose & Drexl, 2003), self-configuration in wireless sensor networks (Frank & Romer, 2007), computational biology (Dueck et al., 2008) and computer vision (Li, 2007; Lazic et al., 2009). The UFLP can be stated as follows: given a set of customers C, a set of facilities F, cost f_j of opening each facility $j \in F$, and cost c_{ij} of connecting customer i to facility j, open a subset of facilities and assign each customer to exactly one facility at minimal total cost. The corresponding integer program (IP) is:

$$\min_{\mathbf{x}} \quad \sum_i \sum_j c_{ij} x_{ij} + \sum_j f_j y_j \tag{1}$$

$$\text{s.t.} \quad \sum_j x_{ij} = 1 \qquad \forall i \in C \tag{2}$$

$$y_j \geq x_{ij} \qquad \forall i \in C, j \in F \tag{3}$$

$$x_{ij}, y_j \in \{0, 1\} \quad \forall i \in C, j \in F \tag{4}$$

In the *metric* problem variant, the connection costs satisfy the following version of the triangle inequality:

$$c_{ij} \leq c_{ij'} + c_{i'j'} + c_{i'j} \quad \forall i, i' \in C \quad \forall j, j' \in F \tag{5}$$

One important special case of the UFLP is the exemplar-based clustering problem, where the customer and facility sets are the same. Furthermore, discrete model selection problems can be described as UFLP instances, where facility costs reflect model complexities and connection costs reflect the goodness-of-fit of data to each model - a framework previously used to identify multiple low-dimensional subspaces in high-dimensional data (Li, 2007; Lazic et al., 2009).

We formulate the UFLP as a maximum-a-posteriori (MAP) inference problem, by treating x_{ij}'s as hidden random variables whose joint log-likelihood corresponds to the UFLP objective. The MAP assignment of the variables then corresponds to the optimal assignment of customers to facilities. We perform inference by running two message passing algorithms on the factor graph problem representation: the standard max-product algorithm (Kschischang et al., 2001) and its convexified variant max-product linear programming (MPLP) (Globerson & Jaakkola, 2007).

The max-product algorithm is guaranteed to converge to the optimal MAP assignment on trees, and has empirically shown excellent performance on graphs with cycles in numerous applications, most notably in the area of error-correcting codes (Bendetto et al., 1996). More recently, it has been used to derive an effective algorithm for exemplar-based clustering problem, un-

Appearing in Proceedings of the 13^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9.

der the name of Affinity Propagation (Frey & Dueck, 2007). However, the theoretical guarantees on max-product convergence and solution optimality for general graphs are still an open area of research.

MPLP (Globerson & Jaakkola, 2007) is one of several recently developed linear programming (LP) based message passing algorithms for graphical models (Kolmogorov, 2006), (Werner, 2007), (Komodakis & Paragios, 2008). The iterative updates of these algorithms are quite similar to those of max-product and correspond to coordinate ascent in the LP dual; the relationship between the different dual problems is described in (Sontag & Jaakkola, 2009). MPLP has several desirable properties: it is guaranteed to converge, the objective function is monotonically non-increasing, and it gives an upper bound on the optimal MAP solution. We use these properties to provide performance guarantees on the solutions MPLP obtains for UFLP. Specifically, we characterize scenarios under which optimality is guaranteed, and augment MPLP with a greedy algorithm that constructs a solution whose cost is at most 3 times the optimal for metric instances.

We evaluate the performance of both standard max-product and MPLP empirically, on synthetic metric data and on the non-metric ORLIB benchmark problem instances. We demonstrate the advantages of the 3-approximation construction and the applicability of message passing algorithms to non-metric instances.

2 RELATED WORK

There exist many ways of approaching NP-hard problems such as the UFLP, including integer programming, approximation algorithms, and various heuristics; our work is most closely related to ρ-approximation algorithms for metric UFLP.

ρ-approximation algorithms are polynomial-time algorithms whose solution is provably at most ρ times worse than optimal, for some constant ρ (called the approximation ratio). For UFLP, the $O(\ln |C|)$-approximation of (Hochbaum, 1982) cannot be improved in general unless $NP \subseteq DTIME(n^{O(\log \log n)})$ (Arora et al., 1998). However, (Guha & Khuller, 1999) have shown that for the *metric* UFLP, constant ρ-approximation algorithms do exist and that $\rho > 1.463$ unless $NP \subseteq DTIME(n^{O(\log \log n)})$.

Most ρ-approximation algorithms for UFLP are based on the standard LP relaxation of the problem, where the integrality constraints $x_{ij}, y_j \in \{0, 1\}$ are replaced by non-negativity constraints $x_{ij}, y_j \geq 0$. These algorithms construct integral solutions whose cost is at most ρ times the LP cost, and hence at most ρ times the optimal IP cost. Two common LP-based approaches are LP rounding and primal-dual methods.

LP rounding algorithms first solve the LP, and then use various techniques to round any fractional solution values to integral. For the UFLP, a popular approach is to construct the solution *support graph* - a bipartite graph in which nodes represent customers and facilities, and weighted edges connect each customer-facility pair (i, j) for which $x_{ij} > 0$ in the LP solution. An integral solution is obtained by greedily clustering the customer nodes, and assigning all cluster members to the cluster center's closest facility. LP-rounding algorithms of (Shmoys et al., 1997), (Chudak & Shmoys, 2003), (Sviridenko, 2002), (Byrka, 2007) differ in the greedy criterion used to choose cluster centers and in graph pre-processing; (Byrka, 2007) obtains best-so-far approximation ratio of 1.5.

Primal-dual approximation algorithms, such as the 3-approximation UFLP algorithm of (Jain & Vazirani, 2001), start with a feasible solution $\mathbf{v}$ to the dual LP, and modify it until it is possible to construct an integral primal feasible solution $\mathbf{x}$ that satisfies certain relaxed complementary slackness conditions with respect to $\mathbf{v}$. The construction is based on those variables for which the complementary slackness conditions are tight. (Jain et al., 2002) use primal-dual analysis to prove that their greedy JMS heuristic guarantees an approximation ratio of 1.61. (Mahdian et al., 2007) further combine the JMS algorithm with the greedy augmentation procedure introduced by (Guha & Khuller, 1999) to obtain the 1.52-approximation MYZ algorithm.

In comparison to existing approximation algorithms, the MPLP-based approach we describe bears the most similarities to primal-dual methods, as it also performs co-ordinate ascent in a dual LP. Similarly to other methods, at convergence we construct a solution support graph and run a greedy variable assignment algorithm to obtain a 3-approximation solution guarantee. The approximation ratio is comparatively high as the greedy algorithm is fairly simple; however, the primary contribution lies in showing how the MPLP fixed point can be used to construct the support graph and provide a bound on the integral solution. The approximation ratio could likely be decreased by applying the more elaborate techniques described in literature.

3 GRAPHICAL MODEL AND THE MAX-PRODUCT ALGORITHM

We formulate UFLP as a MAP inference problem by treating x_{ij} as hidden binary random variables whose joint likelihood is equal to the UFLP objective, and represent it using the factor graph (Kschischang et al.,

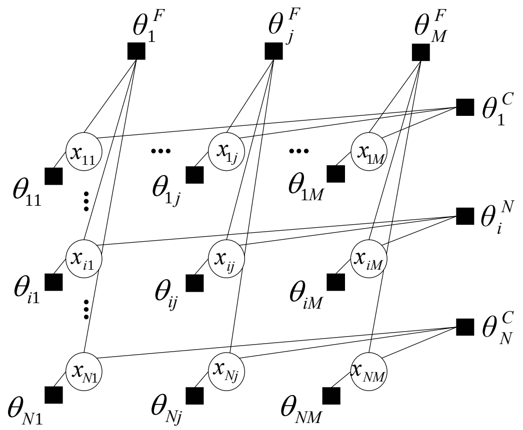

Figure 1: Factor graph representation of UFLP

2001) shown in Fig. 1. Recall that a factor graph is a bipartite graph consisting of variable and factor vertices, where the factors evaluate potential functions over the variables they are connected to. The distribution described by the graph is proportional to the product of all factor potentials, or their sum in the log domain. In Fig. 1, we work in the log-domain and define the following factor functions θ:

$$\theta_{ij}(x_{ij}) = -c_{ij}x_{ij} \tag{6}$$

$$\theta_j^F(x_{:j}) = \begin{cases} -f_j, & \sum_i x_{ij} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

$$\theta_i^C(x_{i:}) = \begin{cases} 0, & \sum_j x_{ij} = 1 \\ -\infty, & \text{otherwise,} \end{cases} \tag{8}$$

where we use the notation $x_{:j} = \{x_{1j}, \ldots, x_{Nj}\}$ and $x_{i:} = \{x_{i1}, \ldots, x_{iM}\}$ with $N = |C|$ and $M = |F|$. The row factors $\theta_i^C(x_{i:})$ enforce the constraint that each customer i is assigned to exactly one facility, while the column factors $\theta_j^F(x_{:j})$ and singleton factors $\theta_{ij}(x_{ij})$ incorporate the facility and connection costs, respectively. The MAP formulation of UFLP is:

$$\text{MAP-UFLP:} \max_{\mathbf{x}} \sum_{ij} \theta_{ij}(x_{ij}) + \sum_j \theta_j^F(x_{:j}) + \sum_i \theta_i^C(x_{i:})$$

We perform MAP inference using the max-sum algorithm, a log-domain equivalent of max-product. Max-sum iteratively updates messages $m_{\theta \to x}(x)$ and $m_{x \to \theta}(x)$ between adjacent factor and variable vertices in the graph. Upon convergence, each variable x is assigned to the value that maximizes the sum of its incoming messages $b(x) = \sum m_{\theta \to x}(x)$, also known as the *belief*.

Max-sum messages are functions of random variables, and for binary variables (as in Fig. 1) they are vectors of length two. In practice, it suffices to only keep track of the difference between the two values, which we will denote as $m \equiv m(1) - m(0)$. Furthermore, we require only the factor-to-variable messages to compute beliefs and make variable assignments. Following the naming convention in Fig. 2, the max-sum message updates for the factor graph of Fig. 1 are:

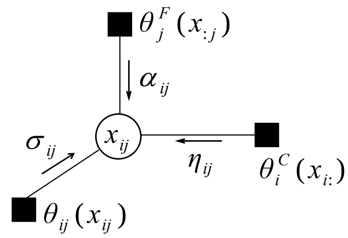

Figure 2: Message naming convention

$$\sigma_{ij} \leftarrow -c_{ij} \tag{9}$$

$$\eta_{ij} \leftarrow -\max_{k \neq j}(\alpha_{ik} - c_{ik}) \tag{10}$$

$$\alpha_{ij} \leftarrow \min[0, -f_j + \sum_{k \neq i} \max(0, \eta_{kj} - c_{kj})] \tag{11}$$

Upon convergence, we calculate beliefs and assign variables according to:

$$b_{ij}(x_{ij}) = -c_{ij} + \alpha_{ij} + \eta_{ij} \tag{12}$$

$$x_{ij}^* = \arg\max_{x_{ij}} b_{ij}(x_{ij}) \tag{13}$$

Finally, we note that although max-sum is not always guaranteed to converge, there exist practical ways of dealing with message oscillations. One common method is to use *damped* messages, whose updates relate to original updates as $m^{new} \leftarrow \lambda m^{old} + (1 - \lambda) m^{update}$, for a constant $\lambda \in [0, 1)$.

4 MPLP UFLP ALGORITHM

MPLP (Globerson & Jaakkola, 2007) is an LP-relaxation based message passing algorithm for MAP inference in graphical models, whose iterative updates are quite similar to those of max-sum. MPLP is guaranteed to converge to a fixed point and its objective gives an upper bound on the optimal MAP objective. In this section, we present the MPLP algorithm for solving the UFLP, describe some cases in which the obtained solution is optimal, and show how the MPLP fixed points can be used to construct a 3-approximation solution for metric UFLP.

4.1 MAP-LP RELAXATION

MPLP is based on a particular LP relaxation of the MAP inference problem, which we now describe. Let $\mu_c(\mathbf{x}_c)$ denote a distribution over variables connected to a factor $\theta_c(\mathbf{x}_c)$, and let $\mathcal{M}$ denote the set of distributions over all factors θ_c, such that (1) each μ_c is a valid distribution and (2) each $\mu_{c_1}(\mathbf{x}_{c_1})$ and $\mu_{c_2}(\mathbf{x}_{c_2})$ agree on their overlap variables $\mathbf{x}_{c_1} \cap \mathbf{x}_{c_2}$. The MAP-LP is:

$$\text{MAP-LP: } \max_{\mu \in \mathcal{M}} \sum_c \sum_{\mathbf{x}_c} \mu_c(\mathbf{x}_c)\theta_c(\mathbf{x}_c) \tag{14}$$

Compared to the MAP problem $\max_{\mathbf{x}} \sum_c \theta_c(\mathbf{x}_c)$, MAP-LP maximizes the weighted sum of potentials evaluated over all of their variable configurations, and the maximization is performed over the weights μ. As in all LP relaxations, the MAP-LP solution is an upper bound on the original problem and MAP-optimal when μ is integral.

For the UFLP factor graph, we require distributions $\mu_{ij}(x_{ij})$, $\mu_j^F(x_{:j})$ and $\mu_i^C(x_{i:})$ over variable sets x_{ij}, $x_{:j}$ and $x_{i:}$ respectively, corresponding to single-node, column and row factors. $\mathcal{M}$ is the set of all valid distributions that agree on singleton marginals. Letting $x_{-kj} = x_{:j} \setminus x_{kj}$ and $x_{i-k} = x_{i:} \setminus x_{ik}$,

$$\mathcal{M} = \left\{ \mu \middle| \begin{array}{ll} \mu \geq 0 & \\ \sum_{x_{ij}} \mu_{ij}(x_{ij}) = 1 & \forall i \in C, j \in F \\ \sum_{x_{-ij}} \mu_j^F(x_{:j}) = \mu_{ij}(x_{ij}) & \forall i \in C, j \in F \\ \sum_{x_{i-j}} \mu_i^C(x_{:i}) = \mu_{ij}(x_{ij}) & \forall i \in C, j \in F \end{array} \right\}$$

There is a simple relationship between the MAP-LP and the standard LP. If we explicitly add the constraint $\sum_j x_{ij} = 1$ (as opposed to adding it implicitly through the potential function $\theta_i(x_{i:})$) and simplify, we get the following problem:

$$\min_{\mu \in \mathcal{M}} \quad \sum_{i,j} c_{ij}\mu_{ij}(1) + \sum_j f_j[1 - \mu_j^F(0, ..., 0)] \tag{15}$$

$$\text{s.t.} \quad \sum_j \mu_{ij}(1) = 1 \quad \forall i \in C, j \in F \tag{16}$$

Comparing Eq. 21 with the standard LP objective $\sum_{ij} c_{ij}x_{ij} + \sum_j f_j y_j$, we can see that the quantities $\mu_{ij}(1)$ and $1 - \mu_j^F(0, \ldots, 0)$ play the roles of x_{ij} and y_j, respectively. The constraints $y_j \geq x_{ij}$ automatically hold since it is always true that $1 - \mu_j^F(0, ..., 0) \geq \mu_{ij}(1)$ for feasible μ. The constraints in Eq. 16 come from the fact that if $\mu_i^C(x_{i:})$ is feasible, normalized, and nonzero only for configurations with $\sum_j x_{ij} = 1$, then $\mu_{ij}(1) = \mu_{i:}^C(0, ..., 1, ..., 0)$ (with 1 in the j^{th} position) and hence $\sum_j \mu_{ij}(1) = 1$.

4.2 MPLP ALGORITHM FOR UFLP

The MPLP message updates are block co-ordinate descent steps in the dual of the MAP-LP, augmented with dummy variables that are copies of μ. Following (Globerson & Jaakkola, 2007), we can derive the following dual problem:

$$\min \quad \sum_{ij} \max_{x_{ij}} b_{ij}(x_{ij}) \tag{17}$$

$$\begin{aligned} \text{s.t.} \quad & b_{ij}(x_{ij}) = -c_{ij}x_{ij} + \alpha_{ij}(x_{ij}) + \eta_{ij}(x_{ij}) \\ & \alpha_{ij}(x_{ij}) = \max_{x_{-ij}} \beta_j^{Fi}(x_{:j}) \quad (18) \\ & \eta_{ij}(x_{ij}) = \max_{x_{i-j}} \beta_i^{Cj}(x_{i:}) \quad (19) \\ & \sum_i \beta_j^{Fi}(x_{:j}) = \theta_j^F(x_{:j}) \forall j, x_{:j} \\ & \sum_j \beta_i^{Cj}(x_{i:}) = \theta_i^C(x_{i:}) \forall i, x_{i:} \end{aligned}$$

MPLP message updates are block co-ordinate descent steps in the dual variables $\beta_i^{Cj}(x_{i:})$ and $\beta_j^{Fi}(x_{:j})$. In practice, we only need to keep track of $\eta_{ij}(x_{ij})$ and $\alpha_{ij}(x_{ij})$ to compute beliefs and make variable assignments. Performing the maximizations in Eq. 18 and Eq. 19 and using the notation $m = m(1) - m(0)$ as before, we get the following message updates:

$$\begin{aligned} \eta_{ij} \quad & \leftarrow \quad -\frac{1}{M} \max_{k \neq j}(\alpha_{ik} - c_{ik}) - \frac{M-1}{M}(\alpha_{ij} - c_{ij}) \\ \alpha_{ij} \quad & \leftarrow \quad \frac{1}{N} \min\left[0, -f_j + \sum_{k \neq i} \max(0, \eta_{kj} - c_{kj})\right] \\ & \quad\quad -\frac{N-1}{N}(\eta_{ij} - c_{ij}) \end{aligned}$$

As in the regular max-sum algorithm, variables are assigned according to $x_{ij}^* = \arg\max_{x_{ij}} b_{ij}(x_{ij})$ at convergence, which is now guaranteed.

4.3 MPLP-UFLP FIXED POINT PROPERTIES

Once the MPLP messages converge, we can easily obtain an integral solution $\mathbf{x}^*$ if each belief $b(x_{ij})$ is uniquely maximized at some x_{ij}^*. However, it is also possible to have tied beliefs $b_{ij}(1) = b_{ij}(0)$, or equivalently $b_{ij} = 0$. Before proceeding, we will describe some message properties for each case. At convergence, we have that

$$\begin{aligned} b_{ij} \quad & = \quad -c_{ij} + \alpha_{ij} + \eta_{ij} \\ & = \quad \frac{1}{M}(\alpha_{ij} - c_{ij}) - \frac{1}{M} \max_{k \neq j}(\alpha_{ik} - c_{ik}) \quad (20) \\ & = \quad \frac{1}{N} \min[\eta_{ij} - c_{ij}, -f_j + \sum_k \max(0, \eta_{kj} - c_{kj})] \end{aligned}$$

From Eq. 20, each customer i can either be uniquely assigned to a single facility j^* with $b_{ij^*} > 0$ or tied between several facilities with $b_{ij} = 0$. We will use C_1 and C_0 to denote the sets of uniquely assigned and tied customers, respectively.

We will also distinguish between sets of open facilities F_1, tied facilities F_0 and closed facilities F_{-1}. From Eq. 20, the quantity $-f_j + \sum_i \max(0, \eta_{ij} - c_{ij})$ will also be greater than, equal to, or less than zero for these three types of facilities.

Let $\hat{\eta}_i$ denote any message η_{ij} to customer i for which $b_{ij} \geq 0$. We do not need to indicate the corresponding facility j for the following reason: when $i \in C_1$, there exists only one such message, and when $i \in C_0$, the η_{ij} messages from all tied facilities are equal, which follows from Eq. 20. We also note that $b_{ij} \geq 0$ always implies that $\eta_{ij} = \hat{\eta}_i \geq c_{ij}$.

Let $\mathbf{x}^*$ be an integral solution constructed by assigning customers to open facilities in F_1 whenever possible, and assigning tied customers $i \in C_0$ arbitrarily to any facility $j \in F_0$ for which $b_{ij} = 0$. With some algebraic manipulation, we can express the negative dual objective $g^D(\alpha, \eta) = -\sum_{ij} b_{ij}(x^*_{ij})$, a lower bound on the optimal IP cost, as:

$$
\begin{aligned}
g^D(\alpha,\eta) &= \sum_{j\in F_1}(f_j + \sum_{i\in C}\min(0, c_{ij} - \eta_{ij})) + \sum_{i\in C}\hat{\eta}_i \\
&= \sum_{j\in F_1} f_j + \sum_{j\in F_1}\sum_{i\in C_1} c_{ij}x^*_{ij} + \sum_{i\in C_0}\hat{\eta}_i \quad (21)
\end{aligned}
$$

where the second equality follows from the fact that whenever $x^*_{ij} = 1$, $c_{ij} - \eta_{ij} < 0$ and $\eta_{ij} = \hat{\eta}_i$. From Eq. 21, the negative dual objective can be split into components corresponding to uniquely assigned and tied customers. When there are no ties, the LP and IP costs are equal and the solution $\mathbf{x}^*$ is optimal. As shown by (Globerson & Jaakkola, 2007) and (Sontag & Jaakkola, 2009), this is true in general for MPLP: when all beliefs have unique maximizers, the solution is guaranteed to be optimal.

4.4 MPLP SOLUTION OPTIMALITY

An integral solution $\mathbf{x}^*$ obtained from an MPLP fixed point, where $x^*_{ij} = \arg\max_{x_{ij}} b_{ij}(x_{ij})$, is guaranteed to be optimal if the LP relaxation is tight. One case in which this is true is when all beliefs have unique maximizers, as shown in the previous section. In general, the LP relaxation will be tight if there exists an integral $\mu \in \mathcal{M}$ such that the following complementary slackness conditions hold $\forall i \in C$, $j \in F$:

$$\sum_{x_{ij}} \mu_{ij}(x_{ij})\big[b_{ij}(x_{ij}) - \max_{x_{ij}} b_{ij}(x_{ij})\big] = 0 \quad (22)$$

$$\sum_{x_{i:}} \mu_i^{Cj}(x_{i:})\big[\beta_i^{Cj}(x_{i:}) - \max_{x_{i-j}} \beta_i^{Cj}(x_{i:})\big] = 0 \quad (23)$$

$$\sum_{x_{:j}} \mu_j^{Fi}(x_{:j})\big[\beta_j^{Fi}(x_{:j}) - \max_{x_{-ij}} \beta_j^{Fi}(x_{:j})\big] = 0 \quad (24)$$

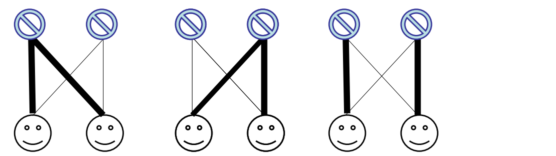

Figure 3: An illustrative example showing a case where the LP relaxation is tight, but multiple optima exist. All costs are equal; smileys represent customers, crossed circles represent facilities, and bold edges show possible solutions. The first two solutions are optimal, while the last one is not.

If some variables are tied with $b_{ij}(1) = b_{ij}(0)$, either the relaxation is tight and there exist multiple integral optima, or the relaxation is not tight. In the former case, the manner in which we decode tied variables may be important, as illustrated in Fig. 3.

We can construct an integral primal feasible solution μ^* corresponding to some $\mathbf{x}^*$ such that $\mu^*(\mathbf{x}^*) = 1$, and $\mu^*(\mathbf{x}) = 0$ whenever $\mathbf{x} \neq \mathbf{x}^*$. Such μ^* automatically satisfies Eq. 22, as all singleton beliefs are maximized. Eq. 23 is satisfied if each customer i is assigned to exactly one facility j for which $b_{ij} \geq 0$. For Eq. 24, there are two cases to consider for any given facility j. If $j \in F_1$, any configuration of tied variables suffices. If $j \in F_0$, Eq. 24 holds only if the corresponding tied variables are either all set to 0, or all set to 1.

To summarize, the LP relaxation is tight and $\mathbf{x}^*$ optimal if (1) all singleton beliefs are maximized at $\mathbf{x}^*$, (2) each customer is assigned to exactly one facility, and (3) each tied facility $j \in F_0$ serves either all or none of its customers for which $b_{ij} = 0$. An intuitive interpretation is that each facility must serve all of the nearby customers to justify its opening cost.

4.5 A 3-APPROXIMATION ALGORITHM FROM THE MPLP FIXED POINTS

We now describe an algorithm for constructing a 3-approximation integral solution from an MPLP fixed point, for metric UFLP instances. We first construct a bipartite support graph $G = (V, E)$, whose vertices are customers and facilities, and whose edges $(i, j) \in E$ connect each customer i and facility j with $b_{ij} \geq 0$.

Similarly to the method of (Jain & Vazirani, 2001), we use the greedy Alg. 1 to open facilities so that (1) no two open facilities are within a path of length 2 in G, (2) each customer is at a path of length 1 or 3 of an open facility to which it gets assigned. The algorithm steps are illustrated in Fig. 4. We will denote the obtained solution by $\bar{\mathbf{x}}^*$, as it may not correspond to an MPLP solution $\mathbf{x}^*$ (specifically, the two solutions will disagree on the customers assigned at path length 3).

Algorithm 1 3-Approximation Variable Assignment

initialize
  $\bar{F}_1 \leftarrow F_1$, $\bar{F}_0 \leftarrow F_0$
  $\bar{C}_1 \leftarrow C_1$, $\bar{C}_0 \leftarrow C_0$, $\bar{C}_3 \leftarrow \{\emptyset\}$
  assign customers in $\bar{C}_1$ to neighbors in $\bar{F}_1$
while $|\bar{F}_0| > 0$ and $|\bar{C}_0| > 0$ **do**
  $\hat{i} \leftarrow \arg\min_{i \in \bar{C}_0} \hat{\eta}_i$
  if $\nexists j \in \bar{F}_0$ s.t. $(\hat{i}, j) \in E$ **then**
    $\bar{C}_3 \leftarrow \bar{C}_3 \cup \hat{i}$, $\bar{C}_0 \leftarrow \bar{C}_0 \setminus \hat{i}$
  else
    open any facility $k \in \bar{F}_0$ s.t. $(\hat{i}, k) \in E$:
    $\bar{F}_1 \leftarrow \bar{F}_1 \cup k$, $\bar{F}_0 \leftarrow \bar{F}_0 \setminus k$
    for all $i \in ne(k)$ **do**
      assign i to k: $x^*_{ik} = 1$, $x^*_{ij} = 0 \forall j \neq k$
      remove path length 2 facilities: $\bar{F}_0 \setminus ne(i)$
    end for
    $\bar{C}_1 \leftarrow \bar{C}_1 \cup ne(\hat{j})$, $\bar{C}_0 \leftarrow \bar{C}_0 \setminus ne(\hat{j})$
  end if
end while
assign customers in $\bar{C}_3$ to closest facilities in $\bar{F}_1$

After running Alg. 1, $\bar{F}_1$ is the set of all open facilities, $\bar{C}_1$ is the set of customers assigned to facilities at path length 1, and $\bar{C}_3$ is the set of customers assigned suboptimally to facilities at path length 3 in the solution $\bar{\mathbf{x}}^*$. We will now compare the IP costs of $\bar{\mathbf{x}}^*$ to the dual objective at convergence. The negative dual objective is equal to:

$$
\begin{aligned}
g^D(\alpha, \eta) &= \sum_{j \in F_1} f_j + \sum_{j \in F_1} \sum_{i \in C_1} c_{ij} x^*_{ij} + \sum_{i \in C_0} \hat{\eta}_i \\
&= \sum_{j \in \bar{F}_1} f_j + \sum_{j \in \bar{F}_1} \sum_{i \in \bar{C}_1} c_{ij} \bar{x}^*_{ij} + \sum_{i \in \bar{C}_3} \hat{\eta}_i
\end{aligned}
$$

The second equality follows from the facts that for each tied facility $j \in F_0$, $f_j + \sum_{i,(i,j)\in E} c_{ij} = \sum_{i,(i,j)\in E} \hat{\eta}_i$, and that facilities in $\bar{F}_1$ serve all their neighbors in G. The IP cost of the solution $\bar{\mathbf{x}}^*$ will be

$$
g^{IP}(\bar{\mathbf{x}}^*) = \sum_{j \in \bar{F}_1} f_j + \sum_{j \in \bar{F}_1} \sum_{i \in \bar{C}_1} c_{ij} \bar{x}^*_{ij} + \sum_{i \in \bar{C}_3} \min_{j \in \bar{F}_1} c_{ij}
$$

Hence, the cost of customers in $\bar{C}_3$ changes from $\hat{\eta}_i$ to $\min_{j \in \bar{F}_1} c_{ij}$, and we can show that $\min_{j \in \bar{F}_1} c_{ij} < 3\hat{\eta}_i$.

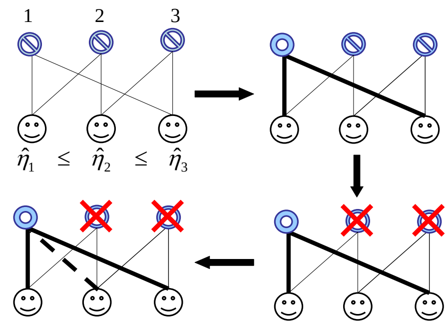

Figure 4: An example run of Alg. 1, where smileys represent customers, crossed circles represent tied facilities, and donuts represent opened facilities. Solid bold lines represent customer assignments at path length 1, while dashed bold lines represent assignments at path length 3.

To see this, consider the example in Fig. 4, where facility 1 is open due to customer 1 with minimal $\hat{\eta}_i$, and customer 2 ends up suboptimally assigned to facility 1. Assigning customer 2 to facility 1 will contribute a total of c_{21} to the IP cost, and we can show that $c_{21} \leq 3\hat{\eta}_2$:

$$
\begin{aligned}
c_{21} &\leq c_{22} + c_{12} + c_{11} && \text{(triangle inequality)} \\
&\leq \eta_{22} + \eta_{12} + \eta_{11} && (\eta_{ij} \geq c_{ij} \ \forall (i,j) \in E) \\
&\leq 3\max(\hat{\eta}_1, \hat{\eta}_2) && (\eta_{ij} = \hat{\eta}_i \ \forall (i,j) \in E\ ) \\
&= 3\hat{\eta}_2 && \text{(greedy order)}
\end{aligned}
$$

In summary, the IP costs of customers and facilities in $\bar{C}_1$ and $\bar{F}_1$ are equal to their LP costs, and customers in $\bar{C}_3$ cost at most 3 times their LP cost. Since the LP objective is a lower bound on the optimal IP cost, it follows that $g^{IP}(\bar{\mathbf{x}}^*) \leq 3g^D(\alpha, \eta) \leq 3g^{IP}(\mathbf{x}^{OPT})$. When there are customers assigned at path length 3, the solution $\bar{\mathbf{x}}^*$ will be different from any solution $\mathbf{x}^*$ that maximizes individual variable beliefs.

Finally, we note that our main contribution here is in showing the construction a support graph from an MPLP fixed point and relating it to the dual MPLP objective; the approximation ratio of 3 could likely be decreased using a more elaborate decoding scheme.

5 EXPERIMENTS

We evaluate the performance of damped max-sum and MPLP for UFLP on several metric and non-metric data sets. For both algorithms, we resolve variable ties by (1) using Alg. 1, and (2) arbitrarily assigning a tied customer i to the next facility j for which $b_{ij} = 0$.

We first evaluate the algorithms on the non-metric ORLIB dataset, one of the most widely used UFLP

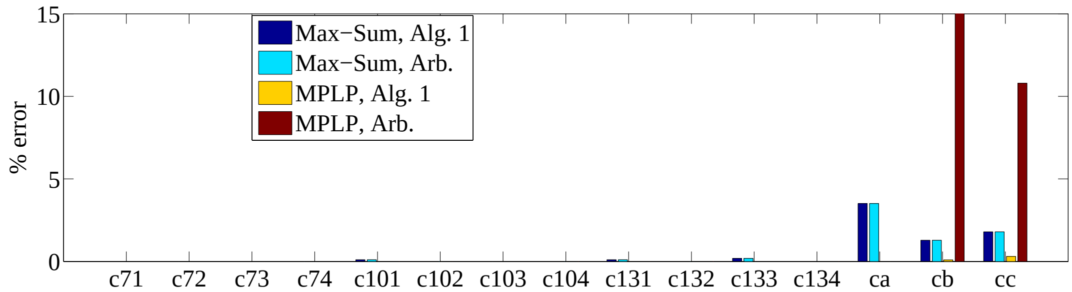

Figure 5: Experimental results on the non-metric ORLIB data set. The error indicates the percentage by which the obtained cost exceeds the optimal.

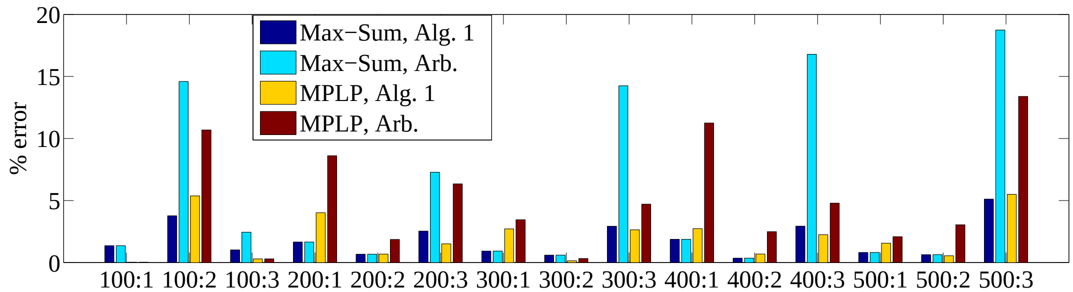

Figure 6: Experimental results on synthetic metric clustering problems, where N:k indicates the data set with N customers, and facility costs set to $\sqrt{N}/10^k$ The error indicates the percentage by which the obtained cost exceeds the LP lower bound.

benchmarks. The ORLIB instances *ca*, *cb* and *cc* are of size $N = 1000$ and $M = 100$, while the $c7*$, $c10*$ and $c13*$ instances have $N = 50$ customers and $M = 16$, $M = 25$, $M = 50$ facilities, respectively.

We then evaluate the algorithms on exemplar-based clustering problems, where data points are also potential facilities, i.e. $F = C$. In this case, max-sum corresponds to the Affinity Propagation algorithm. We first synthetically generate a metric data set, by randomly uniformly sampling $N \in \{100, 200, 300, 400, 500\}$ points in a unit square, setting connection costs to Euclidean distances, and setting all facility costs to either $\sqrt{N}/10$, $\sqrt{N}/100$, or $\sqrt{N}/1000$. We also run the algorithms on two non-metric data sets previously used to evaluate Affinity Propagation in (Dueck et al., 2008): images derived from the Olivetti Face database (Samaria & Harter, 1994), and Document Summarization.[1]

The results are shown in Fig. 5, Fig. 6 and Fig. 7, where the error measures the percentage by which the solution cost exceeds the optimal cost for ORLIB, and the LP lower bound for clustering problems.

On the ORLIB data set, MPLP with the greedy Alg. 1 decoding outperforms the other methods, finding the optimal solutions for 13 instances and obtaining lower costs on the remaining 2 instances. Using Alg. 1 brings a significant improvement over arbitrary decoding for MPLP. On the metric clustering data, the performance of max-sum and MPLP is comparable, while on non-metric data max-sum outperforms MPLP. In all cases, Alg. 1 decoding results in equal or lower cost than arbitrary decoding. In general, Alg. 1 will result in suboptimally connected customers, while arbitrary decoding will result in suboptimally-opened facilities, and the performance will depend on which operation is more costly.

With respect to run time, we observed that damped max-sum with $\lambda = 0.8$ always converges, and takes fewer iterations to do so than MPLP. Some intuition behind the number of iterations is that the MPLP message updates look relatively similar to those of max-sum with large damping: if we set λ to $1 - 1/N$ or $1 - 1/M$, we can relate the two types of updates as $m_{ij}^{MS} = m_{ij}^{MPLP} + \lambda b_{ij}$. In general, damping makes max-sum more stable and likely to converge, but high damping constants also increase the number of required iterations. With $N = 500$ customers, the damping constant corresponding to the multiplicative factors in MPLP would be $\lambda = 0.998$, which would result

[1] Datasets and descriptions available at http://www.psi.toronto.edu/affinitypropagation/

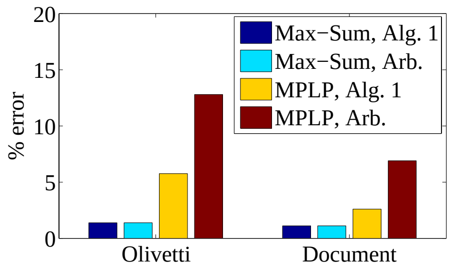

Figure 7: Experimental results on non-metric clustering problems. The error indicates the percentage by which the obtained cost exceeds the LP lower bound.

in more max-sum iterations as well.

6 CONCLUSION

We have described a new approach to solving the UFLP using message passing algorithms. We have also extended the MPLP algorithm with a greedy variable assignment that guarantees a 3-approximation solution for metric UFLP instances. More generally, our approach demonstrates that LP relaxation based message passing algorithms such as MPLP can be used to construct approximations for NP hard problems in a manner similar to primal-dual methods: by performing coordinate ascent in dual variables, and constructing integral primal solutions based on complementary slackness conditions. Relating such algorithms to standard dual LPs and approximation algorithms is an exciting direction of future work.

References

Arora, S., Lund, C., Motwani, R., Sudan, M., & Szegedy, M. (1998). Proof verification and hardness of approximation problems. *Journal of the ACM*, *45*.

Bendetto, S., Montorsi, G., Divsalar, D., & Pollara, F. (1996). *Soft-output decoding algorithms in iterative decoding of turbo codes* (Tech. Rep.). JPL TDA.

Byrka, J. (2007). An optimal bifactor approximation algorithm for the metric uflp. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, *4627*.

Chudak, F., & Shmoys, D. (2003). Improved approximation algorithms for the uncapacitated facility location problem. *SIAM Journal on Computing*, *33*(1).

Dueck, D., Frey, B., Jojic, N., & Jojic, V. (2008). Constructing treatment portfolios using affinity propagation. In *International conference on research in computational molecular biology.*

Frank, C., & Romer, K. (2007). Distributed facility location algorithms for flexible configuration of wireless sensor networks. In *Distributed computing in sensor systems.* Springer.

Frey, B., & Dueck, D. (2007). Clustering by passing messages between data points. *Science.*

Globerson, A., & Jaakkola, T. (2007). Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In *Advances in neural information processing systems.*

Guha, S., & Khuller, S. (1999). Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, *31*.

Hochbaum, D. S. (1982). Heuristics for the fixed cost median problem. *Mathematical Programming*, *22*(2).

Jain, K., Mahdian, M., & Saberi, A. (2002). A new greedy approach for facility location problems. In *Proc. 34th annual acm symposium on theory of computing.*

Jain, K., & Vazirani, V. (2001). Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of the ACM*, *48*.

Klose, A., & Drexl, A. (2003). Facility location models for distribution system design. *European Journal of Operations Research.*

Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *28*(10), 1568–1583.

Komodakis, N., & Paragios, N. (2008). Beyond loose lp-relaxations: optimizing mrfs by repairing cycles. In *European conference on computer vision.*

Kschischang, F., Frey, B., & Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, *47*(2).

Lazic, N., Givoni, I., Frey, B., & Aarabi, P. (2009). Floss: Facility location for subspace segmentation. In *International conference on computer vision.*

Li, H. (2007). Two-view motion segmentation from linear programming relaxation. In *Computer vision and pattern recognition.*

Mahdian, M., Ye, Y., & Zhang, J. (2007). Approximation algorithms for metric facility location problems. *SIAM Journal on Computing*, *36*(2).

Samaria, F., & Harter, A. (1994). Parameterization of a stochastic model for human face identification. In *Proc. 2nd ieee workshop on applications of computer vision.*

Shmoys, D., Tardos, E., , & Aardal, K. (1997). Approximation algorithms for facility location problems. In *Proc. 29th annual acm symposium on theory of computing.*

Sontag, D., & Jaakkola, T. (2009). Tree block coordinate descent for map in graphical models. In *12th international workshop on artificial intelligence and statistics.*

Sviridenko, M. (2002). An improved approximation algorithm for the metric uflp. In *Proc. 9th conf. on integer programming and combinatorial optimization.*

Werner, T. (2007). A linear programming approach to the max-sum problem: a review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *29*(7).