Operations Research Letters 38 (2010) 207–214

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

A generalized Weiszfeld method for the multi-facility location problem

Cem Iyigun [a,*], Adi Ben-Israel [b]

[a] *Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey*

[b] *RUTCOR—Rutgers Center for Operations Research, Rutgers University, NJ, USA*

ARTICLE INFO

Article history:
Received 15 May 2009
Accepted 15 October 2009
Available online 26 November 2009

Keywords:
Fermat–Weber location problem
Multi-facility location problem
Decomposition method
Probabilistic assignments
Weiszfeld method

ABSTRACT

An iterative method is proposed for the K facilities location problem. The problem is relaxed using probabilistic assignments, depending on the distances to the facilities. The probabilities, that decompose the problem into K single-facility location problems, are updated at each iteration together with the facility locations. The proposed method is a natural generalization of the Weiszfeld method to several facilities.

1. Introduction

The **Fermat–Weber location problem** (also **single-facility location problem**) is to locate a **facility** that will serve optimally a set of **customers**, given by their **locations** and **weights**, in the sense of minimizing the weighted sum of distances traveled by the customers. A well known method for solving the problem is the **Weiszfeld method** [30], a gradient method that expresses and updates the sought center as a convex combination of the data points.

The **multi-facility location problem** (**MFLP**) is to locate a (given) number of facilities to serve the customers as above. Each customer is assigned to a single facility, and the problem (also called the **location–allocation problem**) is to determine the optimal locations of the facilities, as well as the optimal assignments of customers (assignment is absent in the single-facility case.)

MFLP is NP hard, [26]. We relax it by replacing rigid assignments with probabilistic assignments, as in [4,3] and [8]. This allows a decomposition of MFLP into single-facility location problems, coupled by the membership probabilities that are updated at each iteration.

2. The problem

Notation. $\overline{1,K} := \{1, 2, \ldots, K\}$. $\|\mathbf{x}\|$ denotes the **Euclidean norm** of a vector $\mathbf{x} \in \mathbb{R}^n$. The **Euclidean distance** $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ is used throughout.

Let $\mathbf{X} := \{\mathbf{x}_i : i \in \overline{1,N}\}$ be a set of N **data points** in $\mathbb{R}^n$, with given **weights** $\{w_i > 0 : i \in \overline{1,N}\}$. Typically, the points $\{\mathbf{x}_i\}$ are the **locations** of **customers**, the weights $\{w_i\}$ are their **demands**.

Given an integer $1 \le K < N$, the **MFLP** is to locate K facilities, and assign each customer to one facility, so as to minimize the sum of weighted distances

$$\min_{\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_K} \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathcal{C}_k} w_i\, d(\mathbf{x}_i, \mathbf{c}_k) \tag{L.K}$$

where $\{\mathbf{c}_k\}$ are the **locations** (or **centers**) of the **facilities**, and $\mathcal{C}_k$ is the **cluster** of customers that are assigned to the kth facility.

For $K = 1$, one gets the **Fermat–Weber location problem**: given $\mathbf{X}$ and $\{w_i : i \in \overline{1,N}\}$ as above, find a point $\mathbf{c} \in \mathbb{R}^n$ minimizing the sum of weighted distances,

$$\min_{\mathbf{c}\in\mathbb{R}^n} \sum_{i=1}^{N} w_i\, d(\mathbf{x}_i, \mathbf{c}), \tag{L.1}$$

see [12,23,24,31] and their references.

If the points $\{\mathbf{x}_i\}$ are not collinear, as is assumed throughout, the objective function of (L.1)

$$f(\mathbf{c}) = \sum_{i=1}^{N} w_i\, d(\mathbf{x}_i, \mathbf{c}) \tag{1}$$

is strictly convex, and (L.1) has a unique optimal solution.

The gradient of (1) is undefined if $\mathbf{c}$ coincides with one of the data points $\{\mathbf{x}_i\}$. For $\mathbf{c} \notin \mathbf{X}$,

$$\nabla f(\mathbf{c}) = -\sum_{i=1}^{N} w_i \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}, \tag{2}$$

* Corresponding author.
E-mail addresses: iyigun@ie.metu.edu.tr (C. Iyigun), adi.benisrael@gmail.com (A. Ben-Israel).

0167-6377/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2009.11.005

and the optimal center $\mathbf{c}^*$, if not in $\mathbf{X}$, is characterized by $\nabla f(\mathbf{c}^*) = \mathbf{0}$, expressing it as a convex combination of the points $\mathbf{x}_i$,

$$\mathbf{c}^* = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i, \text{ with weights } \lambda_i = \frac{w_i/\|\mathbf{x}_i - \mathbf{c}^*\|}{\sum_{m=1}^{N} w_m/\|\mathbf{x}_m - \mathbf{c}^*\|} \text{ that depend on } \mathbf{c}^*.$$

This circular result gives rise to the **Weiszfeld iteration**, [30],

$$\mathbf{c}_+ := T(\mathbf{c}) \tag{3}$$

where $\mathbf{c}_+$ is the updated center, $\mathbf{c}$ is the current center, and

$$T(\mathbf{c}) := \begin{cases} \sum_{i=1}^{N} \left(\dfrac{w_i/\|\mathbf{x}_i - \mathbf{c}\|}{\sum_{m=1}^{N} w_m/\|\mathbf{x}_m - \mathbf{c}\|} \right) \mathbf{x}_i, & \text{if } \mathbf{c} \notin \mathbf{X}; \\ \mathbf{c}, & \text{if } \mathbf{c} \in \mathbf{X}. \end{cases} \tag{4}$$

In order to extend $\nabla f(\mathbf{c})$ to all $\mathbf{c}$, Kuhn [21] modified it as follows: $\nabla f(\mathbf{c}) := -R(\mathbf{c})$, where

$$R(\mathbf{c}) := \begin{cases} -\nabla f(\mathbf{c}), & \text{if } \mathbf{c} \notin \mathbf{X}; \\ \max\{0, \|R^j\| - w_j\} \dfrac{R^j}{\|R^j\|}, & \text{if } \mathbf{c} = \mathbf{x}_j \in \mathbf{X}, \end{cases} \tag{5}$$

$$\text{where} \quad R^j := \sum_{i \neq j} \frac{w_i}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j) \tag{6}$$

is the **resultant force** of $N-1$ forces of magnitude w_i and direction $\mathbf{x}_i - \mathbf{x}_j$, $i \neq j$. The following properties of the mappings $R(\cdot)$, $T(\cdot)$, the optimal center $\mathbf{c}^*$ and any point $\mathbf{x}_j \in \mathbf{X}$ were proved by Kuhn [21]:

$$\mathbf{c} = \mathbf{c}^* \iff R(\mathbf{c}) = \mathbf{0}. \tag{7a}$$

$$\mathbf{c}^* \in \operatorname{conv} \mathbf{X} \text{ (the convex hull of } \mathbf{X}). \tag{7b}$$

If $\mathbf{c} = \mathbf{c}^*$ then $T(\mathbf{c}) = \mathbf{c}$. Conversely, if $\mathbf{c} \notin \mathbf{X}$,

$$T(\mathbf{c}) = \mathbf{c} \text{ then } \mathbf{c} = \mathbf{c}^*. \tag{7c}$$

$$\text{If } T(\mathbf{c}) \neq \mathbf{c} \text{ then } f(T(\mathbf{c})) < f(\mathbf{c}). \tag{7d}$$

$$\mathbf{x}_j = \mathbf{c}^* \iff w_j \geq \|R^j\|. \tag{7e}$$

If $\mathbf{x}_j \neq \mathbf{c}^*$,

$$\text{the direction of steepest descent of } f \text{ at } \mathbf{x}_j \text{ is } R^j/\|R^j\|. \tag{7f}$$

If $\mathbf{x}_j \neq \mathbf{c}^*$ there exists $\delta > 0$ such that

$$0 < \|\mathbf{c} - \mathbf{x}_j\| \implies \|T^s(\mathbf{c}) - \mathbf{x}_j\| > \delta \text{ for some } s. \tag{7g}$$

$$\lim_{\mathbf{c} \to \mathbf{x}_j} \frac{\|T(\mathbf{c}) - \mathbf{x}_j\|}{\|\mathbf{c} - \mathbf{x}_j\|} = \frac{\|R^j\|}{w_j}. \tag{7h}$$

$$\text{For any } \mathbf{c}_0, \text{ if no } \mathbf{c}_r := T^r(\mathbf{c}_0) \in \mathbf{X}, \text{ then } \lim_{r \to \infty} \mathbf{c}_r = \mathbf{c}^*. \tag{7i}$$

These results are generalized in Theorem 1 to the case of several facilities.

Remark 1. Another claim in [21], that

$$T^r(\mathbf{c}_0) \to \mathbf{c}^*$$

for all but a denumerable number of initial centers $\mathbf{c}_0$,

was refuted by Chandrasekaran and Tamir [6]. Convergence can be assured by modifying the algorithm (3) and (4) at a non-optimal center that coincides with a data point $\mathbf{x}_j$. Balas and Yu [1] proposed moving from $\mathbf{x}_j$ in the direction R^j (6) of steepest descent, assuring a decrease of the objective, and non-return to $\mathbf{x}_j$ by (7c) and (7d). Vardi and Zhang [29] guaranteed exit from $\mathbf{x}_j$ by augmenting the objective function with a quadratic of the distances to the other data points. Convergence was also addressed by Ostresh [27], Eckhardt [13], Drezner [11], Brimberg [5], Beck et al. [2], and others.

3. Probabilistic assignments

For $1 < K < N$, the problem (L.K) is NP hard, [26]. It can be solved polynomially in N for $K = 2$, see [10], and possibly for other given K.

We relax the problem by using **probabilistic** (or soft) assignments, with **cluster membership probabilities**,

$$p_k(\mathbf{x}) := \operatorname{Prob}\{\mathbf{x} \in \mathcal{C}_k\}, \quad k \in \overline{1, K},$$

assumed to depend only on the distances $\{d(\mathbf{x}, \mathbf{c}_k) : k \in \overline{1, K}\}$ of the point $\mathbf{x}$ from the K centers. A reasonable assumption is

ASSIGNMENT TO A FACILITY IS MORE PROBABLE THE CLOSER IT IS (A)

and a simple way to model it,

$$p_k(\mathbf{x})\, d(\mathbf{x}, \mathbf{c}_k) = \frac{1}{w} D(\mathbf{x}), \quad k \in \overline{1, K}, \tag{8}$$

where w is the weight of $\mathbf{x}$, and $D(\cdot)$ is a function of $\mathbf{x}$, that does not depend on k. There are other ways to model assumption (A), but (8) works well enough for our purposes.

Model (8) expresses probabilities in terms of distances, displaying neutrality among facilities in the sense of the Choice Axiom of Luce, [25, Axiom 1], see [19, Appendix A]. Other issues such as attractiveness, introduced in the Huff model [15,16], see also [9], are ignored.

Using the fact that probabilities add to one, we get from (8),

$$p_k(\mathbf{x}) = \frac{1/d(\mathbf{x}, \mathbf{c}_k)}{\sum_{\ell=1}^{K} 1/d(\mathbf{x}, \mathbf{c}_\ell)} = \frac{\prod_{j \neq k} d(\mathbf{x}, \mathbf{c}_j)}{\sum_{\ell=1}^{K} \prod_{m \neq \ell} d(\mathbf{x}, \mathbf{c}_m)}, \quad k \in \overline{1, K}, \tag{9}$$

interpreted as $p_k(\mathbf{x}) = 1$ if $d(\mathbf{x}, \mathbf{c}_k) = 0$, i.e., $\mathbf{x} = \mathbf{c}_k$. In the special case $K = 2$,

$$p_1(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{c}_2)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}, \quad p_2(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{c}_1)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}. \tag{10}$$

From (8), we similarly get

$$\frac{D(\mathbf{x})}{w} = \frac{\prod_{j=1}^{K} d(\mathbf{x}, \mathbf{c}_j)}{\sum_{\ell=1}^{K} \prod_{m \neq \ell} d(\mathbf{x}, \mathbf{c}_m)}, \tag{11}$$

which is (up to a constant) the **harmonic mean** of the distances $\{d(\mathbf{x}, \mathbf{c}_j) : j \in \overline{1, K}\}$. In particular,

$$D(\mathbf{x}) = w \frac{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}, \quad \text{for } K = 2. \tag{12}$$

The function (11) is called the **joint distance function** (**JDF**) at $\mathbf{x}$.

Abbreviating $p_k(\mathbf{x})$ by p_k, Eq. (8) is an optimality condition for the extremum problem

$$\min \left\{ w \sum_{k=1}^{K} p_k^2\, d(\mathbf{x}, \mathbf{c}_k) : \sum_{k=1}^{K} p_k = 1,\ p_k \geq 0,\ k \in \overline{1, K} \right\} \tag{13}$$

with variables $\{p_k\}$. The squares of probabilities in (13) are explained as a device for smoothing the underlying objective, min $\{\|\mathbf{x} - \mathbf{c}_k\| : k \in \overline{1, K}\}$, see the seminal article by Teboulle [28].

(L.K) can thus be approximated by the minimization problem

$$\min \sum_{k=1}^{K}\sum_{i=1}^{N} p_k^2(\mathbf{x}_i)\, w_i\, d(\mathbf{x}_i, \mathbf{c}_k) \tag{P.K}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_k(\mathbf{x}_i) = 1, \quad i \in \overline{1,N},$$

$$p_k(\mathbf{x}_i) \ge 0, \quad k \in \overline{1,K},\ i \in \overline{1,N},$$

with two sets of variables, the **centers** $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ and **probabilities** $\{p_k(\mathbf{x}_i) : k \in \overline{1,K},\ i \in \overline{1,N}\}$, corresponding, respectively, to the centers and assignments of the original problem (L.K).

Remark 2. Given a solution of (P.K), if each customer frequents the centers with the corresponding probabilities, then the expected sum of weighted distances is

$$\sum_{k=1}^{K}\sum_{i=1}^{N} p_k(\mathbf{x}_i)\, w_i\, d(\mathbf{x}_i, \mathbf{c}_k), \tag{14}$$

which is no better than if each customer is assigned to the nearest facility, i.e. the one with the highest probability. Therefore, even if an optimal solution of (P.K) is also optimal for (L.K), the value (14) is just an upper bound on the optimal value of (L.K).

Another way to see this is by rewriting (14), using (8) and (11),

$$\sum_{k=1}^{K}\sum_{i=1}^{N} p_k(\mathbf{x}_i)\, w_i\, d(\mathbf{x}_i, \mathbf{c}_k) = \sum_{k=1}^{K}\sum_{i=1}^{N} D(\mathbf{x}_i)$$

$$= K\sum_{i=1}^{N} D(\mathbf{x}_i) \ge \sum_{i=1}^{N} w_i \min_{k\in\overline{1,K}} \{d(\mathbf{x}_i, \mathbf{c}_k)\}. \tag{15}$$

The inequality follows since the harmonic mean is no less than the minimum. The larger K is, the better is the upper bound (14), because the harmonic mean is then closer to the minimum.

4. Probabilities and centers

The objective function of (P.K) is denoted

$$f(\mathbf{c}_1, \dots, \mathbf{c}_K) := \sum_{k=1}^{K}\sum_{i=1}^{N} p_k^2(\mathbf{x}_i)\, w_i\, d(\mathbf{x}_i, \mathbf{c}_k). \tag{16}$$

A natural approach to solving (P.K), see e.g. [7], is to fix one set of variables, and minimize (P.K) with respect to the other set, then fix the other set, etc. We thus alternate between

(1) the **probabilities problem**, i.e. (P.K) with given centers, and
(2) the **centers problem**, (P.K) with given probabilities,

and update their solutions as follows:

Probabilities update. With the centers given, and the distances $d(\mathbf{x}_i, \mathbf{c}_k)$ computed for all centers $\mathbf{c}_k$ and data points $\mathbf{x}_i$, the minimizing probabilities are given explicitly by (9),

$$p_k(\mathbf{x}_i) = \frac{\prod_{j\neq k} d(\mathbf{x}_i, \mathbf{c}_j)}{\sum_{\ell=1}^{K}\prod_{m\neq \ell} d(\mathbf{x}_i, \mathbf{c}_m)}, \quad k \in \overline{1,K}. \tag{17}$$

Centers update. Fixing the probabilities $p_k(\mathbf{x}_i)$ in (P.K), the objective function (16) is a separable function of the cluster centers,

$$f(\mathbf{c}_1, \dots, \mathbf{c}_K) := \sum_{k=1}^{K} f_k(\mathbf{c}_k), \tag{18}$$

where

$$f_k(\mathbf{c}) := \sum_{i=1}^{N} p_k^2(\mathbf{x}_i)\, w_i\, \|\mathbf{x}_i - \mathbf{c}\|, \quad k \in \overline{1,K}. \tag{19}$$

The centers problem thus decomposes into K problems of type (L.1), coupled by the probabilities.

The gradient of (19), wherever it exists, is

$$\nabla f_k(\mathbf{c}) = -\sum_{i=1}^{N} \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|}(\mathbf{x}_i - \mathbf{c}), \quad k \in \overline{1,K}. \tag{20}$$

Zeroing (20), we get the optimal centers

$$\mathbf{c}_k^* = \sum_{i=1}^{N} \lambda_k(\mathbf{x}_i)\, \mathbf{x}_i\ , \quad k \in \overline{1,K}, \tag{21a}$$

as convex combinations of the data points, with weights $\lambda_k(\mathbf{x}_i)$ given by

$$\lambda_k(\mathbf{x}_i) = \frac{p_k^2(\mathbf{x}_i)\, w_i / \|\mathbf{x}_i - \mathbf{c}_k^*\|}{\sum_{j=1}^{N} p_k^2(\mathbf{x}_j)\, w_j / \|\mathbf{x}_j - \mathbf{c}_k^*\|}, \quad k \in \overline{1,K},\ i \in \overline{1,N}. \tag{21b}$$

Eqs. (21a) and (21b) induce K mappings $T_k : \mathbf{c} \to T_k(\mathbf{c})$, $k \in \overline{1,K}$,

$$T_k(\mathbf{c}) := \sum_{i=1}^{N} \left(\frac{p_k^2(\mathbf{x}_i)\, w_i / \|\mathbf{x}_i - \mathbf{c}\|}{\sum_{j=1}^{N} p_k^2(\mathbf{x}_j)\, w_j / \|\mathbf{x}_j - \mathbf{c}\|} \right) \mathbf{x}_i \tag{22a}$$

for $\mathbf{c}$ different than the data points $\{\mathbf{x}_j : j \in \overline{1,N}\}$, and by continuity,

$$T_k(\mathbf{x}_j) := \mathbf{x}_j, \ \text{for all}\ j \in \overline{1,N}. \tag{22b}$$

5. Territories of facilities

In the MFLP (L.K), each facility serves a cluster of customers, in a certain territory (see the discussion by Huff [15]). These territories are given naturally by the membership probabilities of the approximate problem (P.K).

For any two centers $\mathbf{c}_j$, $\mathbf{c}_k$ the locus of the points $\mathbf{x}$ with $p_j(\mathbf{x}) = p_k(\mathbf{x})$ is by (17) represented by the equation

$$\|\mathbf{x} - \mathbf{c}_j\| = \|\mathbf{x} - \mathbf{c}_k\| \tag{23}$$

and is thus a hyperplane (line in $\mathbb{R}^2$). The portion of the hyperplane (23), where

$$p_j(\mathbf{x}) = p_k(\mathbf{x}) \ge p_m(\mathbf{x}), \quad \forall\, m \neq j, k, \tag{24}$$

is either empty, bounded or unbounded. We call it the **common boundary** of the clusters $\mathcal{C}_j$ and $\mathcal{C}_k$. The common boundaries constitute a **Voronoi diagram** of the centers. Each **Voronoi cell** is a polyhedron, and is the natural territory of its facility, see e.g. Figs. 1(b) and 2(b).

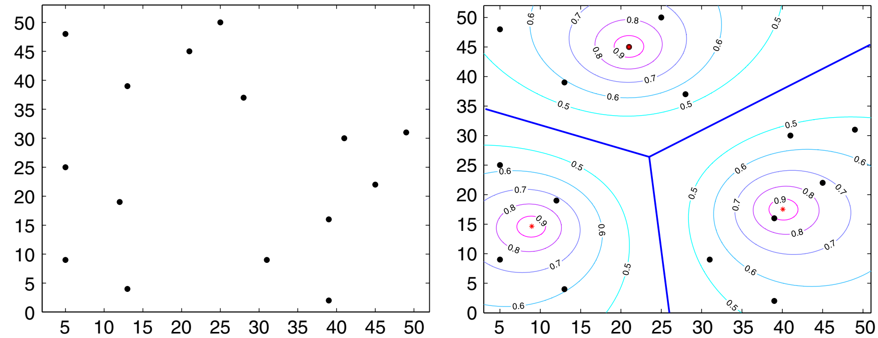
6. Contour approximation and uncertainty

The JDF at a point was given in (11). The JDF of the set $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1,N}\}$ is defined as the sum,

$$D(\mathbf{X}) := \sum_{i=1}^{N} \frac{1}{w_i} D(\mathbf{x}_i). \tag{25}$$

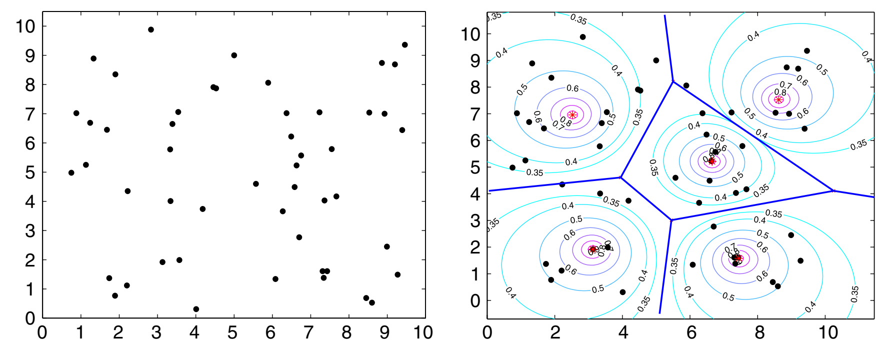
It depends on the K centers, and has the following useful property: for optimal centers, most of the data points are contained in its lower level sets, see [3,18]. We call this property **contour approximation**. The JDF of the set $\mathbf{X}$, (25), is thus a measure of the proximity of customers to their respective facilities: the lower the value of $D(\mathbf{X})$, the better is the set of centers $\{\mathbf{c}_k\}$.

The JDF (11) has the dimension of distance. Normalizing it, we get the dimensionless function



(a) Data points. (b) Probability level sets.

Fig. 1. Illustration of Example 1.



(a) Data points. (b) Probability level sets.

Fig. 2. Illustration of Example 2.

$$E(\mathbf{x}) = \frac{1}{w}\, K\, D(\mathbf{x}) \Bigg/ \left(\prod_{j=1}^{K} d(\mathbf{x}, \mathbf{c}_j) \right)^{1/K}, \tag{26}$$

with 0/0 interpreted as zero. $E(\mathbf{x})$ is the harmonic mean of the distances $\{d(\mathbf{x}, \mathbf{c}_j) : j \in \overline{1, K}\}$ divided by their geometric mean, and can be written, using (9), as the geometric mean of the probabilities (up to a constant),

$$E(\mathbf{x}) = K \left(\prod_{j=1}^{K} p_j(\mathbf{x}) \right)^{1/K}. \tag{27}$$

It follows that $0 \le E(\mathbf{x}) \le 1$, with $E(\mathbf{x}) = 0$ if any $d(\mathbf{x}, \mathbf{c}_j) = 0$, i.e. if $\mathbf{x}$ is a cluster center, and $E(\mathbf{x}) = 1$ if and only if the probabilities $p_j(\mathbf{x})$ are all equal.

In particular, for $K = 2$,

$$E(\mathbf{x}) = 2\, \frac{\sqrt{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2)}}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)} = 2\sqrt{p_1(\mathbf{x}) p_2(\mathbf{x})}\,. \tag{28}$$

$E(\mathbf{x})$ represents the **uncertainty** of classifying the point $\mathbf{x}$, indeed it can be written as

$$E(\mathbf{x}) = \exp\left\{ -I\left(\mathbf{p}(\mathbf{x}), \frac{1}{K}\mathbf{1} \right) \right\}, \tag{29}$$

where $I\left(\mathbf{p}(\mathbf{x}), \frac{1}{K}\mathbf{1}\right)$ is the **Kullback–Leibler distance**, [22], between the distributions

$$\mathbf{p}(\mathbf{x}) = (p_1(\mathbf{x}), p_2(\mathbf{x}), \ldots, p_K(\mathbf{x})) \quad \text{and} \quad \frac{1}{K}\mathbf{1} = \left(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K} \right).$$

The latter distribution, $\frac{1}{K}\mathbf{1}$, is of maximal uncertainty. Therefore, $E(\mathbf{x}) = 1 \iff p(\mathbf{x}) = \frac{1}{K}\mathbf{1}$. We call $E(\mathbf{x})$ the **classification uncertainty function**, abbreviated **CUF**, at $\mathbf{x}$.

The CUF of the data set $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1, N}\}$ is defined as

$$E(\mathbf{X}) := \frac{1}{N} \sum_{i=1}^{N} E(\mathbf{x}_i), \tag{30}$$

and is a monotone decreasing function of K, the number of clusters, decreasing from $E(\mathbf{X}) = 1$ (for $K = 1$), to $E(\mathbf{X}) = 0$ (for $K = N$, the trivial case where every data point is its own cluster).

The "right" number K of facilities to serve the given customers (if not given) is in general determined by economic considerations (operating costs, etc.) An intrinsic criterion for determining the optimal K is provided by the rate of decrease of $E(\mathbf{X})$, see e.g. Fig. 3(d).

7. Optimality conditions and convergence

The gradient (20) is undefined (0/0) if $\mathbf{c}$ coincides with any of the data points. We modify the gradient, following Kuhn [20]–[21], and denote the modified gradient by $-\mathbf{R}_k$.

If a center $\mathbf{c}_k$ is not one of the data points, we copy (20) with a change of sign,

$$\mathbf{R}_k(\mathbf{c}_k) := \sum_{i=1}^{N} \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}_k\|} (\mathbf{x}_i - \mathbf{c}_k). \tag{31}$$

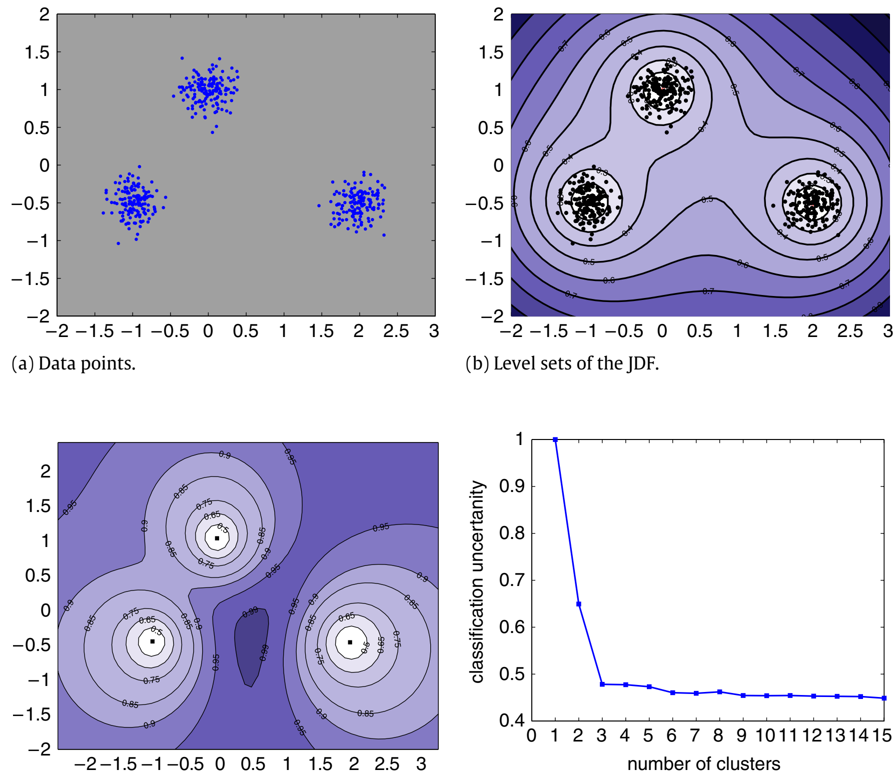
Otherwise, if a center $\mathbf{c}_k$ coincides with a data point $\mathbf{x}_j$ then $\mathbf{x}_j$ belongs with certainty to the kth cluster and, by (17),

$$p_k(\mathbf{x}_j) = 1, \qquad p_m(\mathbf{x}_j) = 0 \quad \text{for all } m \neq k. \tag{32}$$

In this case, define

$$\mathbf{R}_k(\mathbf{x}_j) := \max\{\|\mathbf{R}_k^j\| - w_j,\ 0\} \frac{\mathbf{R}_k^j}{\|\mathbf{R}_k^j\|}, \tag{33a}$$

where



(a) Data points.
(b) Level sets of the JDF.
(c) Level sets of the CUF.
(d) Validation.

Fig. 3. Illustration of Example 3.

$$\mathbf{R}_k^j = \sum_{i \neq j} \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j). \tag{33b}$$

Therefore, if $\|\mathbf{R}_k^j\| < w_j$ then $\mathbf{R}_k(\mathbf{x}_j) = \mathbf{0}$; otherwise, $\mathbf{R}_k(\mathbf{x}_j)$ is a vector with magnitude $\|\mathbf{R}_k^j\| - w_j$ and direction $\mathbf{R}_k^j$.

For $K = 1$, (33a) and (33b) reduce to their single-facility counterparts (5) and (6), respectively.

The results (7a)–(7i) are reproduced, for several facilities, in parts (a)–(i) of the following theorem.

Theorem 1. (a) *Given the data* $\{\mathbf{x}_i : i \in \overline{1,N}\}$ *and* $\{w_i : i \in \overline{1,N}\}$, *let* $\{\mathbf{c}_k : k \in \overline{1,K}\}$ *be arbitrary points, and let the corresponding KN probabilities* $\{p_k(\mathbf{x}_i)\}$ *be given by* (17). *Then, the condition*

$$\mathbf{R}_k(\mathbf{c}_k) = \mathbf{0}, \quad \text{for all } k \in \overline{1,K}, \tag{34}$$

is necessary and sufficient for the points $\{\mathbf{c}_1, \ldots, \mathbf{c}_K\}$ *to minimize* $f(\mathbf{c}_1, \ldots, \mathbf{c}_K)$ *of* (18).

(b) *The optimal centers* $\{\mathbf{c}_k^* : k \in \overline{1,K}\}$ *are in the convex hull of the data points* $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$.

(c) *If a center* $\mathbf{c}$ *is optimal, then it is a fixed point of* T_k *(of* (22a)*),*

$$T_k(\mathbf{c}) = \mathbf{c}, \quad \text{for some } k \in \overline{1,K}. \tag{35}$$

Conversely, if $\mathbf{c} \notin \{\mathbf{x}_j : j \in \overline{1,N}\}$ *and satisfies* (35), *then* $\mathbf{c}$ *is optimal.*

(d) *If* $T_k(\mathbf{c}) \neq \mathbf{c}$, *then*

$$f_k(T_k(\mathbf{c})) < f_k(\mathbf{c}), \quad k \in \overline{1,K}. \tag{36}$$

(e) *If the data point* $\mathbf{x}_j$ *is an optimal center, then* $w_j \geq \|R^j\|$.

(f) *If the data point* $\mathbf{x}_j$ *is not an optimal center, the direction of steepest descent of* f *at* $\mathbf{x}_j$ *is* $\frac{R^j}{\|R^j\|}$.

(g) *Let* $\mathbf{x}_i$ *be a data point that is not an optimal center* $\mathbf{c}_k$, *for a given* $k \in \overline{1,K}$. *Then, there exist* $\delta > 0$ *and an integer* $s > 0$ *such that*

$$0 < \|\mathbf{x}_i - \mathbf{c}\| \leq \delta \quad \text{implies } \|\mathbf{x}_i - T_k^{s-1}(\mathbf{c})\| \leq \delta \quad \text{and} \quad \|\mathbf{x}_i - T_k^s(\mathbf{c})\| > \delta$$

(h)

$$\lim_{\mathbf{c} \to \mathbf{x}_j} \frac{\|\mathbf{x}_j - T_k(\mathbf{c})\|}{\|\mathbf{x}_j - \mathbf{c}\|} = \frac{\|\mathbf{R}_k^j\|}{p_k^2(\mathbf{x}_j)\, w_j} \quad \text{for } k \in \overline{1,K},\ j \in \overline{1,N}.$$

(i) *Given any* $\mathbf{c}_k^0$, $k \in \overline{1,K}$, *define the sequence* $\{\mathbf{c}_k^r = T_k^r(\mathbf{c}_k^0) : r = 1, 2, \ldots\}$. *If no* $\mathbf{c}_k^r$ *is a data point, then* $\lim_{r\to\infty} \mathbf{c}_k^r = \mathbf{c}_k^*$, *for some optimal centers* $\{\mathbf{c}_1^*, \ldots, \mathbf{c}_K^*\}$. $\square$

8. A generalized Weiszfeld method for the multi-facility location problem

The above results are implemented in an algorithm for solving (P.K).

Algorithm 1. A generalized Weiszfeld method for several facilities

---

Data: $\mathcal{D} = \{\mathbf{x}_i : i \in \overline{1,N}\}$ data points (locations of customers),
$\{w_i : i \in \overline{1,N}\}$ weights,
K the number of facilities
$\epsilon > 0$ (stopping criterion)

Initialization: K arbitrary centers $\{\mathbf{c}_k : k \in \overline{1,K}\}$,

Iteration:

Step 1 **compute** distances $\{d(\mathbf{x}, \mathbf{c}_k) : k \in \overline{1,K}\}$ for all $\mathbf{x} \in \mathcal{D}$

Step 2 **compute** probabilities $\{p_k(\mathbf{x}) : \mathbf{x} \in \mathcal{D},\ k \in \overline{1,K}\}$ (using (17))

Step 3 **update** the centers $\{\mathbf{c}_k^+ := T_k(\mathbf{c}_k) : k \in \overline{1,K}\}$

(using (22a) and (22b))

Step 4 **if** $\sum_{k=1}^{K} d(\mathbf{c}_k^+, \mathbf{c}_k) < \epsilon$ **stop**

return to step 1

Writing (22a) as

$$T_k(\mathbf{c}) = \sum_{i=1}^{N} \left(\frac{\frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \right) \mathbf{x}_i - \mathbf{c} + \mathbf{c}$$

$$= \sum_{i=1}^{N} \left(\frac{\frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \right) (\mathbf{x}_i - \mathbf{c}) + \mathbf{c}$$

we get, from (31),

$$T_k(\mathbf{c}) = \mathbf{c} + \mathbf{h}_k(\mathbf{c})\, \mathbf{R}_k(\mathbf{c}) \tag{37}$$

$$\text{with } \mathbf{h}_k(\mathbf{c}) = \frac{1}{\sum_{j=1}^{N} \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \tag{38}$$

showing that Algorithm 1 is a gradient method, following the direction of the resultant $\mathbf{R}_k(\mathbf{c})$ with step of length $\mathbf{h}_k(\mathbf{c})\|\mathbf{R}_k(\mathbf{c})\|$, except for the data points $\{\mathbf{x}_j : j \in \overline{1, N}\}$ which are left fixed by T_k.

Remark 3. (a) Convergence can be established as in the Weiszfeld algorithm, by forcing a move away from non-optimal data points (where the algorithm gets stuck by (22b)), using ideas of Balas and Yu [1], or Vardi and Zhang [29], see Remark 1.

(b) The probabilities $\{p_k(\mathbf{x}_i) : k \in \overline{1, K}, i \in \overline{1, N}\}$ are not needed explicitly throughout the iterations. These probabilities are given by the distances $\{d(\mathbf{x}_i, \mathbf{c}_k)\}$, see (17), and therefore the centers update in Step 3 requires only the distances. Step 2 may thus be omitted.

(c) The final probabilities (following the stop in Step 4) are needed for assigning customers to facilities. In the absence of capacity constraints, each customer is assigned to the nearest center, which by (8) has the highest probability.

(d) The algorithm is robust, since the weights in (21b) are inversely proportional to the distance, and as a result outliers are discounted, having little influence on the facility locations.

(e) Algorithm 1 can be modified to account for capacity constraints on the facilities, as in [17]. This is necessary in **capacitated MFLP** where such constraints may force the customers to split their demands among several facilities, and travel farther than in the unconstrained case. Capacitated MFLP is handled in a sequel article.

(f) Algorithm 1 can handle other models for assumption (A), for example, (8) replaced by,

$$p_k(\mathbf{x})\, \phi(d(\mathbf{x}, \mathbf{c}_k)) = \frac{1}{w} D(\mathbf{x}), \quad k \in \overline{1, K},$$

where $\phi(\cdot)$ is an increasing function, such as d^{α}, $\alpha > 0$ or e^{d}, see [3, Sections 2.8 and 3].

9. Numerical examples

Example 1. This example uses the data of Cooper [7]. It is required to locate 3 facilities to serve the 15 customers shown in Fig. 1(a). The optimal solution has value 143.19, and the value of the worst stable solution is 194.03.

Algorithm 1 was tried 2000 times on this (P.3), using random initial centers, and a tolerance $\epsilon = 0.001$.

The optimal solution was found in 62% of the trials, the worst solution in 0.7% of the trials. The average number of iterations was 23.

The optimal centers are shown in Fig. 1(b), together with some level sets of the membership probability for each cluster. The three half-lines are the common boundaries of the clusters, a Voronoi diagram of the centers, partitioning the plane into territories served by each facility. □

Example 2. This example uses data from Eilon et al. [14, p. 83]. It is required to locate 5 facilities to serve the 50 customers shown in Fig. 2(a).

For the original problem (L.5), different solutions were reported in [14, Table 4.1], with an optimal value of 72.54, and the objective value at the worst solution is 101.99.

We tested Algorithm 1 on these data 2,000 times, with random initial centers and $\epsilon = 0.001$. The average number of iterations was 22. An optimal solution (with value of 71.52, slightly lower than the one reported in [14]) was found in 21% of the 2,000 trials. In 0.35% of the cases, the algorithm found the worst solution.

Fig. 2(b) shows the optimal centers (found by Algorithm 1), some probability level sets for each cluster, and the common boundaries of the centers. The Voronoi diagram has 4 segments and 4 half-lines (2 common boundaries are empty.) □

Example 3. Fig. 3(a) shows a simulated data of 450 customers, organized in 3 equal clusters. Problem (P.3) was solved. Fig. 3(b) shows the centers, and some level sets of the JDF (11), illustrating the contour approximation property, namely that the lower level sets of the JDF capture the data points.

Fig. 3(c) shows some level sets of the CUF (27), with darker shades indicating higher uncertainty. Uncertainty is minimal ($E(\mathbf{x}) = 0$) if $\mathbf{x}$ is one of the centers. Note the patch of maximal uncertainty ($E(\mathbf{x}) \geq 0.99$) in the middle, where one is indifferent between the three centers.

The "right" number of clusters (known a priori in this simulated example) is determined by solving (P.K) for values of $K = 1, 2, \ldots$, calculating the CUF $E(\mathbf{X})$ of the whole dataset (30), and stopping when the marginal decrease of uncertainty is negligible. The results are plotted in Fig. 3(d), confirming $K = 3$ as correct. If the data are amorphous with no clear number of clusters, e.g. the data shown in Fig. 2(a), then the graph of $E(\mathbf{X})$ does not give a clue. □

Acknowledgements

Helpful comments and suggestions by Jack Brimberg, Zvi Drezner, Arie Tamir, Marc Teboulle and a referee are gratefully acknowledged.

Appendix. Proof of Theorem 1

Proof of part (a). If $\mathbf{c}_k$ is not one of the data points, then $-\mathbf{R}_k(\mathbf{c}_k)$ is the gradient (20) at $\mathbf{c}_k$, and (34) is both necessary and sufficient for a minimum, by the convexity of f_k.

If $\mathbf{c}_k$ coincides with a data point $\mathbf{x}_j$, consider the change from $\mathbf{x}_j$ to $\mathbf{x}_j + t\,\mathbf{z}$ where $\|\mathbf{z}\| = 1$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t} f_k(\mathbf{x}_j + t\,\mathbf{z}) \Big|_{t=0} = p_k^2(\mathbf{x}_j)\, w_j - \mathbf{R}_k^j \cdot \mathbf{z}. \tag{39}$$

The greatest decrease of f_k is for $\mathbf{z}$ along $\mathbf{R}_k^j$, i.e., when

$$\mathbf{z} = \frac{\mathbf{R}_k^j}{\|\mathbf{R}_k^j\|}, \quad \text{(proving part (f))}.$$

Therefore, $\mathbf{c}_k$ (that coincides with $\mathbf{x}_j$) is a local minimum if and only if,

$$p_k^2(\mathbf{x}_j)\, w_j - \frac{\mathbf{R}_k^j \cdot \mathbf{R}_k^j}{\|\mathbf{R}_k^j\|} \ge 0,$$

which is equivalent to

$$\|\mathbf{R}_k^j\| \le p_k^2(\mathbf{x}_j)\, w_j,$$

or $\mathbf{R}_k(\mathbf{c}_k) = \mathbf{0}$, by (33a).

Proof of part (b). If $\mathbf{c}_k^*$ is one of the data points, then it is trivially in the convex hull. Otherwise, the condition $\mathbf{R}_k(\mathbf{c}_k^*) = \mathbf{0}$, see (34), results in (21a) and (21b) as above.

Proof of part (c). Follows from part (a), since for $\mathbf{c} \notin \mathbf{X}$,

$$\mathbf{c} = T_k(\mathbf{c}) \Longleftrightarrow R_k(\mathbf{c}) = \mathbf{0},$$

while for $\mathbf{c} \in \mathbf{X}$, $\mathbf{c} = T_k(\mathbf{c})$ for all k.

Proof of part (d). The two functions $f_k(\cdot)$ in (36) are different because the probabilities changed. We prove the inequality (36) for the original probabilities, noting that the updated probabilities (17) are optimal, and result in further decrease.

$\mathbf{c}$ is not one of the data points, since $T_k(\mathbf{c}) \neq \mathbf{c}$. It follows then from (22a) that $T_k(\mathbf{c})$ is the center of gravity of weights $p_k^2(\mathbf{x}_i)\, w_i / \|\mathbf{x}_i - \mathbf{c}\|$ placed at the data points $\mathbf{x}_i$. By elementary calculus, $T_k(\mathbf{c})$ is the unique minimum of the strictly convex function

$$g(\mathbf{y}) = \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \|\mathbf{x}_i - \mathbf{y}\|^2.$$

Since $\mathbf{c} \neq T_k(\mathbf{c})$,

$$\begin{aligned} g(T_k(\mathbf{c})) &= \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \|\mathbf{x}_i - T_k(\mathbf{c})\|^2 < g(\mathbf{c}) \\ &= \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \|\mathbf{x}_i - \mathbf{c}\|^2 = f_k(\mathbf{c}). \end{aligned}$$

On the other hand,

$$\begin{aligned} g(T_k(\mathbf{c})) &= \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \Big[\|\mathbf{x}_i - \mathbf{c}\| + \Big(\|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\|\Big)\Big]^2 \\ &= f_k(\mathbf{c}) + 2\Big(f_k(T_k(\mathbf{c})) - f_k(\mathbf{c})\Big) \\ &\quad + \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \Big[\|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\|\Big]^2 \end{aligned}$$

Combining these results,

$$2 f_k(T_k(\mathbf{c})) + \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \Big[\|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\|\Big]^2 < 2 f_k(\mathbf{c})$$

proving that $f_k(T_k(\mathbf{c})) < f_k(\mathbf{c})$.

Proof of part (e). Follows from part (a), the definition

$$\mathbf{R}_k(\mathbf{x}_j) := \max\{\|\mathbf{R}_k^j\| - p_k^2(\mathbf{x}_j)\, w_j,\ 0\} \frac{\mathbf{R}_k^j}{\|\mathbf{R}_k^j\|}, \tag{33a}$$

and $p_k(\mathbf{x}_j) = 1$, since $\mathbf{x}_j = \mathbf{c}_k$.

Proof of part (f). This was shown in part (a).

Proof of part (g).

$$\begin{aligned} T_k(\mathbf{c}) - \mathbf{x}_i &= \mathbf{c} + h_k(\mathbf{c})\, R_k(\mathbf{c}) - \mathbf{x}_i \\ &= h_k(\mathbf{c}) \sum_{j \neq i}^N \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{c}\|} (\mathbf{x}_j - \mathbf{c}) \\ &\quad + \left(\frac{h_k(\mathbf{c})\, p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} - 1\right)(\mathbf{x}_i - \mathbf{c}). \end{aligned}$$

Since $\mathbf{x}_i$ is not optimal, we have

$$\left\| \sum_{j \neq i}^N \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{x}_i\|} (\mathbf{x}_j - \mathbf{x}_i) \right\| > p_k^2(\mathbf{x}_i)\, w_i.$$

Hence, there exist $\delta' > 0$ and $\epsilon > 0$ such that

$$\left\| \sum_{j \neq i}^N \frac{p_k^2(\mathbf{x}_j)\, w_j}{\|\mathbf{x}_j - \mathbf{c}\|} (\mathbf{x}_j - \mathbf{c}) \right\| \ge (1 + 2\epsilon)\, p_k^2(\mathbf{x}_i)\, w_i$$

for $\|\mathbf{x}_i - \mathbf{c}\| \le \delta'$.

By the definition of h_k, we have

$$\lim_{\mathbf{c} \to \mathbf{x}_i} \frac{h_k(\mathbf{c}) p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} = 1.$$

Hence, there exists $\delta'' > 0$ such that

$$\left| \frac{h_k(\mathbf{c}) p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} - 1 \right| < \frac{\epsilon}{2(1+\epsilon)} \quad \text{for } 0 < \|\mathbf{x}_i - \mathbf{c}\| \le \delta''.$$

Set $\delta = \min(\delta', \delta'')$. For $0 < \|\mathbf{x}_i - \mathbf{c}\| \le \delta$, we have

$$\begin{aligned} \|\mathbf{x}_i - T_k(\mathbf{c})\| &> h_k(\mathbf{c})(1 + 2\epsilon) p_k^2(\mathbf{x}_i)\, w_i - \frac{\epsilon}{2(1+\epsilon)} \|\mathbf{x}_i - \mathbf{c}\| \\ &> \left(1 - \frac{\epsilon}{2(1+\epsilon)}\right)(1 + 2\epsilon)\|\mathbf{x}_i - \mathbf{c}\| \\ &\quad - \frac{\epsilon}{2(1+\epsilon)} \|\mathbf{x}_i - \mathbf{c}\| \\ &= (1 + \epsilon)\, \|\mathbf{x}_i - \mathbf{c}\|. \end{aligned}$$

Since $\|\mathbf{x}_i - \mathbf{c}\| > 0$, $(1+\epsilon)^t \|\mathbf{x}_i - \mathbf{c}\| > \delta$ for some positive integer t and hence $\|\mathbf{x}_i - T_k^s(\mathbf{c})\| > \delta$ for some positive integer s with $\|\mathbf{x}_i - T_k^{s-1}(\mathbf{c})\| \le \delta$.

Proof of part (h). For $\mathbf{c}$ not a data point,

$$\begin{aligned} T_k(\mathbf{c}) &= \sum_{i=1}^N \left(\frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} \Bigg/ \sum_{m=1}^N \frac{p_k^2(\mathbf{x}_m)\, w_m}{\|\mathbf{x}_m - \mathbf{c}\|} \right) \mathbf{x}_i \\ &= \frac{\sum_{i \neq j}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} (\mathbf{x}_i - \mathbf{x}_j) + \mathbf{x}_j \sum_{i=1}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{m=1}^N \frac{p_k^2(\mathbf{x}_m)\, w_m}{\|\mathbf{x}_m - \mathbf{c}\|}} \end{aligned}$$

$$\therefore\ T_k(\mathbf{c}) - \mathbf{x}_j = \sum_{i \neq j}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} (\mathbf{x}_i - \mathbf{x}_j) \Bigg/ \sum_{m=1}^N \frac{p_k^2(\mathbf{x}_m)\, w_m}{\|\mathbf{x}_m - \mathbf{c}\|},$$

$$\begin{aligned} \therefore\ \frac{T_k(\mathbf{c}) - \mathbf{x}_j}{\|\mathbf{x}_j - \mathbf{c}\|} &= \sum_{i \neq j}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|} (\mathbf{x}_i - \mathbf{x}_j) \Bigg/ p_k^2(\mathbf{x}_j) \\ &\quad \times w_j \left(1 + \frac{\|\mathbf{x}_j - \mathbf{c}\|}{p_k^2(\mathbf{x}_j)\, w_j} \sum_{i \neq j}^N \frac{p_k^2(\mathbf{x}_i)\, w_i}{\|\mathbf{x}_i - \mathbf{c}\|}\right). \end{aligned}$$

Taking the limits of the lengths of both sides,

$$\lim_{\mathbf{c} \to \mathbf{x}_j} \frac{\|\mathbf{x}_j - T_k(\mathbf{c})\|}{\|\mathbf{x}_j - \mathbf{c}\|} = \frac{\|\mathbf{R}_k^j\|}{p_k^2(\mathbf{x}_j)\, w_j}.$$

Proof of part (i). With the possible exception of $\mathbf{c}_k^0$, the sequence $\{\mathbf{c}_k^r\}$ lies in the convex hull of vertices, a compact set. By the Bolzano–Weierstrass theorem, there exits a convergent subsequence $\{\mathbf{c}_k^\ell : \ell = 1, 2, \ldots\}$, such that $\lim_{\ell\to\infty} \mathbf{c}_k^\ell = \mathbf{c}_k^\infty$. We prove that $\mathbf{c}_k^\infty = \mathbf{c}_k^*$.

If $\mathbf{c}_k^{\ell+1} = T_k(\mathbf{c}_k^\ell) = \mathbf{c}_k^\ell$ for some ℓ, then the sequence repeats from that point and $\mathbf{c}_k^\infty = \mathbf{c}_k^\ell$. Since $\mathbf{c}_k^\ell$ is not a data point, $\mathbf{c}_k^\infty = \mathbf{c}_k^*$ by Theorem 1(c).

Otherwise, by Theorem 1(d),

$$f_k(\mathbf{c}_k^0) > f_k(\mathbf{c}_k^1) > \cdots > f_k(\mathbf{c}_k^r) > \cdots > f_k(\mathbf{c}_k^*). \tag{40}$$

Hence

$$\lim_{\ell\to\infty} \left(f_k(\mathbf{c}_k^\ell) - f_k(T_k(\mathbf{c}_k^\ell))\right) = 0. \tag{41}$$

The continuity of T_k implies

$$\lim_{\ell\to\infty} T_k(\mathbf{c}_k^\ell) = T_k(\mathbf{c}_k^\infty) \tag{42}$$

we have

$$f_k(\mathbf{c}_k^\infty) - f_k(T_k(\mathbf{c}_k^\infty)) = 0. \tag{43}$$

Therefore, by Theorem 1(d), $\mathbf{c}_k^\infty = T_k(\mathbf{c}_k^\infty)$. If $\mathbf{c}_k^\infty$ is not a data point, then $\mathbf{c}_k^\infty = \mathbf{c}_k^*$ by Theorem 1(c).

In any event, $\mathbf{c}_k^\infty$ lies in the finite set of points $\{\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{c}_k^*\}$, where $\mathbf{c}_k^*$ may be a data point.

The only case that remains is $\mathbf{c}_k^\infty = \mathbf{x}_j$ for some j. If $\mathbf{x}_j \neq \mathbf{c}_k^*$, we first isolate $\mathbf{x}_j$ from the other data points (and $\mathbf{c}_k^*$ if it is not a data point) by a δ-neighborhood that satisfies Theorem 1(g). Then, it is clear that we can choose our subsequence $\mathbf{c}_k^\ell \to \mathbf{x}_j$ such that $\|\mathbf{x}_j - T_k(\mathbf{c}_k^\ell)\| > \delta$ for all ℓ. This means that the ratio $\frac{\|\mathbf{x}_j - T_k(\mathbf{c}_k^\ell)\|}{\|\mathbf{x}_j - \mathbf{c}_k^\ell\|}$ is unbounded. However, this contradicts Theorem 1(h). Hence, $\mathbf{x}_j = \mathbf{c}_k^*$ and the theorem is proved. □

References

[1] E. Balas, C.-S. Yu, A note on the Weiszfeld-Kuhn algorithm for the general Fermat problem, Tech. Report, Carnegie–Melon University, September 1982.
[2] A. Beck, M. Teboulle, Z. Chikishev, Iterative minimization schemes for solving the single source localization problem, SIAM J. Optim. 19 (2008) 1397–1416.
[3] A. Ben-Israel, C. Iyigun, Probabilistic distance clustering, J. Classification 25 (2008) 5–26.
[4] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, 1981.
[5] J. Brimberg, Further notes on convergence of the Weiszfeld algorithm, Yugoslav J. Oper. Res. 13 (2003) 199–206.
[6] R. Chandrasekaran, A. Tamir, Open questions concerning Weiszfeld's algorithm for the Fermat–Weber location problem, Math. Program. 44 (1989) 293–295.
[7] L. Cooper, Heuristic methods for location–allocation problems, SIAM Rev. 6 (1964) 37–53.
[8] T. Drezner, Z. Drezner, The gravity p–median model, European J. Oper. Res. 179 (2007) 1239–1251.
[9] T. Drezner, Z. Drezner, S. Salhi, Solving the multiple competitive facilities location problem, European J. Oper. Res. 142 (2002) 138–151.
[10] Z. Drezner, The planar two-center and two-median problems, Transport. Sci. 18 (1984) 351–361.
[11] Z. Drezner, A note on the Weber location problem, Ann. Oper. Res. 40 (1992) 153–161.
[12] Z. Drezner, K. Klamroth, A. Schöbel, G.O. Wesolowsky, The Weber problem, in: Z. Drezner, H.W. Hamacher (Eds.), Facility Location: Applications and Theory, Springer, 2002.
[13] I.N. Eckhardt, Weber's problem and Weiszfeld's algorithm in general spaces, Math. Program. 18 (1980) 186–196.
[14] S. Eilon, C.D.T. Watson–Gandy, N. Christofides, Distribution Management: Mathematical Modelling and Practical Analysis, Hafner, 1971.
[15] D.L. Huff, Defining and estimating a trading area, J. Marketing 28 (1964) 34–38.
[16] D.L. Huff, A programmed solution for approximating an optimum retail location, Land Econom. 42 (1966) 293–303.
[17] C. Iyigun, A. Ben-Israel, Probabilistic distance clustering adjusted for cluster size, Probab. Engrg. Info. Sci. 22 (2008) 1–19.
[18] C. Iyigun, A. Ben-Israel, Contour approximation of data: a duality theory, Lin. Algeb. Appl. 430 (2009) 2771–2780.
[19] C. Iyigun, A. Ben-Israel, Probabilistic semi-supervised clustering and the uncertainty of classification, in: A. Fink, B. Lausen, W. Seidel, A. Ultsch (Eds.), Advances in Data Analysis, Data Handling and Business Intelligence, Springer, 2009.
[20] H.W. Kuhn, On a pair of dual nonlinear programs, in: J. Abadie (Ed.), Methods of Nonlinear Programming, North-Holland, 1967, pp. 38–54.
[21] H.W. Kuhn, A note on Fermat's problem, Math. Program. 4 (1973) 98–107.
[22] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Statist. 22 (1951) 79–86.
[23] Y. Levin, A. Ben-Israel, A heuristic method for large-scale multi-facility location problems, Comput. Oper. Res. 31 (2004) 257–272.
[24] R.F. Love, J.G. Morris, G.O. Wesolowsky, Facilities Location: Models and Methods, North-Holland, 1988.
[25] R.D. Luce, Individual Choice Behavior, Wiley, 1959.
[26] N. Megiddo, K.J. Supowit, On the complexity of some common geometric location problems, SIAM J. Comput. 13 (1984) 182–196.
[27] L.M. Ostresh, On the convergence of a class of iterative methods for solving the Weber location problem, Oper. Res. 26 (1978) 597–609.
[28] M. Teboulle, A unified continuous optimization framework for center-based clustering methods, J. Machine Learning 8 (2007) 65–102.
[29] Y. Vardi, C.-H. Zhang, A modified Weiszfeld algorithm for the Fermat–Weber location problem, Math. Program. A90 (2001) 559–566.
[30] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnés est minimum, Tohoku Math. J. 43 (1937) 355–386.
[31] G.O. Wesolowsky, The Weber problem: Its history and perspectives, Location Sci. 1 (1993) 5–23.