Math. Program., Ser. A 90: 559–566 (2001)

Digital Object Identifier (DOI) 10.1007/s101070100222

Yehuda Vardi · Cun-Hui Zhang

A modified Weiszfeld algorithm for the Fermat-Weber location problem

Received: September 1999 / Accepted: January 2001
Published online April 12, 2001 –

Abstract. This paper gives a new, simple, monotonically convergent, algorithm for the Fermat-Weber location problem, with extensions covering more general cost functions.

1. Introduction and main results

In this paper we derive a simple, yet nontrivial, modification of the Weiszfeld (1937) iterative algorithm for the computation of the Fermat-Weber location problem in $\mathbb{R}^d$ with the Euclidean distance. Proofs are in Sect. 2, and extensions covering more general cost functions are in Sect. 3.

The Fermat-Weber location problem: Let $\mathbf{x}_1, \dots, \mathbf{x}_m$ be m distinct points in $\mathbb{R}^d$ and $\eta_1, \dots, \eta_m$ be m positive numbers. Think of the η_i's as weights or, better yet, as *"multiplicities"* of the $\mathbf{x}_i$'s, and let $C(\mathbf{y})$ denote the weighted sum of distances of $\mathbf{y}$ from $\mathbf{x}_1, \dots, \mathbf{x}_m$:

$$C(\mathbf{y}) \equiv \sum_i \eta_i d_i(\mathbf{y}), \tag{1}$$

where $d_i(\mathbf{y}) \equiv \|\mathbf{y} - \mathbf{x}_i\|$, the Euclidean distance between $\mathbf{y}$ and $\mathbf{x}_i$ in $\mathbb{R}^d$. The Fermat-Weber location problem is to find a point $\mathbf{y} \in \mathbb{R}^d$ which minimizes the "cost-function" $C(\mathbf{y})$, i.e. to find

$$\mathbf{M} \equiv \mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_m; \eta_1, \dots, \eta_m) \equiv \arg\min \left\{ C(\mathbf{y}) : \ \mathbf{y} \in \mathbb{R}^d \right\}. \tag{2}$$

Note that we consider here only the noncollinear case. When the points $\mathbf{x}_1, \dots, \mathbf{x}_m$ are not collinear, $C(\mathbf{y})$ is positive and strictly convex in $\mathbb{R}^d$, and hence the minimum is achieved at a unique point $\mathbf{M} \in \mathbb{R}^d$. In the collinear case, where $\mathbf{x}_1, \dots, \mathbf{x}_m$ lie in a straight line, the minimum of $C(\mathbf{y})$ is achieved at any one-dimensional median (always exists but may not be unique).

Y. Vardi, C.-H. Zhang: Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08854, USA

Mathematics Subject Classification (2000): Primary 65K05, 90C25; Secondary 62H12, 90B85

The Weiszfeld algorithm: Let $d_i(\mathbf{y}) \equiv \|\mathbf{y} - \mathbf{x}_i\|$ be as in (1) and

$$w_i(\mathbf{y}) \equiv \frac{\eta_i}{d_i(\mathbf{y})} \left\{ \sum_{\mathbf{x}_j \neq \mathbf{y}} \frac{\eta_j}{d_j(\mathbf{y})} \right\}^{-1}, \quad \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}. \tag{3}$$

The Weiszfeld (1937) algorithm for the Fermat-Weber problem has the iteration mapping

$$T_0(\mathbf{y}) \equiv \begin{cases} \sum_{\mathbf{x}_i \neq \mathbf{y}} w_i(\mathbf{y})\mathbf{x}_i, & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}, \\ \mathbf{x}_k, & \text{if } \mathbf{y} = \mathbf{x}_k, k = 1, \ldots, m. \end{cases} \tag{4}$$

Let D_i be the domain of attraction of the data point $\mathbf{x}_i$, $i = 1, \ldots, m$, under the Weiszfeld algorithm. That is, $D_i \equiv \cup_{n=1}^{\infty}\{\mathbf{y} : T_0^{(n)}(\mathbf{y}) = \mathbf{x}_i\}$, where $T_0^{(n)}(\mathbf{y}) \equiv T_0(T_0^{(n-1)}(\mathbf{y}))$ with $T_0^{(0)}$ being the identity mapping. Kuhn (1973) showed that for an initial point $\mathbf{y} \notin \cup_{i=1}^m D_i$, the Weiszfeld algorithm converges monotonically to the unique optimal solution of the Fermat-Weber location problem. He further claimed that $\cup_{i=1}^m D_i$ is a denumerable set. Chandrasekaran and Tamir (1989) pointed out an error in Kuhn's argument and showed that $\cup_{i=1}^m D_i$ could contain a continuum set. Brimberg (1995) considered the cardinality of $\cup_{i=1}^m D_i$ under conditions on the dimensionality of the data set. The cardinality issue was further investigated by Canovas, Canavate and Marin (1998). In this paper, we provide a simple, yet nontrivial, modification of (4), producing an algorithm which always converges monotonically to the unique solution of the Fermat-Weber location problem. In Sect. 2 we prove key arguments of our method, and in Sect. 3 we extend the methodology to more general cost function.

A Modified Weiszfeld algorithm for (2): Define an $\mathbb{R}^d$ to $\mathbb{R}^d$ mapping

$$\widetilde{T} : \mathbf{y} \to \widetilde{T}(\mathbf{y}) \equiv \sum_{\mathbf{x}_i \neq \mathbf{y}} w_i(\mathbf{y})\mathbf{x}_i, \tag{5}$$

where $\{w_i(\mathbf{y}) : i = 1, \ldots, m, \mathbf{x}_i \neq \mathbf{y}\}$ are as in (3). The Weiszfeld algorithm is then

$$\mathbf{y} \to T_0(\mathbf{y}) = \begin{cases} \widetilde{T}(\mathbf{y}), & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}, \\ \mathbf{x}_k, & \text{if } \mathbf{y} = \mathbf{x}_k, k = 1, \ldots, m. \end{cases}$$

Let $\mathbf{y}_0$ be the initialization. If the iterative sequence $\mathbf{y}_n \equiv T_0^{(n)}(\mathbf{y}_0)$ never reaches the data set $\{\mathbf{x}_k, 1 \leq k \leq m\}$, then by Kuhn (1973) the sequence $\{\mathbf{y}_n\}$ converges to $\mathbf{M}$ of (2). The problem is that when $\mathbf{y}_n$ lands on a data point, $\mathbf{x}_k$, the Weiszfeld algorithm gets stuck at $\mathbf{x}_k$, even when $\mathbf{x}_k \neq \mathbf{M}$. Thus, we modify the algorithm for $\mathbf{y} \in \{\mathbf{x}_k : k = 1, \ldots, m\}$. Given $\mathbf{y} \in \mathbb{R}^d$, it is convenient to include $\mathbf{y}$ in the data and define the multiplicity at $\mathbf{y}$ as

$$\eta(\mathbf{y}) \equiv \begin{cases} \eta_k, & \text{if } \mathbf{y} = \mathbf{x}_k, k = 1, \ldots, m, \\ 0, & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}. \end{cases} \tag{6}$$

Our modification of (4) for $\mathbf{y} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ is based on the following observation regarding the Weiszfeld algorithm. For $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, the vector $\mathbf{x} = \widetilde{T}(\mathbf{y})$ in (5) is the unique minimizer of

$$g(\mathbf{x}; \mathbf{y}) \equiv \sum_{i=1}^{m} \frac{\eta_i \|\mathbf{x} - \mathbf{x}_i\|^2}{2d_i(\mathbf{y})}, \tag{7}$$

and in the Weiszfeld algorithm, the problem of arg $\min_{\mathbf{x}} C(\mathbf{x})$ is replaced by the simpler arg $\min_{\mathbf{x}} g(\mathbf{x}; \mathbf{y})$ in each iteration. The argument for the use of $g(\mathbf{x}; \mathbf{y})$ is

$$\frac{\partial}{\partial \mathbf{x}} g(\mathbf{x}; \mathbf{y})\bigg|_{\mathbf{x}=\mathbf{y}} = \frac{\partial}{\partial \mathbf{x}} C(\mathbf{x})\bigg|_{\mathbf{x}=\mathbf{y}}, \quad \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}, \tag{8}$$

so that the two minimization problems are similar in all sufficiently small neighborhoods of $\mathbf{y}$, $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. This would suggest that in the Weiszfeld algorithm, (4), instead of stopping when $\mathbf{y}$ equals a data point, $\mathbf{x}_k$, we should iterate with

$$\mathbf{x}_k \to \arg\min_{\mathbf{x}} \; g(\mathbf{x}; \mathbf{x}_k). \tag{9}$$

However, for this to be meaningful we need to extend the definition of g in (7) to cover $\mathbf{y} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. This is naturally done by defining

$$\begin{aligned} g(\mathbf{x}; \mathbf{y}) &\equiv \eta(\mathbf{y})\|\mathbf{x} - \mathbf{y}\| + \sum_{\mathbf{x}_i \neq \mathbf{y}} \eta_i \|\mathbf{x} - \mathbf{x}_i\|^2/\{2d_i(\mathbf{y})\} \\ &= \begin{cases} \sum_{i=1}^{m} \eta_i \|\mathbf{x} - \mathbf{x}_i\|^2/\{2d_i(\mathbf{x}_k)\}, & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}, \\ \eta_k \|\mathbf{x} - \mathbf{y}\| + \sum_{i \neq k} \eta_i \|\mathbf{x} - \mathbf{x}_i\|^2/\{2d_i(\mathbf{x}_k)\}, & \text{if } \mathbf{y} = \mathbf{x}_k, \; k \leq m. \end{cases} \end{aligned} \tag{10}$$

Although $C(\mathbf{x})$ is not differentiable at $\mathbf{x}_k$, (8) is extended to $\mathbf{y} = \mathbf{x}_k$ in the sense that

$$\lim_{\mathbf{x} \to \mathbf{x}_k, \mathbf{x} \neq \mathbf{x}_k} \left\{ \frac{\partial}{\partial \mathbf{x}} g(\mathbf{x}; \mathbf{x}_k) - \frac{\partial}{\partial \mathbf{x}} C(\mathbf{x}) \right\} = 0. \tag{11}$$

In Sect. 2 we show that the modification (9) of (4) at data points $\mathbf{y} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ results in the following.

The new algorithm:

$$\mathbf{y} \to T(\mathbf{y}) \equiv \left(1 - \frac{\eta(\mathbf{y})}{r(\mathbf{y})}\right)^{+} \widetilde{T}(\mathbf{y}) + \min\left(1, \frac{\eta(\mathbf{y})}{r(\mathbf{y})}\right)\mathbf{y}, \tag{12}$$

with the convention $0/0 = 0$ in the computation of $\eta(\mathbf{y})/r(\mathbf{y})$, where $\widetilde{T}(\mathbf{y})$ is as in (5),

$$r(\mathbf{y}) \equiv \|\widetilde{R}(\mathbf{y})\|, \quad \widetilde{R}(\mathbf{y}) \equiv \sum_{\mathbf{x}_i \neq y} \eta_i \frac{\mathbf{x}_i - \mathbf{y}}{\|\mathbf{x}_i - \mathbf{y}\|}. \tag{13}$$

For $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, $T(\mathbf{y}) = \widetilde{T}(\mathbf{y})$, by (12) with $\eta(\mathbf{y}) = 0$, as in the Weiszfeld algorithm. For $\mathbf{y} = \mathbf{x}_k$, $T(\mathbf{y})$ is between $\widetilde{T}(\mathbf{x}_k)$ and $\mathbf{x}_k$, so that by (5) $T(\mathbf{y})$ is also a weighted average of $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. Moreover, for $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, $\widetilde{R}(\mathbf{y})$ of (13) is the negative of the gradient of $C(\mathbf{y})$. It follows from (5) that

$$\widetilde{R}(\mathbf{y}) = \left\{\widetilde{T}(\mathbf{y}) - \mathbf{y}\right\} \sum_{\mathbf{x}_i \neq \mathbf{y}} \eta_i / d_i(\mathbf{y}). \tag{14}$$

This and (13) imply that $\widetilde{T}(\mathbf{y}) = \mathbf{y} = T(\mathbf{y})$ when $r(\mathbf{y}) = \|\widetilde{R}(\mathbf{y})\| = 0$.

Properties of the solution M in (2) and the algorithm (12):

$$\mathbf{y} = \mathbf{M} \ \text{ iff } \ T(\mathbf{y}) = \mathbf{y} \ \text{ iff } \ r(\mathbf{y}) \leq \eta(\mathbf{y}). \tag{15}$$

In words: $\mathbf{y} \in \mathbb{R}^d$ is the optimal solution of the Fermat-Weber location problem iff it is a fixed-point of our iterative algorithm (12), iff $r(\mathbf{y}) \leq \eta(\mathbf{y})$, where $r(\mathbf{y})$ and $\eta(\mathbf{y})$ are given in (13) and (6) respectively.

$$\textbf{Monotonicity of the algorithm:} \ \text{If } \mathbf{y} \neq \mathbf{M}, \ \text{ then } C\big(T(\mathbf{y})\big) < C(\mathbf{y}). \tag{16}$$

$$\textbf{Convergence Theorem:} \ \lim_{n \to \infty} T^{(n)}(\mathbf{y}) = \mathbf{M} \ \textit{for all } \mathbf{y} \in \mathbb{R}^d. \tag{17}$$

We note that the algorithm is extremely simple to program and our independent simulation results indicate quick convergence. From Katz (1974) the local convergence rate of the new algorithm is linear. Eckhardt (1980) studied the Fermat-Weber problem and the Weiszfeld algorithm in general spaces.

2. Proofs of (15)–(17)

The key to the proofs is the inequality

$$C(T(\mathbf{x}_k)) < C(\mathbf{x}_k) \quad \text{ if } \mathbf{x}_k \neq \mathbf{M}, \ \ k = 1, \ldots, m. \tag{18}$$

This implies that, starting from any initial point $\mathbf{y}$ in $\mathbb{R}^d$, the sequence $T^{(n)}(\mathbf{y})$ in the iterative algorithm (12) visits each $\mathbf{x}_k \neq \mathbf{M}$ at most once and it will not get stuck at $\mathbf{x}_k \neq \mathbf{M}$. After the last visit to the set $\{\mathbf{x}_k, k = 1, \ldots, m, \mathbf{x}_k \neq \mathbf{M}\}$, $T^{(n)}(\mathbf{y})$ converge to the solution $\mathbf{M}$ in (2) by Kuhn (1973). This proves the convergence theorem (17) based on (18). The monotonicity property (16) follows from (18), since $C(T(\mathbf{y})) = C(\widetilde{T}(\mathbf{y}))$ and $C(\widetilde{T}(\mathbf{y})) < C(\mathbf{y})$ for $\mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ (Kuhn's result for the Weiszfeld algorithm). Moreover, Kuhn (1973) proved that

$$\mathbf{y} = \mathbf{M} \quad \text{iff} \quad R(\mathbf{y}) = 0, \qquad R(\mathbf{y}) \equiv \{r(\mathbf{y}) - \eta(\mathbf{y})\}^+ \widetilde{R}(\mathbf{y}) / r(\mathbf{y}), \tag{19}$$

with the convention $R(\mathbf{y}) \equiv 0$ for $r(\mathbf{y}) \equiv \|\widetilde{R}(\mathbf{y})\| = 0$. Thus, $\mathbf{y} = \mathbf{M}$ implies $r(\mathbf{y}) \leq \eta(\mathbf{y})$. Finally, $r(\mathbf{y}) \leq \eta(\mathbf{y})$ implies $T(\mathbf{y}) = \mathbf{y}$ by (12), as $r(\mathbf{y}) = 0$ implies $T(\mathbf{y}) = \widetilde{T}(\mathbf{y}) = \mathbf{y}$, while $T(\mathbf{y}) = \mathbf{y}$ implies $\mathbf{y} = \mathbf{M}$ by the monotonicity (16). Thus, (15) is also proved based on (18).

It remains to prove (18), since the proofs of (15)–(17) have been completed under the assumption that (18) holds true. This is done by first proving

$$T(\mathbf{x}_k) = \arg\min_{\mathbf{x}} g(\mathbf{x}; \mathbf{x}_k), \tag{20}$$

i.e. (9) yields (12) for $\mathbf{y} = \mathbf{x}_k$. Since $\widetilde{T}(\mathbf{x}_k)$ is the weighted average of $\{\mathbf{x}_i, i \neq k\}$ with weights proportional to $\eta_i/d_i(\mathbf{x}_k)$, as in (5) and (3), $\sum_{i\neq k}\{\eta_i/d_i(\mathbf{x}_k)\}\{\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_i\} = 0$. Thus, with $< ., . >$ denoting the Euclidean inner product,

$$\sum_{i\neq k} \frac{\eta_i}{d_i(\mathbf{x}_k)} \left\langle \mathbf{x} - \widetilde{T}(\mathbf{x}_k), \widetilde{T}(\mathbf{x}_k) - \mathbf{x}_i \right\rangle = \left\langle \mathbf{x} - \widetilde{T}(\mathbf{x}_k), \sum_{i\neq k} \frac{\eta_i}{d_i(\mathbf{x}_k)} \{\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_i\} \right\rangle = 0.$$

Using this and simple algebra, $g(\mathbf{x}; \mathbf{x}_k)$ in (10) can be written as

$$\begin{aligned} g(\mathbf{x}; \mathbf{x}_k) &= \eta_k \|\mathbf{x} - \mathbf{x}_k\| + \sum_{i\neq k} \frac{\eta_i}{2d_i(\mathbf{x}_k)} \|\mathbf{x} - \widetilde{T}(\mathbf{x}_k)\|^2 \\ &\quad + \sum_{i\neq k} \frac{\eta_i}{2d_i(\mathbf{x}_k)} \|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_i\|^2 \\ &= \eta_k \|\mathbf{x} - \mathbf{x}_k\| + A_k \|\mathbf{x} - \widetilde{T}(\mathbf{x}_k)\|^2 + B_k, \end{aligned} \tag{21}$$

where $A_k \equiv \sum_{i\neq k} \eta_i/\{2d_i(\mathbf{x}_k)\}$ and $B_k \equiv \sum_{i\neq k}\{\eta_i/(2d_i(\mathbf{x}_k))\}\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_i\|^2$. Since η_k, A_k and B_k in (21) do not depend on $\mathbf{x}$, $g(\mathbf{x}; \mathbf{x}_k) \leq g(\mathbf{z}; \mathbf{x}_k)$ if $\mathbf{x}$ is the projection of $\mathbf{z}$ onto the line segment with endpoints $\mathbf{x}_k$ and $\widetilde{T}(\mathbf{x}_k)$, i.e. $\|\mathbf{x} - \mathbf{x}_k\| \leq \|\mathbf{z} - \mathbf{x}_k\|$ and $\|\mathbf{x} - \widetilde{T}(\mathbf{x}_k)\| \leq \|\mathbf{z} - \widetilde{T}(\mathbf{x}_k)\|$. Thus,

$$\min_{\mathbf{z}} g(\mathbf{z}; \mathbf{x}_k) = \min_{0\leq\gamma\leq 1} g\big(\{(1-\gamma)\widetilde{T}(\mathbf{x}_k) + \gamma\mathbf{x}_k\}\,; \mathbf{x}_k\big). \tag{22}$$

If $\widetilde{T}(\mathbf{x}_k) = \mathbf{x}_k$, then $T(\mathbf{x}_k) = \mathbf{x}_k$ by (12) and (22) implies (20). For $\widetilde{T}(\mathbf{x}_k) \neq \mathbf{x}_k$,

$$g\big((1-\gamma)\widetilde{T}(\mathbf{x}_k) + \gamma\mathbf{x}_k; \mathbf{x}_k\big) = |1-\gamma|\eta_k\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_k\| + \gamma^2 A_k\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_k\|^2 + B_k,$$

as a strictly convex function of γ, is uniquely minimized at

$$\gamma_k \equiv \min\left\{1, \frac{\eta_k\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_k\|}{2A_k\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_k\|^2}\right\} = \min\left\{1, \frac{\eta_k}{r(\mathbf{x}_k)}\right\},$$

since $r(\mathbf{x}_k) = 2A_k\|\widetilde{T}(\mathbf{x}_k) - \mathbf{x}_k\|$ by (13), (14) and the definition of A_k in (21). This and (12) imply that $T(\mathbf{x}_k) = (1-\gamma_k)\widetilde{T}(\mathbf{x}_k) + \gamma_k\mathbf{x}_k$, so that by (22) $T(\mathbf{x}_k)$ is the minimizer of $g(\mathbf{x}; \mathbf{x}_k)$. Thus, (20) also holds for $\widetilde{T}(\mathbf{x}_k) \neq \mathbf{x}_k$.

The final step of the proof is to show (20) $\Rightarrow$ (18). Let $\mathbf{x}_k \neq \mathbf{M}$ in the rest of the proof. By (19) and (14) $r(\mathbf{x}_k) > \eta(\mathbf{x}_k) > 0$ and $\widetilde{T}(\mathbf{x}_k) \neq \mathbf{x}_k$ for $\mathbf{x}_k \neq \mathbf{M}$, so that

$T(\mathbf{x}_k) \neq \mathbf{x}_k$ by (12). This implies $g(T(\mathbf{x}_k); \mathbf{x}_k) < g(\mathbf{x}_k; \mathbf{x}_k)$ by (20) and the strict convexity of $g(\mathbf{x}; \mathbf{x}_k)$ in $\mathbf{x}$. Consequently, by (10) and (1)

$$\begin{aligned} 0 &> g\big(T(\mathbf{x}_k); \mathbf{x}_k\big) - g\big(\mathbf{x}_k; \mathbf{x}_k\big) \\ &= \eta_k \|T(\mathbf{x}_k) - \mathbf{x}_k\| + \sum_{i \neq k} \frac{\eta_i}{2d_i(\mathbf{x}_k)} \Big\{ d_i^2(T(\mathbf{x}_k)) - d_i^2(\mathbf{x}_k) \Big\} \\ &\geq \eta_k \|T(\mathbf{x}_k) - \mathbf{x}_k\| + \sum_{i \neq k} \eta_i \Big\{ d_i(T(\mathbf{x}_k)) - d_i(\mathbf{x}_k) \Big\} \\ &= C(T(\mathbf{x}_k)) - C(\mathbf{x}_k) \end{aligned} \tag{23}$$

due to $(a^2 - b^2)/(2b) \geq a - b$ for $a = d_i(T(\mathbf{x}_k)) \geq 0$ and $b = d_i(\mathbf{x}_k) > 0$. Hence, (18) holds for $\mathbf{x}_k \neq \mathbf{M}$ and the proof is complete.

3. Extensions

In this section, we extend our method to the computation of (2) for more general cost functions of the form

$$C(\mathbf{y}) \equiv \sum_{i=1}^{m} \eta_i(\|\mathbf{y} - \mathbf{x}_i\|) = \sum_{i=1}^{m} \eta_i(d_i(\mathbf{y})), \tag{24}$$

where η_i are continuously differentiable functions in $[0, \infty)$ with $\eta_i'(t) \equiv (d/dt)\eta_i(t) > 0$ for $t > 0$ and $\eta_i(0) = 0$. The cost function (1) is a special case of (24) with linear $\eta_i(t) \equiv \eta_i t$.

Based on the discussion leading to (12), we consider algorithms of the form

$$\mathbf{y} \to T(\mathbf{y}) \equiv \arg\min_{\mathbf{x}} g(\mathbf{x}; \mathbf{y}), \tag{25}$$

with functions $g(\mathbf{x}; \mathbf{y})$ satisfying (8) and (11) for the cost function (24). It can be easily verified by calculus that these conditions hold for

$$g(\mathbf{x}; \mathbf{y}) \equiv \begin{cases} \sum_{i=1}^{m} \{\eta_i'(d_i(\mathbf{y}))/(2d_i(\mathbf{y}))\} d_i^2(\mathbf{x}), & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \\ \eta_k(d_k(\mathbf{x})) + \sum_{i \neq k} \{\eta_i'(d_i(\mathbf{y}))/(2d_i(\mathbf{y}))\} d_i^2(\mathbf{x}), & \text{if } \mathbf{y} = \mathbf{x}_k,\ k \leq m. \end{cases} \tag{26}$$

Since the sums in (26) are quadratic functions of $\mathbf{x}$, it follows directly from the proof of (20) that, for the $g(\mathbf{x}; \mathbf{y})$ in (26), (25) results in the following.

New algorithm for general cost functions (24):

$$\mathbf{y} \to T(\mathbf{y}) \equiv \begin{cases} \widetilde{T}(\mathbf{y}), & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \\ (1 - \gamma_k)\widetilde{T}(\mathbf{x}_k) + \gamma_k \mathbf{x}_k, & \text{if } \mathbf{y} = \mathbf{x}_k, k \leq m, \end{cases} \tag{27}$$

where $\widetilde{T}(\mathbf{y}) \equiv \sum_{\mathbf{x}_i \neq \mathbf{y}} w_i(\mathbf{y})\mathbf{x}_i$ with the weights

$$w_i(\mathbf{y}) \equiv \frac{\eta_i'(d_i(\mathbf{y}))}{d_i(\mathbf{y})W(\mathbf{y})}, \quad \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}, \quad W(\mathbf{y}) \equiv \sum_{\mathbf{x}_j \neq \mathbf{y}} \frac{\eta_j'(d_j(\mathbf{y}))}{d_j(\mathbf{y})}, \tag{28}$$

and with $A_k = \sum_{i \neq k} \eta_i'(d_i(\mathbf{x}_k))/\{2d_i(\mathbf{x}_k)\}$, $\gamma_k \in [0, 1]$ is defined by

$$\gamma_k \equiv \arg\min_{0 \le \gamma \le 1} \left\{ \eta_k\big(|1-\gamma| d_k(\widetilde{T}(\mathbf{x}_k))\big) + \gamma^2 A_k d_k^2(\widetilde{T}(\mathbf{x}_k)) \right\}.$$

Define $\widetilde{R}(\mathbf{y}) \equiv \sum_{\mathbf{x}_i \neq \mathbf{y}} \{\eta_i'(d_i(\mathbf{y}))/d_i(\mathbf{y})\}(\mathbf{x}_i - \mathbf{y})$, and

$$r(\mathbf{y}) \equiv \|\widetilde{R}(\mathbf{y})\|, \quad \eta(\mathbf{y}) \equiv \begin{cases} \eta_k'(0), & \text{if } \mathbf{y} = \mathbf{x}_k \\ 0, & \text{if } \mathbf{y} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}. \end{cases} \tag{29}$$

Then we have the following.

Properties of algorithm (27): Suppose $\eta_i(\sqrt{t})$ in (24) are concave in t. Then,

$$C(T(\mathbf{y})) \le C(\mathbf{y}), \qquad \forall\, \mathbf{y}. \tag{30}$$

If in addition $\eta_i(t)$ are convex in t, then

$$C(T(\mathbf{y})) = C(\mathbf{y}) \text{ iff } T(\mathbf{y}) = \mathbf{y} \text{ iff } C(\mathbf{y}) = C_* \text{ iff } r(\mathbf{y}) \le \eta(\mathbf{y}), \tag{31}$$

with $C_* \equiv \min_{\mathbf{x}} C(\mathbf{x})$, $r(\mathbf{y})$ and $\eta(\mathbf{y})$ defined in (29), and for any initialization $\mathbf{y}_0$,

$$\lim_{n\to\infty} C(\mathbf{y}_n) = C_*, \qquad \text{where } \mathbf{y}_n \equiv T(\mathbf{y}_{n-1}), n \ge 1. \tag{32}$$

Remark. If $\eta_i(t) = a_i t^{p_i}$ for constants $a_i > 0$ and $1 \le p_i \le 2$, then $\eta_i(\sqrt{t})$ are concave and $\eta_i(t)$ are convex. In particular, if $p_i = 2$ for all $1 \le i \le m$, then the cost function is a weighted sum of squares and the algorithm (27) converges in one iteration. If (2) is unique, then (32) implies $\mathbf{y}_n \to \mathbf{M}$ for convex $\eta_i(t)$.

Proof of (30), (31) and (32): The concavity of $\eta_i(\sqrt{t})$ is equivalent to $\eta_i(a) - \eta_i(b) \le \{\eta_i'(b)/(2b)\}(a^2 - b^2)$ for all $a \ge 0$ and $b > 0$, which implies as in (23) that

$$\begin{aligned} C(T(\mathbf{y})) - C(\mathbf{y}) &= \sum_{\mathbf{x}_i = \mathbf{y}} \eta_i(d_i(T(\mathbf{y}))) + \sum_{\mathbf{x}_i \neq \mathbf{y}} \Big\{ \eta_i(d_i(T(\mathbf{y}))) - \eta_i(d_i(\mathbf{y})) \Big\} \\ &\le \sum_{\mathbf{x}_i = \mathbf{y}} \eta_i(d_i(T(\mathbf{y}))) + \sum_{\mathbf{x}_i \neq \mathbf{y}} \frac{\eta_i'(d_i(\mathbf{y}))}{2d_i(\mathbf{y})} \Big\{ d_i^2(T(\mathbf{y})) - d_i^2(\mathbf{y}) \Big\} \\ &= g\big(T(\mathbf{y}); \mathbf{y}\big) - g\big(\mathbf{y}; \mathbf{y}\big) \le 0 \end{aligned}$$

by (24), (25) and (26). Moreover, due to the strict convexity of $g(\mathbf{x}; \mathbf{y})$ in $\mathbf{x}$,

$$C(T(\mathbf{y})) = C(\mathbf{y}) \;\Rightarrow\; g\big(\mathbf{y}; \mathbf{y}\big) = g\big(T(\mathbf{y}); \mathbf{y}\big) = \min_{\mathbf{x}} g(\mathbf{x}; \mathbf{y}) \;\Leftrightarrow\; T(\mathbf{y}) = \mathbf{y}, \tag{33}$$

which implies by (24) and (26) that for all $\mathbf{x}$

$$\liminf_{t\to 0+} \frac{C(\mathbf{y}+t\mathbf{x})-C(\mathbf{y})}{t} = \liminf_{t\to 0+} \frac{g(\mathbf{y}+t\mathbf{x};\mathbf{y})-g(\mathbf{y};\mathbf{y})}{t} \ge 0, \tag{34}$$

in view of (8) and (11). Similar to Kuhn (1973) and directly by calculus, (34) holds iff $r(\mathbf{y}) \le \eta(\mathbf{y})$ iff $C(\mathbf{y}) = C_*$ by the convexity of $C(\mathbf{x})$. Therefore, (31) holds by (33).

The proof (32) is the same as the corresponding parts in Kuhn (1973), so that we omit details. By (30) and (31) we may assume $C(\mathbf{y}_n) > C(\mathbf{y}_{n-1}) > C_*$. Since $\{\mathbf{y}_n\}$ visits each $\mathbf{x}_k$ at most once, we may further assume $\mathbf{y}_n \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. By the convexity and strict monotonicity of $\eta_i(t)$, $\|\mathbf{y}_n\|$ are bounded. Thus, it suffices to show $C(\mathbf{y}_*) = C_*$ for all limit points $\mathbf{y}_*$ of $\{\mathbf{y}_n\}$. By (26) and (30), and by the argument leading to (21),

$$\|\mathbf{y}_n - T(\mathbf{y}_n)\|^2 \sum_{i=1}^{m} \frac{\eta_i'(d_i(\mathbf{y}_n))}{2d_i(\mathbf{y}_n)} = g\big(\mathbf{y}_n;\mathbf{y}_n\big) - g\big(T(\mathbf{y}_n);\mathbf{y}_n\big) \to 0, \tag{35}$$

so that $\|\mathbf{y}_n - \mathbf{y}_{n+1}\| \to 0$. By (27), (28), (29) and (31), for $C(\mathbf{x}_k) > C_*$

$$\frac{d_k(T(\mathbf{y}))}{d_k(\mathbf{y})} = \frac{1}{d_k(\mathbf{y})W(\mathbf{y})} \left\| \sum_{i\ne k} \frac{\eta_i'(d_i(\mathbf{y}))}{d_i(\mathbf{y})}(\mathbf{x}_i - \mathbf{x}_k) \right\| \to \frac{r(\mathbf{x}_k)}{\eta(\mathbf{x}_k)} > 1$$

as $\|\mathbf{y} - \mathbf{x}_k\| \to 0+$; i.e. there exist $\delta_k > 0$ such that $0 < \|\mathbf{y}_n - \mathbf{x}_k\| \le \delta_k$ implies $\|\mathbf{y}_{n+1} - \mathbf{x}_k\| > (1+\delta_k)\|\mathbf{y}_n - \mathbf{x}_k\|$. Thus, the limit point $\mathbf{y}_*$ is outside the set $\{\mathbf{x}_k; k = 1, \ldots, m, C(\mathbf{x}_k) > C_*\}$ due to $\|\mathbf{y}_n - \mathbf{y}_{n+1}\| \to 0$. Now it suffices to consider $y_* \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. In this case, $g_1(\mathbf{x};\mathbf{y}) \equiv (\partial/\partial\mathbf{x})g(\mathbf{x};\mathbf{y})$ is continuous in $(\mathbf{x},\mathbf{y})$ at $(\mathbf{y}_*,\mathbf{y}_*)$. Since $\|\mathbf{y}_n - T(\mathbf{y}_n)\| \to 0$ by (35), $(\mathbf{y}_*,\mathbf{y}_*)$ is a limit point of $(T(\mathbf{y}_n),\mathbf{y}_n)$ in $\mathbb{R}^{2d}$, so that by (8) and (25) and the continuity of $g_1(\mathbf{x};\mathbf{y})$

$$\left\| \frac{\partial C(\mathbf{x})}{\partial\mathbf{x}}\bigg|_{\mathbf{x}=\mathbf{y}_*} \right\| = \big\|g_1\big(\mathbf{y}_*;\mathbf{y}_*\big)\big\| \le \limsup_{n\to\infty} \big\|g_1\big(T(\mathbf{y}_n);\mathbf{y}_n\big)\big\| = 0.$$

Hence, $C(\mathbf{y}_*) = C_*$ by the convexity of $C(\mathbf{x})$.

Acknowledgements. The authors thank J. De Leeuw and R. Vanderbei for references, and NSA and NSF for grant support.

References

1. Brimberg, J. (1995): The Fermat Weber location problem revisited. Math. Program. **71**, 71–76
2. Canovas, L., Canavate, R., Marin, A. (1998): Notas sobre el algoritmo de Weiszfeld. Tech. Rep. No. 9, Faculty of Mathematics, University of Murcia, Spain
3. Chandrasekaran, R., Tamir, A. (1989): Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem. Math. Program. **44**, 293–295
4. Eckhardt, U. (1980): Weber's problem and Weiszfeld's algorithm in general spaces. Math. Program. **18**, 186–196
5. Katz, I.N. (1974): Local convergence in Fermat's problem. Math. Program. **6**, 89–104
6. Kuhn, H.W. (1973): A note on Fermat's problem. Math. Program. **4**, 98–107
7. Weber, A. (1909): Über den Standort der Industrien, Tübingen. [English translation by Freidrich, C.J. (1929): Alfred Weber's Theory of Location of Industries, University of Chicago Press]
8. Weiszfeld, E. (1937): Sur le point pour lequel la somme des distances de *n* points donnés est minimum. Tôhoku Mathematics Journal **43**, 355–386