A GEOMETRIC PERSPECTIVE OF THE WEISZFELD ALGORITHM FOR SOLVING THE FERMAT−WEBER PROBLEM

Helder Manoel Venceslau[1,2], Marilis Bahr Karam Venceslau[1,3], Adilson Elias Xavier[1] and Nelson Maculan[1]

Abstract. The Fermat−Weber problem is a classical location problem that has the Weiszfeld algorithm as its main iterative solution method. This article presents a geometric interpretation of its local convergence for the particular case of three points, with the solution constrained to be an interior point, which is fundamental to the present geometric interpretation. This constraint, on the other hand, implies that the weights associated to each point must obey triangle inequalities. The eigenvalues analysis is developed considering that all weights have the same value, which simplifies calculation and explanation, but the generalization of this analysis is straightforward, as commented in the text. Step-size scaling is also considered for accelerating the convergence rate. The accompanying eigenvalues analysis determines step-size multiplier ranges that ensure convergence. Moreover, the eigenvalues depend on a parameter that is computed based on the sample points configuration.

1. Introduction

The original Fermat's problem was firstly introduced by Pierre de Fermat to challenge Evangelista Torricelli, who found a geometric solution. The problem can be stated as: given three distinct points in a plane, find the point that minimizes the sum of the distances to those points. This point is known as the Fermat point of the triangle formed by the three given points.

The problem can be generalized in diverse ways. The name of the economist Alfred Weber became associated to the problem because of his discussion of the problem, and some generalizations of it, in his 1909 book on industries location. His general version, that considers distance weights and more than three points, became known as the Fermat−Weber problem. There are also multi-source versions of the problem (typically location problems) which specifies more than one "source" points to attend the demands of a set of "sink" points. We are not going to deal with this kind of generalization in this article.

[1] COPPE – PESC, Federal University of Rio de Janeiro, Cidade Universitária, Centro de Tecnologia, Bloco H, zip 21941-972, Rio de Janeiro − RJ, Brazil. `heldermv@cos.ufrj.br`; `marilisbkv@cos.ufrj.br`; `adilson@cos.ufrj.br`; `maculan@cos.ufrj.br`

[2] CEFET/RJ − Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Av. Maracanã, 229, Maracanã, zip 20271-110, Rio de Janeiro − RJ, Brazil. `helder.venceslau@cefet-rj.br`

[3] Colégio Pedro II, Campo de São Cristóvão, 177, São Cristóvão, zip 20921-903, Rio de Janeiro − RJ, Brazil. `marilisvenceslau@cp2.g12.br`

It is not our intention to discuss the long history associated with the Fermat−Weber problem. Instead, it is far more profitable to present some key references. A vast survey on the subject was compiled by Wesolowsky [14]. General discussions regarding location problems are presented by Cooper [5], Drezner [7], Drezner and Hamacher [8], Üster and Love [10], and Smith, Laporte and Harper [12]. Brimberg [1–3] is an active researcher on this subject and a good reference for the present state of art in this field. On the other hand, Weiszfeld [13] created the prevalent algorithm used to solve problems of this class, whether as the main algorithm or as the one used to solve related subproblems.

Despite the fact that there are thousands of scientific articles concerning the Fermat−Weber problem, as well as the Weiszfeld algorithm, this article contemplates aspects and properties never mentioned before. It focus on a geometric interpretation of the local convergence process of the Weber generalization that includes weights to the distances to the three points, but with the solution constrained to be an interior point, which is fundamental to the present geometric interpretation. On the algorithmic side, the effects of changing the step length, as presented by Brimberg [3], is also analyzed on the same basis. Thus, to the best knowledge of the authors, this kind of approach has not been done to date.

Although the present analysis is focused on the Weiszfeld algorithm, we believe that the same methodology can be successfully applied to the analysis of other numerical procedures. The simplicity of the Fermat−Weber problem with only three points and the low complexity of the Weiszfeld algorithm constitute an excellent environment to deeply understand the local convergence of similar methods.

The remainder of this paper is organized as follows. Section 2 reviews the Fermat−Weber problem, the Weiszfeld algorithm and its basic properties. Section 3 presents a geometric interpretation of the local Weiszfeld iteration, the eigenvalues of the associated Jacobian, and the dependence of these eigenvalues on the configuration of the three given points. Section 4 describes the effects on the Jacobian's eigenvalues caused by changing the Weiszfeld algorithm's step-size, and the corresponding consequences on the convergence process. Finally, Section 5 summarizes and presents the conclusions.

2. Statement and basic properties of the Fermat−Weber problem

Let a_i, $i = 1, 2, 3$, be three given points on the plane and $w_i > 0$, $i = 1, 2, 3$, three given weights. The Fermat−Weber location problem can be stated as:

$$x^* = \arg \min_{x \in \mathbb{R}^3} W(x), \tag{2.1}$$

where

$$W(x) = \sum_{i=1}^{3} w_i \, \|x - a_i\|_2 = \sum_{i=1}^{3} w_i \sqrt{\sum_{j=1}^{2} (x_j - a_{ij})^2}. \tag{2.2}$$

This is just the minimization of the w_i weighted sum of the distances from a point x to three given points in an euclidean space of dimension two. The three points are considered not collinear and therefore they are always the vertices of a triangle (denoted by $\triangle a_1 a_2 a_3$), as can be seen in Figure 1. The solution point x^* is simply called "the Fermat−Weber point" in the literature [1, 3, 5–8, 12, 14].

The distance function is a classical example of a convex function and $W(x)$, the positively weighted sum of the distances to a finite set of points, is consequently a convex function. It is easy to verify that $W(x)$ is well defined for all $x \in \mathbb{R}^2$ and never attains negative values, resulting that $W(x)$ pertains to the well behaved class of closed proper convex functions (see [4, 9, 11]). The theory of convex analysis tells us that in such a case a local minimum is actually a global minimum and this is a very comfortable situation because any algorithm used to find local minima will actually find global minima.
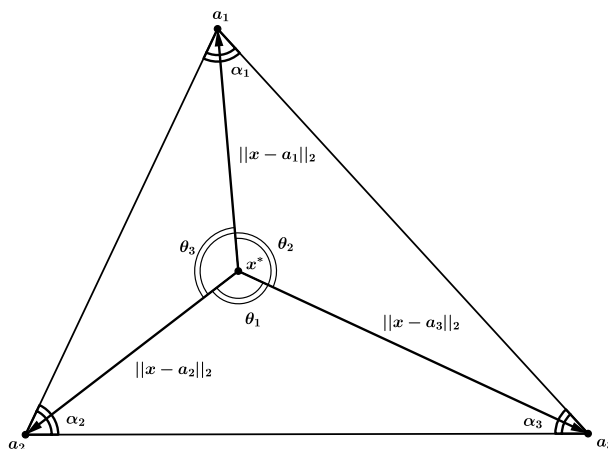
FIGURE 1. Visual reference for the Fermat−Weber problem with three points.

Turning now to a differential analysis, it is straightforward to calculate the tth component of the gradient $\nabla W(x)$:

$$\nabla W_t(x) = \sum_{i=1}^{3} \frac{w_i(x_t - a_{it})}{\sqrt{\sum_{j=1}^{2}(x_j - a_{ij})^2}} = -\sum_{i=1}^{3} w_i \frac{a_{it} - x_t}{\|a_i - x\|_2}. \tag{2.3}$$

This expression clearly shows that $W(x)$ is everywhere differentiable except at the points a_i, where $\nabla W_t(x)$ is not defined.

If we denote ϕ_{it} as the angle between the vector $\overrightarrow{xa_i}$ and the x_t-axis, for $i = 1, 2, 3$ and $t = 0, 1$ (as depicted in Fig. 2), the previous equation can be rewritten as:

$$\nabla W_t(x) = -\sum_{i=1}^{3} w_i \frac{a_{it} - x_t}{\|a_i - x\|_2} = -\sum_{i=1}^{3} w_i \cos \phi_{it}. \tag{2.4}$$

Now, if we consider the weights w_i as the magnitude of three vectors with the same origin in the euclidean plane, the expression (2.4) can be interpreted as the negative of the vector sum of these three vectors. But in a minimization problem the negative of the gradient, in this case the sum of the three vectors, is just the direction used in a descent algorithm. We will return to this geometric interpretation in the next Section.

The Weiszfeld algorithm [13] is the main algorithm used in the search for the Fermat−Weber point. It can be derived from the search for critical points, setting each $\nabla W_t(x)$ to zero:

$$\nabla W_t(x^*) = 0 \;\Rightarrow\; x_t^* = \frac{\sum_{i=1}^{3} \frac{w_i a_{it}}{\sqrt{\sum_{j=1}^{2}(x_j^* - a_{ij})^2}}}{\sum_{i=1}^{3} \frac{w_i}{\sqrt{\sum_{j=1}^{2}(x_j^* - a_{ij})^2}}} = \frac{\sum_{i=1}^{3} y_i(x^*)a_{it}}{\sum_{i=1}^{3} y_i(x^*)} = \sum_{i=1}^{3} \frac{y_i(x^*)}{S(x^*)} a_{it}, \tag{2.5}$$

where

$$y_i(x) = \frac{w_i}{\sqrt{\sum_{j=1}^{2}(x_j - a_{ij})^2}} = \frac{w_i}{\|x - a_i\|_2}$$

and

$$S(x) = \sum_{i=1}^{3} y_i(x).$$

Let's define:

$$\delta_i(x) = \begin{cases} \frac{y_i(x)}{S(x)}, & \forall x \in \mathbb{R}^2 \backslash \{a_1, a_2, a_3\} \\ \delta_{ij}, & x = a_j, \ j = 1, 2, 3 \end{cases}, \ i = 1, 2, 3. \tag{2.6}$$

The apparently artificial values $\delta_i(a_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta, are the easily checked natural limits:

$$\lim_{x \to a_j} \frac{y_i(x)}{S(x)} = \delta_{ij}. \tag{2.7}$$

These limits justify the choice of the notation for these functions.

From the above definitions it follows that x^* is a fixed point of the Weiszfeld iterative algorithm:

$$x^* = \sum_{i=1}^{3} \delta_i(x^*) a_i = \mathbf{A}\delta(x^*), \tag{2.8}$$

where the matrix $\mathbf{A}$ is formed from the juxtaposition of the column vectors a_i, and $\delta(x)$ is the column vector formed from the juxtaposition of the scalars $\delta_i(x)$. Let's define $g(x) = \mathbf{A}\delta(x)$ to simplify the notation of some expressions.

The definition (2.6) of the functions $\delta_i(x)$, $i = 1, 2, 3$, implies that $\forall x \in \mathbb{R}^2$ we have $\sum_{i=1}^{3} \delta_i(x) = 1$ and $0 \le \delta_i(x) \le 1$. Also, $g(x)$ is a continuous function in its domain $\mathbb{R}^2$, since all discontinuities in (2.5) were removed.

For each $x \in \mathbb{R}^2$, $g(x)$ is then a convex combination of the points a_i, which will be denominated $\delta(x)$-convex combination. There are some possible combinations of the $\delta_i(x)$ values and the corresponding positions of $g(x)$ relative to the triangle $\triangle a_1 a_2 a_3$:

- $g(x)$ outside $\triangle a_1 a_2 a_3$: not possible, because $g(x)$ is a $\delta(x)$-convex combination;
- $g(x)$ inside $\triangle a_1 a_2 a_3$: normal case, obtained from $0 < \delta_i(x) < 1$, $i = 1, 2, 3$;
- $g(x)$ is one of the a_i, $i = 1, 2, 3$: possible only if $x = a_i$, because $g(x) = a_i = \delta_1(x)a_1 + \delta_2(x)a_2 + \delta_3(x)a_3$ implies $\delta_i(x) = 1$ and $\delta_j(x) = 0$, $j \ne i$, as its only solution. For example, if $g(x) = a_1$ we have:

$$a_1 = \delta_1(x)a_1 + \delta_2(x)a_2 + \delta_3(x)a_3$$
$$a_1 = (1 - \delta_2(x) - \delta_3(x))a_1 + \delta_2(x)a_2 + \delta_3(x)a_3$$
$$0 = \delta_2(x)(a_2 - a_1) + \delta_3(x)(a_3 - a_1)$$

  and taking into consideration that the vector $a_2 - a_1$ and $a_3 - a_1$ are linearly independent we have $\delta_2(x) = \delta_3(x) = 0$ and of course $\delta_1(x) = 1$. Then by definition of $\delta_1(x)$ we have $x = a_1$;
- $g(x)$ inside one of the edges of $\triangle a_1 a_2 a_3$: not possible, because it would be necessary to have $\delta_i(x) = 0$ for only one specific index i. It does not happen if $x = a_j$, $j = 1, 2, 3$, based on the previous item, and it does not happen if $x \in \mathbb{R}^2 \backslash \{a_1, a_2, a_3\}$ because having $w_i > 0$, $i = 1, 2, 3$, implies that all $y_i(x) > 0$ and therefore all $\delta_i(x) > 0$.

If we denote the interior of $\triangle a_1 a_2 a_3$ as $\overset{\circ}{\triangle} a_1 a_2 a_3$, we can describe the general action of $g(x)$ in two parts:

- $g(a_i) = a_i$, $i = 1, 2, 3$;
- $g(x)$ maps $\mathbb{R}^2 \backslash \{a_1, a_2, a_3\}$ into $\overset{\circ}{\triangle} a_1 a_2 a_3$.

Some interesting conclusions can be traced from the above analysis, if we consider $g(x)$ as the base function for an iterative algorithm. Firstly, as $g(x)$ is a $\delta(x)$-convex combination of the points a_i, the very first iteration is enough to bring an external point $x \in \mathbb{R}^2 \backslash \triangle a_1 a_2 a_3$ to the interior of the triangle.

Secondly, if $x^* \ne a_i$, $i = 1, 2, 3$, then each a_i is a fixed point that can act as a "trap" for the iterative method, preventing it from continuing the search for the actual minimum. This means that the points a_i should be

avoided during the iteration process (but the resulting limit point can still be one of the a_i points). Avoiding the a_i points is something that happens automatically if we don't start at one of these points, because, as shown above, we have $g\left(\mathbb{R}^2\backslash\{a_1, a_2, a_3\}\right) \subset \overset{\circ}{\triangle}a_1a_2a_3$.

Finally, we have the easily checked limit:

$$\lim_{||x||\to\infty} \delta_i(x) = \frac{w_i}{\sum_{i=1}^3 w_i}. \tag{2.9}$$

In case all w_i have the same value, then for a point x located very far away from the triangle we have $\delta_i(x) \cong \frac{1}{3}$. But this means that the first iteration of the iterative method will bring this point to the interior of a very narrow neighborhood of the barycenter of the triangle, a nice property.

From equality (2.8) we have that x^* is a fixed point of the linear function $g(x)$. If we consider using the Banach fixed-point theorem we have that our target now is to find conditions to ensure that the absolute value of each eigenvalue of the Jacobian of $g(x)$ is smaller than one, at least locally to x^*. This fact enables the use of $g(x)$ in an iterative method to find x^*, as shown below:

$$x^{q+1} = g(x^q) = \sum_{i=1}^3 \frac{y_i(x^q)}{S(x^q)}a_i = \mathbf{A}\delta(x^q). \tag{2.10}$$

To calculate the Jacobian of $g(x)$ it is necessary to calculate the gradient of $y_i(x)$ and the gradient of $S(x)$:

$$\nabla_j y_i(x) = \frac{\partial y_i(x)}{\partial x_j} = -\frac{w_i(x_j - a_{ij})}{\|x - a_i\|_2^3} = -y_i(x)\frac{x_j - a_{ij}}{\|x - a_i\|_2^2},$$

$$\nabla_j S(x) = \frac{\partial S(x)}{\partial x_j} = -\sum_{q=1}^3 \frac{w_q(x_j - a_{qj})}{\|x - a_q\|_2^3} = -\sum_{q=1}^3 y_q(x)\frac{x_j - a_{qj}}{\|x - a_q\|_2^2}. \tag{2.11}$$

Thus, the components of the Jacobian of $g(x)$, $J_g(x)_{ij}$, for $i = 1, 2$ and $j = 1, 2$, are:

$$J_g(x)_{ij} = \frac{\partial g_i(x)}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\sum_{p=1}^3 a_{pi}y_p(x)}{S(x)} = \sum_{p=1}^3 \left(\frac{a_{pi}}{S(x)}\frac{\partial y_p(x)}{\partial x_j} - \frac{a_{pi}y_p(x)}{S(x)^2}\frac{\partial S(x)}{\partial x_j}\right)$$

$$= \sum_{p=1}^3 \left(-\frac{y_p(x)}{S(x)}a_{pi}\frac{x_j - a_{pj}}{\|x - a_p\|_2^2} - \frac{y_p(x)}{S(x)^2}a_{pi}\left(\sum_{q=1}^3 y_q(x)\frac{x_j - a_{qj}}{\|x - a_q\|_2^2}\right)\right).$$

Rearranging:

$$J_g(x)_{ij} = \sum_{p=1}^3 \left(\delta_p(x)a_{pi}\sum_{q=1}^3 \delta_q(x)\frac{x_j - a_{qj}}{\|x - a_q\|_2^2} - \delta_p(x)a_{pi}\frac{x_j - a_{pj}}{\|x - a_p\|_2^2}\right)$$

$$= \left(\sum_{p=1}^3 \delta_p(x)a_{pi}\right)\left(\sum_{q=1}^3 \delta_q(x)\frac{x_j - a_{qj}}{\|x - a_q\|_2^2}\right) - \sum_{p=1}^3 \delta_p(x)a_{pi}\frac{x_j - a_{pj}}{\|x - a_p\|_2^2}.$$

As the summation indexes are immaterial, "p" and "q" can be exchanged in the last summation:

$$J_g(x)_{ij} = \left(\sum_{p=1}^3 \delta_p(x)a_{pi}\right)\left(\sum_{q=1}^3 \delta_q(x)\frac{x_j - a_{qj}}{\|x - a_q\|_2^2}\right) - \sum_{q=1}^3 \delta_q(x)a_{qi}\frac{x_j - a_{qj}}{\|x - a_q\|_2^2}.$$

Thus:

$$J_g(x)_{ij} = \sum_{q=1}^{3} \delta_q(x) \frac{\left(\sum_{p=1}^{3} \delta_p(x)a_{pi} - a_{qi}\right)(x_j - a_{qj})}{\|x - a_q\|_2^2}$$

$$= \sum_{q=1}^{3} \delta_q(x) \frac{\left((\mathbf{A}\delta(x))_i - a_{qi}\right)(x_j - a_{qj})}{\|x - a_q\|_2^2}.$$

And finally $J_g(x)$ can be expressed as:

$$J_g(x) = \sum_{i=1}^{3} \delta_i(x) \frac{(\mathbf{A}\delta(x) - a_i)(x - a_i)^T}{\|x - a_i\|_2^2} = \sum_{i=1}^{3} \delta_i(x) \frac{(a_i - \mathbf{A}\delta(x))(a_i - x)^T}{\|a_i - x\|_2^2}. \tag{2.12}$$

The main target of this development is to express the behavior of the Weiszfeld algorithm in a local neighborhood of the Fermat−Weber point, and remembering that $x^* = \mathbf{A}\delta(x^*)$, the expression of $J_g(x)$ at x^* becomes:

$$J_g(x^*) = \sum_{i=1}^{3} \delta_i(x^*) \frac{(a_i - x^*)(a_i - x^*)^T}{\|a_i - x^*\|_2^2}. \tag{2.13}$$

If we define $\mathbf{v}_i = a_i - x^*$, it can be rewritten as:

$$J_g(x^*) = \sum_{i=1}^{3} \delta_i(x^*) \frac{\mathbf{v}_i \mathbf{v}_i^T}{\mathbf{v}_i^T \mathbf{v}_i} = \sum_{i=1}^{3} \delta_i(x^*) \operatorname{Proj}_{\mathbf{v}_i}. \tag{2.14}$$

Considering that in a first order approximation we have $g(x) = x^* + J_g(x^*)(x - x^*)$, it is possible to interpret the action of $J_g(x^*)$ in the following way: $g(x)$ is the result of the sum of x^* (as a base vector) and the $\delta_i(x^*)$ weighted projections of $(x - x^*)$ on each of the vectors $\mathbf{v}_i$, $i = 1, 2, 3$.

As a basic result from linear algebra we know that $\mathbf{v}_i$ is an eigenvector of $\operatorname{Proj}_{\mathbf{v}_i}$ associated with the eigenvalue $\lambda = 1$, and all the other eigenvalues are zero. As a consequence, for any arbitrary vector $\mathbf{u} \in \mathbb{R}^2$ we have $\|\operatorname{Proj}_{\mathbf{v}_i}(\mathbf{u})\|_2 \leq \|\mathbf{u}\|_2$.

If we consider $\mathbf{u}$ as having its origin at x^* (defined as $\mathbf{u} = x - x^*$), we have:

$$\|J_g(x^*)\mathbf{u}\|_2 = \left\|\sum_{i=1}^{3} \delta_i(x^*) \operatorname{Proj}_{\mathbf{v}_i}(\mathbf{u})\right\|_2 < \sum_{i=1}^{3} \delta_i(x^*)\|\operatorname{Proj}_{\mathbf{v}_i}(\mathbf{u})\|_2 \leq \sum_{i=1}^{3} \delta_i(x^*)\|\mathbf{u}\|_2 = \|\mathbf{u}\|_2. \tag{2.15}$$

The inequality above is strict because the vectors $\mathbf{v}_i$ are never perfectly aligned (the points a_i are not collinear). It follows that:

$$\frac{\|J_g(x^*)\mathbf{u}\|_2}{\|\mathbf{u}\|_2} < 1 \ , \forall \mathbf{u} \in \mathbb{R}^2 \Rightarrow \|J_g(x^*)\|_2 < 1, \tag{2.16}$$

where $\|J_g(x^*)\|_2$ means the euclidean norm of the matrix $J_g(x^*)$. Thus $g(x)$ can be used in an iterative method to find x^*, as claimed above, because all the functions considered so far are differentiable (except at the points a_i) and so there is a local neighborhood of x^* where the iterative Weiszfeld method (2.10) converges:

$$x^{q+1} = g(x^q) = \sum_{i=1}^{3} \frac{y_i(x^q)}{S(x^q)} a_i = \mathbf{A}\delta(x^q).$$

It is possible to obtain another expression for $J_g(x^*)$ if we take a closer look at $\nabla W(x)$:

$$\nabla W_t(x) = \sum_{i=1}^{3} \frac{w_i(x_t - a_{it})}{\sqrt{\sum_{j=1}^{2}(x_j - a_{ij})^2}}$$

$$= x_t \sum_{i=1}^{3} \frac{w_i}{\sqrt{\sum_{j=1}^{2}(x_j - a_{ij})^2}} - \sum_{i=1}^{3} \frac{w_i a_{it}}{\sqrt{\sum_{j=1}^{2}(x_j - a_{ij})^2}}, \qquad (2.17)$$

which can be expressed as:

$$\nabla W(x) = x S(x) - \sum_{i=1}^{3} a_i y_i(x). \qquad (2.18)$$

For a given x^q, and using (2.10) we have:

$$\nabla W(x^q) = x^q S(x^q) - \sum_{i=1}^{3} a_i y_i(x^q) = x^q S(x^q) - x^{q+1} S(x^q) \qquad (2.19)$$

and then:

$$x^{q+1} = x^q - \frac{\nabla W(x^q)}{S(x^q)} \Rightarrow g(x) = x - \frac{\nabla W(x)}{S(x)}. \qquad (2.20)$$

This new expression for $g(x)$ permits another way to represent the Jacobian $J_g(x)$ as:

$$J_g(x) = I - \frac{H_W(x)}{S(x)} + \frac{\nabla W(x) \nabla S^T(x)}{S(x)^2}. \qquad (2.21)$$

Remembering that $\nabla W(x^*) = 0$, we have:

$$J_g(x^*) = I - \frac{H_W(x^*)}{S(x^*)}. \qquad (2.22)$$

This is a simple expression for $J_g(x^*)$, which depends only on $S(x^*)$ and on the Hessian of $W(x)$ at x^*.

## 3. Geometric interpretation and convergence

The interesting fact that in the local analysis x^{q+1} is obtained by projections on the vectors $\mathbf{v}_i = a_i - x^*$ gives rise to a geometric interpretation which helps the convergence analysis. We will restrict this interpretation to the cases where $x^* \neq a_i$, $i = 1, 2, 3$, which means that $x^* \in \overset{\circ}{\triangle} a_1 a_2 a_3$.

Returning to equation (2.4) and forcing $\nabla W(x) = 0$ results, for $t = 1, 2$:

$$\nabla W_t(x^*) = - \sum_{i=1}^{3} w_i \cos \phi_{it} = 0. \qquad (3.1)$$

This expression can be easily interpreted as the equilibrium of a system of three forces with magnitudes w_i, $i = 1, 2, 3$.

Figure 2 depicts this system of "forces" and also shows the angles ϕ_{i1} of the $\mathbf{w}_i$ vectors with respect to the horizontal axis x_1. The equilibrium of these three "forces" implies that their sum must be zero, meaning that
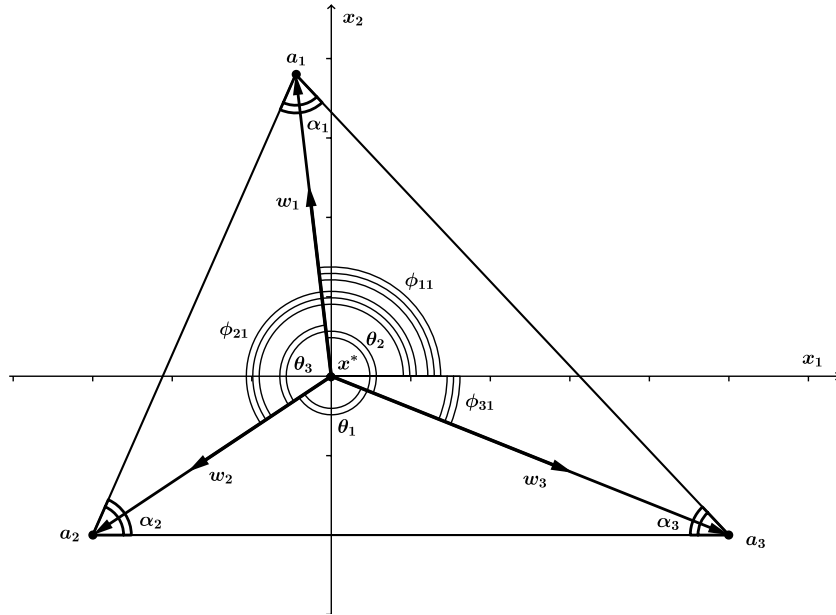
FIGURE 2. Equilibrium of the three "forces" w_1, w_2 and w_3.

they can be considered as the clockwise (or counterclockwise) oriented edges of a triangle. Therefore, the weights w_i must obey the well known triangle inequalities:

$$w_1 < w_2 + w_3, \tag{3.2}$$
$$w_2 < w_1 + w_3, \tag{3.3}$$
$$w_3 < w_1 + w_2. \tag{3.4}$$

To simplify analysis, lets define $u = w_2/w_1 > 0$ and $v = w_3/w_1 > 0$. If we consider (w_2, w_3, w_1) as the homogeneous coordinates of a point in the projective space $\mathbb{RP}^2$, the tuple (u, v) can then be seen as the corresponding coordinates in the affine plane. The previous inequalities may be rearranged and rewritten as:

$$u + v > 1, \tag{3.5}$$
$$u - v < 1, \tag{3.6}$$
$$v - u < 1. \tag{3.7}$$

Figure 3 depicts the uv affine plane, for $u \geq 0$ and $v \geq 0$, divided into four open regions: A, B, C and D. Region B does not satisfy the constraint $v - u < 1$, region C does not satisfy the constraint $u - v < 1$ and region D does not satisfy the constraint $u + v > 1$. Regions B, C and D represent non-equilibrium configurations that are not the focus of our geometric approach. Region A represents the equilibrium region, which satisfies all the constraints. The boundary of region A represents degenerated cases that will not be considered: they would demand parallelism of the "forces" and collinearity of the points a_i as a consequence.

The previous constraints on the values of u and v are enough to guarantee the existence of the equilibrium configuration of forces in the two dimensional euclidean space, but they are not enough to guarantee that this configuration is compatible with the triangle formed by the points a_i, as depicted in Figure 2. This triangle can be characterized by its angles α_i, $i = 1, 2, 3$, which restrict the freedom of the angles θ_i between the "forces" w_i.
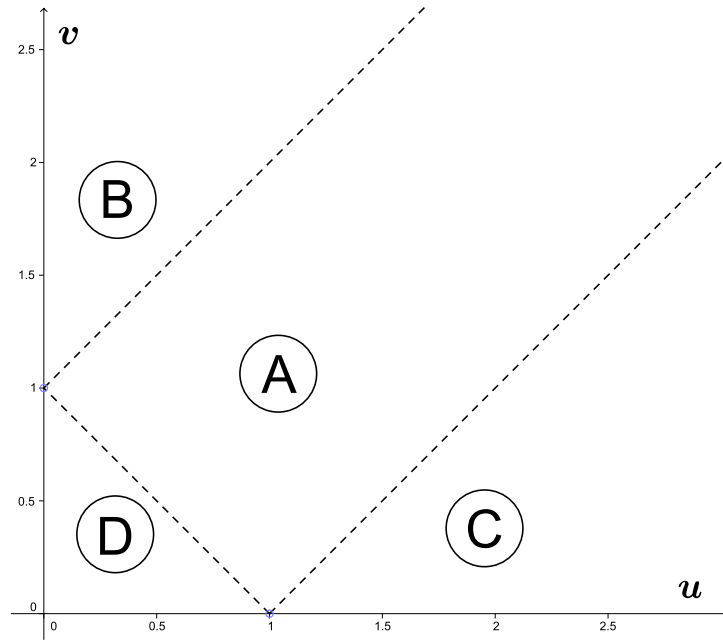
FIGURE 3. The uv affine plane showing the region A of equilibrium.

Applying the law of cosines we obtain:

$$\cos\theta_1 = -\frac{1}{2}\left(\frac{w_2^2 + w_3^2 - w_1^2}{w_2 w_3}\right), \tag{3.8}$$

$$\cos\theta_2 = -\frac{1}{2}\left(\frac{w_1^2 + w_3^2 - w_2^2}{w_1 w_3}\right), \tag{3.9}$$

$$\cos\theta_3 = -\frac{1}{2}\left(\frac{w_1^2 + w_2^2 - w_3^2}{w_1 w_2}\right). \tag{3.10}$$

Of course, if we have $w_i = w$, for $w > 0$, then the solution to the above equations is $\theta_i = 120$ degrees, $i = 1, 2, 3$. Imposing the constraints $-1 \le \cos\theta_i \le 1$, $i = 1, 2, 3$, to these expressions results in the same set of inequalities (3.2), (3.3) and (3.4), which can now be considered redundant.

Expressing now these equations in terms of u and v we have:

$$\cos\theta_1 = -\frac{1}{2}\left(\frac{u^2 + v^2 - 1}{uv}\right), \tag{3.11}$$

$$\cos\theta_2 = -\frac{1}{2}\left(\frac{1 + v^2 - u^2}{v}\right), \tag{3.12}$$

$$\cos\theta_3 = -\frac{1}{2}\left(\frac{1 + u^2 - v^2}{u}\right). \tag{3.13}$$

With help of Figure 2 it is possible to notice that $\theta_i > \alpha_i$, $i = 1, 2, 3$, is a mandatory constraint to make the w_i spatial configuration compatible with the triangle formed by the points a_i. Remembering that the function cosine is decreasing for arguments between 0 and 180 degrees, this constraint becomes $\cos\theta_i < \cos\alpha_i$, $i = 1, 2, 3$.
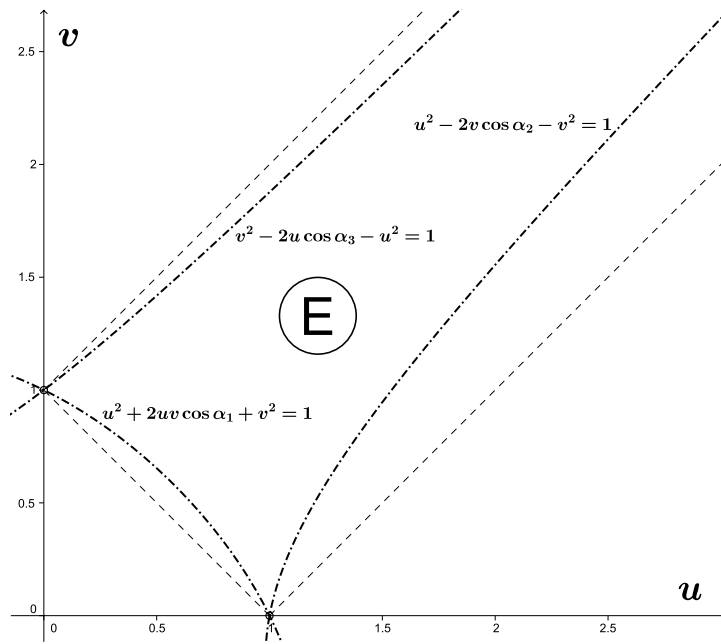
FIGURE 4. uv affine plane showing the restricted region E of equilibrium.

If we substitute our uv definitions for $\cos\theta_i$ in these inequalities we obtain:

$$-\frac{1}{2}\left(\frac{u^2+v^2-1}{uv}\right) < \cos\alpha_1, \qquad (3.14)$$

$$-\frac{1}{2}\left(\frac{1+v^2-u^2}{v}\right) < \cos\alpha_2, \qquad (3.15)$$

$$-\frac{1}{2}\left(\frac{1+u^2-v^2}{u}\right) < \cos\alpha_3, \qquad (3.16)$$

or,

$$u^2 + 2uv\cos\alpha_1 + v^2 > 1, \qquad (3.17)$$

$$u^2 - 2v\cos\alpha_2 - v^2 < 1, \qquad (3.18)$$

$$v^2 - 2u\cos\alpha_3 - u^2 < 1. \qquad (3.19)$$

These additional constraints can also be included in the uv affine plane, and the new acceptable uv (open) region will be denoted as region E, as depicted in Figure 4. The first inequality is the exterior of an ellipse and the other two are the "exterior" of two hyperbolas. Obtaining different curves like ellipses and hyperbolas as boundaries is just a side effect of representing the feasible region in the affine plane, instead of the projective space $\mathbb{RP}^2$, where the tuples (w_2, w_3, w_1) really belong to.

Summarizing, given w_1, w_2 and w_3, $w_i > 0$, $i = 1, 2, 3$, we define $u = w_2/w_1$ and $v = w_3/w_1$. Then the region E can be delimited using the $\cos\alpha_i$ information and the inclusion of the point (u, v) in the region E can be easily tested.

Now we will resume our geometric considerations using $w_i = w$, for $w > 0$, $i = 1, 2, 3$ and assume that all vertices have angles less than 120 degrees. The reason for developing this special situation is its rich geometric appealing mainly caused by the fact that in this case the Fermat−Weber point has the property that the angles
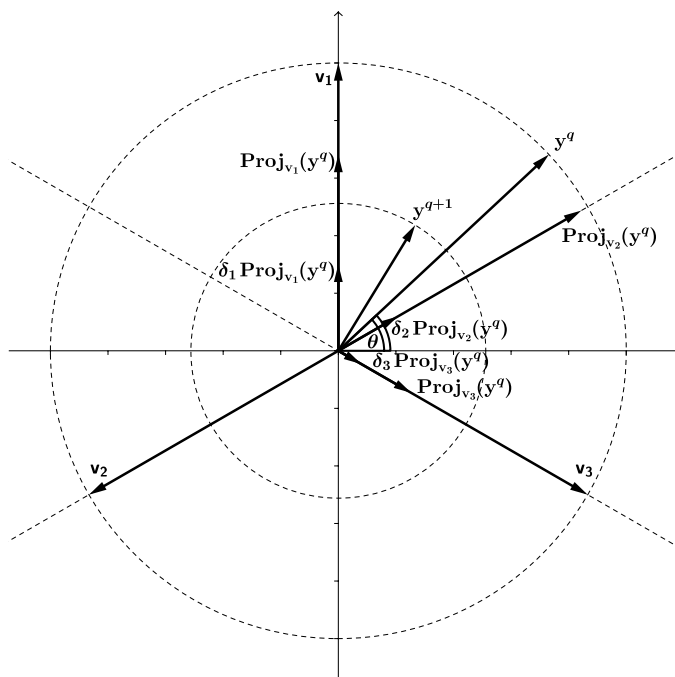
FIGURE 5. Vectors $\mathbf{y}^q = x^q - x^*$, $\mathbf{y}^{q+1} = x^{q+1} - x^*$ and associated projections.

formed by the vectors $\mathbf{v}_i$ are 120 degrees, as depicted in Figure 5 (a similar development can be undertaken using the θ_i angles, but at the cost of more complex expressions). This Figure presents a cartesian plane with x^* translated to the origin, the vector $\mathbf{v}_1$, chosen to point in the upward direction, the vectors $\mathbf{v}_2$ and $\mathbf{v}_3$, located 120 and 240 degrees, respectively, conterclockwise from $\mathbf{v}_1$, the vector $\mathbf{y}^q = x^q - x^*$, the projections on the $\mathbf{v}_i$ directions, its δ_i $(= \delta_i(x^*))$ weighted versions and its sum, the vector $\mathbf{y}^{q+1} = x^{q+1} - x^*$. The angle θ is formed by the vector $\mathbf{y}^q$ and the horizontal axis.

The projection factors on each $\mathbf{v}_i$ are:

$$
\begin{cases}
\pi_{\mathbf{v}_1}: & \cos(90° - \theta) = \sin\theta \\
\pi_{\mathbf{v}_2}: & \cos(120° + 90° - \theta) = -\frac{1}{2}\sin\theta - \frac{\sqrt{3}}{2}\cos\theta \\
\pi_{\mathbf{v}_3}: & \cos(240° + 90° - \theta) = -\frac{1}{2}\sin\theta + \frac{\sqrt{3}}{2}\sin\theta.
\end{cases}
$$

Applying the weights δ_i and calculating the sum of the projection factors of each $\mathbf{v}_i$ component on the h (horizontal) and v (vertical) axes of the plane:

$$
\begin{cases}
\pi_h = & -\frac{\sqrt{3}}{2}\delta_2\left(-\frac{1}{2}\sin\theta - \frac{\sqrt{3}}{2}\cos\theta\right) + \frac{\sqrt{3}}{2}\delta_3\left(-\frac{1}{2}\sin\theta + \frac{\sqrt{3}}{2}\cos\theta\right), \\
\pi_v = & \delta_1\sin\theta - \frac{1}{2}\delta_2\left(-\frac{1}{2}\sin\theta - \frac{\sqrt{3}}{2}\cos\theta\right) - \frac{1}{2}\delta_3\left(-\frac{1}{2}\sin\theta + \frac{\sqrt{3}}{2}\cos\theta\right),
\end{cases}
$$

or,

$$
\begin{cases}
\pi_h = & \frac{3}{4}(\delta_2 + \delta_3)\cos\theta + \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\sin\theta, \\
\pi_v = & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\cos\theta + \delta_1\sin\theta + \frac{1}{4}(\delta_2 + \delta_3)\sin\theta,
\end{cases}
$$

which means that

$$\begin{bmatrix} x_u^{q+1} - x_u^* \\ x_v^{q+1} - x_v^* \end{bmatrix} = \begin{bmatrix} \frac{3}{4}(\delta_2 + \delta_3) & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) \\ \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) & \delta_1 + \frac{1}{4}(\delta_2 + \delta_3) \end{bmatrix} \begin{bmatrix} x_u^q - x_u^* \\ x_v^q - x_v^* \end{bmatrix},$$

or,

$$x^{q+1} - x^* = J_g(x^*)(x^q - x^*), \tag{3.20}$$

where

$$J_g(x^*) = \begin{bmatrix} \frac{3}{4}(\delta_2 + \delta_3) & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) \\ \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) & \delta_1 + \frac{1}{4}(\delta_2 + \delta_3) \end{bmatrix}. \tag{3.21}$$

As a particular case, if $\delta_1 = \delta_2 = \delta_3 = \frac{1}{3}$ then $J_g(x^*) = \frac{1}{2}I$, which implies that $(x^{q+1} - x^*) = \frac{1}{2}(x^q - x^*)$. It means that whenever the points a_i are located at the vertices of an equilateral triangle then the local convergence will be linear with rate $\frac{1}{2}$. The eigenvalues of this matrix, denoted by μ, will define the convergence behavior. It is worth noting that the matrix $J_g(x^*)$ is symmetric (as expected) and that implies that all the eigenvalues are real. They can be found by looking for the roots of:

$$\det\left(\begin{bmatrix} \frac{3}{4}(\delta_2 + \delta_3) - \mu & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) \\ \frac{\sqrt{3}}{4}(\delta_2 - \delta_3) & \delta_1 + \frac{1}{4}(\delta_2 + \delta_3) - \mu \end{bmatrix}\right) = 0. \tag{3.22}$$

Developing this expression:

$$\mu^2 - \left(\frac{3}{4}(\delta_2 + \delta_3) + \delta_1 + \frac{1}{4}(\delta_2 + \delta_3)\right)\mu + \frac{3}{4}(\delta_2 + \delta_3)\left(\delta_1 + \frac{1}{4}(\delta_2 + \delta_3)\right) - \frac{3}{16}(\delta_2 - \delta_3)^2 = 0, \tag{3.23}$$

or,

$$\mu^2 - \mu + \frac{3}{4}K = 0, \tag{3.24}$$

where $K = \delta_1\delta_2 + \delta_1\delta_3 + \delta_2\delta_3$. The roots are given by:

$$\mu_{1,2} = \frac{1}{2} \pm \frac{\sqrt{1 - 3K}}{2}. \tag{3.25}$$

The smallest root will be denoted by $\mu_1 = \frac{1}{2} - \frac{\sqrt{1-3K}}{2}$ and the largest root will be denoted by $\mu_2 = \frac{1}{2} + \frac{\sqrt{1-3K}}{2}$. Focusing on the largest root, it is clear that decreasing K increases μ_2, and it is thus necessary to verify how much K can be reduced. But $\delta_i > 0$, $i = 1, 2, 3$, and this implies that $K > 0$. Also, it is possible to develop another expression for K:

$$1 = (\delta_1 + \delta_2 + \delta_3)^2 = \delta_1^2 + \delta_2^2 + \delta_3^2 + 2(\delta_1\delta_2 + \delta_1\delta_3 + \delta_2\delta_3) = \delta_1^2 + \delta_2^2 + \delta_3^2 + 2K,$$

or,

$$K = \frac{1 - (\delta_1^2 + \delta_2^2 + \delta_3^2)}{2} = \frac{1 - r^2}{2}, \tag{3.26}$$

where $r^2 = \delta_1^2 + \delta_2^2 + \delta_3^2$. This expression can be easily interpreted if we consider r as the radius of a sphere located at the origin of a tridimensional euclidean space $\mathbb{R}^3$ with axes labeled as δ_1, δ_2 and δ_3. As $\delta_1 + \delta_2 + \delta_3 = 1$, the K level curves correspond to the intersection of the spheres defined for each r and the plane $\delta_1 + \delta_2 + \delta_3 = 1$. It is then possible to make $K > 0$ smaller than any $\varepsilon > 0$ simply imposing the constraint $\sqrt{1 - 2\varepsilon} < r < 1$. On the other hand, the largest possible value of K is $\frac{1}{3}$, which can be obtained decreasing r to the smallest possible value, given by $r = \frac{\sqrt{3}}{3}$, or $\delta_1 = \delta_2 = \delta_3 = \frac{1}{3}$ (this is the limit situation when the sphere of radius r is tangent to the plane defined by $\delta_1 + \delta_2 + \delta_3 = 1$).

Then the range of variation of K is $0 < K \leq \frac{1}{3}$ and K can be as small as needed making the angle of one of the vertices of the triangle formed by the points a_i approach 120 degrees. The factor $\Delta = 1 - 3K$ measures

the deviation from the equilateral triangle configuration and will be named *equilaterality coefficient*. Please note that the maximum deviation $\Delta = 1$ $(K = 0)$ in our context means that the triangle has one angle greater than or equal to 120 degrees. On the other hand, minimum deviation $\Delta = 0$ $(K = \frac{1}{3})$ means that the triangle is equilateral.

We have then the following situations:

(1) In the most symmetric situation the points a_i are located at the vertices of an equilateral triangle and $K = \frac{1}{3}$. It results in having $\mu_1 = \mu_2 = \frac{1}{2}$, which means that the rate of convergence is $\frac{1}{2}$ no matter the direction x^q takes when converging to x^*.
(2) In the opposite situation, K is very close to zero $(0 < K \ll \frac{1}{3})$, which implies that μ_1 is very close to zero and μ_2 is very close to one. It means a fast convergence on the direction of the eigenvector associated with μ_1 and a very slow convergence on the direction of the eigenvector associated with μ_2.

In general, we can define the rate of local convergence associated to the Weiszfeld algorithm as the largest eigenvalue associated to the Jacobian:

$$\mu_C = \mu_2 = \frac{1 + \sqrt{1 - 3K}}{2}. \tag{3.27}$$

These results demonstrate that, depending on the geometrical configuration of the points a_i in the plane $\mathbb{R}^2$, one of the eigenvalues will always be greater than or equal to 0.5 and the rate of convergence will follow this worst eigenvalue. If a certain δ_i is too close to one, K is very close to zero and one of the eigenvalues is then very close to one, meaning a very slow convergence to x^*.

4. Step-size generalization

The generalization developed in the article "Accelerating convergence in the Fermat−Weber location problem" [3] includes a parameter $\lambda > 0$ which changes (2.20) in the following way:

$$x^{q+1} = x^q - \lambda \frac{\nabla W(x^q)}{S(x^q)} \Rightarrow g(x) = x - \lambda \frac{\nabla W(x)}{S(x)}. \tag{4.1}$$

The target of this modification is to increase the step-size of the Weiszfeld algorithm, accelerating its convergence. The λ parameter will be called *step-size multiplier*, and $\lambda = 1$ corresponds to the original Weiszfeld algorithm.

But for what values of λ does the method converge? And for these values of λ, which ones improve its convergence rate?

This idea will be developed in this Section using a local expansion in the neighborhood of x^* (the Jacobian without argument will be assumed to be calculated at x^*). The value generated by this modified iteration will be denoted using a subscript λ:

$$\begin{aligned} x_\lambda^{q+1} &= x^q + \lambda(x^{q+1} - x^q) = (1 - \lambda)x^q + \lambda x^{q+1} \\ &= (1 - \lambda)x^q + \lambda \left[x^* + J_g(x^*)(x^q - x^*) \right], \end{aligned} \tag{4.2}$$

or,

$$\begin{aligned} x_\lambda^{q+1} - x^* &= (1 - \lambda)x^q - (1 - \lambda)x^* + \lambda J_g(x^*)(x^q - x^*) \\ &= \left[(1 - \lambda)I + \lambda J_g(x^*) \right] (x^q - x^*), \end{aligned} \tag{4.3}$$

and now it is possible to define the new Jacobian J_g^λ using (2.22):

$$J_g^\lambda = (1 - \lambda)I + \lambda J_g = (1 - \lambda)I + \lambda \left(I - \frac{H_W(x^*)}{S(x^*)} \right) \tag{4.4}$$

Once again, in case $\delta_1 = \delta_2 = \delta_3 = \frac{1}{3}$, then all eigenvalues of J_g are 0.5 or, in other words, $J_g = \frac{1}{2}I$. Using now (4.4) we obtain $J_g^\lambda = (1 - \frac{\lambda}{2})I$, which means that all eigenvalues of J_g^λ are $1 - \frac{\lambda}{2}$. Searching for accelerating the convergence means decreasing the eigenvalues of J_g^λ, which can be obtained by increasing λ, starting from its standard Weiszfeld algorithm value of one. Thus, for $\lambda = 1$ we have linear convergence with rate of convergence 0.5. If we increase λ from 1 to 1.5, we get linear convergence with rate of convergence 0.25. The best result can be obtained letting $\lambda = 2$, what makes $J_g^\lambda = 0$. In a local analysis context, it means that the order of convergence shifts from linear to *superlinear*.

The above mentioned case is a particular case. In general, J_g can have eigenvalues larger than 0.5, and they can attain, in extreme cases, values very close to the value one.

Following the same reasoning developed in Section 3 and applying (4.4), the new matrix that represents the linear transformation J_g^λ is:

$$J_g^\lambda = \begin{bmatrix} 1 - \lambda + \frac{3}{4}(\delta_2 + \delta_3)\lambda & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\lambda \\ \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\lambda & 1 - \lambda + \left(\delta_1 + \frac{1}{4}(\delta_2 + \delta_3)\right)\lambda \end{bmatrix}.$$

The eigenvalues of this matrix, which are functions of λ, will characterize the convergence behavior. Again, they can be found looking for the roots of:

$$\det\left(\begin{bmatrix} 1 - \lambda + \frac{3}{4}(\delta_2 + \delta_3)\lambda - \mu & \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\lambda \\ \frac{\sqrt{3}}{4}(\delta_2 - \delta_3)\lambda & 1 - \lambda + \left(\delta_1 + \frac{1}{4}(\delta_2 + \delta_3)\right)\lambda - \mu \end{bmatrix}\right) = 0. \tag{4.5}$$

Developing this expression:

$$\mu^2 - \left(1 - \lambda + \frac{3}{4}(\delta_2 + \delta_3)\lambda + 1 - \lambda + \left(\delta_1 + \frac{1}{4}(\delta_2 + \delta_3)\right)\lambda\right)\mu$$

$$+ 1 - 2\lambda + \lambda^2 + (\lambda - \lambda^2)\left(\delta_1 + \frac{1}{4}(\delta_2 + \delta_3) + \frac{3}{4}(\delta_2 + \delta_3)\right)$$

$$+ \frac{3}{16}(\delta_2 + \delta_3)(4\delta_1 + \delta_2 + \delta_3) - \frac{3}{16}(\delta_2 - \delta_3)^2\lambda^2 = 0, \quad (4.6)$$

or,

$$\mu^2 - \mu(2 - \lambda) + 1 - \lambda + \frac{3}{4}K\lambda^2 = 0. \tag{4.7}$$

where $K = \delta_1\delta_2 + \delta_1\delta_3 + \delta_2\delta_3$. The roots are given by:

$$\mu = 1 - \frac{\lambda}{2} \pm \lambda\frac{\sqrt{1 - 3K}}{2}. \tag{4.8}$$

Obviously, if we let $\lambda = 1$ then (4.8) becomes (3.25).

As in Section 3, we have then the following situations:

(1) In the most symmetric situation, whenever the points a_i are located at the vertices of an equilateral triangle, we have $K = \frac{1}{3}$. It results in having $\mu_1 = \mu_2 = 1 - \frac{\lambda}{2}$, which means that the rate of convergence starts from 0.5 for $\lambda = 1$, but it can be substantially improved increasing λ until it reaches the value $\lambda = 2$, for which we will have $\mu_1 = \mu_2 = 0$. This corresponds to the already commented *superlinear (local) convergence* situation. It is worth noting that the convergence rate is the same no matter the geometric position of x^q relative to x^*.

If we continue increasing λ, the eigenvalues μ_1 and μ_2 will simultaneously become negative. The convergence can still be sustained if $|\mu_1| = |\mu_2| < 1$, which implies in keeping $\lambda < 4$. On the other side, if $\lambda \geq 4$ the method will diverge.

(2) In a general situation, K is positive but different from $\frac{1}{3}$. Using a previous Section compatible naming scheme, the eigenvalues will be denoted as:

$$\mu_1 = 1 - \frac{\lambda}{2} - \lambda\frac{\sqrt{1-3K}}{2} = 1 - \lambda\left(\frac{1}{2} + \frac{\sqrt{1-3K}}{2}\right), \tag{4.9}$$

$$\mu_2 = 1 - \frac{\lambda}{2} + \lambda\frac{\sqrt{1-3K}}{2} = 1 - \lambda\left(\frac{1}{2} - \frac{\sqrt{1-3K}}{2}\right). \tag{4.10}$$

For a fixed value of K, $0 < K < \frac{1}{3}$, we have:

$$0 < \left(\frac{1}{2} - \frac{\sqrt{1-3K}}{2}\right) < \left(\frac{1}{2} + \frac{\sqrt{1-3K}}{2}\right) < 1.$$

These definitions imply that the coefficient of λ in the expression of μ_1 is larger than the coefficient of λ in the expression of μ_2, which means that μ_1 is more sensitive to the increase of λ than μ_2. For $\lambda = 1$ we have $0 < \mu_1 < \mu_2 < 1$ and increasing λ will decrease both μ_1 and μ_2, but μ_1 will decrease faster than μ_2 and it means that μ_1 can be made zero before μ_2:

$$\mu_1 = 1 - \frac{\lambda}{2} - \lambda\frac{\sqrt{1-3K}}{2} = 0 \Rightarrow \lambda_1 = \frac{2}{1 + \sqrt{1-3K}}. \tag{4.11}$$

So for $\lambda = \lambda_1$ the convergence in the μ_1 associated autovector direction is superlinear. The definition of λ_1 implies that $1 < \lambda_1 \leq 2$, for $0 < K \leq \frac{1}{3}$. On the other hand, it is still possible to reduce μ_2 if we continue increasing λ. However, increasing λ above λ_1 means making μ_1 negative, but this is not a real problem in terms of convergence, as far as we keep $|\mu_1| < 1$. The convergence to x^* will just become oscillatory in the μ_1 associated autovector direction. So, it is possible to increase λ until:

$$\mu_1 = 1 - \frac{\lambda}{2} - \lambda\frac{\sqrt{1-3K}}{2} = -1 \Rightarrow \lambda_S = \frac{4}{1 + \sqrt{1-3K}}. \tag{4.12}$$

Specifically, keeping $\lambda < \lambda_S$ ensures convergence. Notice that $\lambda_S = 2\lambda_1$ and so we have $2 < \lambda_S \leq 4$.

However, if we target the best convergence configuration, the ideal situation could be achieved by increasing λ in such a way as to get $\mu_2 = -\mu_1$, which would imply $|\mu_1| = |\mu_2|$. In this case the convergence rate would be the best possible, considering that the approaching direction is not known (the convergence to x^* will be oscillatory in the μ_1 associated autovector direction, as already commented above). The λ and μ that characterize this situation are:

$$\mu_2 = -\mu_1 \Rightarrow 1 - \frac{\lambda}{2} + \lambda\frac{\sqrt{1-3K}}{2} = -1 + \frac{\lambda}{2} + \lambda\frac{\sqrt{1-3K}}{2} \Rightarrow \lambda_M = 2 \tag{4.13}$$

and

$$\mu_M = 1 - \frac{\lambda_M}{2} + \lambda_M\frac{\sqrt{1-3K}}{2} = \sqrt{1-3K}. \tag{4.14}$$

Thus, $\lambda = \lambda_M = 2$ is (again) the best value to be used for λ because in general we don't know the approaching direction (we don't know, *a priori*, the location of x^*) and μ_M is the best convergence rate that can be achieved in this situation. It is not possible to have a specific value of λ that ensures superlinear convergence in all approaching directions, as in the symmetric situation, but $\lambda = 2$ is still the best alternative (a tradeoff between a good and homogeneous convergence in all directions and the best convergence rate that can be achieved in one specific eigenvector direction).

Another special value of λ is the one that changes the signal of the second eigenvalue:

$$\mu_2 = 1 - \frac{\lambda}{2} + \lambda\frac{\sqrt{1-3K}}{2} = 0 \Rightarrow \lambda_2 = \frac{2}{1 - \sqrt{1-3K}}. \tag{4.15}$$
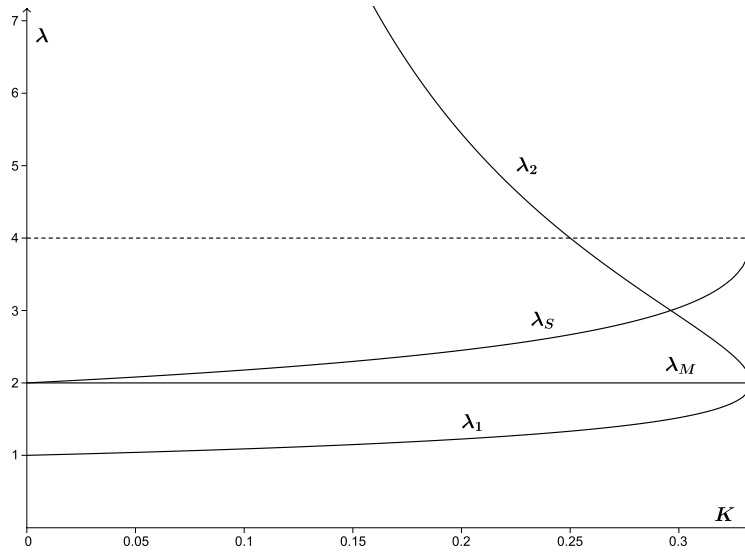
FIGURE 6. λ_1, λ_M, λ_2 and λ_S as functions of K.

So for $\lambda = \lambda_2$ the convergence rate in the direction of the eigenvector associated with μ_2 is superlinear. If $\lambda > \lambda_2$ then $\mu_2 < 0$ and the convergence becomes oscillatory in the direction of the eigenvector associated with μ_2.

An interesting question now is: which one is bigger, λ_S or λ_2? Is there a specific value K_0 for K that implies $\lambda_S = \lambda_2$?

$$\lambda_S = \lambda_2 \Rightarrow \frac{4}{1 + \sqrt{1 - 3K}} = \frac{2}{1 - \sqrt{1 - 3K}} \Rightarrow K_0 = \frac{8}{27} \Rightarrow \lambda = 3. \tag{4.16}$$

Therefore $K = K_0 = \frac{8}{27}$ implies $\lambda_S = \lambda_2 = 3$. A closer examination of the definitions of λ_S and λ_2 leads us to the conclusion that $K < K_0 \Rightarrow \lambda_S < \lambda_2$ and $K > K_0 \Rightarrow \lambda_S > \lambda_2$.

Summarizing: increasing λ from one to λ_1 decreases both μ_1 and μ_2. When λ reaches λ_1 we have $\mu_1 = 0$. Increasing λ from λ_1 to λ_M still decreases μ_1 and μ_2, but μ_1 is now negative, and we start to increase its absolute value. When λ reaches λ_M we have $\mu_2 = -\mu_1 = \mu_M$ and this is what we consider the best convergence case. Still increasing λ beyond λ_M decreases μ_2 but increases the absolute value of μ_1, in such a way that $\mu_2 < \mu_M$, but $|\mu_1| > \mu_M$. Now, depending on which one is bigger, λ_2 or λ_S, μ_2 will change sign (if $\lambda_2 < \lambda_S$), or the algorithm will diverge in the direction associated with the eigenvalue μ_1 (if $\lambda_S < \lambda_2$), or both events will happen simultaneously (if $\lambda_2 = \lambda_S = 3$).

These different cases are depicted in Figure 6.

A figure of merit ρ that measures the benefit of using $\lambda = \lambda_M = 2$ against the standard $\lambda = 1$ can be created to compare both situations. Considering it is expected that $\mu_M < \mu_C$, we define ρ for $\mu_M > 0$ as:

$$\rho = 2\frac{\mu_C}{\mu_M} - 1 = 2\frac{\frac{1}{2} + \frac{\sqrt{1-3K}}{2}}{\sqrt{1 - 3K}} - 1 = \frac{1}{\sqrt{1 - 3K}} - 1 = \frac{1}{\mu_M} - 1,$$

and $\rho = +\infty$ for $\mu_M = 0$. We always have $\rho > 0$ (because $0 < K \leq \frac{1}{3}$ implies that $0 \leq \mu_M < 1$) and this is a direct measurement of the improvement in the linear convergence of the method. For K very close to zero we have μ_M (and μ_C) very close to one and in consequence we have ρ very close to zero. For K approaching $\frac{1}{3}$ the linear convergence rate gets lower (better) and ρ gets bigger. If $K = \frac{1}{3}$ then the convergence is superlinear

and having $\rho = +\infty$ makes a lot of sense. These properties entitles ρ to be named as the *convergence rate improvement index.*

5. Conclusion

The results presented in this article demonstrate that there is an interesting geometric interpretation of the local convergence process for the traditional Weiszfeld algorithm and its step-size modified version when the angles of the triangle formed by the three points a_i, $i = 1, 2, 3$ are less than 120 degrees and the weights are all equal. The same ideas can be applied even if these restrictions are relaxed but the Fermat−Weber point is still interior to the triangle, except that the convergence expressions become more complex.

The eigenvalues analysis also shows that using $\lambda = 2$ in the modified Weiszfeld algorithm results in a substantial improvement in the local convergence rate, but this improvement becomes less noticeable when the points are spread out in a configuration that does not resembles the vertices of an equilateral triangle (in other words, when K is close to 0).

In particular, local superlinear convergence, independent of the geometric position of the iteration points in a neighborhood of the Fermat−Weber point, is obtained only in the ideal configuration of the equilateral triangle, for which $\mu_M = 0$.

It is the aim of future research to analyze other geometric aspects of the Weiszfeld algorithm, extending some results to special configurations that include more than three points. Some of these properties are even valid in the context of global convergence.

References

[1] J. Brimberg, The Fermat−Weber location problem revisited. *Math. Program. Ser. A* **71** (1995) 71–76.

[2] J. Brimberg, Further notes on convergence of the Weiszfeld algorithm. *Yugoslav J. Oper. Res.* **13** (2003) 199–206.

[3] J. Brimberg, R. Chen and D. Chen, Accelerating convergence in the Fermat−Weber location problem. *Oper. Res. Lett.* **22** (1998) 151–157.

[4] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, Cambridge (2004).

[5] L. Cooper, Location-allocation problems. *Oper. Res.* **11** (1963) 331–343.

[6] L. Cooper and I. Katz, The Weber problem revisited. *Comput. Math. Appl.* **7** (1981) 225–234.

[7] Z. Drezner, Facility Location: A Survey of Applications and Methods. *Springer Ser. Oper. Res.* Springer-Verlag, New York (1995).

[8] Z Drezner and H.W. Hamacher, Facility Location: Applications and Theory. Springer, Berlin (2004).

[9] J.B. Hiriart-Urruty and C. Lemarechal, Fundamentals of Convex Analysis. Springer, Berlin (2001).

[10] H. Üster and R.F. Love, The convergence of the Weiszfeld algorithm. *Comput. Math. Appl.* **40** (2000) 443–451.

[11] D.G. Luenberger, Y. Ye, Linear and Nonlinear Programming, third edn. Springer, New York (2008).

[12] H.K. Smith, G. Laporte and P.R. Harper, Locational analysis: highlights of growth to maturity. *J. Oper.Res. Soc.* **60** (2009) 140–148.

[13] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnés est minimum. *Tohoku Math. J.* **43** (1937) 355–386.

[14] G.O. Wesolowsky, The Weber problem: history and perspectives. *Location Sci.* **1** (1993) 5–23.