2-Catalog Segmentation, Facility Location, and the Markov Decision

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/ yyye

Joint work with M. Mahdian, D. Xu and Jiawei Zhang.

Outline

- 2-Catalog Segmentation
- SDP Relaxation and Approximation
- Facility Location
- Greedy Algorithm and Approximation
- The Markov Decision Problem

2-Catalog Segmentation Problem

- **Input:** A ground set I of n items, a family $\{S_1, S_2, \cdots, S_m\}$ of subsets of I and an integer $1 \leq k \leq n$.
- **Output:** Subsets $A_1, A_2 \subset I$ such that $|A_1| = |A_2| = k$.
- **Objective:** To maximize $\sum_{i=1}^{m} \max\{|S_i \cap A_1|, |S_i \cap A_2|\}$.
- **Application:** I is the list of goods; there are m customers where customer i is interested in the goods of S_i; A_1 and A_2 are the two catalogs one of which could be sent to each of the customers such that the (total) satisfaction is maximized.

The Partition Version

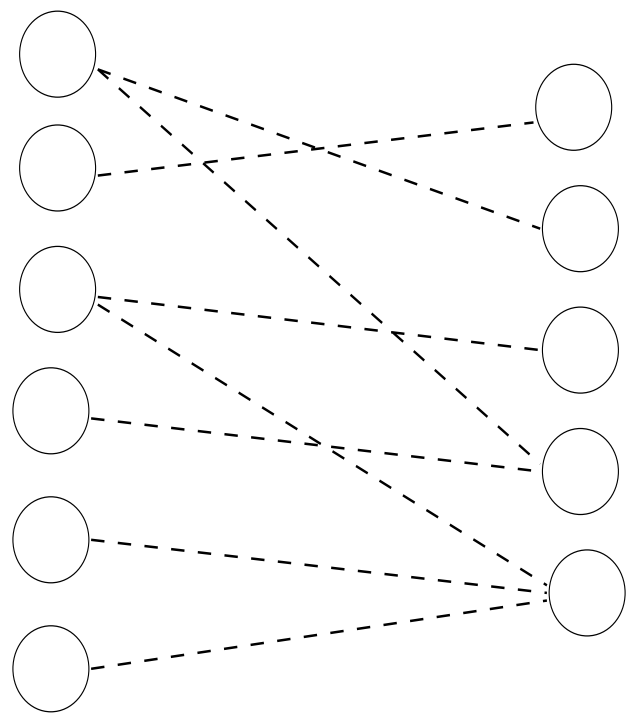
- **Input:** A **bipartite** graph $G = (A, B, E)$ with $|A| = n$ and $|B| = m$,
- **Output:** A partition of $B = B_1 \cup B_2$, and two subsets A_1, A_2 of A **such that** $|A_1| = |A_2| = k$
- **Objective:** To maximize $w(A_1, B_1) + w(A_2, B_2)$.

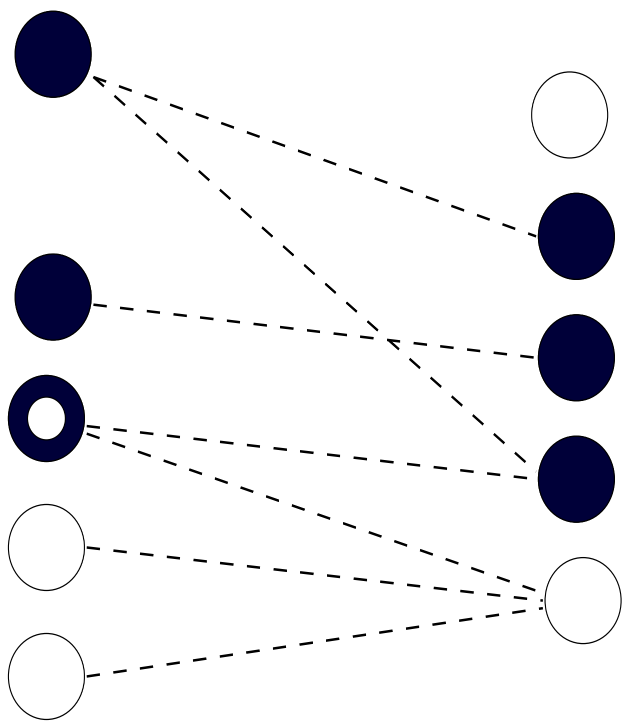
Connections between the two versions:

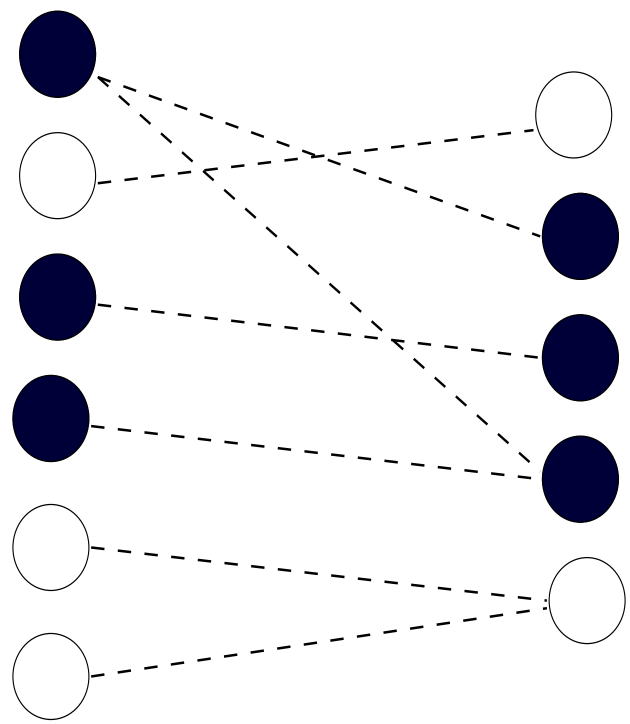
$A \longleftrightarrow I$;

$B \longleftrightarrow$Customers;

the set of edges connected to $i (\in B_i) \longleftrightarrow S_i$.







Previous Results

- The problem was proposed by Kleinberg, Papadimitriou and Raghavan (1998).
- A simple greedy algorithm with performance guarantee (KPR 1998): simply selects $A_1 \in A$ to be the k nodes of largest degrees, $A_2 \in A$ be any k nodes in A, $B_1 = B$, $B_2 = \emptyset$.
- A polynomial time approximation scheme (KPR 1998): for all dense instances of 2-CatSP, in which each node of B has a degree of a constant fraction of n.
- 0.56-approximation algorithm for the case $k = n/2$ by Dodis, Guruswami and Khanna (1999).

Our Results

A randomized approximation algorithm based on the semidefinite programming relaxation which has performance guarantee:

- $\frac{1}{2}$ for all $k \in [1, n]$;
- strictly better than $\frac{1}{2}$ when $k \geq \frac{n}{3}$;
- 0.67 when $k = \frac{n}{2}$.

Key Techniques

- A new integer programming formulation which allows us to obtain a SDP relaxation.

  —The main difficulty was that $A_1 \cap A_2 = \emptyset$ is not required.

  —The new IP model uses two-dimensional vector variables.

- A new variants of the Goemans-Williamson rounding technique which allows us to take the advantage of the underline structure of the problem:

  — the graph is bipartite;

  — the cardinality constraints are on the set of A only

A Special Case

- $k = \frac{n}{2}$ and it is required that $A_1 \cap A_2 = \emptyset$ (Disjoint Case).
- Each of the node in A is either in A_1 or A_2. For $i \in A$, $x_i = 1$ iff $i \in A_1$, otherwise $x_i = -1$; for $j \in B$, $x_j = 1$ iff $j \in B_1$.
- The problem can be formulated as

$$
\begin{array}{rll}
w_* := & \text{Maximize} & \dfrac{1}{2} \displaystyle\sum_{i \in A,\ \ j \in B} w_{ij}(1 + x_i x_j) \\
& \text{subject to} & \sum_{i \in A} x_i = 0, \\
& & x_i^2 = x_j^2 = 1 \quad i \in A, j \in B
\end{array}
\qquad (1)
$$

Relationship with Max-Bisection

- This special case is very similar to the well-known Max-Bisection problem:

$$
\begin{array}{lll}
w_* := & \text{Maximize} & \frac{1}{2} \sum_{i \in A,\ j \in B} w_{ij}(1 - x_i x_j) \\
& \text{subject to} & \sum_i x_i = 0, \\
& & x_i^2 = 1
\end{array}
\qquad (2)
$$

- Is fairly easy to solve since the known SDP based approximation algorithms can also applied to this special case of 2-Catalog problem (DGK 1999).

- 0.65-approximation by the result of Frieze and Jerrum (1995). 0.699-approximation by the results of Ye (1999) in running time $O(n^{3.5})$, and 0.701 by Halperin and Zwick (2001) in $O(n^{9.5})$.

A Harder Special Case

- $k = \frac{n}{2}$ but $A_1 \cap A_2 = \emptyset$ is not required (General Case).
- One dimensional binary variables are not sufficient to model this problem.
- DGK (1999) reduce this problem to the Disjoint Case, and show that the 0.651-approximation for the latter implies a 0.56-approximation for the General Case.
- If we use the 0.70-approximation for Disjoint case, we get a 0.58-approximation for the General Case.

General Case

- Each node i of A has four possible choices: $A1 \setminus A2$, $A2 \setminus A1$, $A1 \cap A2$ and $A \setminus (A1 \cup A2)$.
- One binary variable cannot characterize all choices of node i.
- We design a two-dimensional vector (x_i, y_i) for each node $i \in A$ such that $x_i = 1$ iff $i \in A_1$, otherwise $x_i = -1$; and $y_i = 1$ iff $i \in A_2$, otherwise $y_i = -1$. We assign variable z_j for each node $j \in B$ such that $z_j = 1$ iff $i \in B_1$, otherwise $z_j = -1$.

An Integer Programming Formulation

A quadratic formulation can be obtained by using these vector variables together with an additional reference variable u_0, which makes the objective function and constraints homogeneous:

$$w_* := \quad \text{Maximize} \quad \frac{1}{4} \sum_{i \in A,\ j \in B} w_{ij}(2 + u_0 x_i + u_0 y_i + x_i z_j - y_i z_j)$$

$$\text{subject to} \quad \begin{array}{l} \sum_{i \in A} u_0 x_i = \sum_{i \in A} u_0 y_i = 2k - n, \\ u_0^2 = x_i^2 = y_i^2 = z_j^2 = 1 \quad i \in A, j \in B \end{array} \tag{3}$$

SDP Relaxation

Let vector $u = (u_0;\ x \in R^n;\ y \in R^n;\ z \in R^m)$ and consider matrix $U = uu^T$. Then, it is straightforward to obtain an SDP relaxation of (3):

$$
\begin{aligned}
\text{Max} \quad & \sum_{1 \le i \le n,\ 2n+1 \le j \le 2n+m} w_{ij} \frac{(2 + U_{0i} + U_{0(n+i)} + U_{ij} - U_{(n+i)j})}{4} \\
\text{s.t.} \quad & \textstyle\sum_{i=1}^{n} U_{0i} = \sum_{i=n+1}^{2n} U_{0i} = 2k - n, \\
& \textstyle\sum_{1 \le i,j \le n} U_{ij} = \sum_{n+1 \le i,j \le 2n} U_{ij} = (2k-n)^2, \\
& U_{ii} = 1, \quad i = 0, 1, ..., 2n+m \\
& U \succeq 0.
\end{aligned}
\tag{4}
$$

A High-Level Algorithm

- Solve the SDP relaxation and obtain an optimal matrix $\bar{U}$ which can be write as

$$\bar{U} = \begin{pmatrix} \bar{U}_{00} & \bar{U}_{0x} & \bar{U}_{0y} & \bar{U}_{0z} \\ \bar{U}_{x0} & \bar{U}_{xx} & \bar{U}_{xy} & \bar{U}_{xz} \\ \bar{U}_{y0} & \bar{U}_{yx} & \bar{U}_{yy} & \bar{U}_{yz} \\ \bar{U}_{z0} & \bar{U}_{zx} & \bar{U}_{zy} & \bar{U}_{zz} \end{pmatrix}$$

  according to sub-blocks corresponding to u_0, x, y, and z.

- Use some (randomized) rounding technique to obtain a $\{-1, 1\}$ solution $\hat{x}, \hat{y}, z, u$ which produce $\hat{A}_1, \hat{A}_2, B_1, B_2$. (The solution may not be feasible, i.e, the cardinality constraints may not be satisfied.)

- Use a greedy adjusting procedure to get a feasible solution A_1, A_2, B_1, B_2 from the above solution.

Greedy Procedure

Let A_1, A_2 be the subsets of A with cardinality k such that $w(A_1, B_1)$ is maximized and $w(A_2, B_2)$ is maximized.

$$w(A_1, B_1) + w(A_2, B_2) \geq$$

$$\begin{cases} \frac{k}{|\hat{A}_1|} \cdot w(\hat{A}_1, B_1) + \frac{k}{|\hat{A}_2|} \cdot w(\hat{A}_2, B_2) & \text{if } |\hat{A}_i| \geq k, \ i = 1, 2; \\ w(\hat{A}_1, B_1) + \frac{k}{|\hat{A}_2|} \cdot w(\hat{A}_2, B_2) & \text{if } |\hat{A}_1| \leq k, \ |\hat{A}_2| \geq k; \\ \frac{k}{|\hat{A}_1|} \cdot w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2) & \text{if } |\hat{A}_1| \geq k, \ |\hat{A}_2| \leq k; \\ w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2) & \text{if } |\hat{A}_i| \leq k, \ i = 1, 2, \end{cases} \tag{5}$$

and

$$w(A_1, B_1) + w(A_2, B_2) \geq \max\{\frac{1}{2}, \frac{k}{n}\} w_*.$$

Analysis of the Algorithm

We want to bound the quantity of

$$w(A_1, B_1) + w(A_2, B_2).$$

Assume that the greedy adjusting procedure guarantees that

$$w(A_1, B_1)+w(A_2, B_2) \geq f(k/|\hat{A}_1|, k/|\hat{A}_2|) \left(w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2) \right).$$

Then, (roughly speaking), it is sufficient to bound

- The expected values of $|\hat{A}_1|, |\hat{A}_2|$
- The variance of $|\hat{A}_1|, |\hat{A}_2|$
- The expected value of $w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2)$.

Goemans-Williamson Rounding Technique

- Factorize the optimal matrix $\bar{U} = V^T V$ where $V = (v_0, V_x, V_y, V_z)$ is a $(2n + m + 1) \times (2n + m + 1)$ matrix.
- Randomly choose a unit vector u from S^{2n+m}.
- For each node i, if $u \cdot v_i$ has the same sign as $u \cdot v_0$, then it is 1; otherwise it is -1.
- This randomized rounding produces a solution which has an expected objective value close to the optimal one, and the expected number of nodes in $\hat{A}_1$ or $\hat{A}_2$ are close to k as desired.
- **The problem is that the variances of $|\hat{A}_1|$ or $|\hat{A}_2|$ could be very large.**

A Combined Rounding

- The identity matrix I can be used as the rounding matrix which has a better bound on the expected values and variances of $|\hat{A}_1|$ or $|\hat{A}_2|$, but a worse bound on the objective value. (For simplicity, we only consider $k = \frac{n}{2}$ here).
- We have shown that a careful combination of the optimal matrix $\bar{U}$ and the identity matrix I can improve the overall approximation ratio, i.e., to balance the bounds on the objective value and the sizes.
- We can apply this combined rounding technique to our problem.
- Can we do better by exploiting the structure of the problem?

Structure of the Problem

- Recall that our graph is bipartite and the cardinality constraints are only on the set A.
- For the nodes in A, we can apply combined rounding matrix.
- For the nodes in B, we do not want to use the combined rounding since it has a negative effect on the objective value.
- A new rounding matrix may take advantage of this fact. But note that it must be semi-definite.

New Rounding Matrix

We use the matrix $\bar{U}(\theta) + (1-\theta)P$ as the rounding matrix, where $\theta \in [0,1]$,

$$\bar{U}(\theta) = \begin{pmatrix} \bar{U}_{00} & \sqrt{\theta}\bar{U}_{0x} & \sqrt{\theta}\bar{U}_{0y} & \bar{U}_{0z} \\ \sqrt{\theta}\bar{U}_{x0} & \theta\bar{U}_{xx} & \theta\bar{U}_{xy} & \sqrt{\theta}\bar{U}_{xz} \\ \sqrt{\theta}\bar{U}_{y0} & \theta\bar{U}_{yx} & \theta\bar{U}_{yy} & \sqrt{\theta}\bar{U}_{yz} \\ \bar{U}_{z0} & \sqrt{\theta}\bar{U}_{zx} & \sqrt{\theta}\bar{U}_{zy} & \bar{U}_{zz} \end{pmatrix}$$

and

$$P = \begin{pmatrix} 0 & 0_{1\times n} & 0_{1\times n} & 0_{1\times m} \\ 0_{n\times 1} & I_{n\times n} & 0_{n\times n} & 0_{n\times m} \\ 0_{n\times 1} & 0_{n\times n} & I_{n\times n} & 0_{n\times m} \\ 0_{m\times 1} & 0_{m\times n} & 0_{m\times n} & 0_{m\times m} \end{pmatrix}$$

Analysis of the New Rounding

Recall that we want to bound

- The expected values of $|\hat{A}_1| = \frac{1}{2}\sum_{i\in A}(1+\hat{x}_i\hat{u}_0)$ and $|\hat{A}_2|$.
- The variance of $|\hat{A}_1|$ which can be treated as $|\hat{A}_1|(n-|\hat{A}_1|) = \frac{1}{4}\sum_{i,j\in A}(1-\hat{x}_i\hat{x}_j)$, and that of $|\hat{A}_2|$.
- The expected value of

$$w := w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2)$$

$$= \sum_{i\in A,\ j\in B} w_{ij}\frac{(2+\hat{x}_i\hat{u}_0+\hat{y}_i\hat{u}_0+\hat{x}_i\hat{z}_j-\hat{y}_i\hat{z}_j)}{4}.$$

Analysis of the New Rounding (Continued)

The following are straightforward from Goemans-Williamson and our rounding

$$\begin{aligned}
E[\hat{x}_i \hat{u}_0] &= \frac{2}{\pi} \arcsin(\sqrt{\theta}\bar{U}_{0i}), & i \in A, \\
E[\hat{y}_i \hat{u}_0] &= \frac{2}{\pi} \arcsin(\sqrt{\theta}\bar{U}_{0(m+i)}), & i \in A, \\
E[\hat{z}_j \hat{u}_0] &= \frac{2}{\pi} \arcsin(\bar{U}_{0j}), & j \in B, \\
E[\hat{x}_i \hat{z}_j] &= \frac{2}{\pi} \arcsin(\sqrt{\theta}\bar{U}_{ij}), & i \in A,\ j \in B, \\
E[\hat{y}_i \hat{z}_j] &= \frac{2}{\pi} \arcsin(\sqrt{\theta}\bar{U}_{(m+i)j}), & i \in A,\ j \in B.
\end{aligned}$$

These equations are enough for us for bounding the expected objective value and the sizes.

$$z(\eta,\gamma) := \frac{w}{w_{SDP}} + \gamma\frac{M_1+M_2}{n^2} + \eta\gamma\frac{p_1+p_2}{n}, \qquad (6)$$

where

$$w := w(\hat{A}_1, B_1) + w(\hat{A}_2, B_2)$$

$$p_1 := |\hat{A}_1|,$$

$$p_2 := |\hat{A}_2|,$$

$$M_1 := |\hat{A}_1|(n - |\hat{A}_1|),$$

$$M_2 := |\hat{A}_2|(n - |\hat{A}_2|).$$

Our approximation method yields the partitions $(\hat{A}_1, \hat{A}_2)$ and (B_1, B_2), satisfying the following two inequalities:

$$E[\frac{w}{w_{SDP}}] \geq \alpha,$$

$$E[\frac{p_i}{n}] \geq \beta/2, \quad i = 1, 2,$$

$$E[\frac{M_i}{n^2}] \geq \beta/4 \quad i = 1, 2.$$

$$\mathrm{E}[z(\eta, \gamma)] \geq \alpha + \gamma\beta/2 + \eta\gamma\beta \quad \text{and} \quad z(\eta, \gamma) \leq 1 + \frac{\gamma(1+\eta)^2}{2}.$$

If the random variable $z(\eta, \gamma)$ meets its expectation, then

$$w(A_1, B_1) + w(A_2, B_2) \geq R(\sigma, \theta, \eta, \gamma) \cdot w_*.$$

σ	$\sqrt{\theta}$	α	β	η	γ	R
0.37	0.94	0.7450	0.9115	-0.4103	2.7106	0.5045
0.38	0.94	0.7457	0.9193	-0.3890	2.7648	0.5148
0.40	0.93	0.7386	0.9371	-0.3387	3.0095	0.5345
0.42	0.93	0.7401	0.9481	-0.2976	3.1298	0.5534
0.44	0.93	0.7416	0.9566	-0.2578	3.2408	0.5711
0.46	0.92	0.7352	0.9661	-0.2121	3.5163	0.5876
0.48	0.92	0.7369	0.9697	-0.1749	3.6074	0.6029
0.50	0.92	0.7386	0.9709	-0.1388	3.6719	0.6169

Table 1

Facility Location

In the uncapacitated facility location problem (UFLP), we have

- A set $\mathcal{F}$ of n_f *facilities*, where for every facility $i \in \mathcal{F}$, a nonnegative number f_i is given as the *opening cost* of facility i.
- A set $\mathcal{C}$ of n_c *cities*, where for every city $j \in \mathcal{C}$ and facility $i \in \mathcal{F}$, we have a *connection cost* (a.k.a. service cost) c_{ij} between city j and facility i.
- The objective is to open a subset of the facilities in $\mathcal{F}$, and connect each city to an open facility so that the total cost is minimized.
- We will consider the *metric* version of this problem, i.e., the connection costs satisfy the triangle inequality.

Approximation Results

approx. factor	reference	technique/running time
$O(\ln n_c)$	Hochbaum	greedy algorithm/$O(n^3)$
3.16	Shmoys, Tardos, Aardal et al.	LP rounding
2.41	Guha and Khuller	LP rounding + greedy
1.736	Chudak	LP rounding
$5+\epsilon$	Korupolu, Plaxton, Rajaraman et al.	local search/$O(n^6 \log(n/\epsilon))$
3	Jain and Vazirani	primal-dual method/$O(n^2 \log n)$
1.853	Charikar and Guha	primal-dual method + greedy/$O(n^3)$
1.728	Charikar and Guha	LP rounding + primal-dual method + greedy
1.861	Mahdian et al.	greedy algorithm/$O(n^2 \log n)$
1.61	Jain et al.	greedy algorithm/$O(n^3)$
1.582	Sviridenko	LP rounding
1.517	MYZ	greedy algorithm + greedy/$O(n^3)$

Table 1: Approximation Algorithms for UFLP

Hardness Results

Guha and Khuller proved that it is impossible to get an approximation guarantee of 1.463 for the uncapacitated metric facility location problem, unless NP $\subseteq \mathrm{DTIME}[n^{O(\log \log n)}]$.

Cost-Splitting Approximation

An algorithm is called a (γ_f, γ_c)-approximation algorithm for UFLP, if for every instance $\mathcal{I}$ of UFLP, and for every solution SOL for $\mathcal{I}$ with facility cost F_{SOL} and connection cost C_{SOL}, the cost of the solution found by the algorithm is at most $\gamma_f F_{SOL} + \gamma_c C_{SOL}$.

Let $\gamma_f \geq 1$. Then $\gamma_c \leq \sup_k\{z_k\}$, where z_k

$$
\begin{array}{ll}
\text{max} & \frac{\sum_{i=1}^{k} \alpha_i - \gamma_f f}{\sum_{i=1}^{k} d_i} \\
\text{s.t.} & \forall\, 1 \leq i < k: \ \alpha_i \leq \alpha_{i+1} \\
& \forall\, 1 \leq j < i < k: \ r_{j,i} \geq r_{j,i+1} \\
& \forall\, 1 \leq j < i \leq k: \ \alpha_i \leq r_{j,i} + d_i + d_j \\
& \forall\, 1 \leq i \leq k: \ \sum_{j=1}^{i-1} \max(r_{j,i} - d_j, 0) + \sum_{j=i}^{k} \max(\alpha_i - d_j, 0) \leq f \\
& \forall\, 1 \leq j \leq i \leq k: \ \alpha_j, d_j, f, r_{j,i} \geq 0.
\end{array}
$$

Conservative Opening Factor δ

Theorem 1 *If there is a*

$$(\gamma_f, \gamma_c)$$

approximation algorithm for UFLP, then there is a

$$(\gamma_f + \ln \delta, 1 + \frac{\gamma_c - 1}{\delta})$$

approximation for UFLP.

Prove $(\gamma_f, \gamma_c) = (1.104, 1.7805)$ and select $\delta = 1.5107$.

Extended Results

Problem	Previous ratio	Our ratio	Reduced Problem
SFLP	3	2	LFLP
K-FLP	3		LP
	4.83	3.27	UFLP
2-FLP	4.83	2.17	UFLP
3-FLP	4.83	2.17	UFLP
SMFLP	9	3.98	UFLP
CCFLP	4.59	3.97	UFLP
ANDP	81	34	LBFLP

Table 2: Approximation Algorithms for UFLP

The Markovian Decision Problem

$$
\begin{array}{lccccc}
\text{minimize} & \mathbf{c}_1^T\mathbf{x}_1 & \ldots & +\mathbf{c}_m^T\mathbf{x}_m & & \\
\text{subject to} & (I-\theta P_1)\mathbf{x}_1 & \ldots & +(I-\theta P_m)\mathbf{x}_m & = & \mathbf{e}, \\
 & \mathbf{x}_1 & \ldots & \mathbf{x}_m & \geq & \mathbf{0}.
\end{array}
$$

Column stochastic matrix:

$$\mathbf{e}^T P_i = \mathbf{e}^T, \quad i = 1, ..., m$$

Discount factor:

$$0 \leq \theta \leq 1$$

Complexity Results

Value-Iteration	Policy-Iteration	LP-Algorithm	Combinatorial Interior-Point
$O(n^2 \cdot \frac{L}{1-\theta})$	$O(n^3 \cdot \frac{2^n}{n})$	$O(n^3 L)$	

Complexity Results

Value-Iteration	Policy-Iteration	LP-Algorithm	Combinatorial Interior-Point
$O(n^2 \cdot \frac{L}{1-\theta})$	$O(n^3 \cdot \frac{2^n}{n})$	$O(n^3 L)$	$O(n^3 \cdot n \cdot \ln \frac{1}{1-\theta})$