



Minimizing the Average Distance of Separated Points on the Plane in the L^1 -Distance

Jae-Hoon Kim*, Member, KIICE

Division of Computer Engineering, Pusan University of Foreign Studies, Busan 604-168, Korea

Abstract

Given separated points divided by a line, called a wall, in a plane, we aim to make a gate in the wall to connect the separated points to each other. In this setting, the problem is to find a location for the gate that minimizes the average distance between the points. The problem is a variant of the well-known facility location problem, which is extensively studied in the fields of operations research, location theory, theoretical computer science, and so on. In this paper, we consider the L^1 -distance of the points in the plane. The points are projected onto the wall and so the problem is transformed to a proximity problem of points on a line. Then it is shown that the transformed problem is related to the weighted median problem of points on the line. Therefore, we obtain an $O(n \log n)$ -time algorithm to solve our problem.

Index Terms: Average distance, L^1 -distance, Facility location, Weighted median

I. INTRODUCTION

Let V and W be two sets of separated points in the plane, specifically, the *black* points and the *white* points, respectively. If $|V| = n$ and $|W| = m$, then the *average distance* between V and W , denoted by $\mu(V, W)$, is defined to be the average distance of p and q for all point pairs $(p, q) \in V \times W$:

$$\mu(V, W) = \frac{1}{nm} \sum_{p \in V, q \in W} d(p, q),$$

where $d(p, q)$ denotes the distance of p and q in the plane, in particular, we will only consider the L^1 -distance of the plane in this paper.

In this paper, we consider the case in which V and W are completely separated by a *wall*. This is inspired by the situation where the Palestinian area and the Israeli area are separated by a wall. We cannot go through the wall from a black point to a white point. Thus we wish to make a gate in

the wall in order to connect black points and white points. Then, to go from a black point to a white point, we should go from the black point to the gate and then from the gate to the white point. Therefore, for a gate g , the distance $d(p, q)$ between a black point p and a white point q in this environment is given by

$$d(p, q) = d(p, g) + d(g, q).$$

In this paper, we will investigate a location problem in this environment, which is about where to make a gate in the wall. In particular, we are concerned in locating the gate to minimize the average distance between V and W , which means constructing the gate to make mutual visits between the two separated areas easy.

This is closely related to the well-known *facility location* problem, in which a company wants to open up a number of facilities to serve their customers. In the problem, both the opening of a facility at a specific location and the service of

Received 09 November 2011, Revised 12 December 2011, Accepted 20 December 2011

*Corresponding Author E-mail: jhoon@pufs.ac.kr

Open Access <http://dx.doi.org/10.6109/jicce.2012.10.1.001>

print ISSN:2234-8255 online ISSN:2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

a customer by a facility incurs some cost. The goal is to minimize the overall cost of opening enough facilities to serve all the customers.

There are four primary problems in facility location: the p -median problem, the p -center problem, the uncapacitated facility location problem (UFLP) and the quadratic assignment problem (QAP). These problems are specific variants of the general facility location problem.

Our problem may be considered one of a number of variants of the facility location problem. It corresponds to the case in which there is no cost to open a facility, say a gate, and the service costs are the distances between the pairs of black and white points.

In this paper, we will obtain an efficient algorithm to find a gate in the wall minimizing the average distance between black points and white points.

II. RELATED WORK

The problem of determining the average distance is mostly concentrated on the graph. One method of computing the average distance of a graph with n nodes is simply to compute the distances between all pairs of vertices, which is performed in time $O(n^2)$. In [1], the author asked whether the average distance of a graph is easier to compute than all distances between vertices of the graph. An affirmative answer is given in [2], which presents an algorithm that computes the average distance of an interval graph with m edges of unit weight in time $O(m)$. The author provides an algorithm that computes the average distance of a tree with n nodes and p edges in time $O(n)$.

Our problem is related to variants of the facility location problem in the Euclidean space: the median problem and the center problem. In the median problem, the sum of distances from the clients to their nearest facilities is minimized, whereas the maximum distance between the clients and their nearest facilities is minimized in the center problem.

For the 1-median problem, it is well-known that there is a linear time algorithm in \mathbb{R} . But solving for the exact location of the Euclidean 1-median in two or more dimensions is difficult in general. Indeed, no polynomial-time algorithm is known, nor has the problem been shown to be non-deterministic polynomial-time hard (NP-hard) [3]. There is a randomized algorithm with its running time linear in n and polynomial in $1/e$ [4, 5].

For an arbitrary p , in \mathbb{R} , the p -median can be solved exactly in $O(pn)$ time [6]. However, it is NP-hard even in \mathbb{R}^2 [7]. Arora et al. [8] provide an $O(n^{O(1+1/e)})$ -time e -approximation algorithm.

In the case of the 1-center problem, there is a

deterministic $O(n)$ -time algorithm in \mathbb{R}^2 [9]. This result has been extended to \mathbb{R}^d for any fixed d in $O(d^{O(d)}n)$ time by [10, 11]. The 2-center problem has been solved by a number of works [12-17]. The best recent solution is an $O(n \log^2 n \log^2 \log n)$ -time algorithm given in [18]. Agarwal and Sharir [19] provide a generalization of Drezner's algorithm from \mathbb{R}^2 to \mathbb{R}^d to give an algorithm requiring $O(n^{d+1})$ time.

The Euclidean k -center problem can be solved in linear time in \mathbb{R} using the algorithm of [20] for finding the k -centers in a tree. The problem is shown to be NP-hard in \mathbb{R}^2 [7]. There are many approximation algorithms for the Euclidean k -center problem [21-24].

The problem that we will deal with is one of various variants of the facility location problem and it is first investigated in this paper. We will propose an efficient algorithm running in linear time.

III. ALGORITHMS

Let V and W be the sets of black points and white points, respectively, where $|V| = n$ and $|W| = m$. There is a wall located between V and W and dividing the black points and the white points.

We consider a gate g , a point lying on the wall. Then the distance $d(p, q)$ between $p \in V$ and $q \in W$ is given as $d(p, q) = d(p, g) + d(g, q)$. To compute the distance easily, we will apply the projections of points in $V \cup W$ onto the wall as shown in Fig. 1. Let v be a point in $V \cup W$. Then the projection of v onto the wall is denoted by \bar{v} . Thus the distance from v to the gate g can be given by

$$d(v, g) = d(v, \bar{v}) + d(\bar{v}, g).$$

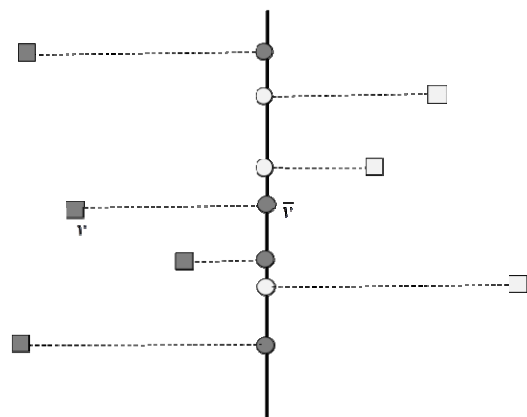


Fig. 1. The projections of points onto the wall.

Here both the point \bar{v} and the gate g are located on the wall. Also note that $d(v, \bar{v})$ is uniquely determined regardless of the location of the gate g .

For $p \in V$ and $q \in W$, the distance $d(p, q)$ is given as

$$\begin{aligned} d(p, q) &= d(p, g) + d(g, q) \\ &= d(p, \bar{p}) + d(\bar{p}, g) + d(g, \bar{q}) + d(\bar{q}, q). \end{aligned}$$

Thus, for any black point $p \in V$, the sum of distances between it and all white points $q \in W$ is obtained by

$$\begin{aligned} \sum_{q \in W} d(p, q) &= m(d(p, \bar{p}) + d(\bar{p}, g)) \\ &\quad + \sum_{q \in W} (d(g, \bar{q}) + d(\bar{q}, q)). \end{aligned}$$

Thus the sum of distances of all pairs of $p \in V$ and $q \in W$ is

$$\begin{aligned} \sum_{p \in V} \sum_{q \in W} d(p, q) &= m \sum_{p \in V} d(p, \bar{p}) + m \sum_{p \in V} d(\bar{p}, g) \\ &\quad + n \sum_{q \in W} d(g, \bar{q}) + n \sum_{q \in W} d(\bar{q}, q). \end{aligned}$$

Let $P = \sum_{p \in V} d(p, \bar{p})$, $Q = \sum_{q \in W} d(\bar{q}, q)$ and $O = m \sum_{p \in V} d(\bar{p}, g) + n \sum_{q \in W} d(g, \bar{q})$. Then both P and Q are fixedly determined. Therefore, the goal is transformed to minimizing the function O . That is, we derive a one-dimensional equivalent problem that finds a gate g to minimize the function O for $(n + m)$ points given on the line. From this point on, we will solve the equivalent problem.

Given points $\bar{p} \in V$ and $\bar{q} \in W$ on the line where $|V| = n$ and $|W| = m$, we want to obtain a (gate) point g that minimizes the objective function O given by

$$O = m \sum_{\bar{p} \in V} d(\bar{p}, g) + n \sum_{\bar{q} \in W} d(g, \bar{q}).$$

Here, we will provide an algorithm \mathcal{A} to solve the above problem. We consider the points $\bar{p} \in V$ and $\bar{q} \in W$ as points p_i with weights w_i on the line. If $p_i \in V$, then $w_i = \frac{1}{2n}$. If $p_i \in W$, then $w_i = \frac{1}{2m}$. Then the objective $O' = \sum_i w_i d(p_i, g)$ is the same as $\frac{1}{2mn} O$. Thus the optimal solution g also minimizes O' . We will show that the weighted median of points p_i minimizes the objective O' .

The weighted median of points p_i is defined to be p_k , one of the points satisfying

$$\sum_{p_i < p_k} w_i \leq \frac{1}{2} \text{ and } \sum_{p_i > p_k} w_i \leq \frac{1}{2},$$

where $\sum_i w_i = 1$.

Thus we consider an algorithm \mathcal{A} to find the weighted

median of points p_i . In \mathcal{A} , first sort the points p_i in non-decreasing order. Then sum up the weights of points until we have found the median, that is, until the sum is more than or equal to $\frac{1}{2}$. Thus \mathcal{A} is run in $O(n \log n)$ time.

Given the validity of the algorithm \mathcal{A} , it remains to prove that the weighted median of points minimizes the objective O' . The next theorem shows that.

Theorem 1: Given points p_i with weights w_i on a line, where $\sum_i w_i = 1$, for the weighted median μ of the points, the objective function $\sum_i w_i d(p_i, p)$ is minimized when $p = \mu$.

Proof: First, sort the points p_i in non-decreasing order. Then let the weighted median μ be the point p_k . Assume to the contrary that the point p minimizing the objective function is not p_k . Then we should consider two cases:

$$p = p_k + \delta \text{ or } p = p_k - \delta.$$

Let $L = \sum_i w_i |p_i - p_k|$, $L_1 = \sum_i w_i |p_i - (p_k + \delta)|$, and $L_2 = \sum_i w_i |p_i - (p_k - \delta)|$. Then,

$$L = \sum_{i=1}^{k-1} w_i (p_k - p_i) + \sum_{i=k+1}^{n+m} w_i (p_i - p_k)$$

because the points p_i are sorted in non-decreasing order. Then we also obtain the following inequalities:

$$L_1 \geq \sum_{i=1}^k w_i (p_k + \delta - p_i) + \sum_{i=k+1}^{n+m} w_i (p_i - p_k - \delta),$$

$$L_2 \geq \sum_{i=1}^{k-1} w_i (p_k - \delta - p_i) + \sum_{i=k}^{n+m} w_i (p_i - p_k + \delta).$$

Then we consider $L_1 - L$ and $L_2 - L$. We can show the terms are both non-negative as follows:

$$L_1 - L = \delta \sum_{i=1}^k w_i - \delta \sum_{i=k+1}^{n+m} w_i \geq 0,$$

$$L_2 - L = \delta \sum_{i=k}^{n+m} w_i - \delta \sum_{i=1}^{k-1} w_i \geq 0,$$

because from the definition of the weighted median,

$$\sum_{i=1}^k w_i \geq \frac{1}{2} \text{ and } \sum_{i=k}^{n+m} w_i \geq \frac{1}{2}.$$

Consequently, we show that $L_1 \geq L$ and $L_2 \geq L$. This completes the proof.

Corollary 1: Let V and W be the sets of black points and white points, respectively, divided by a wall, where $|V| = n$ and $|W| = m$. There is an $O((n + m) \log(n +$

m)))-time algorithm to find a gate minimizing the average distance between V and W for the L^1 -distance of the plane.

IV. CONCLUSIONS

In this paper, we consider a variant of the facility problem in a plane. Given separated points divided by a wall, we find a gate location on the wall to minimize the average distances between the points in L^1 -distance. An $O(n \log n)$ -time algorithm is provided, where n is the total number of points.

For further work, we may consider various objective functions, for example, the longest distance between points, called by the diameter. Also the problem in which the number of gates is more than one is interesting.

REFERENCES

[1] F. R. K. Chung, "The average distance and the independence number," *Journal of Graph Theory*, vol. 12, no. 2, pp. 229-235, 1988.

[2] P. Dankelmann, "Computing the average distance of an interval graph," *Information Processing Letters*, vol. 48, no. 6, pp. 311-314, 1993.

[3] S. L. Hakimi, "Location theory," in *Handbook of Discrete and Combinatorial Mathematics*. Boca Raton: CRC Press, pp. 986-995, 2000.

[4] P. Indyk, "Sublinear time algorithms for metric space problems," *Proceedings of the 31st Annual ACM Symposium on the Theory of Computing*, Atlanta: GA, pp. 428-434, 1999.

[5] P. Bose, A. Maheshwari, and P. Morin, "Fast approximations for sums of distances, clustering and Fermat-Weber problem," *Computational Geometry: Theory and Applications*, vol. 24, no. 3, pp. 135-146, 2003.

[6] R. Hassin and A. Tamir, "Improved complexity bounds for location problems on the real line," *Operations Research Letters*, vol. 10, no. 7, pp. 395-402, 1991.

[7] N. Megiddo and K. J. Supowit, "On the complexity of some common geometric location problems," *SIAM Journal on Computing*, vol. 13, no. 1, pp. 182-196, 1984.

[8] S. Arora, P. Raghavan, and S. Rao, "Approximation schemes for Euclidean k-medians and related problems," *Proceedings of the 30th*

Annual ACM Symposium on Theory of Computing, Dallas: TX, pp. 106-113, 1998.

[9] N. Megiddo, "Linear-time algorithms for linear programming in R^3 and related problems," *SIAM Journal on Computing*, vol. 12, no. 4, pp. 759-776, 1983.

[10] P. K. Agarwal, M. Sharir, and S. Toledo, "An efficient multi-dimensional searching technique and its applications," Duke University, Durham: NC, Technical Report CS-1993-20, 1993.

[11] B. Chazelle and J. Matousek, "On linear-time deterministic algorithms for optimization problems in fixed dimension," *Journal of Algorithms*, vol. 21, no. 3, pp. 579-597, 1996.

[12] Z. Drezner, "The planar two-center and two-median problems," *Transportation Science*, vol. 18, no. 4, pp. 351-361, 1984.

[13] D. Eppstein, "Dynamic three-dimensional linear programming," *ORSA Journal on Computing*, vol. 4, no. 4, pp. 360-368, 1992.

[14] J. Hershberger, "A faster algorithm for the two-center decision problem," *Information Processing Letters*, vol. 47, no. 1, pp. 23-29, 1993.

[15] P. K. Agarwal and M. Sharir, "Planar geometric location problems," *Algorithmica*, vol. 11, pp. 185-195, 1994.

[16] M. J. Katz and M. Sharir, "An expander-based approach to geometric optimization," *Proceedings of the 9th Annual Symposium on Computational Geometry*, San Diego: CA, pp. 198-207, 1993.

[17] M. Sharir, "A near-linear algorithm for the planar 2-center problem," *Discrete and Computational Geometry*, vol. 18, no. 2, pp. 125-134, 1997.

[18] T. M. Chan, "More planar two-center algorithms," *Computational Geometry*, vol. 13, no. 3, pp. 189-198, 1999.

[19] P. K. Agarwal and M. Sharir, "Efficient algorithms for geometric optimization," *ACM Computing Surveys*, vol. 30, no. 4, pp. 412-458, 1998.

[20] G. N. Frederickson, "Parametric search and locating supply centers in trees," *Proceedings of the 2nd Workshop on Algorithms and Data Structure*, Ottawa, pp. 299-319, 1991.

[21] T. F. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theoretical Computer Science*, vol. 38, pp. 293-306, 1985.

[22] D. S. Hochbaum and D. B. Shmoys, "A unified approach to approximate algorithms for bottleneck problems," *Journal of the ACM*, vol. 33, no. 3, pp. 533-550, 1986.

[23] T. Feder and D. Greene, "Optimal algorithms for approximate clustering," *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, Chicago: IL, pp. 434-444, 1988.

[24] P. K. Agarwal and C. M. Procopiuc, "Exact and approximation algorithms for clustering," *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco: CA, pp. 658-667, 1998.



Jae-Hoon Kim

Received his B.S. degree in Mathematics from Sogang University in 1994, his M.S. degree in Mathematics from KAIST in 1996, and his Ph.D. degree in Computer Science from KAIST in 2003. He is currently an associate professor at the Division of Computer Engineering, Busan University of Foreign Studies. His research interests include on-line algorithms, scheduling, and computational geometry.