RAIRO. RECHERCHE OPÉRATIONNELLE

PIERRE HANSEN
MAOLIN ZHENG

An algorithm for the minimum variance point of a network

RAIRO. Recherche opérationnelle, tome 25, nº 1 (1991), p. 119-126

<http://www.numdam.org/item?id=RO_1991__25_1_119_0>

Recherche opérationnelle/Operations Research
(vol. 25, n° 1, 1991, p. 119 à 126)

AN ALGORITHM FOR THE MINIMUM VARIANCE POINT OF A NETWORK (*)

by Pierre HANSEN ([1]) and Maolin ZHENG ([2])

Abstract. – *An $O(mn \log n)$ algorithm is proposed to determine a point of a network with m arcs and n vertices which minimizes the variance of the weighted distances to all vertices.*

Keywords : Network; variance measure; distance.

Résumé. – *On propose un algorithme en $O(mn \log n)$ pour déterminer un point d'un réseau avec m arcs et n sommets qui minimise la variance des distances pondérées à tous les sommets.*

Mots clés : Réseau; variance; distance.

1. INTRODUCTION

Traditionally, single facility location on networks (*see* Handler and Mirchandani [4], Tansel, Francis and Lowe [11], [12], Hansen, Labbé, Peeters and Thisse [5], Brandeau and Chiu [2] for surveys) has been concerned with measures of efficiency (e. g. sum of distances to all vertices) or of effectiveness (e. g. maximum distance to any user). More recently, increasing attention has been given to equity aspects of location. This gives rise to several new location problems, in which an equity criterion based on the dispersion of the distribution of distances from the facility to all users is maximized or minimized. Halpern and Maimon [3] consider the following two criteria: (i) minimize the variance of the distribution of distances, (ii) maximize the Lorenz measure of the distribution of distances. They compare location on trees according to these criteria with location at the median or the center.

(*) Received June 1990.
Acknowledgment: The authors thank the Air Force Office of Scientific Research for its support under grants AFOSR 89-0512 and AFOSR 90-0008 to Rutgers University.
([1]) RUTCOR and School of Business, Rutgers University, New Brunswick, N.J. 08903, U.S.A.
([2]) RUTCOR, Rutgers University, New Brunswick, N.J. 08903, U.S.A.

Maimon [9, 10] proposes an $O(n)$ algorithm to minimize the variance and an $O(n^3 \log n)$ algorithm to maximize the Lorenz measure on a tree network with n vertices. Kincaid and Maimon [7, 8] study variance minimization problems in triangular and in 3-cactus graphs. Hansen and Zheng [6] present $O(n^2 \log n)$ and $O(mn^2 \log n)$ algorithms for maximizing the Lorenz measure on trees and on general networks respectively. In this note, we consider variance minimization for general networks and obtain an $O(mn \log n)$ algorithm; this complexity of course reduces to $O(n^2 \log n)$ for the case of planar networks, which is frequent in practice.

We first recall definitions and introduce notation, following [1], [6] and [9].

A *topological arc* is the image of interval [0, 1] by a continuous mapping f from [0, 1] to R^3 such that $f(\theta) \neq f(\theta')$ for any $\theta \neq \theta'$ in [0, 1]; a *rectifiable arc* is a topological arc of well-defined length. A *network* is then defined as a subset N of R^3 which satisfies the following conditions: (i) N is the union of a finite number of rectifiable arcs; (ii) any two arcs intersect at most at their extremities; (iii) N is connected. A tree is a network without closed curve.

The set of *vertices* of the network consists of the extremities of the arcs defining N, and is denoted by $V=\{v_1, \ldots, v_n\}$. We use E to denote the set of arcs defining N, and assume $|E|=m$. The length of each arc $[v_i, v_j] \in E$ is given, and denoted by l_{v_i, v_j}. Each point $s \in N$ belongs to some arc of N but s may or may not be a vertex. For any two points $s_1, s_2 \in [v_i, v_j]$, let $[s_1, s_2]$ denote the subset of points of $[v_i, v_j]$ between s_1 and s_2 and including them; (s_1, s_2) denote the subset of points of $[v_i, v_j]$ between s_1 and s_2 and not including them. Half-open sets are defined similarly. A *path* $P(s_1, s_2)$ joining $s_1, s_2 \in N$ is a minimal connected subset of points of N containing s_1 and s_2. The length of a path is equal to the sum of the length of all its constituent arcs and subarcs. The *distance* $d(s_1, s_2)$ between $s_1 \in N$ and $s_2 \in N$ is equal to the length of a shortest path joining s_1 and s_2.

Let s be a vertex of N, $[u, v]$ be an arc of N and x be a variable point along $[u, v]$. We assume that x also denotes the length of the subarc $[u, x]$. Thus $x=0$ means x coincides with u. It is easy to see that the distance $d(s, x)$ has the following properties: (*a*) it is continuous and concave; (*b*) let

$$x^* = \frac{d(s, v) - d(s, u) + l_{u, v}}{2}$$

then $d(s, x) = d(s, u) + x$ for $x \in [0, x^*]$ and $d(s, x) = d(s, v) + l_{u, v} - x$ for $x \in [x^*, l_{u, v}]$.

A positive weight $w(v)$ is associated with each $v \in V$. For a subset of vertices $V' \subseteqq V$, let $w(V') = \sum_{v \in V'} w(v)$. Without loss of generality, we may assume that $w(V) = 1$. The median function for $x \in N$ and $V' \subseteqq V$ is defined by $z_m(x, V') = \sum_{v \in V'} w(v) d(x, v)$. For simplicity, we use $z_m(x)$ for $z_m(x, V)$. The variance measure for $x \in N$ is then defined by

$$z_v(x) = \sum_{v \in V} w(v) [d(x, v) - z_m(x)]^2,$$

and we want to locate a point x_v such that

$$z_v(x_v) \leqq z_v(x)$$

for all $x \in N$. We call such a point x_v a *minimum variance point*. Clearly, the variance measure is continuous along an arc.

An interior point x of $[u, v] \in E$ is said to a *bottleneck point* if there is a vertex v_k such that there are two shortest paths from v_k to x, one of them containing $[u, x]$ and the other containing $[x, v]$. We say that v_k and x are *relative* to each other. Since any vertex is relative to at most one bottleneck point on each arc, the number of bottleneck points of N is at most mn. Bottleneck points play an important role in the sequel.

In the next Section, we investigate some basic properties of the variance measure for general networks. An $O(mn \log n)$ algorithm to determine a minimum variance point for a general network is given in Section 3.

2. PROPERTIES OF THE VARIANCE MEASURE

For $x \in [u, v]$, let

$$V(u; x) = \{ t \in V : d(t, u) + x \leqq d(t, v) + l([u, v]) - x \};$$

$$\overline{V(u; x)} = V - V(u; x);$$

$$B(x; [u, v]) = \{ t \in V : d(t, v) + l([u, v]) - x = d(t, u) + x \}.$$

The following properties easily follow from the above definitions:

(i) $t \in V(u; x)$ if and only if $d(t, x) = d(t, u) + x$.

(ii) $t \in \overline{V(u; x)}$ if and only if $d(t, x) = d(t, v) + l([u, v]) - x < d(t, u) + x$.

(iii) $B(x, [u, v]) \cap B(y, [u, v]) = \varnothing$ for $x \neq y$ and $x, y \in (u, v)$.

(iv) $B(x, [u, v]) \subset V(u; x)$ for $x \in [u, v]$.

Let $x_1, x_2, \ldots, x_{k-1}$ denote the bottleneck points of $[u, v]$ ranked by increasing distance from u; let $x_0 = 0$ and $x_k = l([u, v])$, *i.e.*, x_0 is equal to u and x_k is equal to v; $I_i = (x_{i-1}, x_i]$ $(i = 1, 2, \ldots, k)$ and $I_0 = [x_0, x_0]$. For simplicity, we restrict our discussion to given $[u, v]$ and $I_i (i = 1, 2, \ldots, k)$. The propositions and theorems obtained below are true for any arc of N.

The following four propositions and their corollaries are easy and given without proof; they are proved in [6].

PROPOSITION 2.1: *Let x be an interior point of arc $[u, v]$ of N. Then x is a bottleneck point if and only if there is a vertex $t \in V$ such that*

$$x = \frac{d(t, v) + l([u, v]) - d(t, u)}{2}.$$

COROLLARY 2.1: *Let x be an interior point of arc $[u, v]$ of N. Then x is a bottleneck point if and only if $B(x; [u, v]) \neq \varnothing$.*

COROLLARY 2.2: *If x is a bottleneck point of* $[u, v]$, then $\mathrm{B}(x; [u, v])$ *is the set of vertices relative to x.*

Since for $x \in [u, v]$,

$$\begin{aligned} z_m(x) &= \sum_{t \in V} w(t) d(x, t) \\ &= z_m(u, V(u; x)) + z_m(v, \overline{V(u; x)}) \\ &\quad + l([u, v]) w(\overline{V(u; x)}) + (w(V(u; x)) - w(\overline{V(u; x)})) x. \qquad (1) \end{aligned}$$

we need to know how $V(u; x)$ changes when x moves along $[u, v]$. The following propositions describe this.

PROPOSITION 2.2: *For $x \in I_i$ $(1 \leqq i \leqq k)$, $V(u; x)$ is unchanged.*

PROPOSITION 2.3: *Let y be a bottleneck point of $[u, v]$. Then $V(u; y)$ and $V(u; y + \varepsilon)$ are different, where $0 < \varepsilon \leqq \min\{x_i - x_{i-1} : i = 1, \ldots, k\}$.*

In view of these two propositions, we can denote $V(u; x)$ by $V(u; I_i)$ and $\overline{V(u; x)}$ by $\overline{V(u; I_i)}$ when $x \in I_i$. Clearly, for $x \in I_i$ and $v' \in V(u; I_i)$, $d(v', x) = d(v', u) + x$.

PROPOSITION 2.4:

$$V(u; I_0) \supset V(u; I_1) \supset \ldots \supset V(u; I_k)$$

and

$$V(u; I_i) = V(u; I_{i-1}) - B(x_{i-1}, [u, v]) \qquad (i = 1, \ldots, k).$$

COROLLARY 2.3: $\overline{V(u;\, I_0)} \subset \overline{V(u;\, I_1)} \subset \ldots \subset \overline{V(u;\, I_k)}$.

As $z_m(x)$ is continuous for $x \in [u, v]$, from (1) and Proposition 2.2, we have

$$z_m(x) = a_i x + b_i \tag{2}$$

for $x_{i-1} \leqq x \leqq x_i$, where $b_i = z_m(u;\, V(u;\, I_i)) + z_m(\overline{V(u;\, I_i)}) + l([u, v])\, w(\overline{V(u;\, I_i)})$ and $a_i = w(V(u;\, I_i)) - w(\overline{V(u;\, I_i)})$ for $i = 1, 2, \ldots, k$.

So $z_m(x)$ is a piecewise linear function for $x \in [u, v]$. We may therefore express the variance measure as follows.

$$\begin{aligned} z_v(x) &= \sum_{t \in V} w(t)\,(d(t, x) - z_m(x))^2 \\ &= \sum_{t \in V(u;\, I_i)} w(t)\,[(d(t, u) - b_i) + (1 - a_i)\, x]^2 \\ &\quad + \sum_{t \in \overline{V(u;\, I_i)}} w(t)\,[d(t, v) + l([u, v]) - b_i) - (1 + a_i)\, x]^2, \end{aligned}$$

for $x_{i-1} \leqq x \leqq x_i\, (i = 1, 2, \ldots, k)$.

Then, using

(i) $w(V(u;\, I_i)) + w(\overline{V(u;\, I_i)}) = 1$,

(ii) $1 - a_i = 2\, w(\overline{V(u;\, I_i)})$,

(iii) $a_i + 1 = 2\, w(V(u;\, I_i))$,

and the expression of b_i, we expand and simplify (3) as follows:

$$z_v(x) = c_i x^2 + d_i x + e_i \qquad \textit{for} \quad x_{i-1} \leqq x \leqq x_i, \tag{4}$$

where

$$c_i = 4\, w(V(u;\, I_i))\, w(\overline{V(u;\, I_i)}), \tag{5}$$

$$\begin{aligned} d_i = 4\,[w(\overline{V(u;\, I_i)})\, z_m(u;\, V(u;\, I_i)) &- w(V(u;\, I_i))\, z_m(v,\, \overline{V(u;\, I_i)}) \\ &- w(V(u;\, I_i))\, w(\overline{V(u;\, I_i)})\, l([u, v])], \end{aligned} \tag{6}$$

$$\begin{aligned} e_i = \sum_{t \in V(u;\, I_i)} w(t)\, d(t, u)^2 &+ \sum_{t \in \overline{V(u;\, I_i)}} w(t)\, d(t, v)^2 - b_i^2 \\ &+ 2\, l([u, v])\, z_m(v,\, \overline{V(u;\, I_i)}) + w(\overline{V(u;\, I_i)})\, l([u, v])^2, \end{aligned} \tag{7}$$

for $i = 1, 2, \ldots, k$. Let

$$g_i = \sum_{t \in V(u;\, I_i)} w(t)\, d(t, u)^2 + \sum_{t \in \overline{V(u;\, I_i)}} w(t)\, d(t, v)^2. \tag{8}$$

Then

$$e_i = g_i - b_i^2 + 2\,l([u, v])\, z_m(v, \overline{V(u; I_i)}) + w(\overline{V(u; I_i)})\, l([u, v])^2. \qquad (9)$$

It follows from (5)-(9) and the expression of b_i that

PROPOSITION 2.5: *For any given i, if* $z_m(u, V(u; I_i))$, $z_m(v; \overline{V(u; I_i)})$, $w(V(u; I_i))$, *and* g_i *are known, then* $z_v(x)$ *for* $x_{i-1} \leqq x \leqq x_i$, *can be constructed in constant time.*

Moreover (4) implies

PROPOSITION 2.6: *For any given* $i \in \{1, 2, \ldots, k\}$, *the function* $z_v(x)$ *is a polynomial function of degree at most* 2 *along the interval* $[x_{i-1}, x_i]$, *and the coefficient of* x^2 *is greater than or equal to* 0.

Finally Proposition 2.6 leads to the following result:

THEOREM 2.1: $z_v(x)$ *is convex along each* $[x_{i-1}, x_i]$ *Moreover if* $c_i \neq 0$, *then the minimum* $z_v(x_i^*)$ *of* $z_v(x)$ *along* $[x_{i-1}, x_i]$ *occurs at* $d_i/2\,c_i \in (x_{i-1}, x_i)$, otherwise it occurs at x_{i-1} or x_i.

Proof. — The conclusion follows from the convexity of $z_v(x)$ along $[x_{i-1}, x_i]$. The expression of x_i^* when $c_i \neq 0$ is obtained just by setting the first derivative of $z_v(x)$ equal to 0. □

The following proposition shows how to obtain $z_m(u; V(u; I_i))$, $z_m(v, \overline{V(u; I_i)})$, $w(V(u; I_i))$, and g_i from $z_m(u; V(u; I_{i-1}))$, $z_m(v, \overline{V(u; I_{i-1})})$, $w(V(u; I_{i-1}))$, and g_{i-1}. Indeed updating leads to an algorithm with lower complexity than if each I_i is considered repeatedly.

PROPOSITION 2.7:

(i) $z_m(u, V(u; I_i)) = z_m(u, V(u; I_{i-1})) - z_m(u, B(x_{i-1}, [u, v]))$;

(ii) $z_m(v, \overline{V(u; I_i)}) = z_m(v, \overline{V(u; I_{i-1})}) + z_m(v, B(x_{i-1}, [u, v]))$;

(iii) $w(V(u; I_i)) = w(V(u; I_{i-1})) - w(B(x_{i-1}, [u, v]))$, and

$$w(\overline{V(u; I_i)}) = 1 - w(V(u; I_i));$$

(iv) $g_i = g_{i-1} + \sum_{t \in B(x_{i-1}, [u, v])} w(t)\,[d(t, v)^2 - d(t, u)^2]$.

Proof. — From Proposition 2.4, $V(u; I_i) = V(u; I_{i-1}) - B(x_{i-1}, [u, v])$ and $\overline{V(u; I_i)} = \overline{V(u; I_{i-1})} \cup B(x_{i-1}, [u, v])$, we can obtain (i)−(iv) easily. □

3. ALGORITHM

Before giving the main algorithm, we first do some preliminary calculations.

(1) *Calculate the distance matrix $D(d(u, v))$ for all pairs of vertices of N.*

(2) *Rank the bottleneck points of each arc $[u, v]$ as $x_1 < x_2 < \ldots < x_{k-1}$, and calculate $B(x_i, [u, v])$ for $i = 1, 2, \ldots, k-1$.*

These operations can be done in $O(mn \log n)$ time, using n times Dijkstra's algorithm with a heap structure to store temporary labels, and m times Heapsort to rank the bottleneck points.

The principle of the algorithm is to compute $z_v(.)$ for each I_i along an arc $[u, v]$ by updating using Proposition 2.7, then find the minimum of $z_v(.)$ along I_i by Theorem 2.1. Rules of the main algorithm are the following,

Algorithm MVP (Minimum Variance Point)
(1) $z_{opt} \leftarrow M$ (A suitable large number); $x_{opt} \leftarrow \varnothing$;
(2) **For** $[u, v] \in E$ **do**
Let $x_1, x_2, \ldots, x_{k-1}$ be the bottleneck points of $[u, v]$ ranked by increasing distance from $[u, v]$; $x_0 = 0$; $x_k = l([u, v])$; $i \leftarrow 1$;
(*a*) calculate $V(u; I_1)$, $\overline{V(u; I_1)}$, $z_m(u, V(u; I_1))$, $z_m(\overline{V(u; I_1)})$, g_1 and $w(V(u; I_1))$;
$z_v(x) = c_1 x^2 + d_1 x + e_1$ for $0 \leqq x \leqq x_1$; calculate x_1^* and $z_v(x_1^*)$;
if $z_v(x_1^*) < z_{opt}$ then $z_{opt} \leftarrow z(x_1^*)$ and $x_{opt} \leftarrow x_1^*$;
(*b*) **while** $i < k$ **do**
(i) $z_m(u, V(u; I_{i+1})) = z_m(u, V(u; I_i)) - z_m(u, B(x_i, [u, v]))$;
(ii) $z_m(v, \overline{V(u; I_{i+1})}) = z_m(v, \overline{V(u; I_i)}) + z_m(v, B(x_i, [u, v])$;
(iii) $w(V(u; I_{i+1})) = w(V(u; I_i)) - w(B(x_i, [u, v]))$ and

$$w(\overline{V(u; I_{i+1})}) = 1 - w(V(u; I_{i+1}));$$

(iv) $$g_{i+1} = g_i + \sum_{t \in B(x_i, [u, v])} w(t)(d(t, v)^2 - d(t, u)^2);$$

$$b_{i+1} = z_m(u, V(u; I_{i+1})) + z_m(v, \overline{Vu; I_{i+1}})) + l([u, v]) w(\overline{V(u; I_{i+1})});$$

(v) calculate c_{i+1}, d_{i+1} and e_{i+1} by (i)-(iv); $z_v(x) = c_{i+1} x^2 + d_{i+1} x + e_{i+1}$ for $x_i \leqq x \leqq x_{i+1}$; calculate x_i^* and $z_v(x_i^*)$;
if $z_v(x_i^*) < z_{opt}$ then $z_{opt} \leftarrow z(x_i^*)$ and $x_{opt} \leftarrow x_i^*$;
(vi) $i \leftarrow i + 1$
end while
end For;
return z_{opt} and x_{opt}.

THEOREM 3.1: *Algorithm MVP determines a minimum variance point in $O(mn \log n)$ time.*

Proof. – Theorem 2.1, and Proposition 2.7 ensure the correctness of the algorithm. Now we do the complexity analysis. For each arc $[u, v]$, time is $O(n)$. The reason is as follows. In (*a*), since

$$V(u; I_1) = V(u; x) = \{ t \in V : d(t, u) + x \leqq d(t, v) + l([u, v]) - x \}$$

for any given $x \in I_1$, we can obtain $V(u; I_1)$ in $O(n)$ time. Similarly for the remaining calculations in (a), time is $O(n)$. In each iteration of (b), (i)-(iv) require $O(|B(x_i, [u, v])|)$ time, where $|*|$ is the cardinality of $*$; (v) takes constant time. So (b) requires $O(\sum_{x \in \{x_1, \ldots, x_{k-1}\}} |B(x; [u, v])|) = O(n)$ time. Since the preliminary calculations take $O(mn \log n)$ time, MVP determines a minimum variance point in $O(mn \log n)$ time.

REFERENCES

1. C. Berge, Espaces Topologiques. Fonctions multivoques, *Dunod, Paris*, 1966.
2. M. L. Brandeau and S. S. Chiu, An Overview of Representative Problems in Location Research, *Management Science*, 1989, *35*, pp. 645-674.
3. J. Halpern and O. Maimon, Accord and Conflict Among Several Objectives in Locational Decisions on Tree Networks, in "Locational Analysis of Public Facilities" J. F. Thisse and H. G. Zoller Eds., Studies in Mathematical and Managerial Economics, pp. 301-314, Amsterdam, North-Holland, 1983.
4. G. Y. Handler and P. B. Mirchandani, Location on Networks, Cambridge (Mass.), The M.I.T. Press, 1979.
5. P. Hansen, M. Labbe, D. Peeters and J. F. Thisse, Facility Location Analysis, *Fundamentals of Pure and Applied Economics*, 1987, *22*, pp. 1-67.
6. P. Hansen and M. Zheng, Algorithms for Determining a Lorenz Point on Trees and Networks, RUTCOR Research Report #55-89 (submitted for publication).
7. R. K. Kincaid and O. Maimon, Locating a Point of Minimum Variance on Triangular Graphs, *Transportation Science*, 1989, *23*, pp. 216-219.
8. R. K. Kincaid and O. Maimon, A Note on Locating a Central Vertex of a 3-cactus Graph, *Computers and Operations Research*, 1990, *17*, pp. 315-320.
9. O. Maimon, The Variance Equity Measure in Location Decision Theory, *Annals of Operations Research*, 1986, *6*, pp. 147-160.
10. O. Maimon, An Algorithm for The Lorenz Measure in Locational Decisions on Trees, *Journal of Algorithms*, 1988, *9*, pp. 583-596.
11. B. C. Tansel, R. L. Francis and T. J. Lowe, Location on Networks: A Survey. Part I: The p-Center and p-Median Problems, *Management Science*, 1983, *29*, pp. 482-497.
12. B. C. Tansel, R. L. Francis and T. J. Lowe, Location on Networks: A Survey. Part II: Exploiting Tree Network Structure, *Management Science*, 1983, *29*, pp. 498-511.