

Polyhedral Study of the K-Median Problem

Sang-Hyung Ahn

I. Introduction

The past two decades have witnessed a tremendous growth in the literature on location problems. However, among the myriads of formulations the simple plant location problem and the k -median problem have played a central role. This phenomenon is due to the facts that both problems have a wide range of real-world applications and a mathematical formulation of these problems as an integer program has proven very fruitful in the derivation of solution methods.

Consider an index set $I = \{1, 2, \dots, n\}$ of n points, and a positive integer $k \leq n$, and let c_{ij} be the shortest distance between two points $i, j \in I$. The k -median problem consists of identifying a subset S of I , $|S| = k$ so as to minimize $(\sum_{i=1}^n \text{Min}_{j \in S} c_{ij})$. (Here $|S|$ denotes the cardinality of the set S). The k -median problem has the following combinatorial formulation.

Combinatorial Formulation:

$$\text{Min} \left\{ \sum_{i \in I} \text{Min}_{j \in S} c_{ij} \right\}$$

$$\begin{matrix} S \subseteq I \\ |S| = k \end{matrix}$$

We introduce integer variables. Let $y_j = 1$ if a point j is selected as a median, otherwise 0 and $x_{ij} = 1$ if a point j is the closest median to point i , otherwise 0. With x, y variables the k -median problem is formulated as an integer program as follows.

Integer Program Formulation:

$$Z_{IP} = \text{Min} \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \tag{1}$$

subject to

$$\sum_{j=1}^n x_{ij} = 1 \quad i \in I \quad (2)$$

$$\sum_{j=1}^n y_j = k \quad (3)$$

$$0 \leq x_{ij} \leq y_j \leq 1 \quad i, j \in I \quad (4)$$

$$x_{ij}, y_j \text{ integral } i, j \in I \quad (5)$$

A vast number of algorithms were proposed for the k -median problem. We refer readers to ReVelle [19], Francis and White [14], Christofides [7], Jacobsen and Pruzan [16], Handler and Mirchandani [15], Krarup and Pruzan [17], Cornuejols [9] [11] [12], Fisher and Hochbaum [13], Papadimitriou [18], Rosing [20], Beasley and Christofides [8], Boffey [5], Beasley [4].

Most of the successful algorithms for the k -median problem are based on the strong linear programming relaxation. In [1] [2] [3] we presented and explained why the strong linear programming relaxation provides a tight lower bound in the probabilistic sense. In this paper we investigate the phenomenon with a polyhedral approach.

II. Polyhedral Analysis

In this section we investigate the polytope of the extreme solutions to the strong linear program relaxation of k -median problem constraints.

$$\sum_{j=1}^n x_{ij} = 1 \quad i \in I \quad (6)$$

$$\sum_{j=1}^n y_j = k \quad (7)$$

$$x_{ij} \leq y_j \quad i, j \in I \quad (8)$$

$$y_j \leq 1 \quad j \in I \quad (9)$$

$$x_{ij}, y_j \geq 0 \quad i, j \in I \quad (10)$$

Let P_n be the polytope defined by (6)~(10). We present properties of the fractional extreme points (x, y) to P_n below.

Lemma 1:

If (x, y) is a fractional extreme point of the polytope P_n , then for each $i \in I$, there is at most one $j \in I$ with $0 < x_{ij} < y_j$.

Proof:

Let (x, y) be a fractional solution such that above condition does not hold.

Then there exist p, j_1, j_2 such that $x_{pj_1} < y_{j_1}$, $x_{pj_2} < y_{j_2}$.

Let $x_{pj_1}^1 = x_{pj_1} + \epsilon$, $x_{pj_2}^1 = x_{pj_2} - \epsilon$, $x_{pj_1}^2 = x_{pj_1} - \epsilon$, $x_{pj_2}^2 = x_{pj_2} + \epsilon$,

$x_{ij}^1 = x_{ij}^2 = x_{ij}$ for all other i, j and $y_j^1 = y_j^2 = y_j$ for all j .

where $\epsilon = \text{Min} [x_{pj_1}, x_{pj_2}, y_{j_1} - x_{pj_1}, y_{j_2} - x_{pj_2}]$.

Then $(x, y) = (1/2)(x^1, y^1) + (1/2)(x^2, y^2)$. (x^1, y^1) and (x^2, y^2) both are feasible solutions to P_n . This contradicts the assumption that (x, y) is an extreme solution. //

A similar result is known for the simple plant location problem. In fact, Cornuejols et al [10] completely characterized the fractional extreme solutions to the simple plant location problem.

Suppose we are given the shortest distance matrix between all pairs of points. The optimal solution to the 1-median problem is reduced to find a column of the shortest distance matrix with smallest column sum. In fact, when $k=1$, all the extreme solutions to the polytope P_n are integral regardless of n . We show this in the next theorem.

Theorem 2:

The linear programming relaxation of the 1-median problem always has an integer optimal solution.

Proof:

Suppose there exists a fractional extreme solution to the linear programming relaxation of 1-median problem.

Let $J_1 = \{j \in I : 0 < y_j < 1\}$

Then for each $i \in I$, $x_{ij} = y_j$ for all $j \in J_1$ and $x_{ij} = 0$ for all $j \notin J_1$.

Choose any two points $j_1, j_2 \in J_1$, any $p \in I$ and let

$x_{pj_1}^1 = x_{pj_1} + \epsilon$, $x_{pj_2}^1 = x_{pj_2} - \epsilon$, $y_{j_1}^1 = y_{j_1} + \epsilon$, $y_{j_2}^1 = y_{j_2} - \epsilon$

$$x_{pj_1}^2 = x_{pj_1} - \epsilon, \quad x_{pj_2}^2 = x_{pj_2} + \epsilon, \quad y_{j_1}^2 = y_{j_1} - \epsilon, \quad y_{j_2}^2 = y_{j_2} + \epsilon$$

All other x_{ij}, y_j remain unchanged.

Then $(x, y) = \frac{1}{2}(x^1, y^1) + \frac{1}{2}(x^2, y^2)$ and $(x^1, y^1), (x^2, y^2)$ both are feasible solutions to P_n . This contradicts the assumption that (x, y) is an extreme solution. //

In the next theorem, we extend the above result to more general cases.

Theorem 3:

If (x, y) is an extreme solution to the polytope P_n , then $\sum_{j \in J_1} y_j \geq 2$.

Proof:

Suppose (x, y) is an extreme point to P_n with $\sum_{j \in J_1} y_j = 1$.

Let $I_1 = \{i \in I : 0 < x_{ij} < 1 \text{ for some } j \in J_1\}$, $J_2 = \{j \in J : y_j = 1\}$.

We have two cases to consider here.

Case 1: for all $i \in I_1$, $\sum_{j \in J_1} x_{ij} = 1$. (That is, $x_{ij} = 0$ for all $j \in J_2$).

For this case we can derive contradiction in the same way as for Theorem 2.

Case 2: for some $i \in I_1$, $\sum_{j \in J_1} x_{ij} \neq 1$.

Here we have two subcases.

Case 2-1: for all $i \in I_1$, $x_{ij} = y_j$ for only one $j \in J_1$. That is, $x_{ip} = 1 - y_p$ for only one $p \in J_2$ due to Lemma 1.

Case 2-2: for some $i \in I_1$, $x_{ij} = y_j$ for several $j \in J_1$.

Proof:

Choose any two $j_1, j_2 \in J_1$.

Let $I_3 = \{i \in I_1 : x_{ij_1} = y_{j_1}\}$,

$I_4 = \{i \in I_1 : x_{ij_2} = y_{j_2}\}$,

$J_3 = \{j \in J_2 : x_{ij} = 1 - y_{j_1} \text{ and } i \in I_3\}$

$J_4 = \{j \in J_2 : x_{ij} = 1 - y_{j_2} \text{ and } i \in I_4\}$

For case 2-1:

Then we can construct two feasible solutions (x_1, y_1) and (x_2, y_2) as follows.

$$y_{j_1}^1 = y_{j_1} + \epsilon, \quad y_{j_2}^1 = y_{j_2} - \epsilon, \quad y_{j_1}^2 = y_{j_1} - \epsilon, \quad y_{j_2}^2 = y_{j_2} + \epsilon,$$

$$y_j^1 = y_j^2 = y_j \text{ for all } j \neq j_1 \text{ or } j_2.$$

Table 2 : Case 2-1 of Theorem 3

Case 2-1:

(y_j Matrix & x_{ij} Matrix)

	1/3	1/3	1/3	0	1	0	1	$\leftarrow y_j$
	1	2	3	4	5	6	7	
1	1/3	—	—	—	2/3	—	—	
2	—	1/3	—	—	2/3	—	—	$\leftarrow x_{ij}$
3	—	—	1/3	—	—	—	2/3	$J_1 = \{1, 2, 3\}$
4	1/3	—	—	—	—	—	2/3	$J_2 = \{5, 7\}$
5	—	—	—	—	1	—	—	$I_3 = \{1, 4\}$
6	—	—	1/3	—	2/3	—	—	$I_4 = \{2\}$
7	—	—	—	—	—	—	1	$\epsilon = 1/3$

Table 3 : Case 2-2 of Theorem 3

Case 2-2:

(y_j Matrix & x_{ij} Matrix)

	1/3	1/3	1/3	0	1	0	1	0	$\leftarrow y_j$
	1	2	3	4	5	6	7	8	
1	1/3	1/3	—	—	1/3	—	—	—	$\leftarrow x_{ij}$
2	—	1/3	—	—	2/3	—	—	—	$J_1 = \{1, 2, 3\}$
3	—	—	1/3	—	—	—	2/3	—	$J_2 = \{5, 7\}$
4	1/3	—	1/3	—	—	—	1/3	—	$I_3 = \{1, 4\}$
5	—	—	—	—	1	—	—	—	$I_4 = \{2, 8\}$
6	1/3	1/3	1/3	—	—	—	—	—	$i^* = \{6\}$
7	—	—	—	—	—	—	1	—	$\epsilon = 1/3$
8	—	1/3	1/3	—	—	—	1/3	—	

been shown by Cho, Padberg, and Rao [6], Krarup and Pruzan [17] that all the extreme points of $Q_{m,n}$ are integral. The constraint matrix, in fact, is totally unimodular in this case. However, for values as small as $m=n=3$, $Q_{m,n}$ has fractional extreme points. For example, when $c_{13}=c_{21}=c_{32}=1$, all other $c_{ij}=0$ and $f_j=1$ for $j=1, 2, 3$, the unique optimal solution of minimizing $\sum_{i=1}^n \sum_{j=1}^m c_{ij}x_{ij} + \sum_{j=1}^m f_j y_j$ is $y_j=1/2$ for $j=1, 2, 3$ and $x_{11}=x_{12}=x_{22}=x_{23}=x_{31}=x_{33}=1/2$, all other $x_{ij}=0$. The value of this fractional solution is 1.5.

Here we first provide a similar result about P_n and extend it by one more dimension.

Proposition 4:

If (x,y) is an extreme point of P_n , then $|J_1| \geq 3$.

Proof:

Immediate consequence of Theorem 3.

That is, $|J_1|=2$ means $\sum_{j \in J_1} y_j = 1$ and this directly contradicts Theorem 3. //

A direct consequence of the above proposition is that when $n \leq 2$, the k -median problem always has an integer optimal solution. In fact the constraint matrix of the k -median problem is totally unimodular when $n \leq 2$.

Now we extend above results to the case when $n \leq 3$.

Theorem 5:

If (x,y) is an extreme point of P_n , then $|J_1| \geq 4$.

Proof:

Assume (x,y) is a fractional extreme solution to P_n with $|J_1|=3$. Then we must have $\sum_{j \in J_1} y_j \neq 2$. For the case that $\sum_{j \in J_1} y_j = 1$ is eliminated due to theorem 3. Let j_1, j_2, j_3 be the index such that $0 < y_{j_1} < y_{j_2} < y_{j_3} < 1$. We should examine 2 cases.

Case 1: For all $i \in I_1$, $\sum_{j \in J_1} x_{ij} = 1$. That is, $x_{ij} = 0$ for all $j \in J_2$. Note that for each $i \in I_1$, exactly two $x_{ij} \neq 0$ because the sum of any three $y_j, j \in J_1$ is larger than 1.

Case 2: For some $i \in I_1$, $\sum_{j \in J_1} x_{ij} \neq 1$.

Here we have two subcases.

Note that for each $i \in I_1$, exactly two $x_{ij} \neq 0$ due to Lemma 1.

(Case 2-1) For all $i \in I_1$, $x_{ij} = y_j$ for only one $j \in J_1$.

(Case 2-2) For some $i \in I_1$, $x_{ij} = y_j$ for two $j \in J_1$.

For case 1:

Let $J_1 = \{j_1, j_2, j_3\}$ and $\epsilon = \text{Min} [x_{ij}, i \in I_1, \& j \in J_1]$.

Let $I_5 = \{i \in I_1 : 0 < x_{ij_1}, x_{ij_2} < 1\}$

$I_6 = \{i \in I_1 : 0 < x_{ij_1}, x_{ij_3} < 1\}$

$I_7 = \{i \in I_1 : 0 < x_{ij_2}, x_{ij_3} < 1\}$

$I_8 = \{i \in I_1 : x_{ij} = \epsilon\}$

Table 5 : Case 2-1 of Theorem 5

Case 2-1:

(y_j Matrix & x_{ij} Matrix)

	7/12	8/12	9/12	0	1	0	1	$\leftarrow y_j$
	1	2	3	4	5	6	7	
1	7/12	—	—	—	5/12	—	—	$\leftarrow x_{ij}$
2	—	8/12	—	—	4/12	—	—	$j_1=1, j_2=2$
3	—	—	9/12	—	—	—	3/12	$I_3=\{1, 4\}$
4	7/12	—	—	—	—	—	7/12	$I_4=\{2, 6\}$
5	—	—	—	—	1	—	—	$J_3=\{5, 7\}$
6	—	8/12	—	—	—	—	4/12	$J_4=\{5, 7\}$
7	—	—	—	—	—	—	1	$\epsilon=4/12$

Table 6 : Case 2-2 of Theorem 5

Case 2-2:

(y_j Matrix & x_{ij} Matrix)

	7/12	8/12	9/12	0	1	0	1	0	$\leftarrow y_j$
	1	2	3	4	5	6	7	8	
1	7/12	—	—	—	5/12	—	—	—	$\leftarrow x_{ij}$
2	—	8/12	—	—	4/12	—	—	—	
3	—	—	9/12	—	—	—	3/12	—	$I_3=\{1\}$
4	4/12	8/12	—	—	—	—	—	—	$I_4=\{2\}$
5	—	—	—	—	1	—	—	—	$I_6=\{4\}$
6	4/12	—	8/12	—	—	—	—	—	$I_6=\{6\}$
7	—	—	—	—	—	—	1	—	$I_7=\{8\}$
8	3/12	9/12	—	—	—	—	—	—	$I_8=\{3\}$

$$j_1=1, j_2=2.$$

$$J_3=J_4=\{5\}$$

$$\epsilon=3/12.$$

$$x_{i_j}^2 = x_{i_j} - \epsilon, \quad x_{i_j}^1 = x_{i_j} + \epsilon \text{ for all } i \in I_3 \text{ \& } j \in J_3.$$

$$x_{i_j}^2 = x_{i_j} + \epsilon, \quad x_{i_j}^1 = x_{i_j} - \epsilon \text{ for all } i \in I_4 \text{ \& } j \in J_4.$$

$$x_{i_j}^1 = x_{i_j}^2 = x_{i_j} \text{ for all } i \neq I_3 \text{ or } I_4, \quad j \neq J_3 \text{ or } J_4.$$

Expression of $(x, y) = (1/2)(x^1, y^1) + (1/2)(x^2, y^2)$ means contradiction.

For case 2-2:

We can think of case 2-2 as a composite of case 1 and case 2-1, and we derive a contradiction as we did for case 1 and case 2-1. //

Corollary 6:

The k -median problem of $n \leq 3$ always has an integer optimal solution.

Proof: Immediate consequence of Theorem 5. //

References

- [1] 安相炯, Probabilistic Analysis of a Relaxation for the k -median Problem—The Euclidean Model in the Plane, 경영논집 XX 제 3 호, 1986년 9월 46-62.
- [2] 安相炯, Probabilistic Analysis of a Relaxation for the k -Median Problem—A Graphical Model, 경영논집 XXI 제 3 호, 1987년 9월 1-15.
- [3] Ahn, Sang H., Cooper, C., Cornuejols, G. and Frieze, A., "Probabilistic Analysis of a Relaxation for the k -Median Problem," *Math. of O.R.* (1988) vol. 13, No. 1, 1-31.
- [4] Beasley, J.E., A note on solving large p -median problems. technical report SW7 2BX, Imperial College, September, 1984.
- [5] Boffey, T.B. and Karkazis, J., "P-median and multi-medians", *J. Oper. Res. Society* 35 (1984), 57-64.
- [6] Cho, D.C., Padberg, M.W. and Rao, M.R., "On the Uncapacitated Plant Location Problem I: Valid inequality and Facets", *Math. Oper.Res.* 1(1983), 90-100.
- [7] Christofides, N., "*Graph Theory: An Algorithmic Approach*", Academic Press, New York, 1975.
- [8] Christofides, N. and Beasley, J.E., "A tree search algorithm for the p -median problem", *European J. of Oper. Res.* 10 (1982), 196-204.
- [9] Cornuejols, G., Fisher, M.L. and Nemhauser, G.L., "Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms", *Management Sci.* 23 (1977), 789-810.
- [10] Cornuejols, G., Fisher, M.L. and Nemhauser, G.L., "On the uncapacitated location problem", *Ann. Discrete Math.* 1 (1977), 163-177.
- [11] Cornuejols, G., Nemhauser, G.L. and Wolesey, L.A., "Worst-case and probabilistic analysis of algorithms for a location problem", *Operations Res.* 28 (1980), 847-858.
- [12] Cornuejols, G. and J.M. Thizy, "Some Facets of the Simple Plant Location Problem". *SIAM J. Algebraic and Discrete Mathematics* 3 (1982), 504-510.
- [13] Fisher, M.L. and Hochbaum, D.S., "Probabilistic analysis of the planar k -median problem", *Math. of Operations Research* 5, 1 (Feb. 1980), 27-34.

- [14] Francis, R.L. and White, J.A., "*Facility Layout and Location: An Analytic Approach*", Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [15] Handler, G.Y. Mirchandani, P.B., "*Location on Networks: Theory and Algorithms*", MIT-press, Cambridge, MA., 1979.
- [16] Jacobsen, S.K. and Pruzan, P.M., "*Lokalisering-modeller & Losningsmetoder*", Studentlitteratur, Lund, Sweden, 1978.
- [17] Krarup, J. and Pruzan, P.M., "The simple plant location problem: Survey and synthesis", *European Jour. of O.R.* 12 (1983), 36-81.
- [18] Papadimitriou, C., "Worst-case and probabilistic analysis of a geometric location problem", *SIAM Jour. Computing* 10 3 (1981), 542-557.
- [19] ReVelle, C.S., Swain, R.S., "Central facilities location", *Geographical Anal.* 2 (1970), 30-42.
- [20] Rosing, K.E., ReVelle, C.S. and Rosing-Vogelaar, "The p-median and its linear programming relaxation: an approach to large problems", *J. Operational Res. Soc.* 30 (1979), 815-823.