

# THE MULTI-FACILITY LOCATION PROBLEM: A PROBABILISTIC DECOMPOSITION METHOD AND DUALITY

CEM IYIGUN AND ADI BEN-ISRAEL

ABSTRACT. A generalized Weiszfeld method is proposed for the multi-facility location problem. The problem is relaxed using probabilistic assignments, and is decomposed into single facility location problems, that are coupled by these assignments, and can be solved in parallel. The probabilistic assignments are updated at each iteration, using the distances to the current centers. The method thus iterates between assignments and centers updates, with the probabilistic assignments gradually approaching the pure assignments to the computed centers. The method also provides intrinsic criteria for the quality of the solution, and for the optimal number of facilities to serve the given customers. An application to ecology (home range) is given. A duality theorem allows verifying the optimality of any feasible solution (centers and their assignments) by solving a dual problem.

## 1. INTRODUCTION

The **Fermat–Weber** (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, see [16] and [18]. A well known method for solving the problem is the **Weiszfeld method** [46], 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 **location–allocation**) is to determine the optimal locations of the facilities, as well as the optimal assignments of customers (assignment is not an issue in the single facility case.)

MFLP is NP hard, [38]. We approximate it by replacing rigid assignments with probabilistic assignments, as in [5] and [14], and solve it iteratively, in a natural generalization of the Weiszfeld method to several facilities. These results are a continuation of [4], and [25]–[27].

The plan of the paper is as follows. Section 2 describes the Fermat–Weber location problem and the Weiszfeld method for its solution. Section 3 sketches the multi-facility location problem.

Section 4 introduces probabilistic assignments, with cluster membership probabilities that depend on the distances involved. Section 5 introduces the probabilistic location problem (P.K), an approximation of MFLP. The updates of the centers of (P.K) are explained in Section 6.

A generalized Weiszfeld algorithm for solving (P.K) is presented in Section 7. The cluster membership probabilities are used in Section 8 to chart the territories of the facilities, a Voronoi diagram of the centers.

---

*Date:* September 5, 2011.

*2000 Mathematics Subject Classification.* Primary 90B85, 90C46, 62H30; Secondary 90C59, 92D40, 91B14, 94A17.

*Key words and phrases.* Facility location; clustering; Fermat–Weber problem; multi-facility location problem; Weiszfeld method; contour approximation; duality; ecology; home-range; Luce choice axiom; Kullback–Leibler divergence.

The gradient of the objective is modified in Section 9 to guarantee its existence everywhere. The modified gradient is used to characterize optimality, Theorem 1.

Section 10 recalls two useful concepts from probabilistic clustering, the joint distance function (JDF) and the classification uncertainty function (CUF), and applies them to the problem (P.K). An application to ecology is presented in Section 11.

Section 12 gives a duality theory for (P.K), based on the geometric duality of Kuhn ([30], [31]). A duality theorem, Theorem 2, allows verifying the optimality of any feasible solution (centers and their assignments) by solving a dual problem. In Section 13, Algorithm 1 of Section 7 is tested on several data sets from the literature, and the results compared with [7].

Appendix A establishes convergence for Algorithm 1. The duality theorem is proved in Appendix B. Appendix C shows the connection between the CUF (introduced in Section 10) and the Kullback-Leibler entropic distance. The probabilistic model used here is related to Luce's choice axiom [37] in Appendix D.

## 2. THE PROBLEM

Notation: We denote an index set  $\{1, 2, \dots, K\}$  by  $1:K$ . For a vector  $\mathbf{x} = (x_j) \in \mathbb{R}^n$ ,  $\|\mathbf{x}\|$  denotes the **Euclidean norm**,  $\|\mathbf{x}\| := \sqrt{\sum_{j=1}^n |x_j|^2}$ . The **distance** between the vectors  $\mathbf{x}$  and  $\mathbf{y}$  is denoted  $d(\mathbf{x}, \mathbf{y})$ . The **Euclidean distance**

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \quad (1)$$

is used throughout, except in § 11.

The **Fermat–Weber location problem** is: given a set  $\mathbf{X} := \{\mathbf{x}_i : i \in 1:N\}$  of  $N$  **data points** in<sup>1</sup>  $\mathbb{R}^n$ , and their **weights**  $\{w_i > 0 : i \in 1:N\}$ , find a point  $\mathbf{c} \in \mathbb{R}^n$  minimizing the sum of weighted distances,

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

Typically, the points  $\{\mathbf{x}_i\}$  are the **locations** of **customers**, the weights  $\{w_i\}$  are their **demands**, and the sought point  $\mathbf{c}$  is the **location** of the **facility** (or **center**) that serves the customers optimally; see [19], [36], [47] and their references.

If the points  $\{\mathbf{x}_i\}$  are not collinear, as is assumed throughout, the objective function

$$f(\mathbf{c}) = \sum_{i=1}^N w_i d(\mathbf{x}_i, \mathbf{c}) \quad (2)$$

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

The gradient of (2) is undefined if  $\mathbf{c} \in \mathbf{X}$ , i.e. if  $\mathbf{c}$  coincides with one of the data points  $\{\mathbf{x}_i\}$ . For  $\mathbf{c} \notin \mathbf{X}$ ,

$$\nabla f(\mathbf{c}) = - \sum_{i=1}^N w_i \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}, \quad (3)$$

---

<sup>1</sup>Most location problems are planar, i.e.  $n = 2$ , but the results below hold for general  $n$ , and are stated for  $\mathbb{R}^n$ .

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  $\mathbf{x}_i$ ,

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

This circular result gives rise to the **Weiszfeld iteration**, [46],

$$\mathbf{c}_+ := T(\mathbf{c}) \quad (4)$$

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

$$T(\mathbf{c}) := \begin{cases} \sum_{i=1}^N \left( \frac{w_i / \|\mathbf{x}_i - \mathbf{c}\|}{\sum_{m=1}^N w_m / \|\mathbf{x}_m - \mathbf{c}\|} \right) \mathbf{x}_i, & \text{if } \mathbf{c} \notin \mathbf{X}; \\ \mathbf{c}, & \text{if } \mathbf{c} \in \mathbf{X}. \end{cases} \quad (5)$$

In order to extend  $\nabla f(\mathbf{c})$  to all  $\mathbf{c}$ , Kuhn [31] modified its definition as follows:  $\nabla f(\mathbf{c}) := -R(\mathbf{c})$ , where

$$R(\mathbf{c}) := \begin{cases} -\nabla f(\mathbf{c}), & \text{if } \mathbf{c} \notin \mathbf{X}; \\ \max\{0, \|R^j\| - w_j\} \frac{R^j}{\|R^j\|}, & \text{if } \mathbf{c} = \mathbf{x}_j \in \mathbf{X}, \end{cases} \quad (6)$$

and

$$R^j := \sum_{i \neq j} \frac{w_i}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j) \quad (7)$$

is the **resultant force** of  $N - 1$  forces of magnitude  $w_i$  and direction  $\mathbf{x}_i - \mathbf{x}_j$ ,  $i \neq j$ .

**2.1. The results of Kuhn**, [31]. 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 [31]:

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

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

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

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

$$\mathbf{x}_j = \mathbf{c}^* \iff w_j \geq \|R^j\|. \quad (e)$$

$$\text{If } \mathbf{x}_j \neq \mathbf{c}^*, \text{ the direction of steepest descent of } f \text{ at } \mathbf{x}_j \text{ is } R^j / \|R^j\|. \quad (f)$$

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

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

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

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

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

**Remark 1.** Another claim in [31], that

$$T^r(\mathbf{c}_0) \rightarrow \mathbf{c}^* \text{ for all but a denumerable number of initial centers } \mathbf{c}_0,$$

was refuted by Chandrasekaran and Tamir [10]. Convergence can be assured by modifying the algorithm (4)–(5) at a sticky, non-optimal, center that coincides with a data point  $\mathbf{x}_j$ . Balas and Yu [3] suggested moving from  $\mathbf{x}_j$  in the direction  $R^j$  (7) of steepest descent, assuring a decrease of the objective, and non-return to  $\mathbf{x}_j$  by (8c)–(8d). Vardi and Zhang [45] guaranteed leaving  $\mathbf{x}_j$  by adding to the objective function a quadratic term in the distances to the other data points. Convergence was also addressed by Brimberg [8], Drezner [17], Eckhardt [20], Katz [29] and others.  $\square$

### 3. THE MULTI-FACILITY LOCATION PROBLEM

Given  $\mathbf{X} := \{\mathbf{x}_i : i \in 1:N\}$ ,  $\{w_i : i \in 1:N\}$  as above, and an integer  $K \in 1:N$ , the **multi-facility location problem (MFLP)**, or **location-allocation problem**, is to locate  $K$  facilities, and assign customers to these facilities, so as to minimize the sum of weighted distances

$$\min_{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K} \sum_{k=1}^K \sum_{\mathbf{x}_i \in \mathbf{X}_k} w_i d(\mathbf{x}_i, \mathbf{c}_k) \quad (\text{L.K})$$

where  $\{\mathbf{c}_k : k \in 1:K\}$  are the facility locations (or **centers**), and  $\mathbf{X}_k$  is the **cluster** of customers assigned to the  $k^{\text{th}}$  facility. For  $K = 1$  this reduces to the Fermat–Weber location problem, where assignment is absent. The other extreme case,  $K = N$  (every point is a center), is of no interest.

The MFLP (L.K) can be broken into two problems:

(a) the **assignment problem**: given  $K$  centers  $\{\mathbf{c}_k\}$ , assign each of the  $N$  points  $\{\mathbf{x}_i\}$  to the nearest center, and

(b) the **location problem**: given the  $K$  clusters  $\mathbf{X}_k$  formed in (a), find the center of each cluster (using, say, the Weiszfeld method.)

A natural **heuristic** ([12], [34]) for solving (L.K) is to iterate between these two problems: the clusters are updated (assignment), requiring updating the centers (location), then re-assignment of points, etc.

A solution  $\{(\mathbf{X}_k, \mathbf{c}_k) : k \in 1:K\}$  is **stable** if it cannot be improved by re-assigning any point  $\mathbf{x}_i$  to another cluster. The heuristic method described above may terminate in a non-optimal stable solution, declaring it to be optimal, as illustrated in Example 1.

**Example 1.** Given  $N = 4$  points,  $\mathbf{x}_1 = (0, 0)$ ,  $\mathbf{x}_2 = (1, 0)$ ,  $\mathbf{x}_3 = (1, 1)$ ,  $\mathbf{x}_4 = (0, 1)$ , and all weights  $w_i = 1$ , consider the problem (L.2), and three solutions:

**Solution 1:** Each cluster is a pair of adjacent points, see e.g. Fig. 1(a). The centers may lie anywhere on the segments joining the points, and the value of the objective function is 2.

**Solution 2:** Unequal clusters with centers in opposite points, for example,  $\mathbf{c}_1 = (0, 0)$ ,  $\mathbf{c}_2 = (1, 1)$ , see, e.g., Fig. 1(b). The value of the objective is 2.

**Solution 3:** Unequal clusters, see, e.g., Fig. 1(b), with centers  $\mathbf{c}_1 = (\frac{1}{3+\sqrt{3}}, \frac{1}{3+\sqrt{3}})$ ,  $\mathbf{c}_2 = (1, 1)$  and the value of the objective is 1.9318.

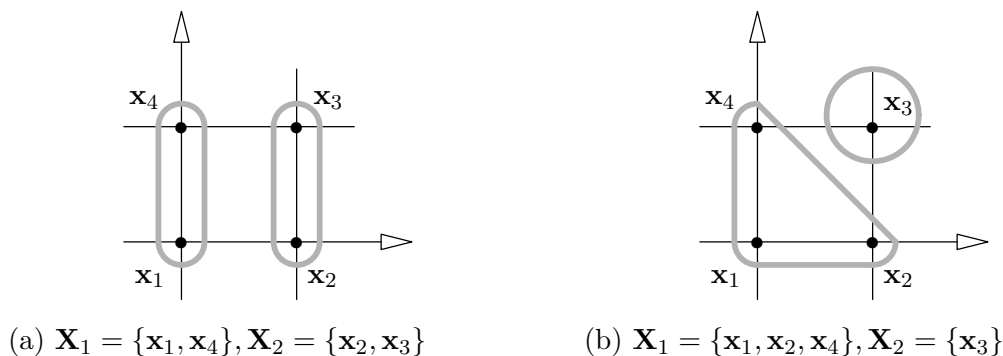


FIGURE 1. Illustration of Example 1

Solution 3 is optimal, but Solution 1 (with centers different than the endpoints) and Solution 2 are stable: for example, in Solution 1, the point  $\mathbf{x}_2$  is not closer to the center of  $\mathbf{X}_1$  than to the center of  $\mathbf{X}_2$ , and should not be re-assigned (to get a better solution) on the basis of distance information alone.

A fourth solution is where each cluster is a pair of opposite points, the centers are anywhere on the diagonals, and the objective function value is  $2\sqrt{2}$ . This solution is unstable, except in the (uninteresting) case where the two centers coincide at  $(\frac{1}{2}, \frac{1}{2})$ .  $\square$

#### 4. CLUSTER MEMBERSHIP PROBABILITIES

For  $1 < K < N$ , the problem (L.K) is NP hard, [38]. It can be solved polynomially in  $N$  for  $K = 2$ , see [15], and possibly for other given  $K$ .

We relax the assignment problem in (L.K) by using probabilistic assignments, see § 5. First we require the concept of **cluster membership probabilities**,

$$p_k(\mathbf{x}) := \text{Prob} \{\mathbf{x} \in \mathbf{X}_k\}, \quad k \in 1:K,$$

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

$$\boxed{\text{membership in a cluster is more likely the closer is its center}} \quad (\text{A})$$

and a simple way to model this assumption is<sup>2</sup>,

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

where  $w$  is the **weight** of  $\mathbf{x}$ , and  $D(\cdot)$  is a function of  $\mathbf{x}$ , that does not depend on  $k$ .  $D(\mathbf{x})$  is called the **joint distance function** (JDF) at  $\mathbf{x}$ .

Equations (8) are optimality conditions for the extremum problem

$$\min \left\{ w \sum_{k=1}^K p_k^2(\mathbf{x}) d(\mathbf{x}, \mathbf{c}_k) : \sum_{k=1}^K p_k(\mathbf{x}) = 1, p_k(\mathbf{x}) > 0, k \in 1:K \right\} \quad (\text{E})$$

<sup>2</sup>There are other ways to model Assumption (A), e.g. [4], but the simple model (8) works well enough for our purposes.

in the probabilities  $\{p_1(\mathbf{x}), \dots, p_K(\mathbf{x})\}$ . The squares of probabilities in the objective of (E) serve to smooth the underlying non-smooth problem, see the seminal paper by Teboulle [43].

Since probabilities add to one we get from (8),

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

and the JDF at  $\mathbf{x}$ ,

$$D(\mathbf{x}) = w \frac{\prod_{j=1}^K d(\mathbf{x}, \mathbf{c}_j)}{\sum_{\ell=1}^K \prod_{m \neq \ell} d(\mathbf{x}, \mathbf{c}_m)}, \quad (10)$$

which is (up to a constant) the **harmonic mean** of the distances  $\{d(\mathbf{x}, \mathbf{c}_k) : k \in 1:K\}$ . The probabilities  $\{p_k(\mathbf{x}) : k \in 1:K\}$  are determined by the centers  $\{\mathbf{c}_k : k \in 1:K\}$ , and the JDF  $D(\mathbf{x})$  also uses the weight  $w$ .

For example, in case  $K = 2$ ,

$$p_1(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{c}_2)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}, \quad p_2(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{c}_1)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}, \quad (11a)$$

$$D(\mathbf{x}) = w \frac{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2)}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)}. \quad (11b)$$

and for  $K = 3$ ,

$$p_1(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{c}_2) d(\mathbf{x}, \mathbf{c}_3)}{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2) + d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_3) + d(\mathbf{x}, \mathbf{c}_2) d(\mathbf{x}, \mathbf{c}_3)}, \quad \text{etc.},$$

$$D(\mathbf{x}) = w \frac{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2) d(\mathbf{x}, \mathbf{c}_3)}{d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_2) + d(\mathbf{x}, \mathbf{c}_1) d(\mathbf{x}, \mathbf{c}_3) + d(\mathbf{x}, \mathbf{c}_2) d(\mathbf{x}, \mathbf{c}_3)}$$

## 5. PROBABILISTIC APPROXIMATION OF (L.K)

Given a point  $\mathbf{x}$  and clusters  $\{\mathbf{X}_k : k \in 1:K\}$ , a **probabilistic assignment** (or **assignment** for short) of  $\mathbf{x}$  is a probability vector  $\mathbf{q}(\mathbf{x}) = (q_1(\mathbf{x}), \dots, q_K(\mathbf{x}))$ , where  $q_k(\mathbf{x})$  is the probability that  $\mathbf{x}$  is assigned to  $\mathbf{X}_k$ . A **pure assignment** is a unit vector, say  $\mathbf{q}(\mathbf{x}) = \mathbf{e}_i$ , indicating that  $\mathbf{x}$  is assigned (deterministically) to  $\mathbf{X}_i$ .

The probabilistic assignments are assumed to have the following property: If a data point  $\mathbf{x}_j$  coincides with a center then  $\mathbf{q}(\mathbf{x}_j)$  is a pure assignment,

$$d(\mathbf{x}_j, \mathbf{c}_k) = 0 \implies q_k(\mathbf{x}_j) = 1, \quad q_m(\mathbf{x}_j) = 0 \text{ for all } m \neq k. \quad (12)$$

We use a probabilistic approximation of the problem (L.K) in the form,

$$\min_{\{\mathbf{c}_1, \dots, \mathbf{c}_K\}} \sum_{k=1}^K \sum_{i=1}^N w_i q_k(\mathbf{x}_i) d(\mathbf{x}_i, \mathbf{c}_k) \quad (\text{P.K})$$

$$\{\{q_1(\mathbf{x}_i), \dots, q_K(\mathbf{x}_i)\} : i \in 1:N\}$$

with two sets of variables, the **centers**  $\{\mathbf{c}_k\}$  and **probabilistic assignments**  $\{q_k(\mathbf{x}_i)\}$ , that are updated iteratively.

Problem (P.K) may have stable solutions that are not optimal for (L.K). If the points  $\{\mathbf{x}_i\}$  are not collinear, as we assume throughout, the optimal solution of (P.K) that is also optimal for (L.K) must have pure assignments. In general, stable solutions of (P.K) may have non-pure assignments. This is illustrated in Example 1, Solution 2.

For any centers and probabilistic assignments, the objective function of (P.K) is an upper bound on the optimal value of (L.K),

$$\sum_{k=1}^K \sum_{i=1}^N w_i q_k(\mathbf{x}_i) d(\mathbf{x}_i, \mathbf{c}_k) \geq \min(\text{L.K}), \quad (13)$$

and therefore so is the optimal value of (P.K),

$$\min(\text{P.K}) \geq \min(\text{L.K}). \quad (14)$$

The probabilistic assignments  $\{q_k(\mathbf{x})\}$  are different than the membership probabilities  $\{p_k(\mathbf{x})\}$  of (9), and in particular may include pure assignments. They are calculated in Algorithm 1 below by raising the membership probabilities to a power  $\nu > 1$  and normalizing<sup>3</sup>,

$$q_k(\mathbf{x}_i) = \frac{p_k^\nu(\mathbf{x}_i)}{\sum_{j=1}^K p_j^\nu(\mathbf{x}_i)}, \quad i \in 1:N, k \in 1:K, \quad (15a)$$

which satisfies (12), since by (9),

$$q_k(\mathbf{x}_i) = \frac{\prod_{j \neq k} d(\mathbf{x}_i, \mathbf{c}_j)^\nu}{\sum_{\ell=1}^K \prod_{m \neq \ell} d(\mathbf{x}_i, \mathbf{c}_m)^\nu}. \quad (15b)$$

To record the dependence on  $\{p_k(\mathbf{x})\}$  and the exponent  $\nu$ , we denote the probabilistic assignments (15a) by  $\{p_k^{(\nu)}(\mathbf{x}_i)\}$ . In previous work ([4] and [25]–[27]) we used the exponent  $\nu = 2$  without renormalization.

For a discussion of the exponent  $\nu$  see § 6.2 below.

**5.1. Bounds for (L.K) with given centers.** Given a set of  $K$  centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ , not necessarily optimal, the minimal value of (L.K) is obtained by assigning each  $\mathbf{x}_i$  to its nearest center,

$$m(\mathbf{c}_1, \dots, \mathbf{c}_K) = \sum_{i=1}^N w_i \min\{d(\mathbf{x}_i, \mathbf{c}_k) : k \in 1:K\}. \quad (16)$$

This minimal value (not necessarily optimal since the centers are arbitrary) can be bounded as follows.

---

<sup>3</sup>This heuristic is one of many possible.

**Corollary 1.** Given a set of centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ , the corresponding membership probabilities  $\{p_k(\mathbf{x}_i) : k \in 1:K, i \in 1:N\}$  of (9), and the values of the JDF of (10),  $\{D(\mathbf{x}_i) : i \in 1:N\}$ , at the points  $\{\mathbf{x}_i : i \in 1:N\}$ ,

$$\sum_{i=1}^N D(\mathbf{x}_i) \leq m(\mathbf{c}_1, \dots, \mathbf{c}_K) \leq \sum_{i=1}^N \frac{D(\mathbf{x}_i)}{\sum_{j=1}^K p_j^2(\mathbf{x}_i)}, \quad (17a)$$

$$\leq K \sum_{i=1}^N D(\mathbf{x}_i). \quad (17b)$$

*Proof.* The left inequality in (17a) follows from

$$D(\mathbf{x}_i) \leq w_i \min\{d(\mathbf{x}_i, \mathbf{c}_k) : k \in 1:K\}, \text{ by (10).}$$

To prove the right inequality, note that for any probabilistic assignments  $\{q_k(\mathbf{x}_i) : i \in 1:N, k \in 1:K\}$ ,

$$\sum_{k=1}^K \sum_{i=1}^N w_i q_k(\mathbf{x}_i) d(\mathbf{x}_i, \mathbf{c}_k) \geq m(\mathbf{c}_1, \dots, \mathbf{c}_K) \quad (18)$$

In particular, for  $q_k(\mathbf{x}) = p_k^{(2)}(\mathbf{x})$ ,

$$\begin{aligned} \text{LHS (18)} &= \sum_{k=1}^K \sum_{i=1}^N w_i \frac{p_k^2(\mathbf{x}_i)}{\sum_{j=1}^K p_j^2(\mathbf{x}_i)} d(\mathbf{x}_i, \mathbf{c}_k), \\ &= \sum_{i=1}^N \frac{w_i}{\sum_{j=1}^K p_j^2(\mathbf{x}_i)} \sum_{k=1}^K p_k(\mathbf{x}_i)^2 d(\mathbf{x}_i, \mathbf{c}_k), \\ &= \sum_{i=1}^N \frac{1}{\sum_{j=1}^K p_j^2(\mathbf{x}_i)} \left( \sum_{k=1}^K p_k(\mathbf{x}_i) \right) D(\mathbf{x}_i), \text{ by (8),} \\ &= \sum_{i=1}^N \frac{D(\mathbf{x}_i)}{\sum_{j=1}^K p_j^2(\mathbf{x}_i)} = \text{RHS (17a).} \end{aligned}$$

Finally, (17b) follows from  $\sum_{j=1}^K p_j^2(\mathbf{x}_i) \geq \frac{1}{K}$ , for all  $i$ .  $\square$

## 6. UPDATES OF CENTERS

Given the assignments  $\{q_k(\mathbf{x}_i)\}$ , say (15b) for some  $\nu > 1$ , the objective function of (P.K) is a separable function of the cluster centers,

$$f(\mathbf{c}_1, \dots, \mathbf{c}_K) := \sum_{k=1}^K f_k(\mathbf{c}_k), \quad (19a)$$

$$\text{where } f_k(\mathbf{c}) := \sum_{i=1}^N w_i q_k(\mathbf{x}_i) \|\mathbf{x}_i - \mathbf{c}\|, \quad k \in 1:K. \quad (19b)$$

The centers problem thus separates into  $K$  problems of type (L.1), coupled by the probabilistic assignments  $\{q_k(\mathbf{x}_i)\}$ . If the data points are not collinear, as we assume throughout, the centers problem has a unique optimal solution  $\{\mathbf{c}_k^* : k \in 1:K\}$ .

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

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

Zeroing the gradients (20) we get the optimal centers  $\{\mathbf{c}_1^*, \dots, \mathbf{c}_K^*\}$  as convex combinations of the data points,

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

with coefficients  $\lambda_k^*(\mathbf{x}_i)$  given by

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

Equations (21a)–(21b) induce  $K$  mappings  $T_k : \mathbf{c} \rightarrow T_k(\mathbf{c})$ ,  $k \in 1:K$ ,

$$T_k(\mathbf{c}) := \sum_{i=1}^N \lambda_{ki} \mathbf{x}_i, \quad \text{with } \lambda_{ki} := \frac{w_i q_k(\mathbf{x}_i) / \|\mathbf{x}_i - \mathbf{c}\|}{\sum_{j=1}^N w_j q_k(\mathbf{x}_j) / \|\mathbf{x}_j - \mathbf{c}\|}, \quad (22a)$$

for  $\mathbf{c} \notin \{\mathbf{x}_j : j \in 1:N\}$ , and by continuity,

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

These mappings are used to update the centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$  as follows,

$$\mathbf{c}_k^+ := T_k(\mathbf{c}_k), \quad \text{for } k \in 1:K. \quad (23)$$

**6.1. The modified gradient.** The gradient (20) is undefined (0/0) if  $\mathbf{c}$  coincides with any of the data points. In order to define the gradient everywhere, including data points, we modify it following Kuhn ([30]–[31]), and denote the modified gradient by  $-\mathbf{R}_k$ .

If a center  $\mathbf{c}_k \notin \{\mathbf{x}_j : j \in 1:N\}$ , copy (20) with a change of sign,

$$\mathbf{R}_k(\mathbf{c}_k) := \sum_{i=1}^N \frac{w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{c}_k\|} (\mathbf{x}_i - \mathbf{c}_k). \quad (24a)$$

Otherwise, if  $\mathbf{c}_k$  coincides with  $\mathbf{x}_j$ , define,

$$\mathbf{R}_k(\mathbf{x}_j) := \max \{ \|\mathbf{R}_k^j\| - w_j q_k(\mathbf{x}_j), 0 \} \frac{\mathbf{R}_k^j}{\|\mathbf{R}_k^j\|}, \quad (24b)$$

where

$$\mathbf{R}_k^j = \sum_{i \neq j} \frac{w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j), \quad (24c)$$

and  $q_k(\mathbf{x}_j) = 1$  in (24b), by (12). Therefore,  $\mathbf{R}_k(\mathbf{x}_j) = \mathbf{0}$  if  $\|\mathbf{R}_k^j\| < w_j$ ; otherwise,  $\mathbf{R}_k(\mathbf{x}_j)$  is a vector with magnitude  $\|\mathbf{R}_k^j\| - w_j$  and direction  $\mathbf{R}_k^j$ . As in (7), the vector  $\mathbf{R}_k^j$  is the **resultant** of  $N - 1$  forces of magnitude  $w_i q_k(\mathbf{x}_i)$  and direction  $\mathbf{x}_i - \mathbf{x}_j$ ,  $i \neq j$ .

**6.2. Updating the exponent  $\nu$  in (15b).** As is typically the case in gradient methods, the iterations (23) make big steps at first, approaching their fixed points, then the iterations slow down and movement at each iteration is small. The “fast” iterations are few in number, the number of the “slow” iterations is determined by the stopping rule.

Most progress towards solving the location problem occurs in the first few iterations. The slow iterations at the end deal mainly with the assignment problem.

With a low value of  $\nu$ , such as  $\nu = 2$ , the distributions  $\{p_k^{(\nu)}(\mathbf{x})\}$  may be far from pure assignments, and require rounding to the nearest integer, 0 or 1.

High values of the exponent  $\nu$  in (15b), say  $\nu > 10$ , produce probabilistic assignments  $\{p_k^{(\nu)}(\mathbf{x}_i)\}$  that are close to pure assignments. For this reason, high values of  $\nu$  are useful in the slow iterations at the end. In contrast, there is no advantage of using high values of  $\nu$  in the beginning fast iterations, where  $\nu < 10$  works fine.

This suggests increasing the exponent  $\nu$  at each iteration. We use here the simple update

$$\nu^+ := \nu + \Delta \tag{25}$$

where  $\Delta > 0$  is the increment per iteration. If  $\nu_0$  is the initial exponent, the one in the  $k^{\text{th}}$ -iteration is thus  $\nu_0 + k\Delta$ . This amounts to a simultaneous approximation of the location and assignment problems of § 3, concentrating on location first then switching gradually to assignment.

Increasing the exponent  $\nu$  results in a decrease of the value of the objective function (19), as shown in following lemma.

**Lemma 1.** Let  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$  be any centers,  $\mathbf{x}$  any point,  $\{d_k(\mathbf{x}, \mathbf{c}_k) : k \in 1:K\}$  the distances, not all equal, and  $\{p_k^{(\nu)}(\mathbf{x}) : k \in 1:K\}$  the probabilities in (15a). If  $\nu_2 > \nu_1 \geq 1$  then

$$\sum_{k=1}^K p_k^{(\nu_2)}(\mathbf{x}) d(\mathbf{x}, \mathbf{c}_k) < \sum_{k=1}^K p_k^{(\nu_1)}(\mathbf{x}) d(\mathbf{x}, \mathbf{c}_k). \tag{26}$$

*Proof.* Arrange the distances, say

$$d(\mathbf{x}, \mathbf{c}_1) \leq d(\mathbf{x}, \mathbf{c}_2) \leq \dots \leq d(\mathbf{x}, \mathbf{c}_K)$$

Then the corresponding probabilistic assignments (15a) satisfy, for any  $\nu \geq 1$ ,

$$p_1^{(\nu)}(\mathbf{x}) \geq p_2^{(\nu)}(\mathbf{x}) \geq \dots \geq p_K^{(\nu)}(\mathbf{x})$$

and are not all equal. Raising the exponent from  $\nu = \nu_1$  to  $\nu = \nu_2$ , increases the highest probabilities and decreases the lowest ones, in particular, there is an index  $j$  such that

$$p_k^{(\nu_2)}(\mathbf{x}) > p_k^{(\nu_1)}(\mathbf{x}) \text{ for } k \leq j, \text{ and } p_k^{(\nu_2)}(\mathbf{x}) \leq p_k^{(\nu_1)}(\mathbf{x}), \text{ for } k > j.$$

Let  $\alpha_k := p_k^{(\nu_2)}(\mathbf{x}) - p_k^{(\nu_1)}(\mathbf{x})$ . Then  $\alpha_k > 0$  for  $k \leq j$ , and  $\alpha_k \leq 0$  for  $k > j$ . Also,  $\sum_{k=1}^K \alpha_k = 0$ , and therefore

$$\sum_{k=1}^j \alpha_k = - \sum_{k=j+1}^K \alpha_k = \alpha, \text{ say.}$$

$$\therefore \frac{1}{\alpha} (\text{LHS}(26) - \text{RHS}(26)) = \sum_{k=1}^j \left(\frac{\alpha_k}{\alpha}\right) d(\mathbf{x}, \mathbf{c}_k) - \sum_{k=j+1}^K \left(-\frac{\alpha_k}{\alpha}\right) d(\mathbf{x}, \mathbf{c}_k) < 0$$

since  $\sum_{k=1}^j \left(\frac{\alpha_k}{\alpha}\right) d(\mathbf{x}, \mathbf{c}_k)$  is a convex combination of low values, and  $\sum_{k=j+1}^K \left(-\frac{\alpha_k}{\alpha}\right) d(\mathbf{x}, \mathbf{c}_k)$  is a convex combination of higher values.  $\square$

## 7. A GENERALIZED WEISZFELD METHOD FOR THE MULTI-FACILITY LOCATION PROBLEM

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

**Algorithm 1.** A generalized Weiszfeld method for several facilities

**Data:**  $\mathbf{X} = \{\mathbf{x}_i : i \in 1:N\}$  data points (locations of customers),  
 $\{w_i : i \in 1:N\}$  weights,  
 $K$  the number of facilities  
 $\epsilon > 0$  (stopping criterion)  
 $\nu_0 > 0$  (initial value of the exponent  $\nu$ )  
 $\Delta > 0$  (the increment in (25))

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

**Iteration:**

- Step 1 **compute** distances  $\{d(\mathbf{x}, \mathbf{c}_k) : k \in 1:K\}$  for all  $\mathbf{x} \in \mathbf{X}$   
Step 2 **update** the exponent  $\nu$  (using (25))  
and the assignments  $\{p_k^{(\nu)}(\mathbf{x}) : \mathbf{x} \in \mathbf{X}, k \in 1:K\}$  (using (15b))  
Step 3 **update** the centers  $\{\mathbf{c}_k^+ := T_k(\mathbf{c}_k) : k \in 1:K\}$  (using (22a)–(22b))  
Step 4 **if**  $\sum_{k=1}^K d(\mathbf{c}_k^+, \mathbf{c}_k) < \epsilon$  **stop**  
**return** to step 1

**Remark 2.**

(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 [3], Vardi and Zhang [45], and others, see Remark 1.

(b) As shown in [4], the above results hold if the Euclidean distance (1) is replaced by an **elliptic distance**

$$d(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x} - \mathbf{y}, Q(\mathbf{x} - \mathbf{y}) \rangle, \quad (27)$$

where  $\langle \cdot, \cdot \rangle$  is the standard inner product, and  $Q$  is a positive definite matrix. Algorithm 1 can thus be used with elliptic distances in Steps 1–4. We illustrate this in § 11.

(c) The algorithm is robust since the coefficients (21b) are directly proportional to the weights, and inversely proportional to the distances. As a result, outliers or low weights are discounted, having little influence on the centers, see, e.g., Example 8 below.  $\square$

**Example 2.** The dependence of Algorithm 1 on the exponent  $\nu$  used in (15b) is illustrated here for the simple Example 1.

Fig. 2 shows three trajectories, starting at the same initial solution (a pair of close points at the center of the unit square  $[0, 1]^2$ ), using a fixed exponent  $\nu$  throughout the iterations, and converging to the three solutions of Example 1. For  $\nu = 6$ ,  $\nu = 12$ , and  $\nu = 18$ , the iterates converged to Solutions 2, 3 and 1, respectively.

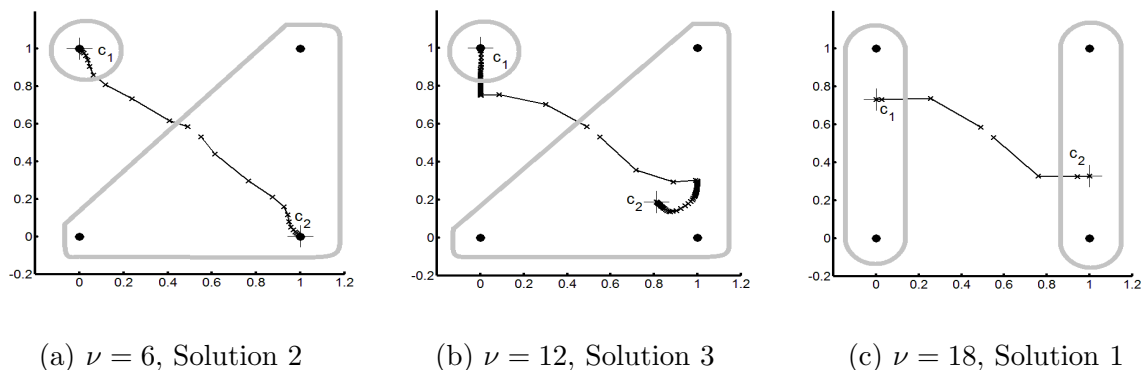


FIGURE 2. Trajectories of iterates in Example 1, for 3 different  $\nu$  values

Table 1 shows the results for 1,000 initial solutions (each a pair of points chosen randomly in  $[0, 1]^2$ ), using different values of  $\nu$  (that are held constant throughout), and a stopping criterion of  $\epsilon = 0.0001$ . The table shows the percentage of cases ending with Solutions 1, 2 or 3 of Example 1, and the average number of iterations. The non-optimal Solution 2 is present only for small  $\nu$ , and occurs in 100% of the time for  $\nu = 6$ . The optimal Solution 3 dominates for intermediate values of  $\nu$  (it occurs 83% of the time for  $\nu = 13$ ), and is present for all larger  $\nu$ . The non-optimal Solution 1 dominates for large  $\nu$ .

## 8. TERRITORIAL BOUNDARIES OF FACILITIES

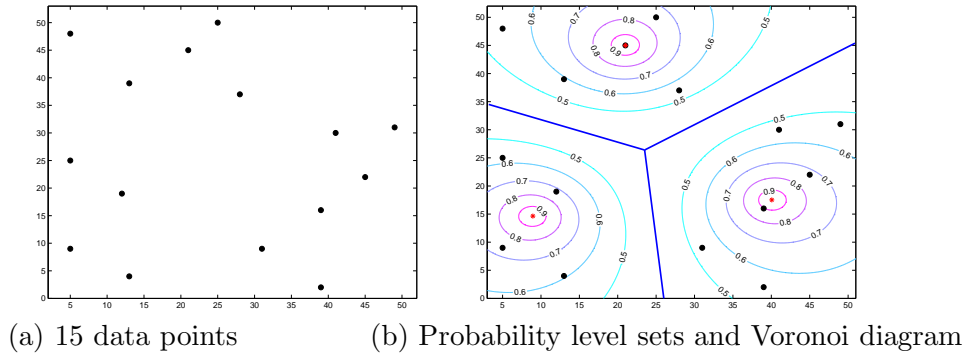
Each center  $\mathbf{c}_k$  serves the customers in a territory that is closer to it than to any other center, see, e.g., [23]–[24]. This territory is expressed naturally using the membership probabilities (9).

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 (9) represented by the equation

$$\|\mathbf{x} - \mathbf{c}_j\| = \|\mathbf{x} - \mathbf{c}_k\| \tag{28}$$

TABLE 1. Dependence of convergence on the exponent  $\nu$  ( $\epsilon = 0.0001$ )

$\nu$	Percentage of initial solutions resulting in			Average # of Iterations
	Solution 1	Solution 2	Solution 3	
6	0.0	100.0	0.0	35.88
13	9.9	7.0	83.1	124.62
14	15.7	4.2	80.1	97.51
16	28.8	1.6	69.6	79.34
18	41.3	0.9	57.8	61.56
20	51.0	0.6	48.4	47.34
22	58.8	0.2	41.0	34.39
30	72.1	0.0	27.9	13.82
40	75.9	0.0	24.1	9.00
50	76.8	0.0	23.2	6.97
100	76.9	0.0	23.1	5.91

FIGURE 3. Illustration of Example 3 ( $N = 15$ ,  $K = 3$ )

and is thus a hyperplane<sup>4</sup>, the orthogonal bisector of the segment joining  $\mathbf{c}_j$  and  $\mathbf{c}_k$ . Throughout this hyperplane one is indifferent between the two centers. The portion of the hyperplane (28) where

$$p_j(\mathbf{x}) = p_k(\mathbf{x}) \geq p_m(\mathbf{x}), \quad \forall m \neq j, k \quad (29)$$

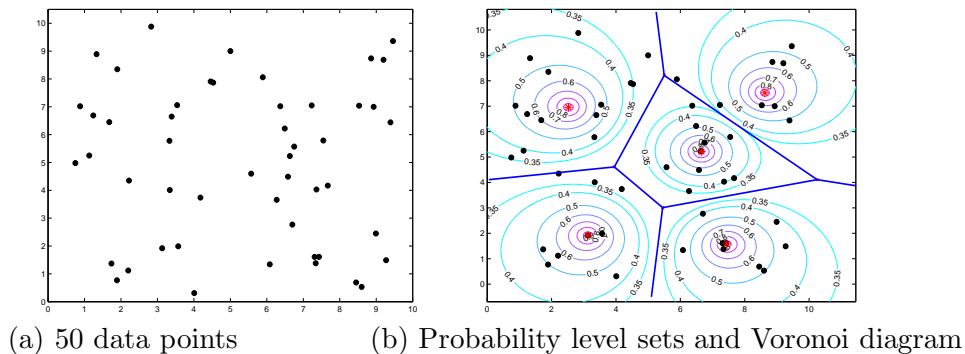
is either empty, bounded or unbounded<sup>5</sup>. We call it the **common boundary** of the clusters  $\mathbf{X}_j$  and  $\mathbf{X}_k$ . The common boundaries constitute a **Voronoi diagram**, [2], of the centers. Each **Voronoi cell** is a polyhedron, and is the natural territory of its facility. □

**Example 3.** Problem 9 in Table 2 below employs  $N = 15$  data points, and  $K = 3$  centers, see Figure 3(a). Figure 3(b) shows the optimal centers, found by Algorithm 1, and 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 4.** Figure 4(a) shows 50 data points of Problem 30 in Table 2 below. This problem, with  $K = 5$ , was solved by Eilon et al in [21, p. 83], reporting an optimal value of 72.54 (later improved to 72.237 in

<sup>4</sup>A line in the typical case  $n = 2$ .

<sup>5</sup>If  $n = 2$  it is either empty, a segment or a half-line.

FIGURE 4. Illustration of Example 4 ( $N = 50$ ,  $K = 5$ )

[7]), and the “worst” value of 101.99. Both values were improved by Algorithm 1, see Table 4, Problem # 30.

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

## 9. CONVERGENCE

Algorithm 1 decomposes the problem (P.K) to  $K$  problems of type (L.1), coupled by the probabilistic assignments. The convergence proof of Kuhn [30]–[31] (with modifications, e.g., [3], [8] or [45]) can be adapted here.

While the algorithm converges from any initial set of centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ , it does not necessarily converge to an optimal solution of the original problem (L.K).

Writing (22a) as

$$T_k(\mathbf{c}) = \sum_{i=1}^N \lambda_{ki} \mathbf{x}_i - \mathbf{c} + \mathbf{c} = \sum_{i=1}^N \lambda_{ki} (\mathbf{x}_i - \mathbf{c}) + \mathbf{c}$$

we get, from (24a),

$$T_k(\mathbf{c}) = \mathbf{c} + \mathbf{h}_k(\mathbf{c}) \mathbf{R}_k(\mathbf{c}) \quad (30)$$

$$\text{with } \mathbf{h}_k(\mathbf{c}) = \frac{1}{\sum_{j=1}^N \frac{w_j q_k(\mathbf{x}_j)}{\|\mathbf{x}_j - \mathbf{c}\|}} \quad (31)$$

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

Results (a)–(i) of § 2.1 are next reproduced for several facilities.

**Theorem 1.** Let the data  $\{\mathbf{x}_i : i \in 1:N\} \subset \mathbb{R}^n$ ,  $\{w_i : i \in 1:N\} \subset \mathbb{R}_+$ , and  $1 \leq K \leq N$  be given. The following statements concern a set of centers  $\{\mathbf{c}_k : k \in 1:K\} \subset \mathbb{R}^n$  and any probabilistic assignments  $\{q_k(\mathbf{x}_i) : i \in 1:N, k \in 1:K\}$  satisfying (12).

(a) If  $\{\mathbf{c}_k : k \in 1:K\}$  are arbitrary centers, then the condition (recall the definition of  $\mathbf{R}_k(\cdot)$  in (24a)–(24b))

$$\mathbf{R}_k(\mathbf{c}_k) = \mathbf{0}, \text{ for all } k \in 1:K, \quad (32)$$

is necessary and sufficient for the points  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$  to minimize  $f(\mathbf{c}_1, \dots, \mathbf{c}_K)$  in (19a).

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

(c) If a center  $\mathbf{c}$  is optimal, then it is a fixed point of  $T_k$  (recall the definition in (22a)–(22b),

$$T_k(\mathbf{c}) = \mathbf{c}, \text{ for some } k \in 1:K. \quad (33)$$

Conversely, if  $\mathbf{c} \notin \{\mathbf{x}_j : j \in 1:N\}$  and satisfies (33), then  $\mathbf{c}$  is optimal.

(d) If  $T_k(\mathbf{c}) \neq \mathbf{c}$ , then the function  $f_k(\cdot)$  of (19b) satisfies

$$f_k(T_k(\mathbf{c})) < f_k(\mathbf{c}), \quad k \in 1:K. \quad (34)$$

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

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

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

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

(h) 
$$\lim_{\mathbf{c} \rightarrow \mathbf{x}_j} \frac{\|\mathbf{x}_j - T_k(\mathbf{c})\|}{\|\mathbf{x}_j - \mathbf{c}\|} = \frac{\|\mathbf{R}_k^j\|}{w_j q_k(\mathbf{x}_j)} \quad \text{for } k \in 1:K, j \in 1:N. \quad (35)$$

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

This theorem is proved in Appendix A.

### Remark 3.

(a) Theorem 1(d) guarantees a decrease of the function  $f_k$  by the iteration (23), but the function  $f_k$  on the left side of (34) may be different than the function  $f_k$  in RHS(34) if the assignments  $\{q_k(\mathbf{x}_i)\}$  are updated, as is the case in Algorithm 1. However, Lemma 1 guarantees a decrease if the assignments  $\{p_k^{(\nu)}(\mathbf{x}_i)\}$  of (15b) are used, with the exponent  $\nu$  raised at each iteration.

(b) Theorem 1(g) states that each non-optimal data point has a neighborhood that eventually repulses the sequence of approximations. Therefore the iterates may not get arbitrarily close to any non-optimal data point.

## 10. CONTOUR APPROXIMATION AND UNCERTAINTY

The **joint distance function (JDF)** at  $\mathbf{x}$ , see (10),

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

depends on the centers  $\{\mathbf{c}_k : k \in 1:K\}$ , and has the following useful property: for optimal centers, most of the data points are contained in the lower level sets of  $D(\mathbf{x})$ , see [1], [26]. We call this property **contour approximation**.

The JDF of the set  $\mathbf{X} = \{\mathbf{x}_i : i \in 1:N\}$  is defined as the sum of the values  $D(\mathbf{x})$  over  $\mathbf{X}$ ,

$$D(\mathbf{X}) := \sum_{i=1}^N D(\mathbf{x}_i). \quad (36)$$

From the contour approximation property it follows that  $D(\mathbf{X})$  is a measure of the proximity of customers to their respective facilities, and the lower the value of  $D(\mathbf{X})$ , the better is the set of centers  $\{\mathbf{c}_k\}$ .

The JDF has the dimension of weighted distance. Normalizing it, we get the dimensionless function

$$E(\mathbf{x}) = \frac{1}{w} K D(\mathbf{x}) / \left( \prod_{j=1}^K d(\mathbf{x}, \mathbf{c}_j) \right)^{1/K}, \quad (37)$$

with  $0/0$  interpreted as zero.  $E(\mathbf{x})$  is the harmonic mean of the distances  $\{d(\mathbf{x}, \mathbf{c}_j) : j \in 1:K\}$  divided by their geometric mean. It follows that  $0 \leq E(\mathbf{x}) \leq 1$ , 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 distances  $d(\mathbf{x}, \mathbf{c}_j)$  (and therefore the probabilities  $p_j(\mathbf{x})$ ) are all equal.

$E(\mathbf{x})$  can be written as the geometric mean of the probabilities (9) (up to a constant),

$$E(\mathbf{x}) = K \left( \prod_{j=1}^K p_j(\mathbf{x}) \right)^{1/K}. \quad (38)$$

In particular, for  $K = 2$ ,

$$E(\mathbf{x}) = 2 \frac{\sqrt{d(\mathbf{x}, \mathbf{c}_1)d(\mathbf{x}, \mathbf{c}_2)}}{d(\mathbf{x}, \mathbf{c}_1) + d(\mathbf{x}, \mathbf{c}_2)} = 2 \sqrt{p_1(\mathbf{x})p_2(\mathbf{x})}. \quad (39)$$

The function  $E(\mathbf{x})$  represents the **uncertainty** of classifying the point  $\mathbf{x}$ , with higher [lower] values of  $E(\mathbf{x})$  corresponding to higher [lower] uncertainty, see Appendix C. We call  $E(\mathbf{x})$  the **classification uncertainty function**, abbreviated **CUF**, at  $\mathbf{x}$ .

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

$$E(\mathbf{X}) := \frac{1}{N} \sum_{i=1}^N E(\mathbf{x}_i). \quad (40)$$

$E(\mathbf{X})$  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 a separate cluster, with no uncertainty.) Indeed, increasing the number of centers from  $K$  to  $K + 1$  allows more customers access to closer centers, with increased probabilities for these centers.

The ‘‘right’’ number  $K$  of facilities to serve the given customers 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. Figure 5(d).

**Example 5.** Figure 5(a) shows a simulated data of  $N = 450$  customers, organized in  $K = 3$  equal clusters. This (P.3) problem was solved, and Figure 5(b) shows the centers, and some level sets of the JDF (10), illustrating the contour approximation property, namely that its lower contours of the JDF capture all the data points.

Figure 5(c) shows some level sets of the CUF (38), 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 the problem (P.K) for values of  $K = 1, 2, \dots$ , calculating the CUF of the whole dataset (40), and stopping when the marginal decrease of uncertainty is negligible. The results are plotted in Figure 5(d), showing that  $K = 3$  is correct. If the data is amorphous with no clear number of clusters, then the graph of  $E(\mathbf{X})$  does not give a clue.  $\square$

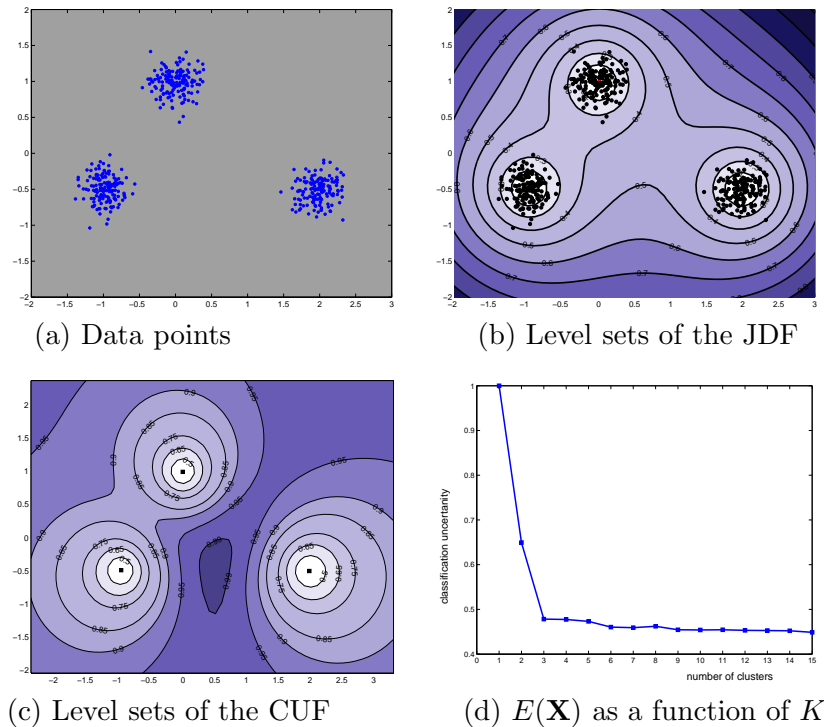


FIGURE 5. Illustration of Example 5 with  $N = 450$  and  $K = 3$

## 11. THE HOME-RANGE

The JDF is related to a useful concept in animal ecology, the **home-range**, that is the geographic territory occupied by an animal population, pursuing its essential and daily activities, see [40] for details. Since the animal activities take place in different locations (drinking near water, etc.), the home-range is roughly made of the activity areas, and “connecting corridors.”

Dixon and Chapman [13] expressed the home-range as a level set of the harmonic mean of the moments of the activity areas, and this was since confirmed for many species.

As an illustration, consider Fig. 6(a), showing 49 observations from a black bear population in south-central Idaho. Fig. 6(b) shows two level sets of the home-range, the core area (inner set) and the 95 percentile area (outer set), see [42].

We similarly express the home-range in terms of the JDF and its level sets, after computing the centers of the activity-areas by Algorithm 1. Since the terrain is not isotropic (certain directions are easier to traverse than others), we use the **Mahalanobis distance** for each cluster, which is the elliptic distance (27) with  $Q = \Sigma^{-1}$ , where  $\Sigma$  is the covariance of the cluster in question. The Mahalanobis distances model the geometry of the clusters, and the harmonic mean of these distances serves to represent the geometry of the home-range. The use of Mahalanobis distances requires an updating of the covariance matrix at each iteration, but no other change in the algorithm.

The problem was solved with Algorithm 1, with 2 and 3 centers, and the level sets of the JDF are shown in Fig. 6(c) and Fig. 6(d), respectively.

For a related, information-theoretic approach to the home-range, see [22]. □

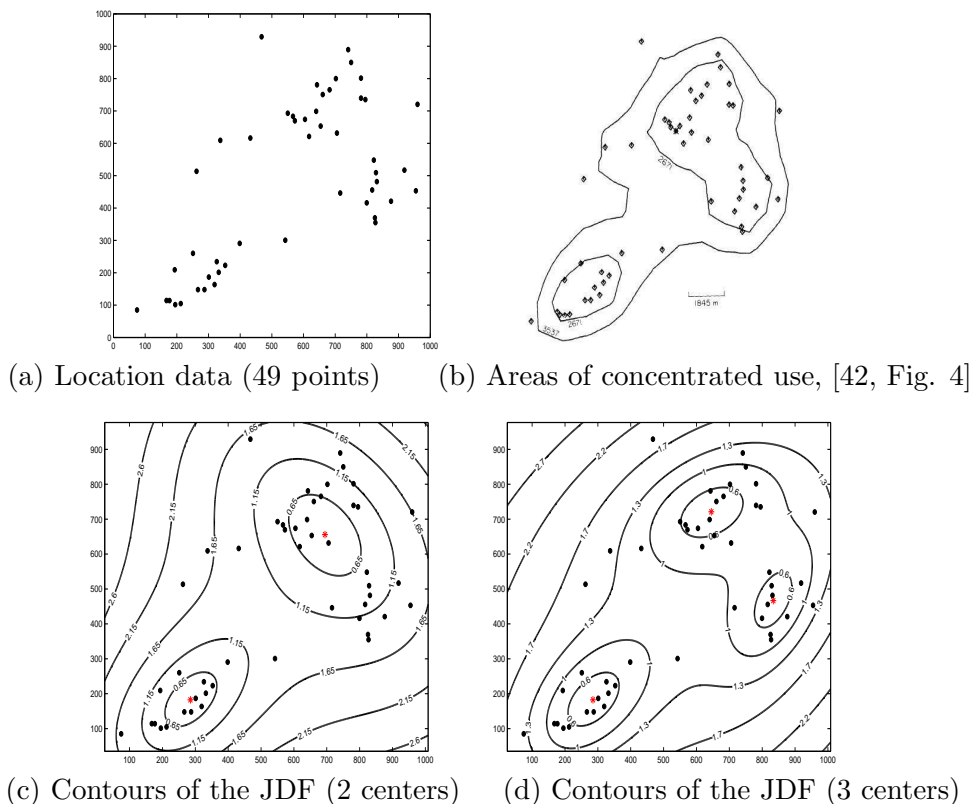


FIGURE 6. Home range of black bear U60, after [42].

## 12. THE DUAL PROBLEM

We abbreviate the probabilistic assignments  $q_k(\mathbf{x}_i)$  by  $q_{ki}$ , for  $k \in 1:K$ ,  $i \in 1:N$ . A dual problem (D) for (P.K) is now given, following [30]. It uses the data

$$\mathcal{S} := \{\{\mathbf{x}_i : i \in 1:N\}, \{w_i : i \in 1:N\}, \{q_{ki} : k \in 1:K, i \in 1:N\}\}, \quad (41)$$

consisting of data points  $\{\mathbf{x}_i\}$ , weights  $\{w_i\}$  and probabilistic assignments  $\{q_{ki}\}$ . The dual variables are  $KN$  vectors  $\{\mathbf{u}_{ki} : k \in 1:K, i \in 1:N\}$ , one for each center and data point. We denote the set of dual variables by  $\mathbf{U}$ .

The **dual problem** is:

$$\max \quad g(\mathbf{U}) = \sum_{k=1}^K \sum_{i=1}^N \mathbf{u}_{ki} \cdot \mathbf{x}_i \quad (\text{D})$$

$$\text{s.t.} \quad \sum_{i=1}^N \mathbf{u}_{ki} = \mathbf{0}, \quad k \in 1:K, \quad (42)$$

$$\|\mathbf{u}_{ki}\| \leq w_i q_{ki}, \quad i \in 1:N, k \in 1:K. \quad (43)$$

Problem (D) is a generalization of the dual problem for the single facility location problem, see also [19, § 1.1.2], [28] and [44].

**Remark 4.**

(a) The dual problem (D) uses the same data as (P.K), namely the points  $\{\mathbf{x}_i\}$ , their weights  $\{w_i\}$ , and the probabilistic assignments  $\{q_{ki}\}$ . The cluster centers are not needed in (D).

(b) Unlike problems (L.K) and (P.K) that have locally optimal solutions, the dual problem (D) has a unique optimal value, and its optimal solutions form a closed convex set.

(c) The dual problem (D) has a natural geometric interpretation. Let the weight  $w_i$  of the  $i^{\text{th}}$ -customer be split among the facilities, so that the  $k^{\text{th}}$ -facility gets the weight  $q_{ki} w_i$ . The vectors  $\mathbf{u}_{ki}$  is interpreted as a force with direction  $(\mathbf{x}_i - \mathbf{c}_k)$ , and magnitude  $\|\mathbf{u}_{ki}\| \leq q_{ki} w_i$ . The location of the  $k^{\text{th}}$ -facility is then the equilibrium point of these forces, as expressed by the constraint (42). The objective of the problem (D) is

$$\sum_{k=1}^K \sum_{i=1}^N \mathbf{u}_{ki} \cdot \mathbf{x}_i = \sum_{k=1}^K \left( \sum_{i=1}^N \mathbf{u}_{ki} \cdot (\mathbf{x}_i - \mathbf{c}_k) \right), \text{ by (42),}$$

the sum of terms  $\mathbf{u}_{ki} \cdot (\mathbf{x}_i - \mathbf{c}_k)$  representing the work done by the force  $\mathbf{u}_{ki}$  along  $(\mathbf{x}_i - \mathbf{c}_k)$ .

**Theorem 2.**

(a) Let the data  $\mathcal{S}$  in (41) be given. Then for any set of centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ , and any set of feasible dual variables  $\mathbf{U} = \{\mathbf{u}_{ki}\}$ ,

$$g(\mathbf{U}) \leq f(\mathbf{c}_1, \dots, \mathbf{c}_K), \quad (44)$$

where  $f(\mathbf{c}_1, \dots, \mathbf{c}_K)$  is the objective function (19a) of the primal problem (P.K).

(b) Given the data  $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , and an optimal solution

$$\{\mathbf{c}_1, \dots, \mathbf{c}_K\}, \{q_{ki} : i \in 1:N, k \in 1:K\},$$

of the primal problem (P.K), there exist feasible dual variables  $\mathbf{U}$  such that

$$g(\mathbf{U}) = f(\mathbf{c}_1, \dots, \mathbf{c}_K). \quad (45)$$

(c) Let  $\mathbf{U}$  be an optimal solution of the dual problem (D). Then there exist  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$  such that (45) holds.  $\square$

A proof of this theorem is given in Appendix B.

**Remark 5.**

(a) Theorem 2 establishes duality between the problems (P.K) and (D). Theorem 2(a) is a **weak duality** theorem in the sense that any feasible solution  $\mathbf{U}$  of (D) gives a lower bound for the optimal value of (P.K), and conversely, any set of centers  $\{\mathbf{c}_k\}$  for (P.K) gives an upper bound on the optimal value of (D). Theorems 2(b)–(c) show that there is no duality gap between (P.K) and (D).

(b) Theorem 2(b) can be used to check the optimality, for the original problem (L.K), of a set of centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ . Indeed, if

$$g(\mathbf{U}) = m(\mathbf{c}_1, \dots, \mathbf{c}_K), \text{ see (16),}$$

then the centers  $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$  are optimal, see, e.g., the following example.

**Example 6.** Consider the 4 points  $\{\mathbf{x}_1, \dots, \mathbf{x}_4\}$  of Example 1, and the 2 clusters  $\mathbf{X}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4\}$ ,  $\mathbf{X}_2 = \{\mathbf{x}_3\}$  shown in Fig. 1(b). We establish the optimality of these clusters using Theorem 2. For  $k \in 1:2$  and  $i \in 1:4$  let  $\mathbf{u}_{ki}$  be the dual variables, and let the assignments  $q_{ki}$  be given by the cluster membership probabilities (9),  $q_{ki} = p_k(\mathbf{x}_i)$ . Then

$$q_{13} = q_{21} = q_{22} = q_{24} = 0$$

( $q_{13} = 0$  since  $\mathbf{x}_3$  does not belong to  $\mathbf{X}_1$ , etc.) and therefore, by the constraints (43),

$$\mathbf{u}_{13} = \mathbf{u}_{21} = \mathbf{u}_{22} = \mathbf{u}_{24} = \mathbf{0}$$

The dual problem (D) can be written

$$\begin{aligned} \max \quad & \mathbf{u}_{11} \cdot \mathbf{x}_1 + \mathbf{u}_{12} \cdot \mathbf{x}_2 + \mathbf{u}_{14} \cdot \mathbf{x}_4 + \mathbf{u}_{23} \cdot \mathbf{x}_3 = \mathbf{u}_{12}[1] + \mathbf{u}_{14}[2] + \mathbf{u}_{23}[1] + \mathbf{u}_{23}[2] \\ \text{s.t.} \quad & \mathbf{u}_{11} + \mathbf{u}_{12} + \mathbf{u}_{14} = \mathbf{0}, \\ & \mathbf{u}_{23} = \mathbf{0}, \\ & \|\mathbf{u}_{ki}\| \leq 1, \text{ for } ki = 11, 12, 14 \text{ and } 23, \end{aligned}$$

where  $\mathbf{u}_{ki}[j]$  is the  $j^{\text{th}}$  component of  $\mathbf{u}_{ki}$ ,  $j = 1, 2$ . This problem has the unique solution

$$\mathbf{u}_{11} = -\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_{12} = \frac{1}{4} \begin{pmatrix} \sqrt{2} + \sqrt{6} \\ \sqrt{2} - \sqrt{6} \end{pmatrix}, \quad \mathbf{u}_{14} = \frac{1}{4} \begin{pmatrix} \sqrt{2} - \sqrt{6} \\ \sqrt{2} + \sqrt{6} \end{pmatrix}, \quad \mathbf{u}_{23} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

and the optimal value is

$$\mathbf{u}_{12}[1] + \mathbf{u}_{14}[2] = \frac{1}{2}(\sqrt{2} + \sqrt{6}) \approx 1.931852$$

which equals the value of Solution 3 in Example 1, whose optimality then follows by Theorem 2.

### 13. NUMERICAL EXAMPLES

Tables 2–6 below present results for the 31 test problems considered in [7]. For each of these problems, Table 2 gives the source (references in bibliography), the number  $N$  of customers, the number  $K$  of facilities, and the coordinates of the  $N$  customers.

Each problem was solved 500 times, using random initializations. The results, for  $\nu_0 = 5$ ,  $\Delta = 0.5$  and  $\epsilon = 0.001$ , are compared in Table 3 with 2 well known methods, the Projection Method and the Alternating Method, see [7].

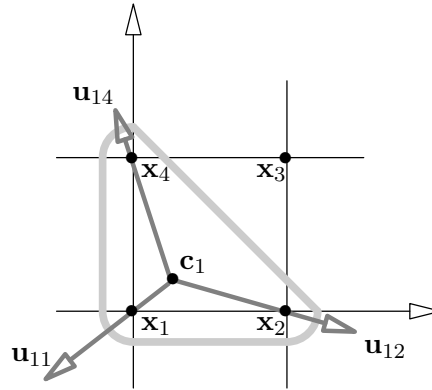


FIGURE 7. Illustration of Example 6: The resultant of the forces  $\mathbf{u}_{11}$ ,  $\mathbf{u}_{12}$  and  $\mathbf{u}_{14}$  is zero

Column 2 (“Best known  $f$ ”) gives the best values of the objective in the literature ([7], [9]), that are assumed optimal. Column 3 (“Best  $f$ ”) gives the best values obtained with Algorithm 1, and Column 4 (“% Error”) the percentage of the error in Column 3 relative to Column 2. For all 31 problems, Algorithm 1 obtained the optimal values, explaining the zeros in Column 4. In problems 27–31, marked by “†” in Column 3, Algorithm 1 gave somewhat better results than the best known values, perhaps because of more precise cluster centers. In problems 19, 20 and 25, marked by a “\*” in Column 3, getting the correct 3<sup>rd</sup>-decimal required a tolerance  $\epsilon = 0.0001$ . The results of the projection method and alternating method, [7], are similarly reported in Columns 5 and 6, and in Columns 7 and 8, respectively.

Table 4 reports, for  $\nu_0 = 5$ ,  $\Delta = 0.5$  and  $\epsilon = 0.001$ , the best values (“Min.  $f$  Value”) and the worst values (“Max.  $f$  Value”) obtained by Algorithm 1, and the percentage of cases (“Frequency (%)”) with these values. In problem 3 the best value was found in all runs.

The average values of all 500 solutions obtained are reported in Column 6 (“Average  $f$  Value”). Columns 7 and 8 give the average number of iterations, and the average CPU time, respectively. (Algorithm 1 is here tested using a program written in MATLAB R2010a and all experiments are conducted on a machine with 2.53 GHz Intel(R) Core(TM)2 Duo processor.)

Table 5 reports the same information as Table 4 but for an increment  $\Delta = 0.1$  and  $\epsilon = 0.0001$ .

**Remark 6.** A comparison of Table 4 (with  $\Delta = 0.5$ ,  $\epsilon = 0.001$ ) and Table 5 ( $\Delta = 0.1$ ,  $\epsilon = 0.0001$ ) shows that the smaller increment  $\Delta$  requires more iterations, but also increases the frequency of optimal values. To prevent early termination, a small tolerance  $\epsilon$  is suggested in conjunction with the small  $\Delta$ . This calls for further research, including other selections of the exponent  $\nu$ , in particular an adaptive control where the exponent  $\nu$  is updated according to the change  $\sum_{k=1}^K d(\mathbf{c}_k^+, \mathbf{c}_k)$  in Step 4 of Algorithm 1.

**Example 7.** This example uses data from Bongartz et al [7, p. 308], consisting of 287 points, with different weights ranging from  $w = 1$  to  $w = 698$ . The data points, shown in Fig. 8(a), are distinguished according to their weights. Customers with weight  $< 20$  (called “minor customers”) are represented by dots. The squares show the “major customers”: the white squares corresponding to customers with weights  $20 \leq w < 100$  and the gray squares to those with weights  $\geq 100$ .

TABLE 2. Data for Problems 1–31

#	Source	$N$	$K$	Coordinates $(x, y)$ of data points				
1	[11, Case 1]	7	2	(15, 15) (31, 23)	(5, 10) (22, 29)	(10, 27)	(16, 8)	(25, 14)
2	[11, Case 2]	7	2	(6, 8) (36, 8)	(6, 32) (36, 32)	(20, 8)	(20, 20)	(20, 32)
3	[11, Case 3]	7	2	(8, 12) (35, 26)	(5, 19) (35, 31)	(5, 26)	(5, 32)	(35, 20)
4	[11, Case 4]	7	2	(5, 23) (31, 32)	(9, 32) (16, 12)	(15, 23)	(21, 32)	(26, 23)
5	[11, Case 5]	7	2	(8, 10) (24, 17)	(8, 26) (31, 19)	(11, 20)	(17, 15)	(17, 22)
6	[11, Case 6]	7	2	(2, 19) (18, 16)	(35, 11) (11, 33)	(31, 33)	(25, 26)	(18, 23)
7	[11, Case 7]	7	2	(28, 6) (39, 33)	(28, 33) (6, 10)	(33, 17)	(33, 23)	(39, 6)
8	[41]	15	2					
9	[12], [41]	15	3	(5, 9)	(5, 25)	(5, 48)	(13, 4)	(12, 19)
10	[41]	15	4	(13, 39)	(28, 37)	(21, 45)	(25, 50)	(31, 9)
11	[41]	15	5	(39, 2)	(39, 16)	(45, 22)	(41, 30)	(49, 31)
12	[41]	15	6					
13	[41]	20	2					
14	[41]	20	3					
15	[41]	20	4	(53, 8)	(1, 34)	(33, 8)	(3, 26)	(17, 9)
16	[41]	20	5					
17	[41]	20	6					
18	[41]	25	2					
19	[41]	25	3					
20	[41]	25	4	(53, 20)	(24, 17)	(40, 22)	(22, 41)	(7, 13)
21	[41]	25	5					
22	[41]	25	6					
23	[41]	30	2					
24	[41]	30	3					
25	[41]	30	4	(5, 17)	(39, 3)	(50, 50)	(16, 40)	(22, 45)
26	[41]	30	5					
27	[21]	50	2	(1.33, 8.89)	(1.89, 0.77)	(9.27, 1.49)	(9.46, 9.36)	(9.20, 8.69)
28	[21]	50	3	(7.43, 1.61)	(6.08, 1.34)	(5.57, 4.60)	(6.70, 2.77)	(8.99, 2.45)
29	[21]	50	4	(8.93, 7.00)	(8.60, 0.53)	(4.01, 0.31)	(3.34, 4.01)	(6.75, 5.57)
30	[21]	50	5	(7.36, 4.03)	(1.24, 6.69)	(3.13, 1.92)	(8.86, 8.74)	(4.18, 3.74)
31	[39]	50	10	(2.22, 4.35)	(0.88, 7.02)	(8.53, 7.04)	(6.49, 6.22)	(4.53, 7.87)
				(4.46, 7.91)	(2.83, 9.88)	(3.39, 5.65)	(0.75, 4.98)	(7.55, 5.79)
				(8.45, 0.69)	(3.33, 5.78)	(6.27, 3.66)	(7.31, 1.61)	(6.37, 7.02)
				(7.23, 7.05)	(1.68, 6.45)	(3.54, 7.06)	(7.67, 4.17)	(2.20, 1.12)
				(3.57, 1.99)	(7.34, 1.38)	(6.58, 4.49)	(5.00, 9.00)	(6.63, 5.23)
				(5.89, 8.06)	(1.13, 5.25)	(1.90, 8.35)	(1.74, 1.37)	(9.37, 6.44)

The problem was solved for  $K = 2, \dots, 5$  using 200 random initializations for each  $K$ . The results are shown in Table 6. The optimal values in the 2nd column are as reported by [9]. The 3rd and 4th column

TABLE 3. Comparison of our method (for  $\nu_0 = 5$ ,  $\Delta = 0.5$  and  $\epsilon = 0.001$ ) with other methods

Problem #	Best known $f$	Algorithm 1		Projection Method [7]		Alternating Method [7]	
		Best $f$	% Error	Best $f$	% Error	Best $f$	% Error
1	50.449	50.449	0	52.001	3.076	50.449	0
2	72.000	72.000	0	72.000	0	72.000	0
3	38.318	38.318	0	38.318	0	38.318	0
4	48.850	48.850	0	48.850	0	48.850	0
5	38.033	38.033	0	38.033	0	38.033	0
6	59.716	59.716	0	59.716	0	61.935	3.716
7	62.203	62.203	0	62.203	0	62.203	0
8	214.281	214.281	0	214.281	0	214.281	0
9	143.196	143.196	0	143.196	0	143.196	0
10	113.568	113.568	0	113.568	0	113.568	0
11	97.289	97.289	0	99.091	1.582	99.091	1.852
12	81.264	81.264	0	81.264	0	83.066	2.217
13	287.606	287.606	0	287.606	0	287.606	0
14	210.195	210.195	0	210.195	0	210.195	0
15	169.434	169.434	0	169.434	0	169.434	0
16	134.263	134.263	0	134.952	0.513	134.952	0.513
17	116.312	116.312	0	116.312	0	116.312	0
18	355.336	355.336	0	355.336	0	355.336	0
19	252.245	252.246*	0	252.245	0	252.245	0
20	207.402	207.403*	0	207.402	0	208.054	0.314
21	169.832	169.832	0	169.832	0	169.832	0
22	152.664	152.664	0	153.673	0.660	153.720	0.692
23	447.727	447.727	0	447.727	0	447.727	0
24	307.372	307.372	0	307.372	0	307.372	0
25	254.146	254.147*	0	254.146	0	254.146	0
26	220.056	220.056	0	220.056	0	220.056	0
27	135.522	135.513†	0	135.546	0.018	135.546	0.018
28	105.214	105.194†	0	105.214	0	105.214	0
29	84.154	84.134†	0	84.154	0	84.154	0
30	72.237	72.225†	0	72.237	0	72.237	0
31	41.685	41.673†	0	42.689	2.409	42.689	2.409

give, respectively, the best and average values found in 200 random initializations. The last two columns give the average number of iterations and the average CPU time.  $\square$

**Example 8.** To illustrate the robustness of Algorithm 1, see Remark 2(c), we solved Example 7 twice for each  $K$ : once with all 287 customers, and again with only the 61 major customers. The same center

TABLE 4. Results for  $\nu_0 = 5$ ,  $\Delta = 0.5$  and  $\epsilon = 0.001$ 

Problem #	Min. $f$ Value	Frequency (%)	Max. $f$ Value	Frequency (%)	Average $f$ Value	Average # of Iterations	Average CPU Time
1	50.449	49.80	58.500	3.40	51.451	53.086	0.0174
2	72.000	71.40	87.122	0.80	74.680	30.978	0.0100
3	38.318	100.00	38.318	n.a.	38.318	341.998	0.2578
4	48.850	93.00	57.650	1.00	49.358	41.158	0.0099
5	38.033	91.20	45.456	5.30	38.611	32.606	0.0079
6	59.716	13.20	65.878	0.20	62.276	42.384	0.0139
7	62.203	72.80	68.968	3.60	62.983	68.300	0.0175
8	214.281	85.80	232.616	0.40	216.833	62.664	0.0281
9	143.196	91.40	185.387	0.40	145.485	105.056	0.0693
10	113.568	58.00	132.522	3.00	119.497	60.202	0.0509
11	97.289	6.80	122.556	0.20	103.448	46.246	0.0488
12	81.264	5.60	110.514	0.10	89.342	39.548	0.0490
13	287.606	61.60	291.390	0.60	288.202	73.440	0.0408
14	210.195	50.20	271.481	0.40	217.033	57.996	0.0449
15	169.434	21.40	218.306	0.60	174.801	62.116	0.0626
16	134.263	23.40	173.498	0.20	146.502	64.622	0.0845
17	116.313	8.20	167.773	0.20	126.854	63.480	0.0971
18	355.336	97.40	408.860	2.60	356.728	72.600	0.0470
19	252.246	77.40	334.403	0.20	257.194	71.084	0.0688
20	207.403	49.60	244.876	0.40	210.585	94.564	0.1251
21	169.832	57.00	217.795	0.80	180.231	109.598	0.1725
22	152.664	3.60	201.719	0.20	160.838	102.720	0.1936
23	447.727	73.20	449.485	6.80	447.898	92.370	0.0694
24	307.372	99.00	307.849	1.00	307.382	85.920	0.0936
25	254.147	62.60	285.587	0.60	260.664	114.000	0.1640
26	220.056	27.80	270.470	0.20	229.815	87.400	0.1565
27	135.513	0.60	144.559	0.20	136.762	52.900	0.0559
28	105.194	1.20	113.798	0.20	106.572	51.950	0.0853
29	84.134	1.80	95.594	0.20	86.705	83.262	0.1844
30	72.225	1.40	89.901	0.40	73.983	45.156	0.1135
31	41.673	1.00	55.552	0.20	44.915	59.574	0.3284

locations were found in both cases. The solution for  $K = 3$  is shown in Fig. 8(b), together with the major customers.  $\square$

## REFERENCES

- [1] M. Arav, Contour approximation of data and the harmonic mean, *J. Math. Inequalities* **2**(2008), 161–167

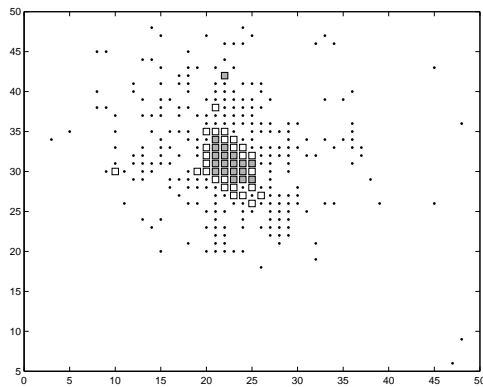
TABLE 5. Results for  $\nu_0 = 5$ ,  $\Delta = 0.1$  and  $\epsilon = 0.0001$ 

Problem #	Min. $f$ Value	Frequency (%)	Max. $f$ Value	Frequency (%)	Average $f$ Value	Average # of Iterations	Average CPU Time
1	50.449	55.40	58.501	3.40	51.410	411.688	0.3341
2	72.000	74.40	81.764	4.00	74.308	61.832	0.0213
3	38.318	100.00	38.318	100.00	38.318	973.155	0.8203
4	48.850	95.40	57.650	4.60	49.255	121.252	0.0442
5	38.033	86.80	48.068	0.20	38.748	31.042	0.0102
6	59.716	17.40	65.871	0.20	61.973	188.906	0.0931
7	62.203	79.60	68.968	3.40	62.798	62.342	0.0209
8	214.281	89.40	232.226	10.60	216.183	362.698	0.3100
9	143.196	97.80	172.188	2.20	143.834	192.270	0.1605
10	113.568	59.00	132.522	6.20	119.426	135.108	0.1252
11	97.289	9.60	123.306	0.40	103.133	172.550	0.2100
12	81.264	10.80	107.289	0.20	88.511	179.946	0.2588
13	287.606	95.80	289.123	4.20	287.670	379.588	0.3659
14	210.195	65.00	234.675	13.40	215.914	217.610	0.2206
15	169.434	27.40	218.306	0.80	173.303	205.930	0.2470
16	134.263	55.40	178.164	0.20	144.668	206.210	0.3063
17	116.312	14.20	153.638	0.20	126.082	197.458	0.3540
18	355.336	99.00	408.860	1.00	355.871	371.790	0.4123
19	252.246	93.20	305.447	0.80	254.077	392.712	0.5595
20	207.403	62.60	241.876	1.20	210.344	483.272	0.8626
21	169.832	63.40	217.193	0.20	179.344	366.894	0.7142
22	152.664	4.60	193.662	0.20	160.075	310.980	0.6723
23	447.727	100.00	447.727	100.00	447.727	567.000	0.8181
24	307.372	100.00	307.372	100.00	307.372	408.000	0.6284
25	254.146	65.80	280.952	8.60	260.910	673.274	1.5900
26	220.056	29.80	271.407	0.20	229.094	471.118	1.2044
27	135.520	100.00	135.520	100.00	135.520	283.514	0.3847
28	105.199	81.60	110.533	18.40	106.177	283.328	0.5437
29	84.134	74.60	94.274	0.60	86.225	537.374	1.5336
30	72.226	4.60	89.898	0.40	73.556	357.688	1.1884
31	41.674	0.40	55.218	0.20	44.725	231.574	1.3875

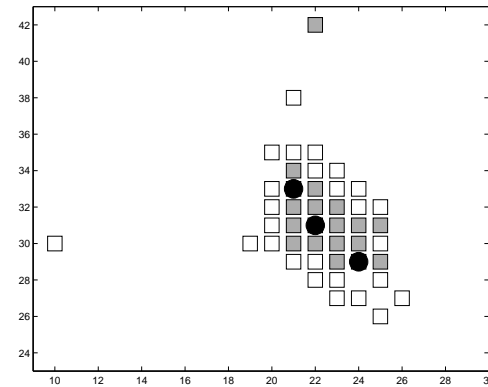
- [2] F. Aurenhammer, Voronoi diagrams - a survey of a fundamental geometric data structure, *ACM Computing Surveys* **23**(1991), 345–405
- [3] E. Balas and C-S Yu, A note on the Weiszfeld-Kuhn algorithm for the general Fermat problem, Tech. Report, Carnegie-Melon University, September 1982
- [4] A. Ben-Israel and C. Iyigun, Probabilistic distance clustering, *J. Classification* **25**(2008), 5–26
- [5] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981

TABLE 6. Results for Example 7,  $\nu_0 = 5$  and  $\Delta = 0.1$ 

$K$	Optimum Value [9]	Best $f$ Value	Error (%)	Average $f$ Value	Error (%)	Average # of Iterations	Average CPU
2	14427.593	14447.970	0.14	14497.313	0.33	297.22	1.9297
3	12095.442	12115.823	0.17	12354.682	2.14	203.02	1.8982
4	10661.477	10681.842	0.19	11066.700	3.80	192.83	2.3884
5	9715.627	9735.993	0.21	10041.572	3.35	173.73	2.7298
10	6705.036	6.714.277	0.14	7275.055	8.50	202.30	7.7752



(a) 287 data points



(b) Major customers and centers

FIGURE 8. Illustration of Example 8 ( $N = 287$ ,  $K = 3$ )

- [6] L. Boitani and T.K. Fuller, *Research techniques in animal ecology: controversies and consequences*, Columbia University Press, New York, 2000
- [7] I. Bongartz, E.H. Calamai and A.R. Conn, A projection method for  $\ell_p$  norm location-allocation problems, *Math. Programming* **66**(1994), 283–312.
- [8] J. Brimberg, Further notes on convergence of the Weiszfeld algorithm, *Yugoslav J. Operations Research* **13**(2003), 199–206
- [9] J. Brimberg, P. Hansen, N. Mladenović and E.D. Taillard, Improvements and comparison of heuristics for solving the uncapacitated multisource Weber problem, *Operations Research* **48**(2000), 444–460
- [10] R. Chandrasekaran and A. Tamir, Open questions concerning Weiszfeld’s algorithm for the Fermat–Weber location problem, *Math. Programming* **44**(1989), 293–295
- [11] L. Cooper, Location-allocation problems, *Operations Research* **11**(1963), 331–343.
- [12] L. Cooper, Heuristic methods for location-allocation problems, *SIAM Review* **6**(1964), 37–53
- [13] K. R. Dixon and J. A. Chapman, Harmonic mean measure of animal activity areas, *Ecology* **61**(1980), 1040–1044
- [14] T. Drezner, Z. Drezner, The gravity  $p$ -median model, *European J. Oper. Res.* **179**(2007), 1239–1251.
- [15] Z. Drezner, The planar two-center and two-median problems, *Transportation Science* **18**(1984), 351–361.
- [16] Z. Drezner (editor), *Location Decisions*, *Annals Oper. Res.* **40**(1992)
- [17] Z. Drezner, A note on the Weber location problem, pp. 153–161 in [16]
- [18] Z. Drezner and H.W. Hamacher (editors), *Facility Location: Applications and Theory*, Springer, 2002
- [19] Z. Drezner, K. Klamroth, A. Schöbel and G.O. Wesolowsky, The Weber problem, Chapter 1 in [18]
- [20] I.N. Eckhardt, Weber’s problem and Weiszfeld’s algorithm in general spaces, *Math. Program.* **18**(1980), 186–196
- [21] S. Eilon, C.D.T. Watson–Gandy and N. Christofides, *Distribution Management: Mathematical Modelling and Practical Analysis*, Hafner, New York, 1971
- [22] J.S. Horne and E.O. Garton, Selecting the best home range model: an information-theoretic approach, *Ecology* **87**(2006), 1146–1152

- [23] D.L. Huff, Defining and estimating a trading area, *J. Marketing* **28**(1964), 34–38
- [24] D.L. Huff, A programmed solution for approximating an optimum retail location, *Land Economics* **42**(1966), 293–303.
- [25] C. Iyigun and A. Ben-Israel, Probabilistic distance clustering adjusted for cluster size, *Probability Engrg. Info. Sci.* **22**(2008), 1–19
- [26] C. Iyigun and A. Ben-Israel, Contour approximation of data: a duality theory, *Lin. Algeb. and Appl.* **430**(2009), 2771–2780
- [27] C. Iyigun and A. Ben-Israel, A generalized Weiszfeld method for the multi-facility location problem, *Operations Research Letters* **38**(2010), 207–214
- [28] W. Kaplan and W.H. Yang, Duality theorem for a generalized Fermat–Weber problem, *Math. Programming* **76**(1997), 285–297
- [29] I.N. Katz, Local convergence in Fermat’s problem, *Math. Program.* **6**(1974), 89–104
- [30] H.W. Kuhn, On a pair of dual nonlinear programs, in J. Abadie (ed.), *Methods of Nonlinear Programming*, Amsterdam, North-Holland, (1967), 38–54
- [31] H.W. Kuhn, A note on Fermat’s problem, *Math. Programming* **4**(1973), 98–107
- [32] S. Kullback. *Information Theory and Statistics*. J. Wiley, New York, 1959.
- [33] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals Math. Statist.*, **22**(1951), 79–86
- [34] Y. Levin and A. Ben-Israel, A heuristic method for large-scale multi-facility location problems, *Computers and Oper. Res.* **31**(2004), 257–272
- [35] R.F. Love and H. Juel, Properties and solution methods for large location–allocation problems, *J. Operational Research Society* **33**(1982), 443–452
- [36] R.F. Love, J.G. Morris and G.O. Wesolowsky, *Facilities Location: Models and Methods*, North-Holland, 1988
- [37] R.D. Luce, *Individual Choice Behavior*, Wiley, New York, 1959
- [38] N. Megiddo and K.J. Supowit, On the complexity of some common geometric location problems, *SIAM Journal on Computing* **13**(1984), 182–196
- [39] B.A. Murtagh and S.R. Niwattisyawong, An efficient method for the multi-depot location–allocation problem, *Journal of Operational Research Society* **33** (1982) 629–634.
- [40] R.A. Powell, *Animal home ranges and territories and home range estimators*, Chapter 3 in [6]
- [41] K.E. Rosing, An optimal method for solving the multi-Weber problems, *European Journal of Operational Research* **58** (1992) 414–426.
- [42] M.D. Samuel, D.J. Pierce and E.O. Garton, Identifying areas of concentrated use within the home range, *Journal of Animal Ecology* **54**(1985), 711–719
- [43] M. Teboulle, A unified continuous optimization framework for center-based clustering methods, *J. Machine Learning* **8**(2007), 65–102
- [44] H. Üster and R.F. Love, Duality in constrained multi-facility location models, *Naval Research Logistics Quart.* **49**(2002), 410–421
- [45] Y. Vardi and C-H Zhang, A modified Weiszfeld algorithm for the Fermat–Weber location problem, *Math. Programming* **A90**(2001), 559–566
- [46] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnés est minimum, *Tohoku Math. J.* **43** (1937), 355–386
- [47] G.O. Wesolowsky, The Weber problem: its history and perspectives, *Location Science* **1**(1993) 5–23
- [48] J.I. Yellott, Jr., Luce’s Choice Axiom, pp. 9094–9097 in *International Encyclopedia of the Social & Behavioral Sciences* (N.J. Smelser and P.B. Baltes, editors), ISBN 0-08-043076-7, 2001.

## Appendix A: 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 (32) 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,

$$\frac{d}{dt} f_k(\mathbf{x}_j + t \mathbf{z}) \Big|_{t=0} = w_j q_k(\mathbf{x}_j) - \mathbf{R}_k^j \cdot \mathbf{z} . \quad (\text{A1})$$

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

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

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

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

which is equivalent to

$$\|\mathbf{R}_k^j\| \leq w_j q_k(\mathbf{x}_j),$$

or  $\mathbf{R}_k(\mathbf{c}_k) = \mathbf{0}$ , by (24b).

**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 (32), results in (21a)–(21b) as above.

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

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

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

**Proof of part (d).** If  $T_k(\mathbf{c}) \neq \mathbf{c}$  then  $\mathbf{c}$  is not one of the data points. It follows then from (22a) that  $T_k(\mathbf{c})$  is the center of gravity of weights

$$\theta_{ki} := \frac{w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{c}\|}, \text{ placed at the data points } \mathbf{x}_i, i \in 1:N. \quad (\text{A2})$$

By elementary calculus,  $T_k(\mathbf{c})$  is the unique minimum of the strictly convex function

$$g(\mathbf{y}) = \sum_{i=1}^N \theta_{ki} \|\mathbf{x}_i - \mathbf{y}\|^2.$$

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

$$g(T_k(\mathbf{c})) = \sum_{i=1}^N \theta_{ki} \|\mathbf{x}_i - T_k(\mathbf{c})\|^2 < g(\mathbf{c}) = \sum_{i=1}^N \theta_{ki} \|\mathbf{x}_i - \mathbf{c}\|^2 = f_k(\mathbf{c}).$$

On the other hand,

$$\begin{aligned} g(T_k(\mathbf{c})) &= \sum_{i=1}^N \theta_{ki} \left[ \|\mathbf{x}_i - \mathbf{c}\| + \left( \|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\| \right) \right]^2 \\ &= f_k(\mathbf{c}) + 2 \left( f_k(T_k(\mathbf{c})) - f_k(\mathbf{c}) \right) + \sum_{i=1}^N \theta_{ki} \left[ \|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\| \right]^2 \end{aligned}$$

Combining these results,

$$2f_k(T_k(\mathbf{c})) + \sum_{i=1}^N \theta_{ki} \left[ \|\mathbf{x}_i - T_k(\mathbf{c})\| - \|\mathbf{x}_i - \mathbf{c}\| \right]^2 < 2f_k(\mathbf{c})$$

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

**Proof of part (e).** Follows from part (a), the definition (24b)

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

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

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

**Proof of part (g).**

$$\begin{aligned} T_k(\mathbf{c}) - \mathbf{x}_i &= \mathbf{c} + h_k(\mathbf{c}) R_k(\mathbf{c}) - \mathbf{x}_i \\ &= h_k(\mathbf{c}) \sum_{j \neq i}^N \theta_{kj} (\mathbf{x}_j - \mathbf{c}) + \left( h_k(\mathbf{c}) \theta_{ki} - 1 \right) (\mathbf{x}_i - \mathbf{c}), \end{aligned}$$

with  $h_k(\mathbf{c})$  as in (31) and  $\theta_{ki}$  in (A2). Since  $\mathbf{x}_i$  is not optimal, we have

$$\left\| \sum_{j \neq i}^N \theta_{kj} (\mathbf{x}_j - \mathbf{x}_i) \right\| > w_i q_k(\mathbf{x}_i)$$

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

$$\left\| \sum_{j \neq i}^N \theta_{kj} (\mathbf{x}_j - \mathbf{c}) \right\| \geq (1 + 2\epsilon) w_i q_k(\mathbf{x}_i) \quad \text{for } \|\mathbf{x}_i - \mathbf{c}\| \leq \delta'.$$

By the definition of  $h_k$ , we have

$$\lim_{\mathbf{c} \rightarrow \mathbf{x}_i} \frac{h_k(\mathbf{c}) w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{c}\|} = 1.$$

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

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

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

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

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  $\|\mathbf{x}_i - T_k^{s-1}(\mathbf{c})\| \leq \delta$ .

**Proof of part (h).** For  $\mathbf{c} \notin \{\mathbf{x}_i : i \in 1:N\}$ ,

$$\begin{aligned} T_k(\mathbf{c}) &= \sum_{i=1}^N \lambda_{ki} \mathbf{x}_i, \quad \text{with } \lambda_{ki} \text{ as in (22a),} \\ &= \sum_{i \neq j}^N \lambda_{ki} (\mathbf{x}_i - \mathbf{x}_j) + \mathbf{x}_j \sum_{i=1}^N \lambda_{ki}. \\ \therefore T_k(\mathbf{c}) - \mathbf{x}_j &= \sum_{i \neq j}^N \lambda_{ki} (\mathbf{x}_i - \mathbf{x}_j), \quad \text{since } \sum_{i=1}^N \lambda_{ki} = 1. \\ \therefore \frac{T_k(\mathbf{c}) - \mathbf{x}_j}{\|\mathbf{x}_j - \mathbf{c}\|} &= \sum_{i \neq j}^N \frac{\lambda_{ki}}{\|\mathbf{x}_j - \mathbf{c}\|} (\mathbf{x}_i - \mathbf{x}_j). \end{aligned} \tag{A3}$$

Using (22a),

$$\begin{aligned} \frac{\lambda_{ki}}{\|\mathbf{x}_j - \mathbf{c}\|} &= \frac{\frac{w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{m=1}^N w_m q_k(\mathbf{x}_m) \frac{\|\mathbf{x}_j - \mathbf{c}\|}{\|\mathbf{x}_m - \mathbf{c}\|}} = \frac{\frac{w_i q_k(\mathbf{x}_i)}{\|\mathbf{x}_i - \mathbf{c}\|}}{w_j q_k(\mathbf{x}_j) + \sum_{m \neq j} w_m q_k(\mathbf{x}_m) \frac{\|\mathbf{x}_j - \mathbf{c}\|}{\|\mathbf{x}_m - \mathbf{c}\|}} \\ \therefore \lim_{\mathbf{c} \rightarrow \mathbf{x}_j} \frac{\lambda_{ki}}{\|\mathbf{x}_j - \mathbf{c}\|} &= \frac{w_i q_k(\mathbf{x}_i)}{w_j q_k(\mathbf{x}_j) \|\mathbf{x}_i - \mathbf{c}\|}. \end{aligned}$$

Finally, by (A3),

$$\begin{aligned} \lim_{\mathbf{c} \rightarrow \mathbf{x}_j} \frac{\|T_k(\mathbf{c}) - \mathbf{x}_j\|}{\|\mathbf{x}_j - \mathbf{c}\|} &= \frac{1}{w_j q_k(\mathbf{x}_j)} \left\| \sum_{i \neq j}^N w_i q_k(\mathbf{x}_i) \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|} \right\| \\ &= \frac{\|\mathbf{R}_k^j\|}{w_j q_k(\mathbf{x}_j)}, \text{ by (24c)}. \end{aligned}$$

**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 exists a convergent subsequence  $\{\mathbf{c}_k^\ell : \ell = 1, 2, \dots\}$ , such that  $\lim_{\ell \rightarrow \infty} \mathbf{c}_k^\ell = \mathbf{c}_k^\infty$ . 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  $\ell$ , 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),

$$f_k(\mathbf{c}_k^0) > f_k(\mathbf{c}_k^1) > \dots > f_k(\mathbf{c}_k^r) > \dots > f_k(\mathbf{c}_k^*) \quad (\text{A4})$$

Hence

$$\lim_{\ell \rightarrow \infty} (f_k(\mathbf{c}_k^\ell) - f_k(T_k(\mathbf{c}_k^\ell))) = 0. \quad (\text{A5})$$

The continuity of  $T_k$  implies

$$\lim_{\ell \rightarrow \infty} T_k(\mathbf{c}_k^\ell) = T_k(\mathbf{c}_k^\infty) \quad (\text{A6})$$

we have

$$f_k(\mathbf{c}_k^\infty) - f_k(T_k(\mathbf{c}_k^\infty)) = 0. \quad (\text{A7})$$

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  $\{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{c}_k^*\}$ , 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  $\delta$ -neighborhood that satisfies Theorem 1(g). Then it is clear that we can choose our subsequence  $\mathbf{c}_k^\ell \rightarrow \mathbf{x}_j$  such that  $\|\mathbf{x}_j - T_k(\mathbf{c}_k^\ell)\| > \delta$  for all  $\ell$ . 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.  $\square$

## Appendix B: Proof of Theorem 2

**Proof of part (a).** The objective function  $g(\mathbf{U})$  can be written, using (42), as

$$g(\mathbf{U}) = \sum_{k=1}^K \left( \sum_{i=1}^N \mathbf{u}_{ki} \cdot \mathbf{x}_i - \left( \sum_{i=1}^N \mathbf{u}_{ki} \right) \cdot \mathbf{c}_k \right) = \sum_{k=1}^K \sum_{i=1}^N \mathbf{u}_{ki} \cdot (\mathbf{x}_i - \mathbf{c}_k). \quad (\text{B1})$$

$$\leq \sum_{k=1}^K \sum_{i=1}^N \|\mathbf{u}_{ki}\| \|\mathbf{x}_i - \mathbf{c}_k\| \quad (\text{B2})$$

$$\leq \sum_{k=1}^K \sum_{i=1}^N w_i q_{ki} \|\mathbf{x}_i - \mathbf{c}_k\|, \text{ by (43),} \quad (\text{B3})$$

$$= f(\mathbf{c}_1, \dots, \mathbf{c}_K).$$

**Proof of part (b).** We distinguish two cases.

**Case 1.** None of the centers  $\{\mathbf{c}_k\}$  coincides with any of the data points  $\{\mathbf{x}_i\}$ .

For  $k \in 1:K$  and  $i \in 1:N$  define

$$\mathbf{u}_{ki} := \frac{w_i q_{ki}}{\|\mathbf{x}_i - \mathbf{c}_k\|} (\mathbf{x}_i - \mathbf{c}_k). \quad (\text{B4})$$

Then from (24a) and Theorem 1 it follows that,

$$\sum_{i=1}^N \mathbf{u}_{ki} = \mathbf{R}_k(\mathbf{c}_k) = \mathbf{0}, \text{ verifying (42).}$$

Also from (B4), we get, for all  $k, i$ ,

$$\|\mathbf{u}_{ki}\| = w_i q_{ki}, \quad (\text{B5})$$

proving that the inequalities (43) hold as equalities, and that  $\{\mathbf{u}_{ki}\}$ , defined by (B4), are feasible.

Finally, from (B4) it follows that

$$\mathbf{u}_{ki} \cdot (\mathbf{x}_i - \mathbf{c}_k) = w_i q_{ki} \|\mathbf{x}_i - \mathbf{c}_k\|, \quad (\text{B6})$$

and (45) follows from (B1).

**Case 2.** A center coincides with one of the data points, say

$$\mathbf{c}_k = \mathbf{x}_j, \text{ for some } k \in 1:K, j \in 1:N, \quad (\text{B7})$$

in which case (12) holds. Define

$$\mathbf{u}_{ki} := \frac{w_i q_{ki}}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j), \quad \text{for } i \neq j, \quad (\text{B8a})$$

$$\mathbf{u}_{kj} := - \sum_{i \neq j} \mathbf{u}_{ki}. \quad (\text{B8b})$$

Then  $\sum_i \mathbf{u}_{ki} = \mathbf{0}$  by definition, and  $\|\mathbf{u}_{ki}\| = w_i q_{ki}$  for all  $i \neq j$ , as in (B5). Next,

$$\mathbf{u}_{kj} = -\mathbf{R}_k^j \text{ by (24c),}$$

and therefore by (24b),

$$\mathbf{R}_k(\mathbf{x}_j) = \mathbf{0} \text{ implies } w_j q_{kj} \geq \|\mathbf{R}_k^j\| = \|\mathbf{u}_{kj}\|,$$

proving that the variables  $\mathbf{U}$  defined by (B8a)–(B8b) are feasible.

From Case 1, equality (B6) holds for all  $\mathbf{x}_i$  that are not centers. Also, for any center  $\mathbf{c}_m \neq \mathbf{c}_k$ ,  $q_{mj} = 0$  (by (12)), and therefore  $\mathbf{u}_{mj} = \mathbf{0}$ . The inequalities (B2)–(B3) therefore reduce to

$$\mathbf{u}_{kj} \cdot (\mathbf{x}_j - \mathbf{c}_k) \leq \|\mathbf{u}_{kj}\| \|\mathbf{x}_j - \mathbf{c}_k\| \leq w_j q_{kj} \|\mathbf{x}_j - \mathbf{c}_k\|,$$

which become trivial equalities, since by (B7) all three terms are zero.

The inequalities (B2)–(B3) therefore hold as equalities, proving (45).

**Proof of part (c).** Writing the objective function  $g(\mathbf{U})$  as in (B1), the Lagrangian of (D) is

$$\sum_{k=1}^K \sum_{i=1}^N \mathbf{u}_{ki} \cdot (\mathbf{x}_i - \mathbf{c}_k) - \sum_{k=1}^K \sum_{i=1}^N t_{ki} (\|\mathbf{u}_{ki}\| - w_i q_{ki})$$

with Lagrange multipliers  $\{t_{ki}\}$ . The Karush–Kuhn–Tucker conditions are

$$(\mathbf{x}_i - \mathbf{c}_k) - t_{ki} \frac{\mathbf{u}_{ki}}{\|\mathbf{u}_{ki}\|} = \mathbf{0}, \quad (\text{B9a})$$

$$\sum_{i=1}^N \mathbf{u}_{ki} = \mathbf{0}, \quad (\text{B9b})$$

$$\|\mathbf{u}_{ki}\| \leq w_i q_{ki}, \quad (\text{B9c})$$

$$t_{ki} \geq 0, \quad (\text{B9d})$$

$$t_{ki} (\|\mathbf{u}_{ki}\| - w_i q_{ki}) = 0, \quad (\text{B9e})$$

for all  $k \in 1:K, i \in 1:N$ . Again we distinguish two cases.

**Case 1.** All  $t_{ki} > 0$ . Then (B5) follows from (B9e) for all  $k, i$ , and from (B9a),

$$\|\mathbf{x}_i - \mathbf{c}_k\| = t_{ki}, \quad (\text{B10})$$

and by substituting (B5) and (B10) in (B9a), we get

$$\mathbf{u}_{ki} := \frac{w_i q_{ki}}{\|\mathbf{x}_i - \mathbf{c}_k\|} (\mathbf{x}_i - \mathbf{c}_k),$$

which is (B4), and the equality (45) follows as in the proof of Theorem 2(b), Case 1.

**Case 2.** Some Lagrange multipliers are zero, say  $t_{kj} = 0$ . Then by (B9a),  $\mathbf{c}_k = \mathbf{x}_j$ , and  $t_{ki} > 0$  for  $i \neq j$ , by (B9a) and (B9d). Therefore, by (B9e),

$$\|\mathbf{u}_{ki}\| = w_i q_{ki}, \quad \text{for all } i \neq j.$$

From  $\mathbf{c}_k = \mathbf{x}_j$  and (B9a) it follows that

$$(\mathbf{x}_i - \mathbf{x}_j) = t_{ki} \frac{\mathbf{u}_{ki}}{\|\mathbf{u}_{ki}\|}$$

and by taking norms,

$$\|\mathbf{x}_i - \mathbf{x}_j\| = t_{ki}, \quad \text{for all } i \neq j.$$

Substituting  $t_{ki}$  and  $\|\mathbf{u}_{ki}\|$  in (B9a) gives,

$$\mathbf{u}_{ki} := \frac{w_i q_{ki}}{\|\mathbf{x}_i - \mathbf{x}_j\|} (\mathbf{x}_i - \mathbf{x}_j), \quad \text{for } i \neq j,$$

and from (B9b),

$$\mathbf{u}_{kj} := - \sum_{i \neq j} \mathbf{u}_{ki},$$

reproducing (B8a)–(B8b), and equality in (45) follows as in the proof of Theorem 2(b), Case 2.  $\square$

### Appendix C: The classification uncertainty function

Let  $\mathbb{P}^K$  be the set of  $K$ -dimensional probability vectors, denoted  $\mathbf{p} = (p_i)$ ,  $\mathbf{q} = (q_i)$ . The **Kullback–Leibler divergence**,

$$I(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^K p_i \log \left( \frac{p_i}{q_i} \right), \quad (\text{C1})$$

is a well-known and useful distance function on  $\mathbb{P}^K$ , [32], [33].

Rewriting the CUF (38) as

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

and taking logarithms, we get

$$-\log E(\mathbf{x}) = \sum_{i \in \mathcal{K}} (1/K) \log \left( \frac{1/K}{p_i(\mathbf{x})} \right) = I(\mathbf{p}(\mathbf{x}), \frac{1}{K} \mathbf{1}), \quad (\text{C2})$$

the Kullback–Leibler distance between the distributions

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

The latter distribution,  $\frac{1}{K} \mathbf{1}$ , is of maximal uncertainty in  $\mathbb{P}^K$ , and consequently the divergence  $I(\mathbf{p}(\mathbf{x}), \frac{1}{K} \mathbf{1})$  is a measure of the uncertainty of the distribution  $\mathbf{p}(\mathbf{x})$ , with smaller values corresponding to greater uncertainty.

Writing (C2) as

$$E(\mathbf{x}) = \exp \left\{ -I(\mathbf{p}(\mathbf{x}), \frac{1}{K} \mathbf{1}) \right\} \quad (\text{C3})$$

shows that  $E(\mathbf{x})$  is an entropic measure of the uncertainty of classification, a monotone increasing function of the uncertainty.

### Appendix D: Cluster membership probabilities

In this appendix,  $d_k(\mathbf{x})$  stands for  $d_k(\mathbf{x}, \mathbf{c}_k)$ , the distance of  $\mathbf{x}$  to the center  $\mathbf{c}_k$  of the  $k^{\text{th}}$ -cluster.

The **membership probabilities**  $\{p_k(\mathbf{x}) : k \in 1:K\}$  of a point  $\mathbf{x}$  are assumed to depend only on the **distances**  $\{d_k(\mathbf{x}) : k \in 1:K\}$ ,

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \quad (\text{D-1})$$

where  $\mathbf{p}(\mathbf{x}) \in \mathbb{R}^K$  is the vector of probabilities ( $p_k(\mathbf{x})$ ), and  $\mathbf{d}(\mathbf{x})$  is the vector of distances ( $d_k(\mathbf{x})$ ). Desirable properties of the relation (D-1) include

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \text{ for all } i, j \in 1:K, \quad (\text{D-2a})$$

$$\mathbf{f}(\lambda \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0, \quad (\text{D-2b})$$

$$Q \mathbf{p}(\mathbf{x}) = \mathbf{f}(Q \mathbf{d}(\mathbf{x})), \text{ for any permutation matrices } Q. \quad (\text{D-2c})$$

Condition (D-2a) states that membership in a cluster is more probable the closer it is, which is Assumption (A) of § 4. Assuming continuity of  $\mathbf{f}$  it follows from (D-2a) that

$$d_i(\mathbf{x}) = d_j(\mathbf{x}) \implies p_i(\mathbf{x}) = p_j(\mathbf{x}),$$

for any  $i, j \in 1:K$ . In particular, the probabilities  $p_k(\mathbf{x})$  are all equal only if so are the distances  $d_k(\mathbf{x})$ .

The meaning of (D-2b) is that the probabilities  $p_k(\mathbf{x})$  do not depend on the scale of measurement, i.e., the function  $\mathbf{f}$  is homogeneous of degree 0. It follows that the probabilities  $p_k(\mathbf{x})$  depend only on the ratios of the distances  $\{d_k(\mathbf{x}) : k \in 1:K\}$ , as is assumed in what follows.

The symmetry of  $\mathbf{f}$ , expressed by (D-2c), guarantees that the probabilities  $\{p_k(\mathbf{x})\}$  do not depend on the numbering of the clusters.

For any nonempty subset  $\mathcal{S} \subset 1:K$ , let

$$p_{\mathcal{S}}(\mathbf{x}) = \sum_{s \in \mathcal{S}} p_s(\mathbf{x}),$$

the probability that  $\mathbf{x}$  belongs to one of the clusters  $\{\mathbf{X}_s : s \in \mathcal{S}\}$ , and let  $p_k(\mathbf{x}|\mathcal{S})$  denote the **conditional probability** that  $\mathbf{x}$  belongs to the cluster  $\mathbf{X}_k$ , given that it belongs to one of the clusters  $\{\mathbf{X}_s : s \in \mathcal{S}\}$ .

Since the probabilities  $p_k(\mathbf{x})$  depend only on the ratios of the distances  $\{d_k(\mathbf{x}) : k \in 1:K\}$ , and these ratios are unchanged in subsets  $\mathcal{S}$  of the index set  $1:K$ , it follows that for all  $k \in \mathcal{S} \subset 1:K$ ,

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|\mathcal{S}) p_{\mathcal{S}}(\mathbf{x}) \quad (\text{D-3})$$

which is the **choice axiom** of Luce, [37, Axiom 1], and therefore, [48],

$$p_k(\mathbf{x}|\mathcal{S}) = \frac{v_k(\mathbf{x})}{\sum_{s \in \mathcal{S}} v_s(\mathbf{x})} \quad (\text{D-4})$$

where  $v_k(\mathbf{x})$  is a scale function, in particular,

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum_{s \in \mathbf{K}} v_s(\mathbf{x})}. \quad (\text{D-5})$$

Assuming  $v_k(\mathbf{x}) \neq 0$  for all  $k$ , it follows that

$$p_k(\mathbf{x}) v_k(\mathbf{x})^{-1} = \frac{1}{\sum_{s \in \mathbf{K}} v_s(\mathbf{x})}, \quad (\text{D-6})$$

where the right hand side is a function of  $\mathbf{x}$ , and does not depend on  $k$ .

Property (D-2a) implies that the function  $v_k(\cdot)$  is monotone decreasing. A simple choice is

$$v_k(\mathbf{x}) = \frac{1}{w d_k(\mathbf{x})}, \quad (\text{D-7})$$

for which (D-6) gives

$$w p_k(\mathbf{x}) d_k(\mathbf{x}) = \frac{1}{\sum_{s \in \mathbf{K}} \left( \frac{1}{d_s(\mathbf{x})} \right)}, \quad (\text{D-8})$$

in agreement with (8)–(9).

CEM IYIGUN, DEPARTMENT OF INDUSTRIAL ENGINEERING, MIDDLE EAST TECHNICAL UNIVERSITY, ANKARA, TURKEY

*E-mail address:* iyigun@ie.metu.edu.tr

ADI BEN-ISRAEL, RUTCOR–RUTGERS CENTER FOR OPERATIONS RESEARCH, RUTGERS UNIVERSITY, 640 BARTHOLOMEW RD., PISCATAWAY, NJ 08854-8003, USA

*E-mail address:* adi.benIsrael@gmail.com