MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A
A NOTE ON THE WEISZFELD-KUHN ALGORITHM
FOR THE GENERAL FERMAT PROBLEM

by
Egon Balas and Chang-Sung Yu

Management Science Research Group
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213
ABSTRACT

The General Fermat Problem, or Weber Problem, asks for a point that minimizes a weighted sum of the distances to m given points. The Weiszfeld-Kuhn algorithm is an iterative procedure that converges to an optimal point for all but a denumerable number of starting points. We give an amended version of the algorithm, that guarantees convergence.
A Note on the Weiszfeld-Kuhn Algorithm for the General Fermat Problem

Egon Balas and Chang-Sung Yu

Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213

The General Fermat Problem, or Weber Problem, asks for a point that minimizes a weighted sum of the distances to m given points. The Weiszfeld-Kuhn algorithm is an iterative procedure that converges to an optimal point for all but a denumerable number of starting points. We give an amended version of the algorithm, that guarantees convergence.

Location theory, Fermat problem, Weber problem, Weiszfeld's algorithm
A NOTE ON THE WEISZFELD-KUHN ALGORITHM

FOR THE GENERAL FERMAT PROBLEM

by

Egon Balas and Chang-Sung Yu

1. Introduction

The problem of finding a point in the plane that minimizes the sum of distances to three given points was posed by Fermat at the beginning of the 17th century, and solved shortly thereafter by Toricelli. A generalized version of this problem, with important applications in location theory, and usually referred to as the General Fermat Problem, or Weber Problem, is as follows:

Given m points $a_i \in \mathbb{R}^n$, called vertices, and associated weights $w_i > 0$, $i = 1, \ldots, m$, find a point $x^* \in \mathbb{R}^n$ that minimizes

$$f(x) = \sum_{i=1}^{m} w_i d_i(x),$$

where $d_i(x)$ is the Euclidean distance between $x$ and $a_i$, i.e.,

$$d_i(x) = \sqrt{\sum_{j=1}^{n} (a_{ij} - x_j)^2}.$$

If the $n$ vertices are collinear, the problem is a special case of the absolute median problem in a tree, solvable in linear time. So we assume the vertices are not collinear. In this case $f$ is strictly convex and $x^*$ is unique. Further, $f$ is differentiable everywhere except at the vertices $a_i$, $i = 1, \ldots, m$, and the gradient of $f$ at $x \neq a_i$, $i = 1, \ldots, m$, is
\[ \nabla f(x) = \sum_{i=1}^{m} \frac{w_i}{d_i(x)} (x - a_i). \]

Imposing the optimality condition \( \nabla f(x^*) = 0 \) then yields

\[ x^* = \left( \sum_{i=1}^{m} \frac{w_i}{d_i(x^*)} a_i \right) / \left( \sum_{i=1}^{m} \frac{w_i}{d_i(x^*)} \right), \]

which is well defined for \( x^* \neq a_i, i = 1, \ldots, m. \)

In 1937, Weiszfeld [2] gave an iterative algorithm for the General Fermat Problem based on the above expression for \( x^* \). He claimed that the algorithm converges to \( x^* \) whenever the starting point is not a vertex.

More recently, Kuhn [1] showed this claim to be false; gave a rigorous analysis of the problem and the algorithm, and proved that a slightly amended version of the latter converges to \( x^* \) for all but a denumerable number of starting points.

Although the probability of choosing a bad starting point is small, no fail-safe method is known for avoiding such points. The purpose of this note is to fill in this gap, and provide a version of the algorithm for which the sequence of iterates is guaranteed to converge to \( x^* \).

2. The Weiszfeld-Kuhn algorithm and its properties

The Weiszfeld-Kuhn algorithm consists of iterating the mapping

\[ T: x \to T(x), \text{ where} \]

\[ T(x) = \begin{cases} 
\sum_{i=1}^{m} \frac{w_i}{d_i(x)} a_i / \sum_{i=1}^{m} \frac{w_i}{d_i(x)}, & x \neq a_i, i = 1, \ldots, m \\
\frac{w_k}{d_k(x)}, & x = a_k, k \in \{1, \ldots, m\}. 
\end{cases} \]
The essential properties of the algorithm are summarized below.

For $k = 1, \ldots, m$, define

$$R_k = \sum_{i=1}^{m} \frac{w_i}{d_i(a_k)} (a_i - a_k)$$

and

$$R(x) = \begin{cases} -\nabla f(x), & x \neq a_i, i = 1, \ldots, m \\ \max\{0, \|R_k\| - w_k\} \cdot R_k / \|R_k\|, & x = a_k, k \in \{1, \ldots, m\} \end{cases}$$

where $\| \cdot \|$ denotes the Euclidean norm.

While the gradient of $f(x)$ exists only for $x \neq a_i, i = 1, \ldots, m$, $R(x)$ is defined for all $x \in \mathbb{R}^n$.

The main properties of $x^*$ and $T$ are as follows (Kuhn [1]):

1. $x = x^*$ if and only if $R(x) = 0$.
2. $x^*$ is in the convex hull of the vertices $a_i$.
3. If $x = x^*$, then $T(x) = x$. If $x$ is not a vertex and $T(x) = x$, then $x = x^*$.
4. If $T(x) \neq x$, then $f(T(x)) < f(x)$.
5. $a_k = x^*$ if and only if $w_k \geq \|R_k\|$.
6. If $a_k \neq x^*$, the direction of steepest descent of $f$ from $a_k$ is $R_k / \|R_k\|$.
7. If $a_k \neq x^*$, there exists $\delta > 0$ such that $0 < d_k(x) \leq \delta$ implies $d_k(T^s(x)) > \delta$ for some $s$.
8. $\lim_{x \to a_k} [d_k(T(x))/d_k(x)] = \|R_k\|/w_k$, $k = 1, \ldots, m$. 
4

(9) For any starting point $x^0$, let $x^r = T^r(x^0), r = 1, 2, \ldots$

If no $x^r$ is a vertex, then $\lim_{r \to \infty} x^r = x^*$.

(10) For all but a denumerable number of starting points $x^0, x^r = T^r(x^0)$ converges to $x^*$.

3. Discussion

From (3) and (4), $f(T(x)) < f(x)$ for any nonoptimal $x$ that is not a vertex. So as long as the sequence $\{x^r\}$ does not reach a vertex, $f(x^r)$ is monotone decreasing with $r$. On the other hand, since $T(x) = x$ whenever $x$ is a vertex, if the sequence $\{x^r\}$ ever reaches a vertex, it never leaves it. Thus convergence to $x^*$ requires the avoidance of vertices. Now (8) in conjunction with (5) implies that for any $x$ different from, but close enough to, some vertex $a_k$, $d_k(T(x)) > d_k(x)$, i.e., $T(x)$ is further from $a_k$ than $x$.

Furthermore, (7) essentially says that every vertex $a_k$ has a $\delta$-neighborhood $N(a_k, \delta)$ such that, if the sequence $\{x^r\}$ enters $N(a_k, \delta)$ with a term $x^r \neq a_k$, then it also leaves $N(a_k, \delta)$. It was on these grounds that Weiszfeld claimed convergence for his algorithm whenever the starting point was not a vertex.

In pointing out the error, Kuhn showed that (7) and (8) do not preclude the occurrence of $T(x) = a_k$ for some $x \notin N(a_k, \delta)$. In other words, although the sequence $\{x^r\}$ cannot reach any vertex $a_k$ from within its neighborhood $N(a_k, \delta)$, it can hit upon $a_k$ from outside its neighborhood, and when that happens, it gets stuck.

At first sight, this problem seems easy to correct. If $x^r = a_k$, it would seem sufficient to replace $a_k$ by any point $x^{r+1} \in N(a_k, \delta), x^{r+1} \neq a_k$. Then by property (7), there is an integer $s > r + 1$ such that $x^s \notin N(a_k, \delta)$. 

\[\ldots\]
However, if \( x^{r+1} \) is chosen arbitrarily, there is no guarantee that once the sequence leaves \( N(a_k, \delta) \), it does not again hit upon \( a_k \). Although the probability of this is very low, the possibility of repeatedly hitting the same vertex \( a_k \) cannot be excluded. And, of course, the sequence could "visit" more than one vertex.

Thus some special rule is called for if one is to avoid such occurrences. This we give next.

4. A convergent version of the algorithm

0. Let

\[
x_j^0 = \frac{\sum_{i=1}^{m} w_i a_{ij}}{i=1} / \sum_{i=1}^{m} w_i , \quad j = 1, \ldots, n,
\]

set \( r = 0 \), and go to 1.

1. If \( x_r = a_k \) for some \( k \in \{1, \ldots, m\} \), go to 2. Otherwise let

\[
x_r^{+1} = \frac{m \sum_{i=1}^{m} w_i}{\sum_{i=1}^{m} \sum_{i=1}^{m} w_i} a_i / \sum_{i=1}^{m} d_i(x_r)
\]

Then if \( x_r^{+1} = x_r \), stop: \( x^* = x_r \). Otherwise set \( r \leftarrow r + 1 \) and go to 1.

2. If \( w_k \geq \|R_k\| \), stop: \( x^* = a_k \). Otherwise find (for instance, by bisection) a scalar \( \lambda > 0 \) such that \( f(a_k + \lambda R_k) < f(a_k) \), let

\[
x_r^{+1} = a_k + \lambda R_k
\]

set \( r \leftarrow r + 1 \) and go to 1.

Proposition. The above algorithm converges to \( x^* \).
Proof. Step 0 chooses the center of gravity as a reasonable starting point. Step 1 is like in the original algorithm; and if no $x^r$ is a vertex, i.e., if step 2 is never used, then from (4) $f(x^{r+1}) < f(x^r)$ whenever $x^{r+1} \neq x^r$, and from (9) the sequence $\{x^r\}$ converges to $x^\ast$.

If $x^r = a_k$ for some $r$ and $k$, and $w_k \geq ||R_k||$, then from (5) $x^\ast = a_k$. Otherwise, from (6) $R_k/||R_k||$ is the direction of steepest descent of $f$ from $a_k$; hence for $\lambda$ sufficiently small, $f(a_k + \lambda R_k) < f(a_k)$. Thus if $x^r = a_k$ is not optimal, $f(x^s) < f(a_k)$ for all $s > r$, and a given vertex $a_k$ can appear in the sequence $\{x^r\}$ at most once. This, together with (9), proves that the sequence $\{x^r\}$ converges to $x^\ast$.

A possible alternative is to use as starting point $x^0 = a_k$, where $k$ is defined by

$$f(a_k) = \min_i f(a_i).$$

In this case step 2 is used only once, and for the rest of the time the algorithm iterates step 1. This starting solution has a higher cost, but may sometimes save on the number of subsequent iterations.

References

