

Weiszfeld's Method: Old and New Results

Amir Beck*

Shoham Sabach†

April 27, 2013

Abstract

In 1937 the 16 years old Hungarian mathematician Endre Weiszfeld devised a method for solving the Fermat-Weber location problem – a problem whose origins can be traced back to the 17th century. This seminal paper of Weiszfeld [51] stirred up an enormous amount of research in the optimization and location communities, and is also being discussed and used till these days. In this paper we review both the past and the ongoing research on Weiszfeld's method. The existing results are presented in a self contained and concise manner – some are derived by new and simplified techniques. We also establish two new results using modern tools of optimization. First, we establish a nonasymptotic $O(1/k)$ rate of convergence of Weiszfeld's method, and second, using an exact smoothing technique, we present a modification of the method with a proven $O(1/k^2)$ rate of convergence.

1 Introduction - The Method

One of the most fundamental location problems is the so-called Fermat-Weber problem that consists of finding an $\mathbf{x} \in \mathbb{R}^n$ that solves the problem

$$(\text{FW}) \quad \min_{\mathbf{x}} \left\{ f(\mathbf{x}) = \sum_{i=1}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| \right\},$$

where $\omega_i > 0$, $i = 1, 2, \dots, m$, are given *weights* and the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$ are given *anchors*. The problem is credited to the well known French mathematician Pierre de Fermat who at the beginning of the 17th century asked the following question:

Given three points in a plane, find a fourth point such that the sum of its distances to the three given points is as small as possible.

*Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: becka@ie.technion.ac.il.

†School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: ss-abach@gmail.com

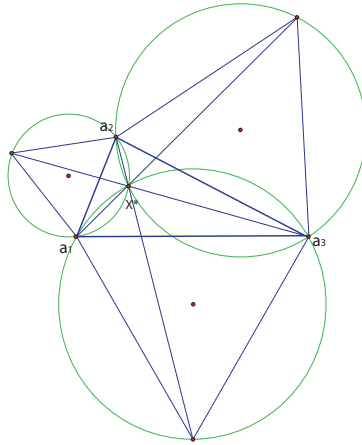


Figure 1: Given a triangle formed by three points \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 , construct three equilateral triangles such that each contains one of the edges from the triangle $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$. Then circumscribe each equilateral triangle. The point of intersection of these three circles is the point that yields the minimum distance to the points \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 is called “the Torricelli Point” and is denoted in the image by \mathbf{x}^* .

This is of course a restricted version of problem (FW) in which $n = 2$, $m = 3$ and all the weights are 1. The Italian physicist and mathematician Evangelista Torricelli (mostly known for inventing the barometer) found a construction method of this point by ruler and compass and it is therefore also called “the Toricelli point”, see Figure 1 for an illustration of Torricelli’s construction. At the beginning of the 20th century, the German economist Alfred Weber incorporated weights, and was thus able to treat facility location problems via the formulation (FW), which was consequently called “the Fermat-Weber” problem. Other names for problem (FW) are “the Fermat problem”, “the Weber problem”, “the Fermat-Torricelli problem”, “the Steiner problem” and many more variants. More details on the history of the Fermat-Weber problem can be found for example in [23] as well as in the survey paper [22].

The purpose of this paper is *not* to present another review on the Fermat-Weber problem, but rather to focus on a very simple algorithm designed for solving it, suggested in 1937 by the 16 years old Hungarian mathematician Endre Vaszhonyi Weiszfeld [51]. Quite interestingly, as indicated by Weiszfeld himself in [49], the focus of Weiszfeld’s paper was *not* to design an algorithm for solving problem (FW), but rather to prove a mathematical theorem, and in fact the all notion of an “algorithm” was unfamiliar to him, as he himself indicated in [49]: “*the word algorithm was unknown to me and to most mathematicians*”. The theorem itself was not new and was already established by Sturm in 1884 [46]. To understand the result Weiszfeld wished to show, let us first write down the expression of

the gradient of the objective function of problem (FW):

$$\nabla f(\mathbf{x}) = \sum_{i=1}^m \omega_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}, \quad \mathbf{x} \notin \mathcal{A},$$

where $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ denotes the set of anchors. Note that the gradient is only defined on points different from the anchors. The theorem Weiszfeld re-established was the following.

Theorem 1.1 (cf. [46, 51]). *Suppose that the anchors are not collinear. Then*

- (a) *Problem (FW) has a unique optimal solution.*
- (b) *Let \mathbf{x}^* be the optimal solution of (FW). If $\mathbf{x}^* \notin \mathcal{A}$, then*

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^m \omega_i \frac{\mathbf{x}^* - \mathbf{a}_i}{\|\mathbf{x}^* - \mathbf{a}_i\|} = \mathbf{0}. \quad (1.1)$$

If $\mathbf{x}^ = \mathbf{a}_i$, for some $i \in \{1, 2, \dots, m\}$, then the following inequality holds*

$$\left\| \sum_{j=1, j \neq i}^m \omega_j \frac{\mathbf{x}^* - \mathbf{a}_j}{\|\mathbf{x}^* - \mathbf{a}_j\|} \right\| \leq \omega_i. \quad (1.2)$$

Some comments about this theorem are required. The anchors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ are said to be *collinear* if they reside on the same line, meaning that there exist $\mathbf{y}, \mathbf{d} \in \mathbb{R}^n$ and $t_1, t_2, \dots, t_m \in \mathbb{R}$ such that $\mathbf{a}_i = \mathbf{y} + t_i \mathbf{d}$, $i = 1, 2, \dots, m$. In the collinear case, it can be shown that the optimal solution of the problem (FW) is the median of the anchors, meaning that the optimal solution is attained at (at least) one of the anchor points, see [39] for further details. It is important to note that Weiszfeld did not analyze the weighted problem, but rather assumed that all the weights are equal to 1. We present the method and results in the weighted scenario, whose analysis is almost identical. Nowadays, this theorem seems as an elementary result, and is a direct consequence of basic convex analysis. Indeed, when the anchors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ are not collinear, the objective function f is strictly convex and thus the optimal solution is unique [45]. The optimality condition (1.1) is just the necessary and sufficient optimality condition $\nabla f(\mathbf{x}^*) = \mathbf{0}$ for unconstrained convex minimization problems at points for which the objective function is differentiable, and the condition (1.2) is the necessary and sufficient optimality condition $\mathbf{0} \in \partial f(\mathbf{a}_i)$ at points of nondifferentiability. These optimality conditions are satisfied also in the collinear case.

Weiszfeld's paper provided three different proofs for Theorem 1.1 and in the first proof he defined a sequence that was supposed to converge to the optimal solution. To present the method, assume that the anchors are not collinear and that \mathbf{x}^* is the unique optimal

solution. We begin by writing explicitly the optimality condition $\nabla f(\mathbf{x}^*) = \mathbf{0}$ under the assumption that $\mathbf{x}^* \notin \mathcal{A}$:

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^m \omega_i \frac{\mathbf{x}^* - \mathbf{a}_i}{\|\mathbf{x}^* - \mathbf{a}_i\|} = \mathbf{0}.$$

The next step is to “extract” \mathbf{x}^* (disregarding the dependency of $\|\mathbf{x}^* - \mathbf{a}_i\|$ in \mathbf{x}^* , $i = 1, 2, \dots, m$) and obtain the relation

$$\mathbf{x}^* = \frac{1}{\sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x}^* - \mathbf{a}_i\|}} \sum_{i=1}^m \frac{\omega_i \mathbf{a}_i}{\|\mathbf{x}^* - \mathbf{a}_i\|}, \quad (1.3)$$

or

$$\mathbf{x}^* = T(\mathbf{x}^*),$$

where the operator $T : \mathbb{R}^n \setminus \mathcal{A} \rightarrow \mathbb{R}^n$ is defined by

$$T(\mathbf{x}) \equiv \frac{1}{\sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}} \sum_{i=1}^m \frac{\omega_i \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}. \quad (1.4)$$

We have thus shown that for any $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$,

$$\mathbf{y} = T(\mathbf{y}) \text{ if and only if } \nabla f(\mathbf{y}) = \mathbf{0}. \quad (1.5)$$

Weiszfeld’s method is just a fixed point method for solving the relation (1.3).

Weiszfeld’s Method

Initialization. $\mathbf{x}_0 \in \mathbb{R}^n \setminus \mathcal{A}$.

General Step ($k = 0, 1, \dots$)

$$\mathbf{x}_{k+1} = T(\mathbf{x}_k). \quad (1.6)$$

This elegant algorithm stirred up an enormous amount of research and had impact on researchers from the optimization as well as the location fields. The contribution of this paper is threefold. First, we will review the intriguing story of the algorithm ever since its derivation in 1937 till these days and present the current status of convergence of analysis. The presentation is self-contained, so the main convergence results will be presented with proofs – some of them are new and simplified. Our second contribution is to provide – using modern tools of optimization – non-asymptotic $O(1/k)$ rate of convergence analysis of the method. Finally, noting that Weiszfeld’s method is essentially a gradient method, our third contribution will be to present an accelerated version of the method, based on a combination of an exact smoothing technique and an optimal gradient method; the resulting method is shown to have a rate of convergence of $O(1/k^2)$.

The paper is organized as follows. In Section 2, we present the original paper of Weiszfeld and describe what was actually proven in that seminal paper (monotonicity of

the sequence of function values), as well as pinpoint the critical mistake in the analysis. The method remained mainly unknown until 1962, and in Section 3 we describe several papers that reinvented the method as well as tell the story of how the original Weiszfeld’s paper was discovered by Harlod Kuhn. Kuhn’s paper from 1973 is discussed in Section 4; we provide a different and simplified proof of its convergence theorem without requiring the usual collinearity assumption of the anchors. We also present his example on why the method can potentially “get stuck” at non-optimal anchor points, and recall his incorrect statement that the number of “bad” starting points (those leading to anchor points) is denumerable. Beginning from Kuhn’s statement in 1973 on the number of “bad” starting points, we present in Section 5 the attempts of resolving this issue until its final closure in 2002. Several modifications of Weiszfeld’s method in which the only difference is the way the method operates on anchor points are presented in Section 6 along with a method to pick the starting point in a way that ensures avoiding anchor points; a review of more elaborate modifications of the method ends the section. Section 7 uses the simplified analysis of the previous sections as well as modern analysis of gradient-based methods to establish a non-asymptotic sublinear rate of convergence of the sequence of function values generated by Weiszfeld’s method. In Section 8 we develop an accelerated version of the method which is based on a combination of an exact smoothing technique and the employment of an optimal gradient method leads to an accelerated version with an $O(1/k^2)$ rate of convergence. The paper ends in Section 9 where the impact of Weiszfeld’s method on other problems different than the Fermat-Weber problem is explored. For the convenience of the reader Appendix A contains a list of the most frequent notations used throughout the paper.

2 The Original Paper of Weiszfeld

Weiszfeld’s paper [51] was originally written in French; an English translation of the paper can be found in the 2009 paper [50]. The translation made by Frank Plastria contains in addition many interesting comments and observations. Taking a close look at the algorithm, one apparent default is the fact that it is actually not well defined. Weiszfeld assumed that the initial vector \mathbf{x}_0 is different from any anchor point, that is, $\mathbf{x}_0 \notin \mathcal{A}$. However, this is not enough to ensure that the sequence $\{\mathbf{x}_k\}_{k \geq 0}$ generated by the method is well defined, since it might happen that a certain iterate \mathbf{x}_k will belong to the anchor set \mathcal{A} , resulting with a division by zero in the computation of the next iterate \mathbf{x}_{k+1} . This situation can occur even if the optimal solution does not belong to \mathcal{A} . This error was recognized later by Kuhn and Kuenne in 1962 [29] (see also the discussion in Section 3).

Putting aside the issue of “getting stuck” at non-optimal anchor points, Weiszfeld was able to prove the monotonicity of the sequence of function values. We will repeat his arguments, but will use the notation used by Beck and Teboulle in [4], that will serve us later on in the new analysis of the method. We begin by providing a different presentation

of Weiszfeld's method. Define the auxiliary function $h : \mathbb{R}^n \times \mathbb{R}^n \setminus \mathcal{A} \rightarrow \mathbb{R}$ by

$$h(\mathbf{x}, \mathbf{y}) \equiv \sum_{i=1}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{y} - \mathbf{a}_i\|}, \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}. \quad (2.1)$$

Given an iterate \mathbf{x}_k , it is not difficult to show that the next iterate $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$ is determined as the minimizer of the function

$$s_k(\mathbf{x}) = h(\mathbf{x}, \mathbf{x}_k) = \sum_{i=1}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{x}_k - \mathbf{a}_i\|}$$

over \mathbb{R}^n . Indeed, $s_k(\cdot)$ is a strongly convex function and its unique minimizer is determined by the optimality condition

$$\nabla s_k(\mathbf{x}) = \mathbf{0}.$$

That is,

$$\sum_{i=1}^m \omega_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|} = \mathbf{0},$$

which after some simple algebraic manipulation can be seen to be equivalent to the relation $\mathbf{x} = T(\mathbf{x}_k)$. In other words, what we have shown is that for any $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$,

$$T(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} h(\mathbf{x}, \mathbf{y}).$$

We will now recall a technical lemma containing the latter property along with several other properties connecting the function f and the auxiliary function h , which will be the key in proving the monotonicity of the sequence of function values (see Lemma 1.1 in [4]).

Lemma 2.1 (Properties of the auxiliary function h). *The following properties hold.*

(i) For any $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$,

$$h(\mathbf{y}, \mathbf{y}) = f(\mathbf{y}).$$

(ii) For any $\mathbf{x} \in \mathbb{R}^n$ and all $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$,

$$h(\mathbf{x}, \mathbf{y}) \geq 2f(\mathbf{x}) - f(\mathbf{y}).$$

(iii) For any $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$,

$$T(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} h(\mathbf{x}, \mathbf{y}).$$

Proof. (i) Follows by simple substitution.

(ii) First, note that for every two real numbers $a \in \mathbb{R}$ and $b > 0$, the inequality

$$\frac{a^2}{b} \geq 2a - b,$$

holds true. Therefore, for every $i = 1, 2, \dots, m$, we have

$$\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{y} - \mathbf{a}_i\|} \geq 2\|\mathbf{x} - \mathbf{a}_i\| - \|\mathbf{y} - \mathbf{a}_i\|.$$

Multiplying the latter inequality by ω_i and summing over $i = 1, 2, \dots, m$, the result follows.

(iii) Follows by the discussion prior to the lemma. \square

Using Lemma 2.1, we are now able to prove the monotonicity property of the operator T with respect to f . The proof here relies on the same arguments of Weiszfeld [51] and of Kuhn [28].

Lemma 2.2 (Monotonicity property of T). *For every $\mathbf{y} \in \mathbb{R}^n \setminus \mathcal{A}$ we have*

$$f(T(\mathbf{y})) \leq f(\mathbf{y}), \quad (2.2)$$

and equality holds if and only if $T(\mathbf{y}) = \mathbf{y}$.

Proof. From Lemma 2.1(iii) we have that $T(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} h(\mathbf{x}, \mathbf{y})$, and by the strict convexity of the function $\mathbf{x} \rightarrow h(\mathbf{x}, \mathbf{y})$, one has

$$h(T(\mathbf{y}), \mathbf{y}) < h(\mathbf{x}, \mathbf{y}) \quad (2.3)$$

for every $\mathbf{x} \neq T(\mathbf{y})$. In particular, if $T(\mathbf{y}) \neq \mathbf{y}$, then

$$h(T(\mathbf{y}), \mathbf{y}) < h(\mathbf{y}, \mathbf{y}) = f(\mathbf{y}), \quad (2.4)$$

where the last equality follows from Lemma 2.1(i). Now, from Lemma 2.1(ii), we have

$$h(T(\mathbf{y}), \mathbf{y}) \geq 2f(T(\mathbf{y})) - f(\mathbf{y}),$$

which combined with (2.4), establishes the desired strict monotonicity. \square

Since the general step of Weiszfeld's method is defined by $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$, and since $\nabla f(\mathbf{x}^*) = \mathbf{0}$ if and only if $\mathbf{x}^* = T(\mathbf{x}^*)$ (see eq. (1.5)), we can immediately conclude that under the condition that the iterates of Weiszfeld's method do not belong to the anchor set \mathcal{A} , the method is nonincreasing and “gets stuck” only at optimal points.

Corollary 2.1 (Monotonicity of the sequence of function values). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's method and assume that $\mathbf{x}_k \notin \mathcal{A}$ for any $k \geq 0$. Then, $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$ for any $k \geq 0$, and equality holds if and only if \mathbf{x}_k is an optimal solution of problem (FW).*

3 Reinventing the Wheel

Weiszfeld’s method remained unknown until 1962. One evidence for this is the fact that the algorithm was rediscovered several times without any knowledge of Weiszfeld’s earlier work. As far as we know, the first to rediscover the method was Miehe in 1958 [31] who studied an even more general problem where the location of *several* points are to be determined. The paper is formulated in the two-dimensional plane, but the extension to \mathbb{R}^n is obvious. The derivation of the method in [31] is identical to the isolation process described in Section 1. Miehe, like Weiszfeld, did not treat the situation when the method reaches an anchor point.

Four years later, in 1962, Kuhn and Kuenne rediscovered Weiszfeld’s method for solving problem (FW) in the plane [29]. At the time of the writing of the paper Kuhn and Kuenne were not aware of Weiszfeld’s work, but an appendix called “added in proof” was added reading as follows.

[29, p. 33]: **Added in proof:** *In the period between the submission of this paper and its publication by the journal, the results have been extended and further literature bearing on the problem has been discovered. We shall sketch the nature of the extensions and list only those references that bear directly on the material developed in the body of the paper. First of all, the algorithm of the paper has been considered, independently of the present account, at least three times. The first published version seems to be that of E. Weiszfeld in the Tohoku Mathematical Journal...*

The story of how the authors were made aware of Weiszfeld’s method is known to us since in 2002, Weiszfeld (who changed his name to Andrew Vazsonyi) recalled the following story from the 1960.

[49, p. 12]: *After reading more, I discovered that a well-known mathematician, Harold W. Kuhn of Princeton University, had given a talk in Budapest on his discovery of an algorithm to solve the location problem. After the talk, a former colleague of mine walked to the blackboard and wrote in big letters: “VAZSONYI.”*

“Who is that?” Kuhn asked.

“The name of the Hungarian mathematician who discovered your algorithm thirty years ago”, my old friend said. “He lives in the United States but published his revolutionary approach under the name Weiszfeld.”

In the “Added in proof” section of [29], Kuhn and Kuenne mention that the convergence proof of Weiszfeld contains an error since the iterates may belong to the anchor set \mathcal{A} . They continue to claim that it can be proven that either $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$, then convergence to the optimal solution can be guaranteed, or that the method gets stuck at an anchor point ($\mathbf{x}_k \in \mathcal{A}$ for some $k \geq 0$). In addition, they hypothesise that the latter case may

only occur in “*at most a denumerable number of (starting) points in the convex hull of \mathcal{A}* ”. Later on, we will return to these claims and check their validity (see Section 5).

Another observation which appears in Kuhn and Kuenne [29] is that Weiszfeld’s method is in fact a gradient method. Indeed, a simple computation shows that an alternative representation of the operator T is given by

$$T(\mathbf{x}) = \mathbf{x} - \frac{1}{L(\mathbf{x})} \nabla f(\mathbf{x}) \quad (\mathbf{x} \notin \mathcal{A}), \quad (3.1)$$

where the operator $L : \mathbb{R}^n \setminus \mathcal{A} \rightarrow \mathbb{R}_{++}$ is defined by

$$L(\mathbf{x}) = \sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}. \quad (3.2)$$

Therefore, Weiszfeld’s method can be written as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L(\mathbf{x}_k)} \nabla f(\mathbf{x}_k). \quad (3.3)$$

One year later, in 1963, unaware of Weiszfeld’s contribution, Cooper [17] also reinvented the method, but again for the more general problem of multiple locations in \mathbb{R}^2 . Cooper did not provide a convergence analysis, but mentioned that in his numerical tests, the method works very well in comparison to other methods. It seems that after 1963, researchers from the optimization as well as the location communities were very well aware of the method, and Weiszfeld’s original paper [51] got its rightful credit.

4 The Paper of Kuhn from 1973 - The Beginning of a (Correct) Convergence Analysis

The 1973 paper of Kuhn [28] is a continuation of his joint paper with Kuenne from 1962 [29]. Besides re-establishing the monotonicity property of the sequence of objective function values (see Corollary 2.1), he was concerned with two theoretical questions - both mentioned in his “added in proof” section of the previous paper [29].

- A. Assuming that all the iterates do not belong to the anchor set \mathcal{A} , *i.e.* $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$, can convergence to the optimal solution be proven?
- B. Is the number of starting points of the method for which the method “gets stuck” at non-optimal anchor points denumerable?

Kuhn’s answer to both questions was yes. Unfortunately, as we will see later on, the answer to the second question was wrong. We will now explore in details each of the two theoretical questions A and B.

4.1 The Convergence of the Sequence

The theorem Kuhn proved in [28] is now recalled.

Theorem 4.1 (Convergence of Weiszfeld’s method). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be a sequence generated by the Weiszfeld’s method. If $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$, then the sequence $\{\mathbf{x}_k\}_{k \geq 0}$ converges to an optimal solution of problem (FW).*

Kuhn proved this theorem under the assumption that the anchors are not collinear. The assumption in the statement of Theorem 4.1, namely that the iterates do not belong to the anchor set \mathcal{A} , is a bit problematic since it is not clear how to guarantee that such a condition will hold. Probably the reason for such an assumption is the empirical observation that the method practically does not “get stuck” at non-optimal anchor points. The arguments used in [28] for proving Theorem 4.1 are quite lengthy and technical. We will provide here a different proof that utilizes the relations between the objective function f and the auxiliary function h . This approach will be also rather beneficial since it will be the basis for the rate of convergence analysis which will be discussed in Section 7. In addition, the proof does not require the assumption of collinearity of the anchors. Before proving the theorem, we will establish another property on the objective function f . This result is very similar to the so-called “descent lemma” for continuously differentiable functions (see, *e.g.*, [11]). However, its validity for the non-smooth function f is far from being obvious.

Lemma 4.1 (Descent lemma for the Fermat-Weber objective function). *Suppose that $\mathbf{y} \notin \mathcal{A}$. Then*

$$f(T(\mathbf{y})) \leq f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), T(\mathbf{y}) - \mathbf{y} \rangle + \frac{L(\mathbf{y})}{2} \|T(\mathbf{y}) - \mathbf{y}\|^2. \quad (4.1)$$

Proof. Note that the function $\mathbf{x} \mapsto h(\mathbf{x}, \mathbf{y})$ is quadratic with associated matrix $L(\mathbf{y})\mathbf{I}$. Therefore, its second-order Taylor expansion around \mathbf{y} is exact and can be written as

$$h(\mathbf{x}, \mathbf{y}) = h(\mathbf{y}, \mathbf{y}) + \langle \nabla_{\mathbf{x}} h(\mathbf{y}, \mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + L(\mathbf{y}) \|\mathbf{x} - \mathbf{y}\|^2.$$

Since $h(\mathbf{y}, \mathbf{y}) = f(\mathbf{y})$ (see Lemma 2.1(i)) and $\nabla_{\mathbf{x}} h(\mathbf{y}, \mathbf{y}) = 2\nabla f(\mathbf{y})$ (simple computation), we have that

$$h(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}) + 2 \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + L(\mathbf{y}) \|\mathbf{x} - \mathbf{y}\|^2.$$

Substituting $\mathbf{x} = T(\mathbf{y})$ in the latter identity yields

$$h(T(\mathbf{y}), \mathbf{y}) = f(\mathbf{y}) + 2 \langle \nabla f(\mathbf{y}), T(\mathbf{y}) - \mathbf{y} \rangle + L(\mathbf{y}) \|T(\mathbf{y}) - \mathbf{y}\|^2.$$

Hence, from Lemma 2.1(ii), we obtain

$$2f(T(\mathbf{y})) - f(\mathbf{y}) \leq f(\mathbf{y}) + 2 \langle \nabla f(\mathbf{y}), T(\mathbf{y}) - \mathbf{y} \rangle + L(\mathbf{y}) \|T(\mathbf{y}) - \mathbf{y}\|^2.$$

Therefore

$$2f(T(\mathbf{y})) \leq 2f(\mathbf{y}) + 2 \langle \nabla f(\mathbf{y}), T(\mathbf{y}) - \mathbf{y} \rangle + L(\mathbf{y}) \|T(\mathbf{y}) - \mathbf{y}\|^2,$$

which readily implies (4.1). □

Remark 4.1. It is interesting to note that in a way the latter result mimics known results on continuously differentiable functions. Suppose that g is a continuously differentiable function over \mathbb{R}^n , and assume that its gradient ∇g is Lipschitz continuous with parameter L_g , meaning that

$$\|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq L_g \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Then the “descent lemma” for such function, stating that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$g(\mathbf{y}) \leq g(\mathbf{x}) + \langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L_g}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

This is one of the key inequalities used to analyze the convergence properties of gradient-based methods, see for example [4, 35]. Of course, in our case, f is not differentiable, but nonetheless a version of the descent lemma still holds when \mathbf{x} is specifically chosen as $T(\mathbf{y})$, and when $L(\mathbf{y})$ takes the role of the Lipschitz constant. Note also that as was already mentioned, Weiszfeld’s method is a gradient method with stepsize $1/L(\mathbf{x}_k)$, which is also an indication that $L(\mathbf{x}_k)$ serves as a substitute for the Lipschitz constant since the gradient method for finding the minimizer of g is known to converge when the stepsize is chosen as $1/L_g$.

Using the descent lemma for the function f , we can now prove the following lemma stating an inequality that will be the basis for the convergence of the sequence as well as for the rate of convergence analysis that will be derived in Section 7. Note that we do not require the assumption on the collinearity of the anchor points and hence the optimal set is not necessarily a singleton and it will be denoted from now on by X^* . The optimal value will be denoted by f^* .

Lemma 4.2. *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be a sequence generated by Weiszfeld’s method and assume that $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$. Then, for any $\mathbf{x} \in \mathbb{R}^n$, the following inequality holds*

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}) \leq \frac{L(\mathbf{x}_k)}{2} (\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2). \quad (4.2)$$

Proof. Substituting $\mathbf{y} = \mathbf{x}_k$ in (4.1) and using the fact that $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$, we obtain

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L(\mathbf{x}_k)}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2. \quad (4.3)$$

By the gradient inequality we have that $f(\mathbf{x}_k) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x} \rangle$ for any $\mathbf{x} \in \mathbb{R}^n$, which combined with (4.3) yields

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x} \rangle + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L(\mathbf{x}_k)}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x} \rangle + \frac{L(\mathbf{x}_k)}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &= f(\mathbf{x}) + L(\mathbf{x}_k) \langle \mathbf{x}_k - \mathbf{x}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x} \rangle + \frac{L(\mathbf{x}_k)}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &= f(\mathbf{x}) + \frac{L(\mathbf{x}_k)}{2} (\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2), \end{aligned}$$

where the second equality follows from (3.3) and the last equality follows from the identity that $2 \langle \mathbf{w} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle = \|\mathbf{w} - \mathbf{v}\|^2 - \|\mathbf{w} - \mathbf{u}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2$ for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. This completes the proof. \square

Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's method with initial point \mathbf{x}_0 . In those cases that the left-hand side of (4.2) is positive, that is when $f(\mathbf{x}) \leq f(\mathbf{x}_k)$ for all $k \geq 0$, we get as a direct result of the latter lemma the so-called Fejer monotonicity of $\{\mathbf{x}_k\}_{k \geq 0}$ – a result which seems to be unknown in the literature.

Corollary 4.1 (Fejer monotonicity). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be a sequence generated by Weiszfeld's method and assume that $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$. Then, for any $\mathbf{x} \in \mathbb{R}^n$ which satisfies $f(\mathbf{x}) \leq f(\mathbf{x}_k)$ for all $k \geq 0$, the following inequality holds*

$$\|\mathbf{x}_{k+1} - \mathbf{x}\| \leq \|\mathbf{x}_k - \mathbf{x}\|.$$

In addition, the sequence $\{\mathbf{x}_k\}_{k \geq 0}$ is bounded.

Proof. Follows directly from (4.2) along with the fact that $f(\mathbf{x}_{k+1}) \geq f(\mathbf{x})$. The boundedness of the sequence then readily follows by the fact that $\|\mathbf{x}_k - \mathbf{x}\| \leq \|\mathbf{x}_0 - \mathbf{x}\|$ for any $k \geq 0$. \square

The convergence result of Kuhn, namely Theorem 4.1, can now be easily deduced from the Fejer monotonicity property of the sequence $\{\mathbf{x}_k\}_{k \geq 0}$.

Proof of Theorem 4.1. We will prove this result in two steps. First, we will prove that $\{\mathbf{x}_k\}_{k \geq 0}$ converges, and then we will show that its limit point is an optimal solution of problem (FW).

The sequence $\{\mathbf{x}_k\}_{k \geq 0}$ is bounded by Corollary 4.1. To prove the convergence of $\{\mathbf{x}_k\}_{k \geq 0}$, it only remains to show that all converging subsequences have the same limit. Suppose in contradiction that there exist two subsequences $\{\mathbf{x}_{k_j}\}_{j \geq 0}$ and $\{\mathbf{x}_{n_j}\}_{j \geq 0}$ converging to different limits $\tilde{\mathbf{x}}$ and $\bar{\mathbf{x}}$, respectively.

From Corollary 2.1, it follows that $f(\tilde{\mathbf{x}}) \leq f(\mathbf{x}_k)$ for all $k \geq 0$, and thus from Corollary 4.1, we get that the sequence $\{\|\mathbf{x}_k - \tilde{\mathbf{x}}\|\}_{k \geq 0}$ is nonincreasing. Since this sequence is also bounded from below, it converges to some $\bar{l}_1 \in \mathbb{R}$. It is clear that

$$\lim_{k \rightarrow \infty} \|\mathbf{x}_k - \tilde{\mathbf{x}}\| = \lim_{j \rightarrow \infty} \|\mathbf{x}_{k_j} - \tilde{\mathbf{x}}\| = 0.$$

But, on the other hand,

$$\lim_{k \rightarrow \infty} \|\mathbf{x}_k - \tilde{\mathbf{x}}\| = \lim_{j \rightarrow \infty} \|\mathbf{x}_{n_j} - \tilde{\mathbf{x}}\| = \|\bar{\mathbf{x}} - \tilde{\mathbf{x}}\|.$$

Hence $\bar{l}_1 = 0 = \|\bar{\mathbf{x}} - \tilde{\mathbf{x}}\|$, which is obviously a contradiction. This proves that $\{\mathbf{x}_k\}_{k \geq 0}$ converges.

We denote by $\tilde{\mathbf{x}}$ the limit of $\{\mathbf{x}_k\}_{k \geq 0}$. Now we will prove that $\tilde{\mathbf{x}}$ is an optimal solution of problem (FW). It is clear that if $\tilde{\mathbf{x}} \notin \mathcal{A}$, then taking the limit as $k \rightarrow \infty$ in the equation

$\mathbf{x}_{k+1} = T(\mathbf{x}_k)$, and using the continuity of the operator T at non-anchor points, we obtain that $\tilde{\mathbf{x}} = T(\tilde{\mathbf{x}})$. The optimality of $\tilde{\mathbf{x}}$ now follows immediately from (1.5). On the other hand, if $\tilde{\mathbf{x}} \in \mathcal{A}$ then there exists $j \in \{1, 2, \dots, m\}$ such that $\tilde{\mathbf{x}} = \mathbf{a}_j$. From Lemma 2.1(iii) we have

$$\nabla_{\mathbf{x}} h(\mathbf{x}_{k+1}, \mathbf{x}_k) = \mathbf{0},$$

which can be written explicitly as follows

$$\sum_{i=1}^m \omega_i \frac{\mathbf{x}_{k+1} - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|} = \mathbf{0}.$$

Thus,

$$\sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{x}_{k+1} - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|} = -\omega_j \frac{\mathbf{x}_{k+1} - \mathbf{a}_j}{\|\mathbf{x}_k - \mathbf{a}_j\|},$$

and after taking the norm on both sides we get

$$\left\| \sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{x}_{k+1} - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|} \right\| = \omega_j \frac{\|\mathbf{x}_{k+1} - \mathbf{a}_j\|}{\|\mathbf{x}_k - \mathbf{a}_j\|} \leq \omega_j, \quad (4.4)$$

where the inequality follows from the fact that the sequence $\{\|\mathbf{x}^k - \mathbf{a}_j\|\}_{k \geq 0}$ is nonincreasing (see Corollary 2.1 using the fact that $f(\mathbf{a}_j) \leq f(\mathbf{x}_k)$ for all $k \geq 0$). Taking the limit in (4.4) as $k \rightarrow \infty$, yields the inequality

$$\left\| \sum_{i=1, i \neq j}^m \omega_i \frac{\tilde{\mathbf{x}} - \mathbf{a}_i}{\|\tilde{\mathbf{x}} - \mathbf{a}_i\|} \right\| \leq \omega_j,$$

which by Theorem 1.1, shows that $\tilde{\mathbf{x}} = \mathbf{a}_j$ is an optimal solution of (FW). \square

4.2 “Bad” Starting Points

As was already mentioned, even in the earlier paper of Kuhn and Kuenne [29] from 1962, it was obvious that Weiszfeld’s method can reach a non-optimal anchor point. This situation is described in the literature as “getting stuck”, and starting points of the method leading to this situation are called “bad” starting points. In his 1973 paper, Kuhn gave an example of such a case, as we recall now.

Example 1 (Kuhn’s counterexample). The example is in the 2-dimensional space and the anchors are $\mathbf{a}_1 = (-2, 0)^T$, $\mathbf{a}_2 = (-1, 0)^T$, $\mathbf{a}_3 = (1, 0)^T$, $\mathbf{a}_4 = (2, 0)^T$, $\mathbf{a}_5 = (0, 1)^T$ and $\mathbf{a}_6 = (0, -1)^T$. All the weights are one. It is easy to see that $\nabla f(0, 0) = \mathbf{0}$, so that the optimal solution of the problem is $\mathbf{x}^* = \mathbf{0}$. Kuhn then studied the behavior of the operator

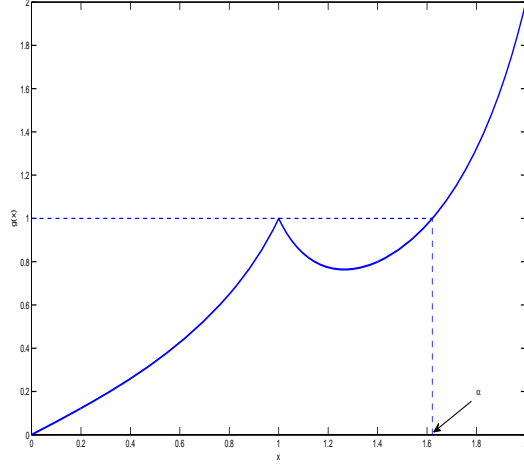


Figure 2: Kuhn’s construction of a “bad” starting point. $T\left((\alpha, 0)^T\right) = \mathbf{a}_3$ where α is the solution of the equation $g(\alpha) = 1$.

T on points on the x -axis given by $(x, 0)^T$. Then

$$\begin{aligned} T\left((x, 0)^T\right) &= \frac{1}{\sum_{i=1}^6 \frac{1}{\|(x, 0)^T - \mathbf{a}_i\|}} \sum_{i=1}^6 \frac{\mathbf{a}_i}{\|(x, 0)^T - \mathbf{a}_i\|} \\ &= \frac{\begin{pmatrix} \frac{-2}{|x+2|} + \frac{-1}{|x+1|} + \frac{1}{|x-1|} + \frac{2}{|x-2|} \\ 0 \end{pmatrix}}{\frac{1}{|x+2|} + \frac{1}{|x+1|} + \frac{1}{|x-1|} + \frac{1}{|x-2|} + \frac{2}{\sqrt{x^2+1}}} \equiv \begin{pmatrix} g(x) \\ 0 \end{pmatrix}. \end{aligned}$$

A value $\alpha \in [0, 2]$ for which $T\left((\alpha, 0)^T\right) = \mathbf{a}_3$ was then found, that is, a value α satisfying

$$g(\alpha) = 1. \quad (4.5)$$

The plot of g over the interval $[0, 2]$ can be found in Figure 2 where the solution of (4.5) which is $\alpha = 1.6213$ (up to three digits of accuracy) is described.

Kuhn claimed that the number of “bad” starting points is always denumerable. The key argument used to show this result is quoted here.

[28, p. 107]: *If we insert T from a vertex \mathbf{a}_i , we must solve algebraic equations. Thus we obtain a finite number of \mathbf{x}_0 such that $T(\mathbf{x}_0) = \mathbf{a}_i$.*

However, this argument is incorrect as will be explained in the following section.

5 Counting the “Bad” Starting Points

The flaw in Kuhn’s argument is that actually algebraic system of equations can have a continuum number of solutions. This was shown by Chandrasekaran and Tamir in their 1989 paper [15], where the following counterexample to Kuhn’s claim in \mathbb{R}^3 was given.

Example 2 (Chandrasekaran and Tamir’s counterexample to Kuhn). Consider the problem in \mathbb{R}^3 with anchors $\mathbf{a}_1 = (1, 0, 0)^T$, $\mathbf{a}_2 = (-1, 0, 0)^T$, $\mathbf{a}_3 = (0, 0, 0)^T$, $\mathbf{a}_4 = (0, 2, 0)^T$ and $\mathbf{a}_5 = (0, -2, 0)^T$. Let $\omega_1 = \omega_2 = \omega_3 = \omega_5 = 1$ and $\omega_4 = 3$. Consider the point \mathbf{a}_3 which is not optimal since $f(\mathbf{a}_3) = 10 > f(0, 1, 0) = 7 + 2\sqrt{2}$. To show that the algebraic system $T(\mathbf{x}) = \mathbf{a}_3$ has an infinite number of solutions, consider the points of the form $\mathbf{x} = (0, y, z)^T$. Then $T(\mathbf{x}) = \mathbf{a}_3$ is equivalent to the system

$$\frac{3\mathbf{a}_4}{\|\mathbf{x} - \mathbf{a}_4\|} + \frac{\mathbf{a}_5}{\|\mathbf{x} - \mathbf{a}_5\|} = 0.$$

That is,

$$\frac{1}{\sqrt{(y-2)^2 + z^2}} (0, 6, 0)^T + \frac{1}{\sqrt{(y+2)^2 + z^2}} (0, -2, 0)^T = (0, 0, 0)^T,$$

which is the same as

$$36(y+2)^2 + 36z^2 - 4((y-2)^2 + z^2) = 0.$$

After some simple algebraic manipulation, we conclude that all the points on the circle $(y+2.5)^2 + z^2 = 2.25$ solve the system $T(\mathbf{x}) = \mathbf{a}_3$.

The latter example is special in the sense that the anchors, although not collinear, reside in a lower dimensional affine subspace. Chandrasekaran and Tamir conjectured that such a situation is the only one in which a continuum of “bad” starting points can occur.

[15, p. 295]: *“In view of the above examples we conjecture that if the non-collinearity is replaced by the stronger assumption that the convex hull of the points $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ is of full dimension, then the algebraic system $T(\mathbf{x}) = \mathbf{a}_i$ has a finite number of solutions for $i = 1, 2, \dots, m$. Phrased differently, the conjecture is that for each $i = 1, 2, \dots, m$, there is a finite number of solutions to $T(\mathbf{x}) = \mathbf{a}_i$ in the minimal affine set containing the points $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$.”*

To write it explicitly, the first conjecture of Chandrasekaran and Tamir is formulated as follows.

Conjecture 1 (Chandrasekaran and Tamir conjecture). *When the affine hull of the anchor set \mathcal{A} is the entire space \mathbb{R}^n , the number of solutions of the system $T(\mathbf{x}) = \mathbf{a}_i$, for any $i = 1, 2, \dots, m$, is finite, and hence there is a denumerable number of “bad” starting points.*

As quoted above, Chandrasekaran and Tamir also had a more general conjecture that the number of “bad” starting points is denumerable when the starting point is restricted a priori to be in the affine hull of the anchor set \mathcal{A} . Conjecture 1 was resolved in 1995 by Brimberg [12]. More precisely, Brimberg proved the following result (written in this paper’s terminology).

Theorem 5.1 (cf. [12, Theorem 1, p. 75]). *The set of starting points \mathbf{x}_0 which will terminate the sequence generated by Weiszfeld’s method at some anchor point \mathbf{a}_i , $i = 1, 2, \dots, m$, after a finite number of iterations is denumerable if the affine hull of \mathcal{A} is \mathbb{R}^n .*

At this point, one would think that the issue of “counting” the number of bad starting points was resolved. However, Brimberg also claimed that in fact the number of “bad” starting points is denumerable *if and only if* the affine hull of \mathcal{A} is \mathbb{R}^n . The “only if” claim is a not part of Chandrasekaran and Tamir’s conjecture. Later on, in 2002, Cánovas, Cañavate and Marín showed in [14] that the “only if” claim is not correct by providing two counterexamples. We will present the first counterexample which shows a setting in which the number of “bad” starting points is denumerable (in fact zero) even though the affine hull of \mathcal{A} is not the entire space \mathbb{R}^n .

Example 3 (Counterexample to Brimberg’s “only if” part). Consider the unweighted Fermat-Weber problem with $m = n = 3$ given by the anchors $\mathbf{a}_1 = (1, 0, 0)^T$, $\mathbf{a}_2 = (0, 1, 0)^T$ and $\mathbf{a}_3 = (0, 0, 0)$. Note that \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 are the extreme points of $\text{conv}(\mathcal{A})$, and that for any $\mathbf{x} \notin \mathcal{A}$,

$$T(\mathbf{x}) = \frac{1}{\sum_{i=1}^3 \frac{1}{\|\mathbf{x} - \mathbf{a}_i\|}} \sum_{i=1}^3 \frac{1}{\|\mathbf{x} - \mathbf{a}_i\|} \mathbf{a}_i.$$

Hence, $T(\mathbf{x})$ is a convex combination of \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 with positive coefficients. However, this means that $T(\mathbf{x})$ cannot be equal to any of the extreme points, that is, to an anchor point. Therefore, the system $T(\mathbf{x}) = \mathbf{a}_i$ has no solutions for any $i = 1, 2, 3$, and in this case Weiszfeld’s method is well defined for any starting point which is not in \mathcal{A} .

6 Bypassing the Anchor Points - Modifying the Method

6.1 The Modified Weiszfeld’s Method

Since the issue of bumping into anchor points is an important issue in the convergence analysis of the method, these instances should be treated. In this section we will begin by reviewing a modification of the method in which the problem of reaching an anchor point is treated in a “surgical” manner. Specifically, the class of methods that we consider coincides with Weiszfeld’s method when the current iterate \mathbf{x}_k , for some $k \geq 0$, is not an anchor point, and when the iterate is equal to some \mathbf{a}_i , then the next iterate will be equal to \mathbf{a}_i if \mathbf{a}_i is optimal and equal to another point, with a smaller function value when \mathbf{a}_i

is not optimal. To test the optimality of the anchor points, we are required to define the following quantities (see the optimality conditions in Theorem 1.1)

$$\mathbf{R}_j \equiv \sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_i - \mathbf{a}_j\|}, \quad j = 1, 2, \dots, m.$$

The anchor point \mathbf{a}_j , $j = 1, 2, \dots, m$, is optimal if and only if

$$\|\mathbf{R}_j\| \leq \omega_j.$$

The general scheme for the modified approach can be written as follows.

Modified Weiszfeld's Method

Initialization. $\mathbf{x}_0 \in \mathbb{R}^n$.

General Step ($k = 0, 1, \dots$)

$$\mathbf{x}_{k+1} = \tilde{T}(\mathbf{x}_k) = \begin{cases} T(\mathbf{x}_k), & \mathbf{x}_k \notin \mathcal{A}, \\ \mathbf{a}_j, & \mathbf{x}_k = \mathbf{a}_j \ (1 \leq j \leq m) \text{ and } \|\mathbf{R}_j\| \leq \omega_j \\ S(\mathbf{a}_j), & \mathbf{x}_k = \mathbf{a}_j \ (1 \leq j \leq m) \text{ and } \|\mathbf{R}_j\| > \omega_j. \end{cases} \quad (6.1)$$

The new operator \tilde{T} coincides with the usual Weiszfeld's operator T at non-anchor points. At non-optimal anchor points, another operator, which is denoted by S is invoked. The question that arises is of course how to define the operator S on a specific non-optimal anchor point \mathbf{a}_j , $j = 1, 2, \dots, m$. However, at this point, we will just require that for a non-optimal anchor point \mathbf{a}_j , the point $S(\mathbf{a}_j)$ will have a smaller function value than \mathbf{a}_j . Under this condition, based on Kuhn's convergence result (see Theorem 4.1), we can prove convergence of the sequence generated by the modified Weiszfeld's method. Note that we do not require the non-collinearity assumption that is always assumed in the literature.

Theorem 6.1 (Convergence of the modified Weiszfeld's method). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be a sequence generated by the modified Weiszfeld's method. Assume that $f(S(\mathbf{a}_j)) < f(\mathbf{a}_j)$ for any $\mathbf{a}_j \in \mathcal{A}$ for which $\|\mathbf{R}_j\| > \omega_j$. Then, the sequence $\{\mathbf{x}_k\}_{k \geq 0}$ converges to an optimal solution of problem (FW).*

Proof. There are two options. Either the sequence “gets stuck” at a fixed point of the method, *i.e.*, at a point $\tilde{\mathbf{x}}$ for which $\tilde{T}(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}$. If $\tilde{\mathbf{x}} \notin \mathcal{A}$, then by (1.5), it follows that $\nabla f(\tilde{\mathbf{x}}) = 0$ and hence the method converges to the optimal solution $\tilde{\mathbf{x}}$. If $\tilde{\mathbf{x}} = \mathbf{a}_j$ for some $j \in \{1, 2, \dots, m\}$, then since $S(\mathbf{a}_j) \neq \mathbf{a}_j$ if \mathbf{a}_j is not optimal, it follows that \mathbf{a}_j must be optimal, and we conclude that once again the sequence generated by the method converges to the optimal solution $\tilde{\mathbf{x}}$. The second option is when the sequence does not get stuck at a fixed point. In this case the sequence generated by the method is strictly monotone: $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ for all $k \geq 0$ (follows from Corollary 2.1 and the assumption that $f(S(\mathbf{a}_j)) < f(\mathbf{a}_j)$ for any $\mathbf{a}_j \in \mathcal{A}$ for which $\|\mathbf{R}_j\| > \omega_j$). Therefore, the iterates are

different from each other and thus, since there are only a finite number of anchor points, it follows that there exists a positive integer K such that $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq K$. The sequence $\{\mathbf{x}_k\}_{k \geq K}$ is the sequence generated by Weiszfeld's method with initial point \mathbf{x}_K , and since none of its elements is in \mathcal{A} , it follows from Theorem 4.1 that it converges to an optimal solution of problem (FW). \square

Note that since in practice Weiszfeld's method does not actually reach anchor points, the modified method is actually the same as Weiszfeld's method for all practical purposes.

6.2 Choosing the Operator S

The most natural way to define S on an anchor point \mathbf{a}_j , $j = 1, 2, \dots, m$, is to find a descent direction of f at \mathbf{a}_j , and take a step along this direction. To find such a descent direction, note that the objective function f can be written as

$$f(\mathbf{x}) = \omega_j \|\mathbf{x} - \mathbf{a}_j\| + f_j(\mathbf{x}),$$

where

$$f_j(\mathbf{x}) \equiv \sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\|.$$

Therefore, taking a direction $\mathbf{d} \in \mathbb{R}^n$ satisfying $\|\mathbf{d}\| = 1$, we can define the function

$$\alpha(t) = f(\mathbf{a}_j + t\mathbf{d}) = \omega_j t \|\mathbf{d}\| + f_j(\mathbf{a}_j + t\mathbf{d}) = \omega_j t + f_j(\mathbf{a}_j + t\mathbf{d}).$$

The directional derivative of f at \mathbf{a}_j in the direction of \mathbf{d} is given by (note that f_j is differentiable at \mathbf{a}_j):

$$f'(\mathbf{a}_j; \mathbf{d}) = \alpha'(0) = \omega_j + f'_j(\mathbf{a}_j; \mathbf{d}) = \omega_j + \langle \nabla f_j(\mathbf{a}_j), \mathbf{d} \rangle.$$

The smallest directional derivative is attained at $\mathbf{d} = \mathbf{d}_j$ where $\mathbf{d}_j = -\nabla f_j(\mathbf{a}_j) / \|\nabla f_j(\mathbf{a}_j)\|$, which is the steepest descent direction. Since $\nabla f_j(\mathbf{a}_j) = \mathbf{R}_j$, we can summarize by writing that the steepest descent direction of f at \mathbf{a}_j is

$$\mathbf{d}_j = -\frac{\mathbf{R}_j}{\|\mathbf{R}_j\|}.$$

In all of the papers that deal with this slightly modified Weiszfeld's method, the operator S is chosen as

$$S(\mathbf{a}_j) = \mathbf{a}_j + t_j \mathbf{d}_j,$$

where t_j is some well chosen stepsize. Several choices of the stepsize were discussed in the literature. Ostresh [39] (1978) considered the following stepsize

$$t_j = t_O \equiv c \frac{\|\mathbf{R}_j\| - \omega_j}{L(\mathbf{a}_j)}, \quad [\text{Ostresh [39]}]$$

where $c \in [1, 2]$, and the definition of the operator L (originally given in (3.2)) is extended to include also anchor points

$$L(\mathbf{x}) = \begin{cases} \sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}, & \mathbf{x} \notin \mathcal{A}, \\ \sum_{i=1, i \neq j}^m \frac{\omega_i}{\|\mathbf{a}_j - \mathbf{a}_i\|}, & \mathbf{x} = \mathbf{a}_j \ (1 \leq j \leq m). \end{cases} \quad (6.2)$$

A totally different analysis of Vardi and Zhang [48] (2001) results with the stepsize

$$t_j = t_{VZ} \equiv \frac{\|\mathbf{R}_j\| - \omega_j}{L(\mathbf{a}_j)}, \text{ [Vardi and Zhang [48]]}$$

meaning that this is the stepsize of Ostresh with $c = 1$. The following different stepsize was considered by Rautenbach et. al. [43] (2004),

$$t_j = t_R \equiv \min \left\{ \frac{s_j}{2}, \frac{\|\mathbf{R}_j\| - \omega_j}{4L(\mathbf{a}_j)} \right\} \text{ [Rautenbach et. al. [43]]}$$

where $s_j := \min \{\|\mathbf{a}_j - \mathbf{a}_i\| : i \neq j, 1 \leq i \leq m\}$. Obviously, the largest step is the one given by Ostresh when c is taken to be 2, and the smallest step is the one given by Rautenbach. We will now give a proof that indeed the stepsize given by Vardi and Zhang results with a decrease in the function value. We also give an explicit expression for a lower bound on the amount of decrease resulting from taking the Vardi-Zhang stepsize. This new result will be important later on in establishing the complexity results of the method (see Section 7). The technical and lengthy proof of the lemma can be found in Appendix B.

Lemma 6.1. *Suppose that \mathbf{a}_j , for some $j \in \{1, 2, \dots, m\}$, is not an optimal solution of problem (FW), meaning that $\|\mathbf{R}_j\| > \omega_j$. Then*

$$f(\mathbf{a}_j) - f(\mathbf{a}_j + t_j \mathbf{d}_j) \geq \frac{(\|\mathbf{R}_j\| - \omega_j)^2}{2L(\mathbf{a}_j)},$$

where $t_j = t_{VZ} = (\|\mathbf{R}_j\| - \omega_j) / L(\mathbf{a}_j)$.

6.3 Choosing the Starting Point

The modified method was devised in order to relax the assumption that $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$ in Kuhn's convergence result. However, we can use the operator S to carefully pick a starting point that will guarantee that the iterates will not coincide with anchor points. The simple procedure for choosing the starting point, which we call "the SP method", is now described.

SP method

(a) Let

$$p \in \operatorname{argmin} \{f(\mathbf{a}_i) : 1 \leq i \leq p\}.$$

(b) If \mathbf{a}_p is optimal (easily checked by verifying that $\|\mathbf{R}_p\| \leq \omega_p$), then the output is \mathbf{a}_p with an indication that it is the optimal solution. Otherwise, if \mathbf{a}_p is not optimal, the output is the point $S(\mathbf{a}_p)$.

The important property of the SP method is that if there is no optimal anchor point, the output of the method is a starting point \mathbf{x}_0 which satisfies

$$f(\mathbf{x}_0) < \min \{f(\mathbf{a}_1), f(\mathbf{a}_2), \dots, f(\mathbf{a}_p)\}.$$

The latter inequality, along with the monotonicity property of the sequence of function values (see Corollary 2.1) implies that $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$ and hence the convergence of the sequence to an optimal solution is assured by Theorem 4.1. We summarize this discussion in the following corollary. The underlying assumption is that there are no optimal anchor points (otherwise, the SP method will produce the optimal solution). In addition, under this assumption the anchors are necessarily not collinear, implying that the optimal solution is unique.

Corollary 6.1. *Suppose that there is no optimal anchor point for problem (FW). Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's method with starting point \mathbf{x}_0 produced by the SP method. Then, $\mathbf{x}_k \notin \mathcal{A}$ for all $k \geq 0$ and $\mathbf{x}_k \rightarrow \mathbf{x}^*$ as $k \rightarrow \infty$.*

6.4 Further Modifications of Weiszfeld's Method

Aside of the rather local modifications of Weiszfeld's method mentioned in the previous subsections, many more modifications were suggested in the literature. In the 1978 paper [39], Ostresh suggested to accelerate the Weiszfeld's method by using the following idea. Let $\lambda \in [1, 2]$ and consider the operator

$$T_\lambda(\mathbf{y}) = \mathbf{y} + \lambda(T(\mathbf{y}) - \mathbf{y}).$$

Ostresh proved the convergence of the modified method defined by $\mathbf{x}_{k+1} = T_\lambda(\mathbf{x}_k)$ for any choice of $\lambda \in [1, 2]$, and under a suitable treatment of the anchor points. The use of a stepsize can also be found in [16] (1984). In [20] (1992), Drezner uses an adaptive method in order to choose a different λ at each iteration, so that the method reads as $\mathbf{x}_{k+1} = T_{\lambda_k}(\mathbf{x}_k)$. Another acceleration technique was proposed by Drezner in [21] (1995) where Steffensen's method, which is a general acceleration scheme for fixed point method, was tested numerically against Weiszfeld's method with several stepsize strategies for choosing λ .

Deviating from the main focus of this paper, which is Weiszfeld's method, we note that the Fermat-Weber problem can also be solved by more sophisticated methods than Weiszfeld's method. For example, Calamai and Conn [13] and Overton [40] solved the more general problem of minimizing a sum of norms of affine functions by Newton methods combined with an active set approach. In addition, the problem can be recast as a second order cone programming and solved via interior point methods [7, 37]. A specially devised primal-dual interior point method for the minimization of the sum of Euclidean norms was analyzed by Andersen et al. in [1].

There are of course many variations of Weiszfeld's method when the problem to be solved is not the Fermat-Weber problem, but this will be the subject of Section 9.

7 Rate of Convergence of Weiszfeld's Method

Local and asymptotic rate of convergence of Weiszfeld's method was discussed in the 1974 paper of Katz [26]. However, there does not seem to be global non-asymptotic rate of convergence analysis of Weiszfeld's method in the literature. The main objective of this section is to derive such rate of convergence, but before, we would like to recall several related results on global non-asymptotic rate of convergence for gradient-based methods. This type of results are usually of the form

$$f(\mathbf{x}_k) - f^* \leq \frac{C}{k^\theta},$$

where $C, \theta > 0$ are constants. For example, for nonsmooth convex problems we can employ under some unrestrictive conditions the subgradient method to solve problem (FW). This will result with a rate of convergence of $O(1/\sqrt{k})$ (that is, $\theta = \frac{1}{2}$), see *e.g.*, [35, 41]. In this case, since the method is not monotone, the bound is on $\min_{n=1,2,\dots,k} f(\mathbf{x}_n) - f^*$, rather than on $f(\mathbf{x}_k) - f^*$. If the problem is smooth, then the gradient method can be employed and the convergence rate will accelerate to $O(1/k)$ (corresponding to $\theta = 1$). Another option is to use an “optimal” gradient method which uses the memory of the last *two* iterations. These methods have a rate of convergence of $O(1/k^2)$ (corresponding to $\theta = 2$). For a wealth of fundamental results on these issues, see for instance, [3, 4, 34, 35].

In principal, since problem (FW) is nonsmooth, and since Weiszfeld's method is a type of a gradient method, it seems logical to assume that only the inferior $O(1/\sqrt{k})$ can be established. However, as we shall see, when choosing the starting point carefully, we can actually prove a rate of convergence of $O(1/k)$ even though the problem is nonsmooth. Later on, we will even show that by modifying the method we can establish the fast $O(1/k^2)$ rate of convergence.

Throughout this section we assume that the starting point is chosen according to the SP method with the Vardi-Zhang stepsize strategy. That is,

$$\mathbf{x}_0 = \mathbf{a}_p + t_p \mathbf{d}_p, \tag{7.1}$$

where

$$p \in \operatorname{argmin} \{f(\mathbf{a}_i) : 1 \leq i \leq p\}, \tag{7.2}$$

and

$$\mathbf{d}_p = -\frac{\mathbf{R}_p}{\|\mathbf{R}_p\|}, \quad t_p = \frac{\|\mathbf{R}_p\| - \omega_p}{L(\mathbf{a}_p)}. \tag{7.3}$$

The only assumption on the data is that there is no optimal anchor point. Otherwise, the SP method will produce an optimal solution. Note also that in this setting the anchors are necessarily not collinear and hence there exists a unique optimal solution \mathbf{x}^* . Note that in the premise of the following theorem, which will be the basis for the main convergence

analysis, we assume that the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$ is upper bounded. Later on, in Lemma 7.2, we will show the validity of this assumption, as well as find an explicit expression for the upper bound.

Theorem 7.1 (Sublinear rate of convergence of Weiszfeld's method). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's with \mathbf{x}_0 chosen by (7.1)-(7.3). Then, for any $k \geq 0$, we have*

$$f(\mathbf{x}_k) - f^* \leq \frac{\bar{L}}{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|^2, \quad (7.4)$$

where \bar{L} is an upper bound of the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$.

Proof. From (4.2) with $k = n$, we have

$$f(\mathbf{x}_{n+1}) - f^* \leq \frac{L(\mathbf{x}_n)}{2} (\|\mathbf{x}_n - \mathbf{x}^*\|^2 - \|\mathbf{x}_{n+1} - \mathbf{x}^*\|^2).$$

Therefore, by the Fejer monotonicity of the sequence $\{\mathbf{x}_k\}_{k \geq 0}$ (see Corollary 4.1) and the boundedness of the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$, we get

$$f(\mathbf{x}_{n+1}) - f^* \leq \frac{\bar{L}}{2} (\|\mathbf{x}_n - \mathbf{x}^*\|^2 - \|\mathbf{x}_{n+1} - \mathbf{x}^*\|^2).$$

Summing this inequality over $n = 0, \dots, k-1$ gives

$$\begin{aligned} \sum_{n=0}^{k-1} (f(\mathbf{x}_{n+1}) - f^*) &\leq \frac{\bar{L}}{2} \sum_{n=0}^{k-1} (\|\mathbf{x}_n - \mathbf{x}^*\|^2 - \|\mathbf{x}_{n+1} - \mathbf{x}^*\|^2) \\ &= \frac{\bar{L}}{2} (\|\mathbf{x}_0 - \mathbf{x}^*\|^2 - \|\mathbf{x}_k - \mathbf{x}^*\|^2) \\ &\leq \frac{\bar{L}}{2} \|\mathbf{x}_0 - \mathbf{x}^*\|^2. \end{aligned}$$

Since the sequence $\{f(\mathbf{x}_k)\}_{k \geq 0}$ is nonincreasing (see Corollary 2.1) we have

$$k(f(\mathbf{x}_k) - f^*) \leq \sum_{n=0}^{k-1} (f(\mathbf{x}_{n+1}) - f^*) \leq \frac{\bar{L}}{2} \|\mathbf{x}_0 - \mathbf{x}^*\|^2,$$

that is,

$$f(\mathbf{x}_k) - f^* \leq \frac{\bar{L}}{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|^2.$$

This proves the desired result. \square

All that is left to find an explicit upper bound of the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$. For that, we will use the following result which establishes a lower bound on the distance between the anchor points to those points with smaller function values than $f(\mathbf{x}_0)$ where \mathbf{x}_0 is chosen by (7.1)-(7.3).

Lemma 7.1 (Lower bound on $\|\mathbf{a}_i - \mathbf{x}\|$). *Let \mathbf{x}_0 be chosen by (7.1)-(7.3). Then, for any $i = 1, 2, \dots, m$ and any \mathbf{x} satisfying $f(\mathbf{x}) \leq f(\mathbf{x}_0)$, the inequality*

$$\|\mathbf{x} - \mathbf{a}_i\| \geq \frac{f(\mathbf{a}_i) - f(\mathbf{x}_0)}{\omega},$$

holds with $\omega = \sum_{i=1}^m \omega_i$.

Proof. From the fact that $f(\mathbf{x}) \leq f(\mathbf{x}_0)$, the convexity of f and the Cauchy-Schwarz inequality it follows, for any $i = 1, 2, \dots, m$ and any \mathbf{x} , that

$$f(\mathbf{a}_i) - f(\mathbf{x}_0) \leq f(\mathbf{a}_i) - f(\mathbf{x}) \leq \langle \mathbf{v}_i, \mathbf{a}_i - \mathbf{x} \rangle \leq \|\mathbf{v}_i\| \|\mathbf{a}_i - \mathbf{x}\|, \quad (7.5)$$

where $\mathbf{v}_i \in \partial f(\mathbf{a}_i)$. Note that the subdifferential set $\partial f(\mathbf{a}_i)$ can be written as

$$\partial f(\mathbf{a}_i) = \left\{ \sum_{j=1, j \neq i}^m \omega_j \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_i - \mathbf{a}_j\|} + \omega_i \mathbf{z}_i : \|\mathbf{z}_i\| \leq 1 \right\}.$$

Therefore, there exists $\tilde{\mathbf{z}}_i$ such that $\|\tilde{\mathbf{z}}_i\| \leq 1$ and

$$\mathbf{v}_i = \sum_{j=1, j \neq i}^m \omega_j \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_i - \mathbf{a}_j\|} + \omega_i \tilde{\mathbf{z}}_i.$$

Hence from the triangle inequality we get

$$\|\mathbf{v}_i\| = \left\| \sum_{j=1, j \neq i}^m \omega_j \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_i - \mathbf{a}_j\|} + \omega_i \tilde{\mathbf{z}}_i \right\| \leq \sum_{j=1, j \neq i}^m \omega_j \left\| \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_i - \mathbf{a}_j\|} \right\| + \omega_i \|\tilde{\mathbf{z}}_i\| \leq \sum_{i=1}^m \omega_i = \omega,$$

which combined with (7.5) yields the inequality

$$f(\mathbf{a}_i) - f(\mathbf{x}_0) \leq \omega \|\mathbf{a}_i - \mathbf{x}\|.$$

The desired result now follows by dividing the last inequality by ω . □

We can now find an upper bound of the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$. This is done in the next lemma.

Lemma 7.2 (Upper bound of the sequence $\{L(\mathbf{x}_k)\}_{k \geq 0}$). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's method with \mathbf{x}_0 chosen by (7.1)-(7.3). Then, for any $k \geq 0$, we have*

$$L(\mathbf{x}_k) \leq \frac{2L(\mathbf{a}_p) \omega^2}{(\|\mathbf{R}_p\| - \omega_p)^2}, \quad (7.6)$$

where $\omega = \sum_{i=1}^m \omega_i$.

Proof. By the monotonicity of the sequence $\{f(\mathbf{x}_k)\}_{k \geq 0}$ (see Corollary 2.1) it follows that $f(\mathbf{x}_k) \leq f(\mathbf{x}_0)$ for all $k \geq 0$. Therefore, from Lemma 7.1 it follows that for any $i \in \{1, 2, \dots, m\}$, we have

$$\frac{1}{\|\mathbf{x}_k - \mathbf{a}_i\|} \leq \frac{\omega}{f(\mathbf{a}_i) - f(\mathbf{x}_0)}.$$

Since $f(\mathbf{x}_0) < f(\mathbf{a}_p) \leq f(\mathbf{a}_i)$ for all $1 \leq i \leq m$, we deduce that

$$\frac{1}{\|\mathbf{x}_k - \mathbf{a}_i\|} \leq \frac{\omega}{f(\mathbf{a}_p) - f(\mathbf{x}_0)}.$$

Thus,

$$L(\mathbf{x}_k) = \sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x}_k - \mathbf{a}_i\|} \leq \sum_{i=1}^m \frac{\omega \omega_i}{f(\mathbf{a}_p) - f(\mathbf{x}_0)} = \frac{\omega^2}{f(\mathbf{a}_p) - f(\mathbf{x}_0)}$$

which along with Lemma 6.1 implies the desired result. \square

Combining Theorem 7.1 and Lemma 7.2, we finally obtain the following rate of convergence result of Weiszfeld's method.

Theorem 7.2 (Sublinear rate of convergence of the Weiszfeld's method - Explicit version). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by Weiszfeld's with \mathbf{x}_0 chosen by (7.1)-(7.3). Then, for any $k \geq 0$, we have*

$$f(\mathbf{x}_k) - f^* \leq \frac{M}{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|^2, \quad (7.7)$$

where $M = 2L(\mathbf{a}_p) \omega^2 / (\|\mathbf{R}_p\| - \omega_p)^2$.

8 Acceleration via Optimal Schemes

The question that now arises is whether we can find an acceleration of Weiszfeld's method with a better theoretical rate of convergence than that of the original Weiszfeld's method. Recall that in Section 7, we showed that Weiszfeld's method with a specific choice of a starting point converges in terms of function values to the optimal value in a rate of $O(1/k)$. The natural idea is now to use an accelerated gradient-based scheme in order to assure convergence with the faster rate of $O(1/k^2)$. Unfortunately, accelerated schemes such as Nesterov's optimal method [34] and FISTA [3] are not monotone, meaning that the sequence of function values is not necessarily nonincreasing. This causes a genuine theoretical difficulty to employ the accelerated schemes on the Fermat-Weber problem since the monotonicity was a crucial argument in showing that Weiszfeld's method does not get stuck in anchor points.

We are therefore led to consider an additional and different idea. Instead of bypassing the anchor points, which are the points of non-differentiability, we will simply eliminate them, by using the idea of *smoothing*.

Given a minimization problem

$$(M) \quad \min \{q(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\},$$

where the objective function q is nonsmooth, the idea of smoothing is to replace the objective with a smooth approximation q_μ and solve the problem

$$(M_\mu) \quad \min \{q_\mu(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\},$$

where μ is the so-called “smoothing parameter” that controls the level of smoothness and the proximity of the approximation q_μ to q . A more precise definition can be found for example in [5]. Some earlier works on smoothing techniques are [8, 9, 10, 32]. The more recent works [5, 36] show that complexity results can be obtained by employing a fast gradient-based method with a rate of $O(1/k^2)$ on the smooth counterpart (M_μ) . In these works the smooth problem (M_μ) is not equivalent to the original problem (M) , but only serves as an approximation. The complexity result states that by employing a fast gradient-based method on the smooth problem with a special choice of the smoothing parameter, an ε -optimal solution can be obtained after $O(1/\varepsilon)$ iterations, which corresponds to a method with an $O(1/k)$ rate of convergence. This result is not impressive in the context of the Fermat-Weber problem for which we have already shown in Section 7 that Weiszfeld’s method with a specially chosen starting point shares this rate of convergence. The challenge is therefore to present a method with an $O(1/k^2)$ rate of convergence. For that, we present here an *exact smoothing* scheme in which the original problem (FW) is replaced by a different problem that is equivalent to problem (FW) in the sense that its minimizer is exactly the same as the minimizer of the original problem.

We assume that there is no optimal anchor point (otherwise, as usual, it is trivial to find it). In this case, anchors are necessarily not collinear, implying that the objective function is strictly convex and there exists a unique optimizer \mathbf{x}^* .

We begin by combining two simple but essential results. First, from Lemma 7.1 we can find a lower bound on the distance of \mathbf{x}^* from each of the anchor points:

$$\|\mathbf{x}^* - \mathbf{a}_i\| \geq \frac{f(\mathbf{a}_i) - f(\mathbf{w})}{\omega}, \quad \forall i \in \{1, 2, \dots, m\}, \quad (8.1)$$

where \mathbf{w} chosen in the same way as the starting point (7.1)-(7.3) is picked. More precisely, we take

$$\mathbf{w} = \mathbf{a}_p + t_p \mathbf{d}_p, \quad (8.2)$$

where

$$p \in \operatorname{argmin} \{f(\mathbf{a}_i) : 1 \leq i \leq m\}, \quad (8.3)$$

and

$$\mathbf{d}_p = -\frac{\mathbf{R}_p}{\|\mathbf{R}_p\|}, \quad t_p = \frac{\|\mathbf{R}_p\| - \omega_p}{L(\mathbf{a}_p)}. \quad (8.4)$$

Second, we can find a smooth function upper bounding the norm function that coincides with the norm function outside a specified ball. Indeed, let us denote the norm function by $g(\mathbf{x}) \equiv \|\mathbf{x}\|$ and let $r > 0$. Then we define the following function, which we refer to as a *smooth approximation*

$$g_r(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|, & \|\mathbf{x}\| \geq r, \\ \frac{\|\mathbf{x}\|^2}{2r} + \frac{r}{2}, & \|\mathbf{x}\| < r. \end{cases}$$

The smooth approximation function g_r enjoys two essential properties: (i) it is continuously differentiable with Lipschitz gradient with constant $1/r$ and (ii) it is an upper bound on g .

Lemma 8.1 (Properties of g_r). *Let $r > 0$. Then*

- (i) $g_r(\mathbf{x}) \geq g(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$.
- (ii) g_r is continuously differentiable and its gradient is Lipschitz continuous with constant $1/r$.

Proof. (i) It is clear from the definition of g_r that we have to show the result only when $\|\mathbf{x}\| < r$. Indeed, in this case we have the following identity

$$g_r(\mathbf{x}) - g(\mathbf{x}) = \frac{\|\mathbf{x}\|^2}{2r} + \frac{r}{2} - \|\mathbf{x}\| = \left(\frac{\|\mathbf{x}\|}{\sqrt{2r}} - \sqrt{\frac{r}{2}} \right)^2.$$

Hence $g_r(\mathbf{x}) \geq g(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$.

(ii) First, we will write the gradient of g_r :

$$\nabla g_r(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| \geq r \\ \frac{\mathbf{x}}{r}, & \|\mathbf{x}\| < r. \end{cases}$$

Thus, it is clear that g_r is continuously differentiable. In order to prove the Lipschitz continuity of ∇g_r , we will use the fact that ∇g_r can be written as

$$\nabla g_r = \frac{1}{r} P_B,$$

where P_B is the orthogonal projection operator onto the closed and convex ball $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq r\}$. We can now use the fact the orthogonal projection operator P_B is nonexpensive, meaning that it is Lipschitz continuous with constant 1 (see [11, Proposition 2.1.3(c), p. 201]). This proves the desired result. \square

Remark 8.1. The function g_r is closely related to the so-called Huber function [24] given by

$$H_r(\mathbf{x}) = \begin{cases} \|\mathbf{x}\| - \frac{r}{2}, & \|\mathbf{x}\| \geq r, \\ \frac{\|\mathbf{x}\|^2}{2r}, & \|\mathbf{x}\| < r. \end{cases}$$

In fact, $g_r(\mathbf{x}) + \frac{r}{2} = H_r(\mathbf{x})$. We can deduce part (ii) of Lemma 8.1 from the known properties of the Huber function (see, e.g., [5, 36]), but we have chosen to provide a self-contained proof. In addition, since the Huber function is convex, the convexity of g_r follows.

Now, motivated by the inequalities (8.1), we will define the following smooth approximation of the Fermat-Weber objective function

$$f_s(\mathbf{x}) \equiv \sum_{i=1}^m \omega_i g_{b_i}(\mathbf{x} - \mathbf{a}_i),$$

where

$$b_i = \frac{f(\mathbf{a}_i) - f(\mathbf{w})}{\omega}, \quad i = 1, 2, \dots, m.$$

Now we have the following lemma stating that f_s is an exact convex smoothing counterpart of f .

Lemma 8.2 (Properties of the exact smoothing function f_s). *Let \mathbf{w} be defined by (8.2)-(8.3). The following properties hold:*

- (i) f_s is convex over \mathbb{R}^n .
- (ii) $f_s(\mathbf{x}) \geq f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$.
- (iii) The optimal solution of the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f_s(\mathbf{x})$$

is \mathbf{x}^* - the optimal solution of problem (FW).

- (iv) f_s is continuously differentiable and its gradient is Lipschitz continuous with constant

$$L_s = \omega \sum_{i=1}^m \frac{\omega_i}{f(\mathbf{a}_i) - f(\mathbf{w})}.$$

Moreover, the Lipschitz constant L_s can be upper bounded by

$$L_s \leq \frac{2L(\mathbf{a}_p)\omega^2}{(\|\mathbf{R}_p\| - \omega_p)^2}.$$

Proof. (i) Follows by the fact that f_s is the weighted sum of the convex functions g_{b_i} , $i = 1, 2, \dots, m$.

(ii) First, from Lemma 8.1(i), it follows that $g_{b_i}(\mathbf{x}) \geq g(\mathbf{x}) (= \|\mathbf{x}\|)$ for all $\mathbf{x} \in \mathbb{R}^n$. Hence, for any $\mathbf{x} \in \mathbb{R}^n$:

$$f_s(\mathbf{x}) = \sum_{i=1}^m \omega_i g_{b_i}(\mathbf{x} - \mathbf{a}_i) \geq \sum_{i=1}^m \omega_i g(\mathbf{x} - \mathbf{a}_i) = \sum_{i=1}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| = f(\mathbf{x}).$$

(iii) From (8.1) the inequality $\|\mathbf{x}^* - \mathbf{a}_i\| \geq b_i$ holds for any $1 \leq i \leq m$ and hence $g_{b_i}(\mathbf{x}^* - \mathbf{a}_i) = \|\mathbf{x}^* - \mathbf{a}_i\|$ for any $1 \leq i \leq m$. Consequently,

$$f_s(\mathbf{x}^*) = \sum_{i=1}^m \omega_i g_{b_i}(\mathbf{x}^* - \mathbf{a}_i) = \sum_{i=1}^m \omega_i \|\mathbf{x}^* - \mathbf{a}_i\| = f(\mathbf{x}^*).$$

Let $\mathbf{x} \in \mathbb{R}^n$ be different than \mathbf{x}^* . Since \mathbf{x}^* is the strict global minimum of f over \mathbb{R}^n , it follows that $f(\mathbf{x}^*) < f(\mathbf{x})$. Hence,

$$f_s(\mathbf{x}^*) = f(\mathbf{x}^*) < f(\mathbf{x}) \leq f_s(\mathbf{x}),$$

where the last inequality follows from part (ii). We therefore conclude that \mathbf{x}^* is also the unique minimizer of f_s over \mathbb{R}^n .

(iv) For each $1 \leq i \leq m$, by Lemma 8.1(ii) it follows that g_{b_i} is continuously differentiable with gradient which is Lipschitz with constant $1/b_i$. Therefore, by its definition, the function f_s is continuously differentiable with gradient which is Lipschitz continuous with constant

$$\sum_{i=1}^m \frac{\omega_i}{b_i} = \omega \sum_{i=1}^m \frac{\omega_i}{f(\mathbf{a}_i) - f(\mathbf{w})}.$$

Since $f(\mathbf{a}_i) \geq f(\mathbf{a}_p)$ for all $1 \leq i \leq m$ (follows from (8.3)), we can also upper bound the Lipschitz constant by

$$\omega \sum_{i=1}^m \frac{\omega_i}{f(\mathbf{a}_i) - f(\mathbf{w})} \leq \omega \sum_{i=1}^m \frac{\omega_i}{f(\mathbf{a}_p) - f(\mathbf{w})} = \frac{\omega^2}{f(\mathbf{a}_p) - f(\mathbf{w})} \leq \frac{2L(\mathbf{a}_p)\omega^2}{(\|\mathbf{R}_p\| - \omega_p)^2},$$

where the last inequality follows from Lemma 6.1. □

The effect of the smoothing operation is illustrated in Figure 3.

Now that we know that the nonsmooth Fermat-Weber problem can be replaced by the smooth counterpart

$$(\text{FW}_s) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f_s(\mathbf{x}),$$

which has the same optimal solution. We can invoke one of the optimal gradient-based methods for solving smooth convex optimization problems. One of the simplest option is Nesterov's method from 1983 [34]. We begin by explicitly writing the accelerated method with a constant stepsize version.

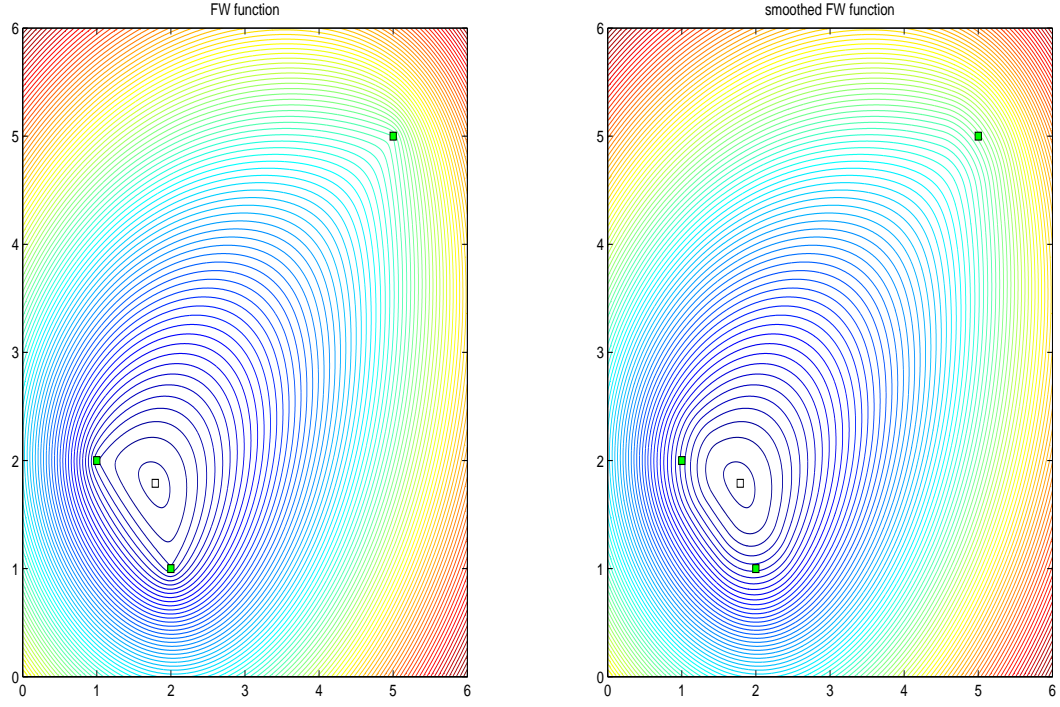


Figure 3: The left image describes the contour lines of the unweighted Fermat-Weber function with anchors $\mathbf{a}_1 = (1, 2)^T$, $\mathbf{a}_2 = (2, 1)^T$ and $\mathbf{a}_3 = (5, 5)^T$. The anchors are denoted by the filled squares, while the empty square stands for the optimal solution. In the right image the contour lines of a smoothed function are given. Here we have chosen to replace each of the norm functions $g(\mathbf{x}) = \|\mathbf{x}\|$ by the smooth counterpart $g_{1/2}(\mathbf{x})$.

Fast Weiszfeld Method (constant stepsize)**Initialization.** $\mathbf{y}_1 = \mathbf{x}_0 \in \mathbb{R}^n$ and $t_1 = 1$.**General Step** ($k = 0, 1, \dots$) Compute

$$\begin{aligned}\mathbf{x}_k &= \mathbf{y}_k - \frac{1}{L_s} \nabla f_s(\mathbf{y}_k), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \mathbf{y}_{k+1} &= \mathbf{x}_k + \left(\frac{t_k - 1}{t_{k+1}} \right) (\mathbf{x}_k - \mathbf{x}_{k-1}).\end{aligned}$$

A possible drawback of this basic fast scheme is that the Lipschitz constant L_s can be too conservative (*i.e.*, large). We therefore also consider a version with a backtracking stepsize rule. This version was considered in the context of the more general composite model [3].

Fast Weiszfeld Method (backtracking stepsize)**Initialization.** $L_0 > 0$, some $\eta > 1$, and $\mathbf{x}_0 \in \mathbb{R}^n$. Set $\mathbf{y}_1 = \mathbf{x}_0$ and $t_1 = 1$.**General Step** ($k = 0, 1, \dots$) Find the smallest nonnegative integer i_k such that with $\bar{L} = \eta^{i_k} L_{k-1}$

$$f_s\left(\mathbf{y} - \frac{1}{\bar{L}} \nabla f_s(\mathbf{y})\right) \leq f_s(\mathbf{y}) - \frac{1}{2\bar{L}} \|\nabla f_s(\mathbf{y})\|^2.$$

Set $L_k = \eta^{i_k} L_{k-1}$ and compute

$$\begin{aligned}\mathbf{x}_k &= \mathbf{y}_k - \frac{1}{L_k} \nabla f_s(\mathbf{y}_k), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \mathbf{y}_{k+1} &= \mathbf{x}_k + \left(\frac{t_k - 1}{t_{k+1}} \right) (\mathbf{x}_k - \mathbf{x}_{k-1}).\end{aligned}$$

The following convergence result for both the constant stepsize scheme as well as the backtracking version was proved in [3].

Theorem 8.1 (Convergence of the fast Weiszfeld method). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ and $\{\mathbf{y}_k\}_{k \geq 0}$ be two sequences which generated by the fast Weiszfeld method. Then, for any $k \geq 1$, we have*

$$f_s(\mathbf{x}_k) - f^* \leq \frac{2\alpha L_s \|\mathbf{x}_0 - \mathbf{x}^*\|}{(k+1)^2},$$

where $\alpha = 1$ for the constant stepsize setting and $\alpha = \eta$ for the backtracking stepsize setting.

Since $f_s(\mathbf{x}) \geq f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$, we can immediately conclude an $O(1/k^2)$ rate of convergence result of the original objective function to the optimal value.

Corollary 8.1 (Convergence of the fast Weiszfeld's method - Original function values). *Let $\{\mathbf{x}_k\}_{k \geq 0}$ and $\{\mathbf{y}_k\}_{k \geq 0}$ be two sequences which generated by the fast Weiszfeld method. Then, for any $k \geq 1$, we have*

$$f(\mathbf{x}_k) - f^* \leq \frac{2\alpha L_s \|\mathbf{x}_0 - \mathbf{x}^*\|}{(k+1)^2},$$

where $\alpha = 1$ for the constant stepsize setting and $\alpha = \eta$ for the backtracking stepsize setting.

9 Weiszfeld's Method Beyond the Fermat-Weber Problem

Weiszfeld's method had an impact on the development of many numerical methods for solving various problems, and its influence was not restricted to the Fermat-Weber problem. Perhaps the most natural generalization of the Fermat-Weber problem (and Weiszfeld's method as well) is to the problem of multifacility location. Actually one of the first papers that dealt with Weiszfeld's method (without knowing it...) was Miehle's paper from 1958 [31] where he considered a problem of finding the locations of *two* points in \mathbb{R}^2 . The general multi-facility location problem was considered by Radó in [42] (1988), where the problem was formulated as

$$\min \sum_{j=1}^n \sum_{i=1}^m w_{ji} \|\mathbf{x}_j - \mathbf{a}_i\| + \sum_{j=1}^n \sum_{\ell=1}^n v_{j\ell} \|\mathbf{x}_j - \mathbf{x}_\ell\|.$$

Here $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^d$ are the fixed anchors and $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^d$ are the n locations that we seek to find. The nonnegative numbers w_{ji} and v_{jk} are given weights. Radó considered the following generalization of Weiszfeld's method. To construct the $(k+1)$ -th iterate from the k -th iterate, like in Weiszfeld's method, each of the norm expressions $\|\mathbf{x}\|$ is replaced by the term $\|\mathbf{x}\|^2 / \|\mathbf{x}^k\|$. That is, at the k -th iteration the following minimization problem is solved:

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n} \sum_{j=1}^n \sum_{i=1}^m w_{ji} \frac{\|\mathbf{x}_j - \mathbf{a}_i\|^2}{\|\mathbf{x}_i^k - \mathbf{a}_j\|} + \sum_{j=1}^n \sum_{\ell=1}^n v_{j\ell} \frac{\|\mathbf{x}_j - \mathbf{x}_\ell\|^2}{\|\mathbf{x}_j^k - \mathbf{x}_\ell^k\|}.$$

The solution to the above convex problem, which is the next iterate $\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \dots, \mathbf{x}_n^{k+1}$, is attained at its unique stationary point which is the solution of following linear system of equations:

$$\sum_{i=1}^m w_{ji} \frac{\mathbf{x}_j - \mathbf{a}_i}{\|\mathbf{x}_i^k - \mathbf{a}_j\|} + \sum_{j=1}^n \sum_{\ell=1}^n v_{j\ell} \frac{\mathbf{x}_j - \mathbf{x}_\ell}{\|\mathbf{x}_j^k - \mathbf{x}_\ell^k\|} = 0, \quad j = 1, 2, \dots, n.$$

Some convergence properties of this generalized Weiszfeld's method were studied in [42]. Another type of a multi-facility location problem that involves also clustering was considered by Iyigun and Ben-Israel in [25] (2010), and also by Teboulle in [47] (2007). The authors constructed a generalization of Weiszfeld's method, and found several properties such as monotonicity; of course, since the problem is nonconvex, it does not seem possible to construct a method that is guaranteed to converge to the optimal global optimal solution.

Another type of generalizations of the Fermat-Weber problem leading to corresponding generalizations of Weiszfeld's method are concerned with replacing the Euclidean norm in the objective function by another norm. For example, in [27] (2010) Katz and Vogl considered a Weiszfeld method which solves a version of the Fermat-Weber problem in which the Euclidean norms are replaced with weighted Euclidean norms. Cooper extended Weiszfeld's method in [18] (1968) to consider a Fermat-Weber type problem in which the objective is to minimize a weighted sum of *powers* of the Euclidean norms:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^n \omega_i \|\mathbf{x} - \mathbf{a}_i\|^K,$$

where $K > 0$. Of course, the problem is convex only when $K \geq 1$. This extension of the problem was also studied by Chen in [16] (1984). An extension of both the Fermat-Weber problem and the Weiszfeld's method for p -norms was considered by Morris in [33] (1981).

A totally different location problem in which Weiszfeld-type ideas were used is the *source localization problem*. In this problem we are given measurements of the distances of a source in an unknown location \mathbf{x} from the m anchors:

$$\|\mathbf{x} - \mathbf{a}_i\| \approx d_i, \quad i = 1, 2, \dots, m.$$

One formulation of the problem consists of finding the \mathbf{x} resulting with the minimum sum of squared errors:

$$(\text{SL}) \quad \min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^m (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2.$$

Despite its apparent resemblance to Fermat-Weber, problem (SL) is quite different in the sense that it is a difficult nonconvex problem. One of the methods studied by Beck et al in [6] (2008) uses the idea of Weiszfeld to replace the norm terms of the form $\|\mathbf{x} - \mathbf{a}_i\|$ with the expressions $\|\mathbf{x} - \mathbf{a}_i\|^2 / \|\mathbf{x}_k - \mathbf{a}_i\|$, thus resulting with the following iterative scheme:

$$\mathbf{x}_{k+1} \in \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^m \left(\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{x}_k - \mathbf{a}_i\|} - d_i \right)^2.$$

The convergence of this scheme (to stationary points) was studied in [6].

We conclude this section by noting that *the iteratively reweighted least squares method*, which is one of the most popular optimization algorithms for solving a wide variety of

problems involving norms, is essentially based on ideas from Weiszfeld's method. As an illustration of the method, consider the problem of robust regression which consists of solving the l_1 -norm problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 = \sum_{i=1}^m |\mathbf{a}_i^T \mathbf{x} - b_i| \right\},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_m^T$ are the rows of the matrix \mathbf{A} . At each iteration, the absolute values $|\mathbf{a}_i^T \mathbf{x} - b_i|$ are replaced by $(\mathbf{a}_i^T \mathbf{x} - b_i)^2 / |\mathbf{a}_i^T \mathbf{x}_k - b_i|$, and the $(k+1)$ -th is determined from the k -th iteration by the update formulas:

$$\mathbf{x}_{k+1} \in \operatorname{argmin} \left\{ \sum_{i=1}^m \frac{(\mathbf{a}_i^T \mathbf{x} - b_i)^2}{|\mathbf{a}_i^T \mathbf{x}_k - b_i|} \right\}.$$

This scheme was studied in in the context of various types of models and the literature of this method covers hundreds of paper that were written in the past 60 years; the list of references [2, 19, 30, 38, 44] is just a very small representative sample of works dealing with various applications and theoretical properties.

A Notations

Following is a list of notations that are used throughout the paper.

- $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}$ - the set of anchors.
- $\omega_1, \omega_2, \dots, \omega_m$ - given positive weights.
- $\omega = \sum_{i=1}^m \omega_i$ - sum of weights.
- $f(\mathbf{x}) = \sum_{i=1}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\|$ - the Fermat-Weber objective function.
- \mathbf{x}^* - an optimal solution of the Fermat-Weber problem. If the anchors are not collinear, \mathbf{x}^* is the *unique* optimal solution.
- X^* - the optimal solution set of the Fermat-Weber problem. When the anchors are not collinear, X^* is the singleton $\{\mathbf{x}^*\}$.
- f^* - the optimal value of the Fermat-Weber problem.
- $T(\mathbf{x}) = \frac{1}{\sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}} \sum_{i=1}^m \frac{\omega_i \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}$ - the operator defining Weiszfeld's method.
- $h(\mathbf{x}, \mathbf{y}) \equiv \sum_{i=1}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{y} - \mathbf{a}_i\|}$ - an auxiliary function used to analyze Weiszfeld's method.
- $L(\mathbf{x}) = \sum_{i=1}^m \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}$ - serves as a kind of "Lipschitz" constant and the operator T can be written as taking a gradient step with stepsize $1/L(\mathbf{x})$: $T(\mathbf{x}) = \mathbf{x} - 1/L(\mathbf{x}) \nabla f(\mathbf{x})$.

- $\mathbf{R}_j = \sum_{i=1, i \neq j}^m \omega_i (\mathbf{a}_i - \mathbf{a}_j) / \|\mathbf{a}_i - \mathbf{a}_j\|$, $j = 1, 2, \dots, m$. An important property related to \mathbf{R}_j is that \mathbf{a}_j is optimal if and only if $\|\mathbf{R}_j\| \leq w_j$.
- $\mathbf{d}_j = -\mathbf{R}_j / \|\mathbf{R}_j\|$ - the steepest descent direction of f at \mathbf{a}_j .

B Proof of Lemma 6.1

Let $\mathbf{x} := \mathbf{a}_j + t_j \mathbf{d}_j$. Then, from the definition of \mathbf{x} , we have that $\mathbf{d}_j = (1/t_j)(\mathbf{x} - \mathbf{a}_j)$ and hence

$$\begin{aligned}
-\frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 &= \frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 - 2 \left\langle \mathbf{x} - \mathbf{a}_j, \frac{\mathbf{x} - \mathbf{a}_j}{t_j} \right\rangle \\
&= \frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 - 2 \langle \mathbf{x} - \mathbf{a}_j, \mathbf{d}_j \rangle \\
&= \frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 - 2 \left\langle \mathbf{x} - \mathbf{a}_j, -\frac{\mathbf{R}_j}{\|\mathbf{R}_j\|} \right\rangle \\
&= \frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 + \frac{2}{\|\mathbf{R}_j\|} \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle. \tag{B.1}
\end{aligned}$$

Now, we will expand the first term of the right-hand side of (B.1)

$$\begin{aligned}
\frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 &= \frac{L(\mathbf{a}_j)}{\|\mathbf{R}_j\| - \omega_j} \|\mathbf{x} - \mathbf{a}_j\|^2 \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \frac{\omega_i}{\|\mathbf{a}_j - \mathbf{a}_i\|} \right) \|\mathbf{x} - \mathbf{a}_j\|^2 \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \sum_{i=1, i \neq j}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_j\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \sum_{i=1, i \neq j}^m \omega_i \left(\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} + 2 \frac{\langle \mathbf{x} - \mathbf{a}_i, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} + \frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} \right) \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \sum_{i=1, i \neq j}^m \omega_i \left(\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} + 2 \frac{\langle \mathbf{x} - \mathbf{a}_i, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} + \|\mathbf{a}_j - \mathbf{a}_i\| \right).
\end{aligned}$$

The mid-term can be also written as follows

$$\begin{aligned}
2 \frac{\langle \mathbf{x} - \mathbf{a}_i, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} &= 2 \frac{\langle \mathbf{x} - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} + 2 \frac{\langle \mathbf{a}_j - \mathbf{a}_i, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} \\
&= 2 \frac{\langle \mathbf{x} - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} - 2 \|\mathbf{a}_j - \mathbf{a}_i\|.
\end{aligned}$$

Thus, from the definition of \mathbf{R}_j , we have

$$\begin{aligned}
\frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 &= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \sum_{i=1, i \neq j}^m \omega_i \left(\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} + 2 \frac{\langle \mathbf{x} - \mathbf{a}_j, \mathbf{a}_i - \mathbf{a}_j \rangle}{\|\mathbf{a}_j - \mathbf{a}_i\|} - \|\mathbf{a}_j - \mathbf{a}_i\| \right) \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} + 2 \left\langle \mathbf{x} - \mathbf{a}_j, \sum_{i=1, i \neq j}^m \omega_i \frac{\mathbf{a}_i - \mathbf{a}_j}{\|\mathbf{a}_j - \mathbf{a}_i\|} \right\rangle - f(\mathbf{a}_j) \right) \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} - 2 \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle - f(\mathbf{a}_j) \right).
\end{aligned}$$

Note that by the fact that $\frac{a^2}{b} \geq 2a - b$ for any $a \in \mathbb{R}$ and $b \in \mathbb{R}_{++}$ we have

$$\frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{a}_j - \mathbf{a}_i\|} \geq 2 \|\mathbf{x} - \mathbf{a}_i\| - \|\mathbf{a}_j - \mathbf{a}_i\|.$$

Thence

$$\begin{aligned}
\frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 &\geq \frac{1}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i (2 \|\mathbf{x} - \mathbf{a}_i\| - \|\mathbf{a}_j - \mathbf{a}_i\|) - 2 \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle - f(\mathbf{a}_j) \right) \\
&= \frac{1}{\|\mathbf{R}_j\| - \omega_j} \left(2 \sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - 2 \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle - 2f(\mathbf{a}_j) \right) \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle - f(\mathbf{a}_j) \right).
\end{aligned}$$

Plugging the last inequality in (B.1) yields

$$\begin{aligned}
-\frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 &= \frac{1}{t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 + \frac{2}{\|\mathbf{R}_j\|} \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle \\
&\geq \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle - f(\mathbf{a}_j) \right) + \frac{2}{\|\mathbf{R}_j\|} \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - f(\mathbf{a}_j) \right) - \left(\frac{2}{\|\mathbf{R}_j\| - \omega_j} - \frac{2}{\|\mathbf{R}_j\|} \right) \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - f(\mathbf{a}_j) \right) - \frac{2\omega_j}{\|\mathbf{R}_j\| (\|\mathbf{R}_j\| - \omega_j)} \langle \mathbf{x} - \mathbf{a}_j, \mathbf{R}_j \rangle \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - f(\mathbf{a}_j) \right) + \frac{2\omega_j}{\|\mathbf{R}_j\| - \omega_j} \langle \mathbf{x} - \mathbf{a}_j, \mathbf{d}_j \rangle \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - f(\mathbf{a}_j) + \omega_j \langle \mathbf{x} - \mathbf{a}_j, \mathbf{d}_j \rangle \right) \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} \left(\sum_{i=1, i \neq j}^m \omega_i \|\mathbf{x} - \mathbf{a}_i\| - f(\mathbf{a}_j) + \omega_j \|\mathbf{x} - \mathbf{a}_j\| \right) \\
&= \frac{2}{\|\mathbf{R}_j\| - \omega_j} (f(\mathbf{x}) - f(\mathbf{a}_j)),
\end{aligned}$$

where the second equality from below follows from the fact that $1 = \|\mathbf{d}_j\| = \|\mathbf{x} - \mathbf{a}_j\|/t_j$. Hence,

$$f(\mathbf{a}_j) - f(\mathbf{x}) \geq \frac{\|\mathbf{R}_j\| - \omega_j}{2t_j} \|\mathbf{x} - \mathbf{a}_j\|^2 = t_j \frac{\|\mathbf{R}_j\| - \omega_j}{2} = \frac{(\|\mathbf{R}_j\| - \omega_j)^2}{2L(\mathbf{a}_j)}$$

where the first equality follows from the fact that $\|\mathbf{x} - \mathbf{a}_j\| = t_j$ and the last equality follows from the definition of t_j . \square

References

- [1] K. D. Andersen, E. Christiansen A. R. Conn, and M. L. Overton. An efficient primal-dual interior-point method for minimizing a sum of euclidean norms. *SIAM J. Scient. Comp.*, 22:243–262, 2000.
- [2] A. E. Beaton and J. W. Tukey. The fitting of power series, meaning polynomials, illustrated on bandspectroscopic data. *Technometrics*, 16:147–185, 1974.
- [3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

- [4] A. Beck and M. Teboulle. Gradient-based algorithms with applications to signal recovery problems. In D. Palomar and Y. Eldar, editors, *Convex Optimization in Signal Processing and Communications*, pages 139–162. Cambridge University Press, 2009.
- [5] A. Beck and M. Teboulle. Smoothing and first order methods: A unified framework. *SIAM J. Optimization*, 22(2):557–580, 2012.
- [6] A. Beck, M. Teboulle, and Z. Chikishev. Iterative minimization schemes for solving the single source localization problem. *SIAM J. Optim.*, 19(3):1397–1416, 2008.
- [7] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization, 2001.
- [8] A. Ben-Tal and M. Teboulle. A smoothing technique for nondifferentiable optimization problems. In *Optimization, Springer Lecture Notes in Mathematics (S. Dolecki, ed.)*, 1405, pages 1–11, 1989.
- [9] D. P. Bertsekas. Nondifferentiable optimization via approximation. *Math. Programming Stud.*, (3):1–25, 1975.
- [10] D. P. Bertsekas. *Constrained Optimization and Lagrangian Multipliers*. Academic Press, New York, 1982.
- [11] D. P. Bertsekas. *Nonlinear Programming*. Belmont MA: Athena Scientific, second edition, 1999.
- [12] J. Brimberg. The Fermat-Weber location problem revisited. *Math. Programming*, 71(1):71–76, 1995.
- [13] P. H. Calamai and A. R. Conn. A projected newton method for l_p norm location problems. *Math. Programming*, 38:75–109, 1987.
- [14] L. Cánovas, R. Cañavate, and A. Marín. On the convergence of the Weiszfeld algorithm. *Math. Program.*, 93(2):327–330, 2002.
- [15] R. Chandrasekaran and A. Tamir. Open questions concerning Weiszfeld’s algorithm for the Fermat-Weber location problem. *Math. Programming*, 44(3):293–295, 1989.
- [16] R. Chen. Solution of location problems with radial cost functions. *Computers and Mathematics with Applications*, 10(1):87–94, 1984.
- [17] L. Cooper. Location-allocation problems. *Operations Res.*, 11:331–343, 1963.
- [18] L. Cooper. An extension to the generalized weber problem. *Journal of Regional Science*, 8:181–197, 1968.

- [19] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Comm. Pure Appl. Math.*, 63(1):1–38, 2010.
- [20] Z. Drezner. A note on the weber location problem. *Annals of Operations Research*, 40(1):153–161, 1992.
- [21] Z. Drezner. A note on accelerating the weiszfeld procedure. *Location Science*, 3(4):275–279, 1995.
- [22] Z. Drezner, K. Klamroth, A. Schobel, and G. O. Wesolowsky. The weber problem. In Z. Drezner and H. W. Hamacher, editors, *Facility Location: Applications and Theory*, pages 1–36.
- [23] H. A. Eiselt and V. Marianov. Pioneering developments in location analysis. In H. A. Eiselt and V. Marianov, editors, *Foundations of Location Analysis*, International Series in Operations Research & Management Science, pages 3–22. Springer, 2011.
- [24] P. J. Huber. *Robust statistics*. John Wiley & Sons Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.
- [25] C. Iyigun and A. Ben-Israel. A generalized Weiszfeld method for the multi-facility location problem. *Oper. Res. Lett.*, 38(3):207–214, 2010.
- [26] I. N. Katz. Local convergence in Fermat’s problem. *Math. Programming*, 6:89–104, 1974.
- [27] I. N. Katz and S. R. Vogl. A Weiszfeld algorithm for the solution of an asymmetric extension of the generalized Fermat location problem. *Comput. Math. Appl.*, 59(1):399–410, 2010.
- [28] H. W. Kuhn. A note on Fermat’s problem. *Math. Programming*, 4:98–107, 1973.
- [29] H. W. Kuhn and R. E. Kuenne. An efficient algorithm for the numerical solution of the generalized weber problem in spatial economics. *Journal of Regional Science*, 4:21–34, 1962.
- [30] C. L. Lawson. Contributions to the theory of linear least maximum approximations. 1961. PhD thesis, UCLA.
- [31] W. Miehle. Link-length minimization in networks. *Operations Res.*, 6:232–243, 1958.
- [32] J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
- [33] J. G. Morris. Convergence of the Weiszfeld algorithm for Weber problems using a generalized “distance” function. *Oper. Res.*, 29(1):37–48, 1981.

- [34] Y. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269(3):543–547, 1983.
- [35] Y. Nesterov. *Introductory Lectures on Convex Optimization*. Kluwer, Boston, 2004.
- [36] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152, 2005.
- [37] Y. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PE: SIAM, 1994.
- [38] D. P. O’Leary. Robust regression computation using iteratively reweighted least squares. *SIAM J. Matrix Anal. Appl.*, 11(3):466–480, 1990. Sparse matrices (Gleneden Beach, OR, 1989).
- [39] L. M. Ostresh. On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research*, 26(4):597–609, 1978.
- [40] M. L. Overton. A quadratically convergent method for minimizing a sum of euclidean norms. *Math. Programming*, 27:34–63, 1983.
- [41] B. T. Polyak. *Introduction to optimization*. Translations Series in Mathematics and Engineering. Optimization Software Inc. Publications Division, New York, 1987. Translated from the Russian, With a foreword by Dimitri P. Bertsekas.
- [42] F. Radó. The Euclidean multifacility location problem. *Oper. Res.*, 36(3):485–492, 1988.
- [43] D. Rautenbach, M. Struzyna, C. Szegedy, and J. Vygen. *Weiszfeld’s algorithm revisited once again*. Report Forschungsinstitut für Diskrete Mathematik. Forschungsinstitut für Diskrete Mathematik, Rheinische Friedrich-Wilhelms-Universität.
- [44] J. R. Rice and K. H. Usow. The Lawson algorithm and extensions. *Math. Comp.*, 22:118–127, 1968.
- [45] R. T. Rockafellar. *Convex Analysis*. Princeton NJ: Princeton Univ. Press, 1970.
- [46] R. Sturm. Ueber den punkt kleinster entfernungssumme von gegebenen punkten. *Journal fur die reine und angewandte Mathematik*, 97:49–61, 1884. (In german - On the point of smallest distance sum from given points).
- [47] M. Teboulle. A unified continuous optimization framework for center-based clustering methods. *Journal of Machine Learning Research*, 8, 2007.
- [48] Y. Vardi and C. H. Zhang. A modified Weiszfeld algorithm for the Fermat-Weber location problem. *Math. Program.*, 90(3):559–566, 2001.

- [49] A. Vazsonyi. Pure mathematics and the weiszfeld algorithm. *Decision Line*, 33:12–13, 2002. Available at http://www.decisionsciences.org/DecisionLine/Vol33/33_3/index.htm.
- [50] E. Weiszfeld. On the point for which the sum of the distances to n given points is minimum. *Ann. Oper. Res.*, 167:7–41, 2009. Translated from the French original [Tohoku Math. J. **43** (1937), 355–386] and annotated by Frank Plastria.
- [51] E. V. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnees est minimum. *The Tohoku Mathematical Journal*, 43:355–386, 1937.