Universitat de Barcelona
Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

El mètode de Weiszfeld i la resolució del problema de Weber

Autor: Vinyet Costa Garcia

Director: Dr. Miquel Bosch
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 13 de juny de 2022

Abstract

An historic view of the nonlinear optimization problem, the Weber problem, has been given, as well as its mathematical formulation. To solve the problem the Weiszfeld's method has been explained and its convergence proofed. The Newton method applied to the Weber problem has been exposed.

Resum

S'ha estudiat el problema de Weber, un problema d'optimització no lineal. S'ha vist una trajectòria històrica del problema i la seva formulació matemàtica. Per resoldre'l s'ha exposat el mètode de Weiszfeld i s'ha demostrat la seva validesa i convergència. També, s'ha vist el mètode de Newton aplicat al problema de Weber.

2020 Mathematics Subject Classification. 11G05, 11G10, 14G10

Agraïments

Vull agrair al meu tutor, Dr. Miquel Bosch, a la meva família i a les meves amigues pel recolzament al llarg de la carrera.

Índex

1 Introducció

1.1 Context i justificació del treball

L'optimització és un camp de les matemàtiques molt ampli que en les darreres dècades ha donat peu a un munt de recerca degut als avenços en computació que permeten l'ús de mètodes complexos per resoldre un problema. Tot i la evident intersecció de l'optimització en altres assignatures del grau de matemàtiques, no s'ha cursat cap assignatura centrada en aquest tema i, per això, resulta interessant. El problema de Weber va captar l'interès de l'autora gràcies al marc de Varignon. Fullejant un llibre d'optimització vaig veure el següent dibuix.

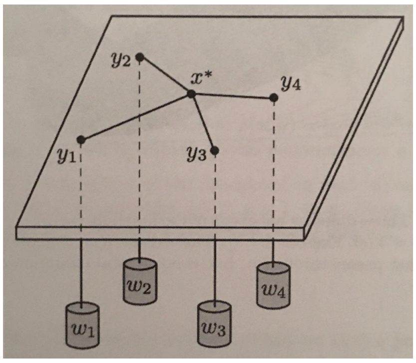

Figura 1: Marc de Varignon. Font: [4]

Aquesta és una construcció mecànica del problema de Weber que, essencialment, consisteix a trobar un punt x^* tal que minimitzi la seva distància ponderada amb la d'un seguit de vèrtexs donats, y_1, y_2, y_3, y_4 a la figura Figura 1. La ponderació de la distància ve donada pels pesos w_1, w_2, w_3, w_4 a la figura Figura 1. Al començar a recaptar informació sobre el problema de Weber seguida vaig veure que el mètode de Weiszfeld hi estava molt entrelligat, doncs segueix sent un dels mètodes més populars per tractar el problema de Weber. Per això mateix, s'ha decidit estudiar aquest problema d'optimització i el mètode més utilitzat per resoldre'l.

1.2 Objectius

Els objectius d'aquest treball de final de grau són estudiar el problema de Weber, entendre el mètode de Weiszfeld i el procés que se segueix abans de poder aplicar un algorisme (demostrar validesa i convergència, tenir en compte els problemes, consideracions i limitacions del mètode), ser capaç d'implementar el mètode i usar coneixements previs per donar una bona reflexió sobre els resultats obtinguts.

1.3 Estructura de la memòria

La memòria està dividida en tres parts. La primera part correspon a la secció 2: El problema de Weber. Aquí s'ha recollit la trajectòria històrica del problema de Weber. S'han vist els seus orígens al segle XVII com un problema geomètric en triangles i les seves primeres resolucions: la construcció amb regla i compàs del punt de Torricelli o la solució de Steiner. A continuació, s'ha introduït el plantejament que va donar Weber i la rellevància d'aquest problema en la teoria de localització, així com algun exemple d'aplicacions del problema.

Abans de passar a formalitzar el problema de Weber, s'ha volgut donar una mica de teoria sobre optimització. S'ha tractat el concepte d'optimitzar i s'ha vist com es classifiquen els problemes d'optimització (restriccions, linealitat, continuïtat,...) a través de generalitzacions del problema de Weber. També s'han definit alguns conceptes que tenen un paper clau a l'hora d'estudiar problemes d'optimització com mínim global o convexitat, seguit d'un conjunt de proposicions demostrades de convexitat i derivabilitat que ens permetran provar existència, unicitat i caracterització del mínim de la funció que volem minimitzar per resoldre el problema de Weber. Finalment, s'ha donat la formalització del problema de Weber que s'estudiarà en el treball i s'han aplicat els conceptes introduïts en l'apartat d'optimització per provar l'existència i unicitat del mínim, sota bones condicions.

La segona part del treball queda recollida en la secció 3: El mètode de Weiszfeld. Primer de tot, s'ha volgut donar una mica de context històric sobre els orígens de l'algorisme. S'ha parlat dels resultats obtinguts per Weiszfeld, per Kuhn i per Kuenne a l'hora de desenvolupar el mètode. Per explicar en què consisteix el mètode, s'ha vist la seva construcció des de una condició necessària i suficient per ser mínim. Seguidament, s'han comentat alguns problemes en la definició del mètode i la seva representació com a mètode del gradient. Aleshores, s'ha dedicat un apartat a demostrar la convergència del mètode a la solució del problema: s'ha estudiat la monotonia de la successió que genera el mètode, una representació alternativa de la funció d'iteració , un seguit de proposicions prèvies i, finalment, s'ha pogut provar la convergència. En el darrer apartat d'aquesta secció, s'ha presentat una modificació del mètode que permetrà fer una bona implementació ja que soluciona els problemes en la definició del mètode de Weiszfeld comentats anteriorment.

Per acabar, en la última part de la memòria, secció 3: Aplicació, s'ha recollit la implementació l'algorisme modificat del mètode de Weiszfeld. En el primer apartat s'ha presentat el mètode de Newton, típic mètode per trobar els zeros d'una funció, i com es pot aplicar al problema de Weber. S'han explicat les consideracions a fer per assegurar una bona definició i la convergència a la solució del problema. A continuació, s'han descrit els comentaris sobre la implementació dels mètodes que es pot trobar als annexos. Finalment, s'han donat els resultats obtinguts de la implementació i les conclusions del treball.

1.4 Principals referències

Les principals referències que s'han usat per la redacció d'aquest treball són les següents. De l'article [2] de Beck i Sabach l'any 2015 s'ha extret la major part d'informació de l'algorisme de Weiszfeld. Dona una repassada històrica al seu descobriment i la demostració de convergència que s'ha seguit en aquest treball. El llibre d'optimització no lineal de Bertsekas [4] d'on s'han après les bases necessàries d'optimització i s'ha usat de recolzament per anar verificant la major part de resultats de convexitat i derivabilitat. Finalment, l'article de Görner i Kanzow [8] en el qual s'ha basat el mètode de Newton aplicat al problema de Weber. Evidentment, s'han usat altres fonts per completar, comparar i entendre els procediments i resultats del treball i que es poden consultar al final de la memòria.

2 El problema de Weber

La història del problema de Weber té més de tres cents anys d'història i implica a matemàtics com Fermat, Torricelli, Steiner o Kuhn. Per això mateix, aquest problema s'ha treballat des de molts angles diferents i n'han sorgit diverses variacions. En aquesta primera secció, busquem repassar la seva trajectòria i explorar les primeres solucions geomètriques, així com entendre la rellevància del problema avui en dia. Donarem una mica de base teòrica que ens servirà per entendre millor el problema de Weber i, per acabar, dedicarem la última secció a formalitzar el problema i estudiar-ne algunes propietats importants.

2.1 Els orígens: Fermat, Torricelli i Steiner

La primera constància que es té del que actualment es coneix com el problema de Weber es remunta al segle XVII quan el famós matemàtic francès Pierre de Fermat (1601-1665) va proposar el següent repte: *Donats tres punts en un pla, troba un quart punt de manera que la suma de les distàncies d'aquest punt als tres altres sigui el menor possible*. Una primera solució geomètrica és sovint atribuïda al físic i matemàtic italià Evangelista Torricelli (1598-1647) que va respondre al repte amb una construcció amb regla i compàs.

Tal i com podem veure en la figura 2, la proposta de Torricelli va ser la següent. Partint dels tres punts donats (A, B, C en la figura 2), construïm tres triangles equilàters a partir de cada un dels costats del triangle ABC. Aleshores, dibuixem, per a cada triangle equilàter, la circumferència que inscriu el triangle. Observem que les tres circumferències es tallen en un punt. Aquest punt es coneix en geometria com el punt de Torricelli, punt de Fermat-Torricelli o punt de Fermat i Torricelli va afirmar que és el punt que es troba a distància mínima del triangle donat.

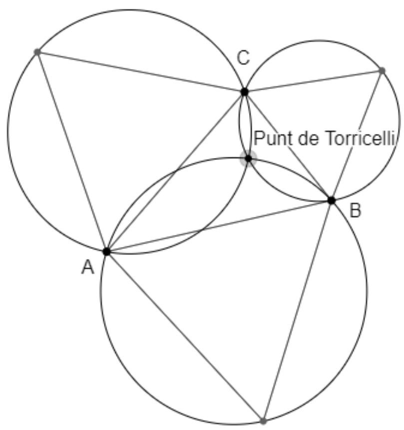

Figura 2: Construcció del punt de Torricelli

Una primera observació que podem fer respecte aquesta construcció és que no es va demostrar que el punt de Torricelli fos el de mínima distància als vèrtexs. Ens haurem d'esperar fins a principis del segle XIX per veure una demostració i per veure que prenent els tres punts inicials en unes disposicions concretes, el punt de Torricelli deixa de ser el punt de distància mínima als vèrtexs.

Jacob Steiner (1796-1863), conegut geòmetra a la Universitat de Berlín, va plantejar el mateix problema geomètric de la manera següent: *Volem connectar tres pobles A, B i C per un sistema de carreteres de manera que la longitud total de carreteres sigui el mínim possible.* Anem a formular-lo matemàticament i a donar una mica de notació per poder resoldre el problema com va fer Steiner.

Problema 2.1. Donats tres punts A, B i C, busquem un quart punt P de manera que la suma $a + b + c$ sigui mínima, on a, b i c són les distàncies entre P i A, B, C respectivament.

Steiner va donar la següent solució.

Solució. Si algun dels angles del triangle ABC és igual o major a 120°, aleshores el punt P coincideix amb el vèrtex amb major angle. Si els tres angles del triangle ABC són menors de 120°, aleshores P és el punt que forma 120° entre qualsevol parella de A, B, C, tal i com es veu en la següent figura 3.

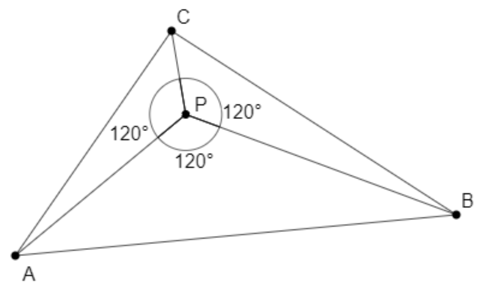

Figura 3: Solució de Steiner. Cas: tres angles d'ABC menors a 120°.

A continuació, provarem la solució de Steiner seguint els arguments fets en el llibre [6]. Abans és interessant comentar la relació entre la solució de Steiner i la construcció de Torricelli. Efectivament, el punt P que descriu Steiner en el cas on els angles del triangle ABC són menors a 120 graus és el punt de Torricelli. En el cas en què algun angle del triangle és superior a 120 graus, el punt de Torricelli deixarà de ser el punt de mínima distància i ho serà el vèrtex amb aquest angle. Si un dels angles és exactament 120 graus, aleshores el punt de Torricelli i el vèrtex amb aquest angle coincideixen.

Abans de començar amb la prova, demostrem la següent propietat geomètrica, que ens servirà per demostrar la solució de Steiner més fàcilment.

Proposició 2.2. *En el pla, donada una recta L i dos punts P, Q que es troben a la mateixa banda de la recta L, si R és el punt de la recta L que minimitza la suma de distàncies $PR + QR$, aleshores l'angle que formen PR i L i l'angle que formen QR i L són iguals (veure figura 4).*

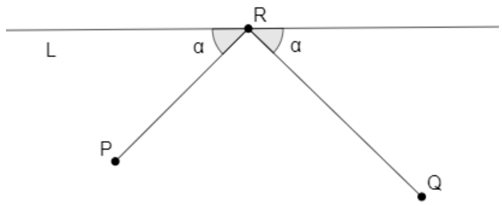

Figura 4:

Demostració. Primer de tot, construïm el punt R. Prenem el punt P' com el punt simètric de P respecte la recta L. Observem que ara podem considerar L com el lloc geomètric dels punts que estan a la mateixa distància de P i de P'. Ara tracem la recta que passa pels punts P' i Q. Anem a veure que el punt on tallen aquesta recta i la recta L és R, el punt de L que minimitza $PR + QR$.

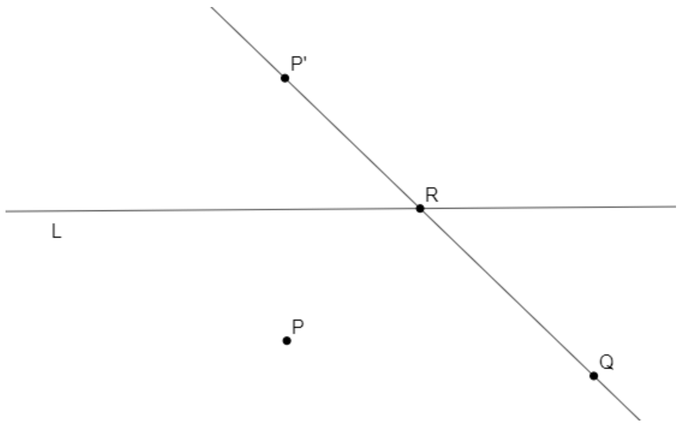

Figura 5: Construcció del punt R

Volem provar que $PR + QR < PR' + QR'$, $\forall R' \in L, R' \neq R$. Sigui R' un punt qualsevol de L i diferent de R. Sabem que $PR = P'R$ i $PR' = P'R'$. Per tant, $PR + QR = P'R + RQ = P'Q$, ja que P', Q i R estan alineats. Pel que fa a la suma amb R', tenim $PR' + QR' = P'R' + QR' > P'Q$, ja que P', R i Q formen un triangle i la suma de dos costats d'un triangle sempre és major que el tercer costat (si els tres vèrtexs del triangle no estan alineats, que sabem que no ho estan perquè R' està en L i és diferent de R). Així doncs, hem vist que $PR + QR = P'Q < PR' + QR'$, és a dir, R és el punt que volíem construir.

Observem que l'angle que formen $P'R$ i L i l'angle que formen QR i L són angles oposats de dues rectes que es tallen, per tant, són angles iguals. Per acabar la demostració doncs, només ens falta veure que l'angle que formen $P'R$ i L és igual a l'angle que formen PR i L. Això és evident ja que per construcció, tenim que L és la bisectriu de l'angle $\widehat{PRP'}$. $\square$

Demostrem que la solució proposada per Steiner és correcta.

Demostració. Sigui P el punt que satisfà que la suma $a+b+c$ és mínima. Dividim el problema en dos casos: o bé P coincideix amb algun vèrtex del triangle o bé no hi coincideix. Primer determinarem que P és el punt definit a la solució en els dos casos.

Cas 1: P és un dels vèrtex

En aquest cas, tenim que el problema esdevé triar dos costats del triangle de manera que la seva suma sigui la mínima possible. Hem de triar els dos costats més curts i, per tant, P ha de ser el vèrtex oposat al costat més llarg, és a dir, el vèrtex amb el major angle del triangle.

Cas 2: P no és un dels vèrtex

Volem veure que els angles $\widehat{APB}$, $\widehat{BPC}$ i $\widehat{CPA}$ mesuren tots 120 graus. Considerem una circumferència K de radi c amb centre el vèrtex C. P serà el punt sobre la circumferència que faci mínima la suma $AP+BP$. Suposem que ni A ni B es troben ni sobre la circumferència ni en el seu interior, és a dir, es troben com a la figura. Aleshores podem aplicar la proposició prèvia 2.2 considerant la recta tangent a la circumferència K que passa per P. En el nostre cas, tenim que A i B són els dos punts al mateix costat d'una recta i P és el punt de la recta que minimitza la suma de distàncies. Per tant, obtenim que l'angle que formen el segment AP i K i el que formen BP i K son iguals. Conseqüentment, com CP és un radi de la circumferència, CP és perpendicular a la circumferència, el que implica que l'angle que formen els segments AP i CP i el que formen BP i CP també son iguals. Podem seguir el mateix raonament canviant de vèrtexs i , per tant, tenim que els tres $\widehat{APB}$, $\widehat{BPC}$ i $\widehat{CPA}$ són iguals i han de mesurar 120. Per veure això, hem suposat que A i B es troben fora de la circumferència. Si suposem que, per exemple, A està dins o a sobre la circumferència és fàcil veure la següent desigualtat:

$$a+b+c = AP+BP+CP \geqslant AB+AC,$$

ja que $AC \leqslant c$ i, per tant, tindríem que si $P=A$ la distància seria menor. Això contradiu la nostra hipòtesi.

Per veure quan ocorre cadascun dels casos es pot estudiar la construcció del punt P.[2] □

2.2 El problema de Weber i la teoria de localització

A principis del segle XX, l'economista alemany Alfred Weber canvia l'enfoc del problema i el planteja considerant fàbriques, costs de transports, proveïdors i clients... Es planteja un problema de la vida real i li dóna a un previ problema de geometria una aplicació, donant-li una perspectiva més industrial. Per escriure aquest apartat s'han consultat les referències [7] i [7].

Problema 2.3. Una empresa ha d'establir la ubicació d'una fàbrica que proporcioni matèria primera a dos magatzems i un cert producte a un mercat. La distància des

[2]Consultar [6] per discussió de la construcció del punt P.

de la fàbrica als magatzems i al mercat suposen un cost econòmic per l'empresa que es pot estimar i que considerem proporcional a la distància entre la fàbrica i aquests punts d'interès. Quina és la localització de la fàbrica que genera el menor cost econòmic a l'empresa?

El principal canvi entre el problema que va formular Weber i els problemes que hem resolt geomètricament, més enllà d'una formulació més pràctica, és l'aparició d'una ponderació a les distàncies que formen la suma que volem minimitzar. Un altre canvi important i que és el que se sol fer en aplicacions pràctiques és augmentar el nombre de vèrtexs. Fins ara, hem estat estudiant triangles però la generalització més important és considerar un conjunt de punts donats (els vèrtexs) i buscar un punt que faci que la suma de les distàncies d'aquest punt a cadascun dels vèrtexs sigui mínima. A més, podem o no considerar que les distàncies són ponderades. Altres generalitzacions inclouen haver d'afegir més d'una fàbrica o considerar que en comptes de voler apropar-nos als vèrtexs, ens en volem allunyar. També fins ara ens hem restringit al pla, però quan parlem d'aplicacions del problema més enllà de col·locar una fàbrica en una ciutat, veiem que sovint és útil augmentar les dimensions i treballar en $\mathbb{R}^n$.

El que és evident és que aquest problema ha sigut objecte d'estudi de matemàtics i altres durant segles i per això també ha rebut molts noms: Problema de Fermat, Problema de Weber, Problema de Steiner, Problema de Fermat-Weber, ... Nosaltres l'anomenarem Problema de Weber ja que va ser gràcies a Weber que el problema es va popularitzar i és com és més conegut en la literatura.

Aquests tipus de problemes es coneixen com a problemes de localització i són la base de la teoria de localització d'instal·lacions (o *Facility Location Theory* en anglès). Aquesta àrea de les matemàtiques va tenir molts descobriments al llarg del segle XX i el problema de Weber en va ser un dels orígens. Un problema de localització és un problema amb la següent estructura: donat un espai mètric i un conjunt de punts donats, determina un nombre de punts addicionals que minimitzin una funció de la distància entre punts nous i existents.

Quan parlem de col·locar més d'una instal·lació apareixen els problemes d'assignació, que estan molt entrelligats amb els problemes de localització. Suposem que tenim un conjunt de clients (el nostres vèrtexs) i volem obrir dos centres de distribució de productes des d'on s'envien les comandes que han realitzat els clients per internet. No només es tracta de col·locar els dos centres, sinó que també s'ha de decidir a quin dels dos centres assignarem cada client. S'ha de tenir en compte la distància del client al centre però també la capacitat de cada centre, un centre amb masses comandes que no és capaç d'assumir no ens interessa. S'ha d'optimitzar la localitació i l'assignació (*location-allocation problems*).

La teoria de localització i el problema de Weber tenen un munt d'aplicacions i no només a l'hora de col·locar fàbriques. Si deixem de considerar l'espai on es troba el nostre problema com una regió geogràfica podem trobar exemples molt interessants. Imaginem que estem en un partit polític i volem decidir d'entre un conjunt de candidats quin volem que representi el nostre partit. En aquest cas, els votants seran el nostre conjunt de vèrtexs i la nostra funció de distància és una estimació de com de probable és que un cert votant voti a un candidat en concret. Veiem que els problemes poden ser molt complexos i tenir aplicacions molt útils i interessants. També tenim aplicacions en taxonomia, en algoritmes de *clustering*, i més!

La teoria de localització ha tingut molts avanços en els darrers anys però continua sent una àrea d'estudi vigent on queden moltes preguntes obertes. A més, parlem de problemes als quals sovint no podem trobar una solució exacta, sinó que hem de trobar les millors aproximacions possibles, així que és necessari l'ajustament de variables per aconseguir bons resultats.

2.3 Optimització: generalitzant el problema de Weber

La teoria de localització és una part petita d'un camp més gran de les matemàtiques, l'optimització. L'optimització s'encarrega de d'aquells problemes que consisteixen a trobar la millor solució possible. Sovint resoldre problemes d'optimització és molt complicat pel que s'han desenvolupat un gran nombre de mètodes, algorismes i heurístiques per estudiar aquests problemes.

En un problema d'optimització tenim una funció donada $f : \mathbb{R}^n \to \mathbb{R}$ i el nostre objectiu és minimitzar o maximitzar $f(x)$ en Ω, $\Omega \subseteq \mathbb{R}^n$. A la funció f l'anomenem funció objectiu, o funció de cost, i ens indica com de bo és un candidat $x \in \mathbb{R}^n$ a solució pel nostre problema. Observem que trobar un màxim d'una funció f és equivalent a minimitzar la funció $-f$. Per tant, a partir d'ara parlarem només de minimitzar. I quan diem minimitzar ens referim a trobar un mínim.

Definició 2.4. *Diem que $x^* \in \mathbb{R}^n$ és un mínim local de $f : \mathbb{R}^n \to \mathbb{R}$ si existeix $\epsilon > 0$ tal que $\forall x \in \mathbb{R}^d$ on $\|x - x^*\| < \epsilon$ se satisfà $f(x^*) \leqslant f(x)$.*

Per trobar una bona solució d'un problema d'optimització, no només ens val a trobar un mínim local, el que ens interessa és trobar un mínim global.

Definició 2.5. *Diem que $x^* \in \mathbb{R}^n$ és un mínim global si $\forall x \in \mathbb{R}^n, f(x^*) \leqslant f(x)$.*

El problema de Weber doncs, és un problema d'optimització: la nostra funció objectiu és la funció de la distància i la solució és el lloc on volem col·locar la fàbrica.

Dins l'optimització podem diferenciar entre dos tipus de problemes en funció de l'espai d'on podem triar els candidats a solució, x. Si considerem tot l'espai i prenem $\Omega = \mathbb{R}^n$ tenim un problema sense restriccions. Per exemple, si considerem que podem col·locar la fàbrica en qualsevol punt de l'espai, tindrem un problema sense restriccions. Si, en canvi, decidim que la fàbrica ha d'estar dins dels límits de la ciutat, per exemple, tindrem un conjunt de restriccions que haurà de satisfer la nostra solució x^*, a part de minimitzar la funció objectiu. Principalment, ens centrarem en el cas del problema de Weber sense restriccions, prenent $\Omega = \mathbb{R}^n$.

Si ens trobem en problemes sense restriccions o en problemes amb restriccions però on les condicions que caracteritzen Ω són restriccions d'igualtat, $h(x) = 0$, o restriccions de desigualtat, $h(x) \leqslant 0$, on $h : \mathbb{R}^n \to \mathbb{R}^m (m \leqslant n)$, aleshores parlem de problemes continus. Si l'espai de candidats no és continu, acostuma a ser un conjunt de punts finits. Per exemple, si prenem un conjunt de fàbriques ja existents i volem triar quina d'elles comprar. Aleshores, estaríem treballant amb un problema discret.

Una altra classificació dels problemes d'optimització ve donada per la funció objectiu. Si f i h són funcions lineals, parlem d'optimització lineal i, si no són lineals, aleshores ens trobem dins l'optimització no lineal. La funcions que mesuren distàncies a l'espai n-dimensional acostumen a ser no lineals, així que, com el problema de Weber utilitza una funció que incorpora mesura de distàncies, és un problema d'optimització no lineal.

La derivabilitat i la convexitat de la funció objectiu juguen un paper important a l'hora de resoldre un problema d'optimització, inclòs el problema de Weber. Si la funció objectiu és diferenciable o dues vegades diferenciable, podrem aplicar eines de càlcul i, si no és diferenciable, però és convexa, podrem usar anàlisis de convexitat. Més endavant estudiarem aquestes propietats en la funció objectiu del problema de Weber, però primer, anem a definir convexitat i a demostrar algunes propietats bàsiques de funcions convexes i diferenciables usades en optimització i que ens serviran més endavant, seguint les demostracions en [4].

Pel que fa a convexitat, primer donem la definició de funció convexa i funció estrictament convexa.

Definició 2.6 (Convexitat). *Una funció $f : \mathbb{R}^n \to \mathbb{R}$ és convexa si per tot $t \in [0,1]$ i per tots $x_1, x_2 \in \mathbb{R}$, se satisfà la següent desigualtat:*

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2)$$

Diem que la funció és estrictament convexa quan se satisfà la desigualtat estrictament:

$$f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$$

La proposició següent ens permet obtenir informació sobre si els mínims d'una funció són locals o globals segons la seva convexitat.

Proposició 2.7. *Sigui $f : \mathbb{R}^n \to \mathbb{R}$ una funció.*

a) Si f és convexa i x^* és un mínim de f, aleshores x^* és un mínim global de f.

b) Si f és una funció estrictament convexa, aleshores f té com a molt un mínim global.

Demostració. Demostrarem els dos apartats per contradicció.

a) Suposem que $f : \mathbb{R}^n \to \mathbb{R}$ és una funció convexa i que $x^* \in \mathbb{R}^n$ és un mínim local de f, però no un mínim global. Aleshores, existeix $y \in \mathbb{R}^n$, $y \neq x$ tal que $f(y) < f(x)$. Això és equivalent a dir que $f(x) - f(y) > 0$ i que $tf(x) - tf(y) > 0$, $\forall t \in [0,1)$. Ara usant la convexitat de f en aquests dos punts observem que

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) = tf(x) - tf(y) + f(y) < f(x), \forall t \in [0,1)$$

Hem arribat a una contradicció: x^* és un mínim local!

b) Suposem que $f : \mathbb{R}^n \to \mathbb{R}$ és una funció estrictament convexa que té dos mínims globals, $x, y \in \mathbb{R}^n$. Aleshores podem considerar la seva mitjana $(x+y)/2$.

$$f(\frac{x}{2} + \frac{y}{2}) < \frac{f(x)}{2} + \frac{f(y)}{2} = f(x) = f(y)$$

Hem arribat a contradicció: y i x són mínims globals! □

La proposició a continuació ens dóna una condició necessària d'optimització sobre punts on la funció objectiu és diferenciable i el típic mètode per trobar els mínims d'una funció: anul·lar la derivada o, en cas de diverses variables, el gradient $\nabla f(x)$.

Proposició 2.8. *Sigui $f : \mathbb{R}^n \to \mathbb{R}$ una funció diferenciable amb derivades contínues en un conjunt obert U. Si x^* un mínim local de f en U, aleshores $\nabla f(x^*) = 0$.*

La demostració d'aquesta proposició segueix la demostració en [4].

Demostració. Primer de tot, sota les hipòtesis de la proposició, suposem que x^* és un mínim local de f en U i definim la funció $g : \mathbb{R} \to \mathbb{R}$ com $g(\alpha) := f(x^* + \alpha y)$, on $y \in \mathbb{R}^n$. Fixant el valor de y, podem derivar g utilitzant la regla de la cadena.

$$\frac{dg(\alpha)}{d\alpha} = y^T \nabla f(x^* + \alpha y)$$

Ara, prenent la derivada per $\alpha = 0$ i utilitzant que, com f és diferenciable, $y^T \nabla f(x) = f'(x; y)$, on $f'(x; y)$ denota la derivada direccional de f en x en la direcció y, obtenim que

$$\frac{dg(0)}{d\alpha} = y^T \nabla f(x^*) = \lim_{\alpha \searrow 0} \frac{f(x^* + \alpha y) - f(x^*)}{\alpha} \geqslant 0$$

L'expressió és positiva ja que x^* és un mínim local en U i, per tant, el numerador i el denominador del límit són positius. Així doncs, el producte $y^T \nabla f(x^*)$ és positiu per tot $y \in \mathbb{R}^d$, inclosos un cert $y \neq 0$ i el seu oposat, $-y$. Per tant, obtenim que $\nabla f(x^*) = 0$. □

Per resoldre problemes de minimització sense restriccions normalment s'usen algorismes basats en el descens iteratiu. El descens iteratiu consisteix en partir d'un punt inicial x_0 i generar una successió de manera que satisfaci $f(x_{k+1}) < f(x_k)$. És a dir, en cada iteració el valor de f decreix i millorem la nostra solució actual fins a, si tot va bé, arribar al mínim. Un tipus important de mètodes que utilitzen el descens iteratiu són els mètodes del gradient. Aquests mètodes, com el seu nom indica, usen el gradient de la funció $\nabla f(x)$ que ens marca en quina direcció la funció augmenta i disminueix, per prendre un pas en la direcció desitjada. La forma general d'una iteració d'un mètode del gradient es pot escriure com:

$$x_{k+1} = x_k + \alpha_k d_k,$$

on α_k és un pas positiu i d_k és una direcció de descens calculada a partir del gradient.

2.4 El problema de Weber: formalització i estudi

Fins ara hem vist una mica d'història del problema així com diferents variacions i aplicacions. Ara anem a concretar el problema que estudiarem a partir d'ara.

Problema 2.9 (Problema de Weber)**.** Donats un conjunt de punts $A = \{a_1, ..., a_m\}$ en $\mathbb{R}^n$ i un conjunt de nombres reals positius $w_i > 0, i = 1, \ldots, m$ volem minimitzar la funció

$$f(x) := \sum_{i=1}^{m} w_i \|x - a_i\|, \; x \in \mathbb{R}^n$$

on $\|x - a_i\|$ denota la distància euclidiana entre x i a_i. Als punts $a_1, ..., a_m$ els anomenem vèrtexs i als nombres w_i, els seus pesos associats. Els punts $x \in \mathbb{R}^n$ que minimitzen la funció f són solucions del problema de Weber.

Alguna primera observació sobre aquesta formulació del problema. Hem considerat la distància euclidiana perquè és la que habitualment es fa servir. Un cas especial que es dóna en aquest problema és quan tots els vèrtexs són col·lineals.

Definició 2.10. *Diem que els punts $a_1, \ldots, a_m \in \mathbb{R}^n$ són col·lineals si es troben en una mateixa recta. És a dir, si existeixen $y, d \in \mathbb{R}^n$ i $t_1, \ldots, t_m \in \mathbb{R}$ tals que $a_i = y + t_i d$, $i = 1, \ldots, m$.*

És evident que si els vèrtexs estan en una mateixa recta, aleshores els punts que minimitzin la nostra funció objectiu també ho estaran. Així que si els vèrtexs són col·lineals, podem reduir el problema al cas d'una dimensió i prendre $a_1, \ldots, a_m \in \mathbb{R}$. Anem a veure el cas sense ponderar, és a dir, $w_i = 1, \forall i$.

Problema 2.11 (Cas vèrtexs col·lineals sense pesos)**.** Donats els nombres reals $a_1, \ldots, a_m$, busquem un real x que minimitzi la funció:

$$g(x) := \sum_{i=1}^{m} |a_i - x|$$

Solució. Primer de tot, sense perdre de generalització, podem assumir que els punts $a_1, \ldots, a_m$ estan ordenats de menor a major, és a dir, $a_i < a_{i+1}, \forall i$. Aleshores és evident que $x \in [a_1, a_m]$. Prenem $y \in [a_1, a_m]$ i ϵ prou petit i considerem les distàncies respecte y i respecte $y + \epsilon$. Al sumar ϵ a y, ens allunyem en ϵ de tots els k vèrtexs a_i tal que $a_i < y$ i ens apropem en ϵ a la resta de $m - k$ vèrtexs, $0 \leqslant k \leqslant m$. Per tant, tenim que

$$g(y + \epsilon) = g(y) + k\epsilon - (m - k)\epsilon = g(y) - (2k - m)\epsilon$$

Per tant, voldrem seleccionar y de manera que deixi el mateix nombre de vèrtexs a l'esquerra i a la dreta, $k = m/2$. Suposem que tenim un nombre parell de punts. Aleshores el mínim i la solució del problema serà qualsevol dels punts $y \in [a_{\frac{m}{2}}, a_{\frac{m}{2}+1}]$.Si tenim un nombre de punts imparells, el mínim de g es donarà al vèrtex

$a_{\lceil \frac{m}{2} \rceil}$ ja que té el mateix nombre de punts a l'esquerra i a la dreta. I augmentar aquest punt en ϵ farà augmentar en ϵ la distància dels punts a la dreta del vèrtex i disminuir en ϵ la distància dels punts a l'esquerra. Com té el mateix nombre de vèrtexs a cada costat aquests dos valors es cancel·len i només quedarà la distància que s'ha augmentat respecte el vèrtex central- Si disminuïm el vèrtex en ϵ passa exactament el mateix.

En el cas on afegim pesos, observem que podem seguir un argument similar al de la solució anterior. En comptes de comptar en quants termes del sumatori sumarem ϵ i en quants termes restarem ϵ, el que haurem de fer és ponderar ϵ en cada terme i trobar quin punt o entre quins punts s'assoleix l'equilibri. Formalitzem aquesta idea.

Problema 2.12 (Cas vèrtexs col·lineals amb pesos)**.** Donats els nombres reals $a_1, \dots, a_m$ i els nombres reals positius $w_1, \dots, w_m$, busquem un real x que minimitzi la funció:

$$g(x) := \sum_{i=1}^{m} w_i |a_i - x|$$

La solució amb pesos es fa anàlega a la solució sense pesos, però el resultat que s'obté és una mitjana ponderada. A partir d'ara, suposarem que els vèrtexs no són col·lineals.

Com hem dit abans, per estudiar problemes d'optimització és important tenir en compte la convexitat i derivabilitat de la funció objectiu.

Proposició 2.13. *La funció objectiu f és positiva i estrictament convexa.*

Demostració. Clarament, com els pesos són positius i les distàncies també, la funció és positiva.

Primerament, veiem que la funció és convexa. Siguin $x_1, x_2 \in \mathbb{R}^n$ i $t \in [0,1]$ qualssevol. Sumant i restant el mateix terme obtenim la següent igualtat.

$$f(tx_1 + (1-t)x_2) = \sum_{i=1}^{m} w_i \|tx_1 + (1-t)x_2 - a_i\| =$$
$$= \sum_{i=1}^{m} w_i \|tx_1 - ta_i + ta_i + (1-t)x_2 - a_i\| = \sum_{i=1}^{m} w_i \|t(x_1 - a_i) + (1-t)(x_2 - a_i)\|$$

I aplicant les propietats de la desigualtat triangular i la homogeneïtat de la norma a cadascun dels termes del sumatori veiem que la funció és convexa.

$$f(tx_1 + (1-t)x_2) = \sum_{i=1}^{m} w_i \|t(x_1 - a_i) + (1-t)(x_2 - a_i)\| \leqslant$$
$$\leqslant |t| \sum_{i=1}^{m} w_i \|x_1 - a_i\| + |1-t| \sum_{i=1}^{m} w_i \|x_2 - a_i\| = tf(x_1) + (1-t)f(x_2)$$

Per veure que és estrictament convexa, només cal observar que perquè es doni la igualtat de la definició de convexitat, s'ha de donar la igualtat per cada terme del sumatori. És a dir, $\forall i \in \{1, \ldots, m\}$

$$w_i\|tx_1 + (1-t)x_2 - a_i\| = w_i\|t(x_1 - a_i)\| + w_i\|(1-t)(x_2 - a_i)\|$$

I com es tracta d'una desigualtat triangular, la igualtat implica que x_1, x_2 i a_i estan alineats per tot i. Per tant, la desigualtat és estricte en $f(tx_1+(1-t)x_2) \leqslant tf(x_1)+(1-t)f(x_2)$ implica que els vèrtexs estan alineats i la funció serà estrictament convexa quan no ho estiguin. $\square$

I del fet que la funció objectiu és positiva i estrictament convexa i gràcies a la propietat de convexitat 2.7 tenim el següent resultat.

Corol·lari 2.14. *Si els vèrtexs no són col·lineals, la funció f té un únic mínim.*

Demostració. La unicitat és clara. L'existència del mínim és conseqüència del fet que

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \sum_{i=1}^{m} w_i\|x - a_i\| = \infty$$

$\square$

Pel que fa a la derivabilitat, tenim que la funció f és de classe C^1 en $\mathbb{R}^n \setminus A$. Calculem el gradient de la funció, per $x \notin A$, tenim

$$\nabla f(x) = \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|}$$

Observem que l'argument de l'existència del mínim també és vàlid en el cas col·lineal. Per tant, podem usar la proposició 2.8 sobre la funció objectiu f en $\mathbb{R}^n \setminus A$ i obtenim el següent resultat.

Corol·lari 2.15. *Si x^* és un mínim de f en $\mathbb{R}^n \setminus A$, aleshores $\nabla f(x^*) = 0$.*

Observem que usant la convexitat i la derivabilitat de f, hem pogut convertir una condició necessària per ser mínim en una caracterització del mínim global de f.

Així doncs, en aquesta secció, després de resseguir els orígens geomètrics del problema de Weber i el seu paper dins de la teoria de localització, hem definit bé el problema, hem resolt el cas en què els vèrtexs són col·lineals i hem usat algunes propietats bàsiques d'optimització per veure que el problema no col·lineal té solució única.

3 El mètode de Weiszfeld

El mètode de Weiszfeld és el més popular a l'hora de resoldre el problema de Weber. En aquest capítol, veurem els seus orígens, entendrem el seu funcionament i demostrarem la seva validesa. També comentarem els punts fluixos del mètode i propostes que s'han fet per millorar-lo. Finalment, procedirem a presentar la implementació que s'ha programat, juntament amb les decisions preses a l'hora de triar punts inicials, criteris de parada i altres.

3.1 Orígens i algorisme

L'any 1937 el matemàtic hongarès Endre Weiszfeld va publicar l'article [13], on donava tres demostracions d'un teorema ja establert per Sturm l'any 1884. Aquest teorema tracta una versió del que hem anomenat en aquest treball problema de Weber. En la primera demostració del teorema, Weiszfeld defineix una successió que suposadament ha de convergir a la solució del problema de Weber. Aquest article va passar desapercebut per la major part d'investigadors durant anys i el mètode va ser redescobert per altres matemàtics que no tenien constància dels desenvolupaments de Weiszfeld. En aquest treball, hem vist la traducció de anotada de l'article del 2009 [14] i, per redactar aquest apartat, ens hem basat en les referències [2] i [12].

L'article de Weiszfeld considera el cas sense pesos o, el que és el mateix, el cas en què tots els pesos són iguals a 1. També treballa únicament al pla i a l'espai en tres dimensions. Donada la similitud dels arguments entre el cas amb pesos i el cas sense pesos i que en la major part d'estudis posteriors s'ha tractat la versió amb pesos, ens concentrarem en el cas amb pesos, tal i com hem definit en el problema 2.9.

A continuació, presentem una versió del teorema que va demostrar Weiszfeld. En aquest teorema es donen la existència i unicitat de la solució del problema de Weber i dues condicions d'optimitat.

Teorema 3.1. *Siguin $A = \{a_1, ..., a_m\}$ punts de $\mathbb{R}^n$ no col·lineals i f la funció objectiu del problema de Weber definida en 2.9. Aleshores, existeix un únic punt que minimitza f. És a dir, el problema de Weber té una única solució òptima. A més a més, el mínim queda caracteritzat per les dues condicions següents:*

a) El punt $x \in \mathbb{R}^n \setminus A$ és mínim de f si i només si

$$\nabla f(x) = \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|} = 0.$$

b) El vèrtex a_i, $i \in \{1, 2, \ldots, m\}$, és mínim de f si i només si

$$\left\| \sum_{j=1, j\neq i}^{m} w_j \frac{a_i - a_j}{\|a_i - a_j\|} \right\| \leqslant w_i.$$

Demostració. Observem que amb la introducció a l'optimització i l'anàlisi prèvia del problema que hem fet en la secció anterior, aquest resultat té parts força evident. Primerament, tenim l'existència i unicitat del mínim de f pel corol·lari 2.14, és a dir, el problema té una única solució òptima.

Pel que fa a la primera condició, la implicació d'esquerra a dreta l'hem vist en el corol·lari 2.15. Anem a veure la implicació de dreta a esquerra. Suposem que existeix $x \in \mathbb{R}^n \setminus A$ tal que $f(x) = 0$. Sigui $y \in \mathbb{R}^n$ diferent de x.Podem aplicar la desigualtat de Cauchy-Schwarz que ens diu que si u i v són dos vectors de $\mathbb{R}^n$, aleshores $|u \cdot v| \leqslant \|u\|\|v\|$ prenent $u = y - a_i$ i $v = x - a_i$. A més podrem considerar la desigualtat estricta al comparar els sumatoris perquè com els vèrtexs no són col·lineals, almenys per algun $i \in \{1, \ldots, m\}$ u i v seran linealment independents.

$$f(y) = \sum_{i=1}^{m} w_i \|y - a_i\| > \sum_{i=1}^{m} w_i (y - a_i) \frac{x - a_i}{\|x - a_i\|} = \sum_{i=1}^{m} w_i ((y - x) + (x - a_i)) \frac{x - a_i}{\|x - a_i\|} =$$

$$= (y - x) \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|} + \sum_{i=1}^{m} w_i \frac{\|x - a_i\|^2}{\|x - a_i\|} = (y - x) \nabla f(x) + \sum_{i=1}^{m} w_i \|x - a_i\|.$$

Usant que per hipòtesi $\nabla f(x) = 0$, veiem que ens queda que $f(y) > f(x)$. Demostrant la primera condició del teorema.

La segona condició es pot trobar demostrada en [11].

□

Observem que per la unicitat del mínim si existeix un punt $x \in \mathbb{R}^n \setminus A$ que satisfà la primera condició, aleshores no existeix cap vèrtex que satisfaci la segona condició i, el mateix al revés, si existeix un vèrtex que compleixi la segona condició, aleshores no hi haurà cap punt $x \in \mathbb{R}^n \setminus A$ que verifiqui la primera condició.

A continuació, veiem el raonament darrere de l'algorisme iteratiu del mètode de Weiszfeld. Primer de tot, recordem el gradient de la funció objectiu f. Si $x \notin A$,

$$\nabla f(x) = \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|}$$

Suposem que els vèrtexs en A no són col·lineals. Del teorema 3.1, tenim que existeix $x^* \in \mathbb{R}$ que és sol·lució del problema de Weber. Suposant que $x^* \notin A$, tenim per 2.15 que

$$\nabla f(x^*) = 0$$

Podem aïllar x^* de l'expressió del gradient de manera parcial.

$$x^* = \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|x^* - a_i\|}} \sum_{i=1}^{m} \frac{w_i a_i}{\|x^* - a_i\|}$$

A partir d'aquesta expressió, definim el següent operador $T : \mathbb{R}^n \setminus A \to \mathbb{R}^n$:

$$T(x) := \frac{1}{\displaystyle\sum_{i=1}^{m} \frac{w_i}{\|x - a_i\|}} \sum_{i=1}^{m} \frac{w_i a_i}{\|x - a_i\|} \tag{3.1}$$

Observem que l'operador T, a l'igual que el gradient, no està definit per $x \in A$. Per construcció de l'operador, observem que se satisfà que $\forall y \in \mathbb{R}^n \setminus A$,

$$T(y) = y \iff \nabla f(y) = 0 \tag{3.2}$$

De manera que hem transformat un problema de trobar els zeros d'una funció en un problema de trobar el punt fix d'una altra funció. Ara podem utilitzar el mètode del punt fix amb l'operador T com a funció d'iteració i obtenim l'algorisme de Weiszfeld.

Definició 3.2 (Mètode de Weiszfeld). *Definim el mètode de Weiszfeld de la manera següent. Donat un punt inicial $x_0 \in \mathbb{R}^n \setminus A$, $\forall k \geqslant 0$,*

$$x_{k+1} = T(x_k)$$

on $T : \mathbb{R}^n \setminus A \to \mathbb{R}^n$ és l'operador (3.1).

Definició 3.3 (Successió de Weiszfeld). *Donat un punt inicial $x_0 \in \mathbb{R}^n \setminus A$, a la successió $\{x_k\}_{k \geqslant 0}$ generada pel mètode de Weiszfeld l'anomenem successió de Weiszfeld.*

La primera observació que podem fer és que aquest mètode no està ben definit. El domini de T és $\mathbb{R}^n \setminus A$ mentre que la seva imatge és tot $\mathbb{R}^n$. Per tant, ens podríem trobar amb què un valor de la successió de Weiszfeld coincideixi amb un dels vèrtexs i el següent terme de la successió no estigui definit. Weiszfeld no va tenir en compte que podríem obtenir un vèrtex en la successió i, suposant que $\{x_k\}_{k \geqslant 0}$ és una successió de Weiszfeld tal que $x_k \notin A, \forall k$, va demostrar la monotonia de la successió $\{f(x_k)\}_{k \geqslant 0}$, tal i com veurem en el següent apartat.

El mètode es va començar a popularitzar arran de l'article [10] per Kuhn i Kuenne de l'any 1962, que presentava l'algorisme com un descobriment nou. Més endavant, van afegir un apèndix on es donava el crèdit a Weiszfeld i es comentaven les seves demostracions. Precisament en aquest apèndix, Kuhn i Kuenne reconeixen l'error en la definició del mètode esmentat més amunt i afirmen que es pot demostrar que o bé $x_k \notin A, \forall k$ i la convergència a la solució òptima es pot assegurar o bé que el mètode es queda atrapat en un vèrtex. A més a més, van plantejar una hipòtesi dient que el mètode només es queda atrapat en un vèrtex per un conjunt numerable de punts inicials trobats en l'embolcall convex de A. Aquesta hipòtesi va resultar ser falsa, com veurem més endavant.

Una altre apunt interessant sobre el mètode de Weiszfeld fet en [10] és que és un mètode del gradient. Tal i com es fa en [2], això és fàcil de veure donant una representació alternativa de T. Sigui $x \in \mathbb{R}^n \setminus A$,

$$T(x) = x - \frac{1}{L(x)} \nabla f(x)$$

on l'operador $L : \mathbb{R}^n \setminus A \to \mathbb{R}^+$ ve definit per:

$$L(x) := \sum_{i=1}^{m} \frac{w_i}{\|x - a_i\|} \tag{3.3}$$

Efectivament, si substituïm l'operador $L(x)$ i el gradient a l'anterior fórmula i ajuntem els dos termes obtinguts, obtenim que

$$\begin{aligned} x - \frac{1}{L(x)}\nabla f(x) = x - \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|x-a_i\|}} \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|} = \\ = \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|x-a_i\|}} \left(\sum_{i=1}^{m} w_i \frac{x}{\|x - a_i\|} - \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x - a_i\|} \right) = \\ = \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|x-a_i\|}} \sum_{i=1}^{m} \frac{w_i a_i}{\|x - a_i\|} = T(x). \end{aligned}$$

Per tant el mètode de Weiszfeld es pot reescriure com:

$$x_{k+1} = x_k - \frac{1}{L(x_k)} \nabla f(x_k) \tag{3.4}$$

3.2 Convergència

En primer lloc, veurem la monotonia de la successió de valors de la funció objectiu, que va ser primerament vista per Weiszfeld en [13] i refeta per Kuhn en [9]. Nosaltres usarem els mateixos arguments però amb una notació diferent i seguint la demostració en [2].

Primer de tot, cal redefinir el mètode de Weiszfeld. Si ens trobem sota el problema de Weber 2.9, definim la funció auxiliar $h : \mathbb{R}^n \times \mathbb{R}^n \setminus A \to \mathbb{R}$ com:

$$h(x, y) := \sum_{i=1}^{m} w_i \frac{\|x - a_i\|^2}{\|y - a_i\|} \tag{3.5}$$

Volem veure que en una successió de Weiszfeld, donada una iteració $x_k \notin A$, el següent iterat $x_{k+1} = T(x_k)$ queda determinat pel mínim de la següent funció $s_k : \mathbb{R}^n \to \mathbb{R}$,

$$s_k(x) := h(x, x_k) = \sum_{i=1}^{m} w_i \frac{\|x - a_i\|^2}{\|x_k - a_i\|} \tag{3.6}$$

Primer de tot, caracteritzem el mínim de la funció s_k.

Proposició 3.4. *La funció s_k definida en l'equació* (3.6) *és estrictament convexa.*

Demostració. Siguin $y_1, y_2 \in \mathbb{R}^n$ i $t \in [0,1]$ qualssevol. Volem veure que

$$s_k(ty_1 + (1-t)y_2) < ts_k(y_1) + (1-t)s_k(y_2)$$

De manera similar a la demostració de la proposició 2.13, podem sumar i restar el mateix terme i usar la propietat de la norma de homogeneïtat i la desigualtat triangular per veure la convexitat.

$$\begin{aligned}
s_k(ty_1 + (1-t)y_2) &= \sum_{i=1}^{m} w_i \frac{\|ty_1 + (1-t)y_2 - a_i\|^2}{\|x_k - a_i\|} = \\
&= \sum_{i=1}^{m} w_i \frac{\|t(y_1 - a_i) + (1-t)(y_2 - a_i)\|^2}{\|x_k - a_i\|} \leqslant \\
&\leqslant |t| \sum_{i=1}^{m} w_i \frac{\|y_1 - a_i\|^2}{\|x_k - a_i\|} + |1-t| \sum_{i=1}^{m} w_i \frac{\|y_2 - a_i\|^2}{\|x_k - a_i\|} = \\
&= ts_k(y_1) + (1-t)s_k(y_2)
\end{aligned}$$

La desigualtat és estricta ja que hem aplicat la desigualtat triangular a cada terme del sumatori i, per tant, perquè es doni la igualtat, s'ha de donar en cada terme del sumatori. Això implicaria que els vèrtexs són col·lineals, però estem sota la hipòtesi que no ho són. Per tant, la funció s_k és estrictament convexa. $\square$

Com a conseqüència de la convexitat i del fet que s_k és una funció de classe C^1, usant les proposicions 2.7 i 2.8 obtenim que la funció s_k té com a molt un mínim global que queda determinat per la condició:

$$\nabla s_k(x) = 2 \sum_{i=1}^{m} w_i \frac{x - a_i}{\|x_k - a_i\|} = 0$$

De la mateixa manera que hem trobat l'operador T, aïllem la variable x de l'expressió anterior, per obtenir que

$$x = T(x_k) = \frac{1}{\displaystyle\sum_{i=1}^{m} \frac{w_i}{\|x_k - a_i\|}} \sum_{i=1}^{m} \frac{w_i a_i}{\|x_k - a_i\|}$$

Per tant, hem vist que efectivament $T(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}}\, h(x,y)$.

Per poder provar la monotonia, ens cal primer veure les següents propietats que relacionen la funció objectiu f i la funció auxiliar h.

Proposició 3.5 (Propietats de la funció auxiliar h)**.** *Siguin f i h les funcions definides en el problema 2.9 i l'equació* (3.5) *respectivament. Aleshores, les següents propietats es compleixen.*

a) Per tot $y \in \mathbb{R}^n \setminus A$,

$$h(y,y) = f(y).$$

b) Per tot $x \in \mathbb{R}^n$ i per tot $y \in \mathbb{R}^n \setminus A$,

$$h(x, y) \geqslant 2f(x) - f(y).$$

c) Per tot $y \in \mathbb{R}^n \setminus A$,

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

Demostració. a) Per veure la primera propietat només cal substituir a les definicions de les dues funcions:

$$h(y, y) = \sum_{i=1}^{m} w_i \frac{\|y - a_i\|^2}{\|y - a_i\|} = \sum_{i=1}^{m} w_i \|y - a_i\| = f(y)$$

b) Per provar la segona propietat, primer de tot observem que donats dos nombres reals $a \in \mathbb{R}$ i $b \in \mathbb{R}^+$, se satisfà la desigualtat

$$\frac{a^2}{b} \geqslant 2a - b$$

ja que $(a - b)^2 = a^2 + b^2 - 2ab \geqslant 0$. Per tant, prenent $a = \|x - a_i\|$ i $b = \|y - a_i\|$ tenim que per cada $i = 1, 2, \ldots, m$,

$$\frac{\|x - a_i\|^2}{\|y - a_i\|} \geqslant 2\|x - a_i\| - \|y - a_i\| \implies w_i \frac{\|x - a_i\|^2}{\|y - a_i\|} \geqslant 2w_i\|x - a_i\| - w_i\|y - a_i\|$$

Sumant els termes de $i = 1$ a $i = m$ obtenim la segona propietat.

c) La tercera propietat ja l'hem demostrat abans d'enunciar la proposició. □

A continuació, usant les propietats de la funció auxiliar h podem veure la monotonia de l'operador T respecte la funció objectiu f.

Proposició 3.6 (Monotonia de T). *Per cada $y \in \mathbb{R}^n \setminus A$,*

$$f(T(y)) \leqslant f(y)$$

i la igualtat es dóna si i només si $T(y) = y$.

Demostració. En primer lloc, usem la representació alternativa de l'operador T que hem vist en la proposició 3.5c, $T(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}}\ h(x, y)$, és a dir que per cada $x \neq T(y)$,

$$h(T(y), y) < h(x, y).$$

En particular, si $T(y) \neq y$, aleshores de la proposició 3.5a tenim que

$$h(T(y), y) < h(y, y) = f(y).$$

Aplicant la propietat que ens falta, 3.5b, veiem que

$$h(T(y), y) \geqslant 2f(T(y)) - f(y).$$

Ajuntant les dues últimes desigualtats obtenim que si $T(y) \neq y$,

$$2f(T(y)) - f(y) < f(y) \implies f(T(y)) < f(y).$$

Per tant, tenim que si $T(y) \neq y$, aleshores $f(T(y)) < f(y)$ i és evident que si $T(y) = y$, aleshores $f(T(y)) = f(y)$, □

Així doncs, com la successió de Weiszfeld queda determinada per la funció d'iteració T i com hem vist que $T(y) = y \iff \nabla f(y) = 0$ en l'equació (3.2), podem concloure de la monotonia de T que si cap valor de la successió de Weiszfeld coincideix amb algun vèrtex, aleshores el mètode decreix i es *queda aturat* només a punts òptims.

Corol·lari 3.7 (Monotonia de la successió de valors de la funció f)**.** *Sigui* $\{x_k\}_{k\geqslant 0}$ *una successió de Weiszfeld tal que* $x_k \notin A, \forall k \geqslant 0$. *Aleshores, tenim que* $\forall k \geqslant 0$

$$f(x_{k+1}) \leqslant f(x_k),$$

i la igualtat es dóna si i només si x_k *és una solució òptima del problema de Weber.*

El següent pas per veure la validesa del mètode de Weiszfeld és demostrar que la successió de Weiszfeld convergeix a la solució del problema de Weber. El teorema de convergència va ser provat per Kuhn l'any 1973 en [9], assumint que els vèrtexs no són col·lineals. Per demostrar el teorema, com fins ara, seguirem els arguments en [2] que utilitzen la funció auxiliar h prèviament definida en (3.5) i que no requereixen la hipòtesi que els vèrtexs no siguin col·lineals. El fet que no ens calgui fer suposicions sobre col·linealitat implica que no podem assegurar la unicitat del mínim, és a dir, el problema de Weber que estem resolent pot tenir més d'una solució.

Aquest teorema també arrossega el problema en la definició del mètode. Com que el mètode no està ben definit si algun dels iterats coincideix amb un vèrtex, Kuhn va incloure la hipòtesi que tots els valors de la successió de Weiszfeld siguin diferents dels vèrtexs. El problema és, que al contrari del que creia Kuhn, no podem assegurar que no caiguem en un vèrtex a l'usar el mètode. Més endavant, comentarem més a fons aquest problema.

A continuació i abans de veure el teorema de convergència, enunciarem algunes proposicions sobre la nostra funció objectiu f que ens serviran per demostrar el teorema. Aquesta primera proposició és molt semblant al *descent lemma*, una lema que es basa en la derivabilitat de la funció objectiu i que és molt utilitzat en mètodes del gradient. Tot i que hem vist que el mètode de Weiszfeld és un mètode del gradient, la nostra funció objectiu no és de classe C^1 a tot $\mathbb{R}^n$ i, per tant, n'enunciem una versió.

Proposició 3.8. *Sigui* $y \notin A$, *aleshores*

$$f(T(y)) \leqslant f(y) + \langle \nabla f(y), T(y) - y \rangle + \frac{L(y)}{2} \|T(y) - y\|^2, \tag{3.7}$$

on $L : \mathbb{R}^n \setminus A \to \mathbb{R}^+$ és l'operador que hem definit en (3.3) *i* $\langle \cdot , \cdot \rangle$ *és el producte escalar.*

Demostració. Considerem el desenvolupament de Taylor de la funció auxiliar $x \mapsto h(x,y)$ al voltant del punt y. Primer observem que com $x \mapsto h(x,y)$ és una funció quadràtica, tenim que tots els termes a partir del terme de grau 3 inclòs són nuls i, per tant, el desenvolupament de Taylor és exacte. El gradient formarà part del terme de grau 1.

$$\nabla_x h(x,y) = 2\sum_{i=1}^{m} w_i \frac{x - a_i}{\|y - a_i\|}.$$

El terme de grau 2 del desenvolupament ve donat per la matriu Hessiana de la funció, que és igual a $2L(y)I$ on I és la identitat. Aquest terme serà:

$$\frac{1}{2!}(x-y)^T 2L(y)I(x-y) = L(y)(x-y)^T(x-y) = L(y)\|x-y\|^2.$$

I per tant, podem escriure la funció com:

$$h(x,y) = h(y,y) + \langle \nabla_x h(y,y), x - y \rangle + L(y)\|x-y\|^2.$$

Per la proposició 3.5a, tenim que $h(y,y) = f(y)$ i usant que $\nabla_x h(y,y) = 2\sum_{i=1}^{m} w_i \frac{y-a_i}{\|y-a_i\|} = 2\nabla f(y)$, obtenim que

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

Ara podem substituir $x = T(y)$,

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

Finalment, usant la proposició 3.5b,

$$2f(T(y)) - f(y) \leqslant f(y) + 2\langle \nabla f(y), T(y) - y \rangle + L(y)\|T(y)-y\|^2,$$

que implica que

$$2f(T(y)) \leqslant 2f(y) + 2\langle \nabla f(y), T(y) - y \rangle + L(y)\|T(y)-y\|^2.$$

Dividint per 2 a les dues bandes de la desigualtat queda demostrada la proposició.

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

□

Aquesta versió del *descent lemma* ens serveix per demostrar la següent desigualtat.

Proposició 3.9. *Sigui* $\{x_k\}_{k\geqslant 0}$ *una successió de Weiszfeld i suposem que* $x_k \notin A, \forall k \geqslant 0$. *Aleshores, per tot* $x \in \mathbb{R}^n$, *se satisfà la següent desigualtat:*

$$f(x_{k+1}) - f(x) \leqslant \frac{L(x_k)}{2}(\|x_k - x\|^2 - \|x_{x+1} - x\|^2) \tag{3.8}$$

on $L : \mathbb{R}^n \setminus A \to \mathbb{R}^+$ *és l'operador* (3.3).

Demostració. Primer de tot, usem l'anterior proposició. Prenem la desigualtat en (3.7) substituint $y = x_k$ i tenint en compte que $T(x_k) = x_{k+1}$.

$$f(x_{k+1}) \leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L(x_k)}{2}\|x_{k+1} - x_k\|^2 \tag{3.9}$$

Com f és una funció diferenciable i convexa, se satisfà la següent desigualtat del gradient: $f(x_k) \leqslant f(x) + \langle \nabla f(x_k), x_k - x \rangle$, per tot $x \in \mathbb{R}^n$.[3] Aplicant aquesta desigualtat a l'equació anterior (3.9) i recordant la propietat distributiva del producte escalar tenim que

$$\begin{aligned} f(x_{k+1}) &\leqslant f(x) + \langle \nabla f(x_k), x_k - x \rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L(x_k)}{2}\|x_{k+1} - x_k\|^2 = \\ &= f(x) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L(x_k)}{2}\|x_{k+1} - x_k\|^2 \end{aligned}$$

En l'apartat anterior, hem vist que el mètode de Weiszfeld es podia reescriure com a mètode del gradient (3.4):

$$x_{k+1} = x_k - \frac{1}{L(x_k)}\nabla f(x_k) \implies \nabla f(x_k) = L(x_k)(x_k - x_{k+1})$$

Així que continuant amb la desigualtat anterior podem veure que

$$\begin{aligned} f(x_{k+1}) &\leqslant f(x) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L(x_k)}{2}\|x_{k+1} - x_k\|^2 = \\ &= f(x) + L(x_k)\langle x_k - x_{k+1}, x_{k+1} - x_k \rangle + \frac{L(x_k)}{2}\|x_{k+1} - x_k\|^2 = \\ &= f(x) + \frac{L(x_k)}{2}(\|x_k - x\|^2 - \|x_{k+1} - x\|^2) \end{aligned}$$

Aquesta última igualtat ve donada per la següent propietat del producte escalar. Per tot $u, v, w \in \mathbb{R}^n$,

$$2\langle w - v, u - v \rangle = \|w - v\|^2 - \|w - u\|^2 + \|u - v\|^2$$

Per tant,

$$f(x_{k+1}) - f(x) \leqslant \frac{L(x_k)}{2}(\|x_k - x\|^2 - \|x_{x+1} - x\|^2)$$

□

[3] Demostració en la proposició 1.1.7 de [3]

I com a conseqüència tenim un corol·lari clau en la prova del teorema de convergència.

Corol·lari 3.10 (Monotonia de Fejér). *Sigui* $\{x_k\}_{k\geqslant 0}$ *una successió de Weiszfeld i suposem que* $x_k \notin A, \forall k \geqslant 0$. *Aleshores, per tot* $x \in \mathbb{R}^n$ *tal que* $f(x) \leqslant f(x_k)$, *per tot* $k \geqslant 0$ *se satisfà:*

$$\|x_{k+1} - x\| \leqslant \|x_k - x\|$$

I la successió $\{x_k\}_{k\geqslant 0}$ *és fitada.*

Demostració. Com $f(x) \leqslant f(x_k)$, $\forall k \geqslant 0$, obtenim que la banda esquerra de la desigualtat (3.8) és no negativa i, per tant, la banda dreta també ho ha de ser. Conseqüentment, obtenim $\|x_{k+1} - x\| \leqslant \|x_k - x\|$. La successió és fitada ja que $\|x_{k+1} - x\| \leqslant \|x_0 - x\|$, $\forall k \geqslant 0$. □

Seguidament, exposem una propietat de convergència de les successions.

Proposició 3.11. *Sigui* $\{a_k\}_{k\geqslant 0}$ *un successió fitada en* $\mathbb{R}^n$ *tal que totes les seves successions parcials convergents tenen un mateix límit* $a \in \mathbb{R}^n$. *Aleshores, la successió* $\{a_k\}_{k\geqslant 0}$ *també convergeix al límit* a.

Demostració. Suposem que $\{a_k\}$ no convergeix al límit $a \in \mathbb{R}^n$. Això vol dir que $\exists \epsilon > 0$ tal que $\forall K \in \mathbb{N}$, $\exists k \in \mathbb{N}, k > K$ tal que $\|a_k - a\| > \epsilon$. Agafem aquest ϵ i definim una successió parcial $\{a_{k_l}\}$ de $\{a_k\}$ de la següent manera. Prenent $K \in \mathbb{N}$ sabem que $\exists k_1 > K$ tal que $\|a_{k_1} - a\| > \epsilon$. Agafem a_{k_1} el primer terme de la successió parcial. Ara prenem $K = k_1$ i tenim que $\exists k_2 > k_1$ tal que $\|a_{k_2} - a\| > \epsilon$. Agafem a_{k_2} el segon terme de la successió parcial. I així successivament construïm $\{a_{k_l}\}$ prenent $a_{k_{l+1}}$ de manera que $k_{l+1} > k_l$ i $\|a_{k_{l+1}} - a\| > \epsilon$.

No podem garantir que aquesta successió parcial és convergent però com $\{a_k\}$ és una successió fitada, també ho és $\{a_{k_l}\}$ i podem aplicar el teorema de Bolzano-Weierstrass: tota successió fitada conté una successió parcial convergent.[4] Per definició, és evident que aquesta segona successió parcial no convergeix al límit a, contradient la hipòtesi i demostrant que la successió $\{a_k\}$ és convergent i el seu límit ha de ser a. □

A partir d'aquesta propietat i de la monotonia de Fejér podrem demostrar el resultat de convergència donat per Kuhn.

Teorema 3.12 (Convergència del mètode de Weiszfeld). *Sigui* $\{x_k\}_{k\geqslant 0}$ *una successió de Weiszfeld. Si* $x_k \notin A, \forall k \geqslant 0$, *aleshores la successió convergeix a una solució òptima del problema de Weber.*

Demostració. Primer veurem que la successió convergeix i, després, que el seu límit és solució del problema de Weber 2.9.

[4] Demostració del teorema de Bolzano-Weierstrass en el teorema 2.5.5 en [1].

Per la monotonia de Fejér (Corol·lari 3.10), hem vist que la successió $\{x_k\}_{k\geqslant 0}$ és fitada i, demostrant que totes les seves successions parcials convergents tenen el mateix límit, podem aplicar la proposició 3.11 i obtenir la convergència de la successió. Suposem doncs que existeixen dues successions parcials $\{x_{n_j}\}_{j\geqslant 0}$ i $\{x_{m_j}\}_{j\geqslant 0}$ que convergeixen a límits diferents ℓ_n i ℓ_m, respectivament. Anem a veure que arribem a contradicció.

De la monotonia de la successió de valors de la funció f (Corol·lari 3.7), tenim que $f(\ell_n) \leqslant f(x_k)$, $\forall k \geqslant 0$ i, per tant, podem aplicar la monotonia de Fejér (Corol·lari 3.10) en ℓ_n i obtenim que $\|x_{k+1}-\ell_n\| \leqslant \|x_k-\ell_n\|$, $\forall k \geqslant 0$. Conseqüentment, considerem la successió monòtona decreixent $\{\|x_k - \ell_n\|\}_{k\geqslant 0}$. Aquesta successió és fitada superiorment ja que $\{x_k\}_{k\geqslant 0}$ ho és i és fitada inferiorment per 0. Per tant, com és una successió monòtona i fitada, convergeix pel teorema de convergència monòtona.[5] Clarament, com tota successió parcial convergent d'una successió convergent convergeix al mateix límit que la successió original[6], obtenim que

$$\lim_{k\to\infty} \|x_k - \ell_n\| = \lim_{j\to\infty} \|x_{n_j} - \ell_n\| = 0.$$

Amb el mateix argument també observem que

$$\lim_{k\to\infty} \|x_k - \ell_n\| = \lim_{j\to\infty} \|x_{m_j} - \ell_n\| = \|\ell_m - \ell_n\|.$$

Per tant, $\|\ell_m-\ell_n\| = 0$ i hem trobat una contradicció demostrant que la successió $\{x_k\}_{k\geqslant 0}$ convergeix a un límit $\ell = \ell_n = \ell_m$.

Ara volem veure que ℓ, el límit de $\{x_k\}_{k\geqslant 0}$, és solució òptima del problema de Weber 2.9. Si $\ell \notin A$, aleshores $\ell = T(\ell)$ ja que T és una funció continua en $\mathbb{R}^n \setminus A$ i podem prendre el límit quan $k \to \infty$ en l'equació $x_{k+1} = T(x_k)$. Per tant, en aquest cas, per (3.2) tenim que com ℓ és un punt fix de T, aleshores és solució del problema de Weber. Si $\ell \in A$, prenem $j \in \{1, 2, \ldots, m\}$ de manera que $\ell = a_j$. De la proposició 3.5c, tenim que $\forall x_k \in \mathbb{R}^n \setminus A$,

$$x_{k+1} = T(x_k) = \operatorname*{argmin}_{x\in\mathbb{R}^n} h(x, x_k).$$

Per tant,

$$\nabla_x h(x_{k+1}, x_k) = \nabla s_k(x_{k+1}) = \sum_{i=1}^{m} w_i \frac{x_{k+1} - a_i}{\|x_k - a_i\|} = 0.$$

Aïllant el terme $i = j$, obtenim

$$\sum_{i=1, i\neq j}^{m} w_i \frac{x_{k+1} - a_i}{\|x_k - a_i\|} = -w_j \frac{x_{k+1} - a_j}{\|x_k - a_j\|},$$

i prenent normes a les dues bandes de la igualtat i tenint en compte que la successió $\{\|x_k-a_j\|\}_{k\geqslant 0}$ és decreixent (conseqüència de la monotonia de Fejér (Corol·lari 3.10)

[5] Demostració del teorema de convergència monòtona en el teorema 2.4.2 de [1].

[6] Demostració en el teorema 2.5.2 de [1]

usant la desigualtat vista en aquesta demostració $f(a_j) = f(\ell) \leqslant f(x_k)$, $\forall k \geqslant 0)$, tenim que

$$\left\| \sum_{i=1, i \neq j}^{m} w_i \frac{x_{k+1} - a_i}{\|x_k - a_i\|} \right\| = w_j \frac{\|x_{k+1} - a_j\|}{\|x_k - a_j\|} \leqslant w_j$$

Finalment, prenent el límit quan $k \to \infty$ tenim la següent desigualtat:

$$\left\| \sum_{i=1, i \neq j}^{m} w_i \frac{\ell - a_i}{\|\ell - a_i\|} \right\| \leqslant w_j,$$

que segons el teorema 3.1b, implica que ℓ ha de ser solució del problema de Weber. □

Així doncs hem demostrat que el mètode de Weiszfeld, si no cau en cap vèrtex, convergeix a una solució del problema de Weber 2.9 i, en el cas de no col·linealitat, convergeix a la única solució.

3.3 Modificació del mètode

El principal problema del mètode de Weber és que no podem assegurar que cap dels iterats coincideixi amb un dels vèrtexs i haguem d'aturar l'algorisme. En aquest apartat, aprofundirem en aquest detall i proposarem una modificació del mètode que ens permetrà aplicar l'algorisme fins a trobar el mínim en tots els casos. L'estudi del problema en el vèrtexs s'ha redactat usant les referències [2] i[5] i la modificació de l'algorisme es troba en [2].

Tal i com ja s'ha comentat amb anterioritat, quan Weiszfeld va primer exposar el seu mètode no va tenir en compte que un dels vèrtexs podia formar part de la successió generada. De fet, a la pràctica s'observa que aquest cas no es dóna pràcticament mai. Tot i això, no està provat matemàticament que no es pugui donar i s'han construït exemples que arriben a un vèrtex [5]. Kuhn en el seu article [9] va afirmar que el conjunt de Kuhn, el conjunt de punts inicials que generen una successió de Weiszfeld que conté algun vèrtex a_i, és numerable i que, escollint a l'atzar un punt inicial, la probabilitat de quedar-se atrapat en un vèrtex és nul·la. D'aquesta manera garantint la validesa de l'algorisme.

També hem comentat ja que aquesta afirmació és falsa i això ho van veure Chandrasekaran i Tamir l'any 1989. Van publicar un article amb diversos contraexemples i es van fixar que en tots els exemples que contradeien l'afirmació de Kuhn el conjunt de vèrtexs es trobava contingut en un hiperplà de R^n. Propiciant la següent conjectura.

Conjectura 3.13 (Chandrasekaran i Tamir)**.** *Si els vèrtexs del poblema de Weber 2.9 no estan continguts en cap hiperplà, aleshores el conjunt de Kuhn és numerable.*

Aquesta conjectura va ser resolta per Brimberg l'any 1995 qui també va afegir una caracterització de la numerabilitat del conjunt de Kuhn.

Teorema 3.14. *El conjunt de Kuhn és numerable si i només si el conjunt de vèrtexs no està contingut en cap hiperplà de* $\mathbb{R}^n$.

Malauradament, aquest teorema no és correcte tal i com van mostrar Cánovas, Cañavate i Marín més endavant. Van exposar contraexemples a la implicació $\Longrightarrow$ i van trobar errors en alguns arguments clau de l'altra implicació. Aixi doncs, la conjectura no ha estat demostrada. També és interessant comentar que els contraexemples que refuten la demostració de Brimberg no rebutgen el contingut de la conjectura així que continua estan oberta.

Per tant, busquem modificar l'algorisme en cas que ens trobéssim amb un vèrtex a l iterar. La idea és mantenir la funció d'iteració T definida en (3.1) tal com fins ara i definir un nou operador, S, en cas que ens trobem un vèrtex. En primer lloc, haurem de comprovar si aquest vèrtex és el mínim buscat amb la segona condició del teorema 3.1 que ens diu que el vèrtex a_i, $i \in \{1, 2, \ldots, m\}$ és mínim de f si i només si

$$\|R_i\| := \left\| \sum_{j=1, j\neq i}^{m} w_j \frac{a_i - a_j}{\|a_i - a_j\|} \right\| \leqslant w_i.$$

Definim el mètode de Weiszfeld modificat de la manera següent.

Definició 3.15. *Donat un punt inicial* $x_0 \in \mathbb{R}$, $\forall k \geqslant 0$, *definim la funció d'iteració com:*

$$x_{k+1} = \bar{T}(x_k) = \begin{cases} T(x_k) & si \quad x_k \notin A \\ a_i & si \quad x_k = a_i \quad i \quad \|R_i\| \leqslant w_i \\ S(a_i) & si \quad x_k = a_i \quad i \quad \|R_i\| > w_i \end{cases}$$

Observem que si el mètode no arriba a un vèrtex en cap moment, el mètode modificat és exactament el mateix que el mètode de Weiszfeld. L'operador S encara no l'hem definit, però perquè el mètode trobi un mínim és lògic imposar sobre S la condició: $f(S(a_i)) < f(a_i)$ per tot $a_i \in A$ tal que $\|R_i\| > w_i$. Sota aquesta hipòtesi i usant la convergència del mètode de Weiszfeld podem demostrar la convergència del mètode de Weiszfeld modificat [2].

Teorema 3.16. *Sigui* $\{x_k\}_{k\geqslant 0}$ *una successió generada pel mètode de Weiszfeld modificat. Si* $f(S(a_i)) < f(a_i)$ *per tot* $a_i \in A$ *tal que* $\|R_i\| > w_i$, *aleshores la successió convergeix a una solució del problema de Weber 2.9.*

Demostració. Tenim dos casos. El primer cas es dóna quan la successió arriba a un punt fix, és a dir, $\bar{T}(x_k) = x_k$. Si $x_k \notin A$, aleshores tenim que $T(x_k) = x_k$ i per (3.2) x_k és un mínim de f. Si $x_k = a_i \in A$ per algun $i \in \{1, \ldots, m\}$, aleshores, com la hipòtesi sobre S implica que $S(a_i) \neq a_i$ si a_i no és el mínim, tenim que a_i haurà de ser el mínim i veiem que en el primer cas la successió convergeix a la solució.

El segon cas es dóna quan la successió no arriba a un punt fix. De la monotonia de la successió de valors (corol·lari 3.7) i del fet que $f(S(a_i)) < f(a_i)$ per tot $a_i \in A$ tal que $\|R_i\| > w_i$, obtenim que la successió generada és estrictament monòtona: $f(x_{k+1}) < f(x_k)$, $\forall k \geqslant 0$. Per tant, tots els iterats del mètode són diferents i, com que el conjunt de vèrtexs és finit, podem assegurar que existeix un enter positiu K tal que $a_k \notin A$, $\forall k \geqslant K$. Aleshores, la successió $\{x_k\}_{k \geqslant K}$ esdevé una successió de Weiszfeld amb punt incicial x_K i com que cap dels seus elements és un vèrtex, pel teorema de convergència del mètode de Weiszfeld (teorema 3.12) tenim que convergeix a la solució del problema de Weber. $\square$

Per poder definir S en un vèrtex a_i de manera que satisfaci la condició desitjada, volem trobar una direcció de descens de f en a_i i prendre un pas en aquesta direcció. Primer de tot, considerem la nostra funció objectiu f i observem que es pot reescriure de la següent manera.

$$f(x) = w_i\|a - a_i\| + f_i(x),$$

on

$$f_i(x) := \sum_{j=1, j\neq i}^{m} w_j\|x - a_j\|.$$

Si fixem una direcció, un vector unitari, $d \in \mathbb{R}^n$, podem definir la següent funció.

$$\alpha(t) = f(a_i + td) = w_i t\|d\| + f_i(a_i + td) = w_i t + f_i(a_i + td).$$

Com f_i és diferenciable en a_i, podem veure que si tenim en compte la definició de derivada

$$\alpha'(0) = \lim_{t\to 0} \frac{\alpha(t) - \alpha(0)}{t} = \lim_{t\to 0} \frac{f(a_i + td) - f(a_i)}{t},$$

i la definició de la derivada direccional de f en a_i en la direcció d,

$$f'(a_i; d) = \lim_{t\to 0} \frac{f(a_i + td) - f(a_i)}{t},$$

aleshores tenim que

$$f'(a_i; d) = \alpha'(0) = w_i + f_i'(a_i; d) = w_i + \langle \nabla f_i(a_i), d\rangle.$$

El menor valor de la derivada direccional[7] s'assoleix quan $d = d_i = -\nabla f_i(a_i)/\|\nabla f_i(a_i)\|$ i com $\nabla f_i(a_i) = R_i$ prendrem la direcció com

$$d_i = -\frac{R_i}{\|R_i\|} \qquad on \qquad R_i = \sum_{j=1, j\neq i}^{m} w_j \frac{a_i - a_j}{\|a_i - a_j\|} \tag{3.10}$$

[7] [2]

Generalment, a l'usar aquest mètode modificat s'escull el següent operador S.

$$S(a_i) = a_i + t_i d_i,$$

on t_i és un pas adient.

El pas t_i es pot escollir de diverses maneres mentre sigui prou petit i verifiqui $f(S(a_i)) = f(a_i + t_i d_i) < f(a_i)$. Nosaltres considerem el següent pas tal i com es fa en [2] i perquè ens servirà també al definir el mètode de Newton al següent apartat.

$$t_i = \frac{\|R_i\| - w_i}{L(a_i)}, \tag{3.11}$$

prenent L com una variació de l'operador definit en (3.3) que estigui definit en els vèrtexs,

$$L(x) := \begin{cases} \sum_{i=1}^{m} \dfrac{w_i}{\|x - a_i\|}, & x \notin A, \\ \sum_{i=1, j\neq i}^{m} \dfrac{w_i}{\|a_j - a_i\|}, & x = a_j (1 \leqslant j \leqslant m). \end{cases}$$

Així doncs, hem vist una mica la problemàtica del mètode en els vèrtexs i hem donat un modificació de l'algorisme de Wreiszfeld vàlid i convergent.

4 Aplicació

4.1 El mètode de Newton: algorisme i convergència

És difícil valorar l'eficàcia d'un algorisme sense poder comparar els resultats, així que abans de passar a l'implementació del mètode de Weiszfeld, veurem el mètode de Newton. El mètode de Newton és un dels mètodes del gradient més complexos i alhora més ràpids i s'utilitza per trobar els zeros d'una funció. En aquesta secció presentarem el mètode de Newton, comentarem com podem aplicar-lo al problema de Weber i compararem els resultats obtinguts amb els resultats del mètode de Weiszfeld. Les principals referències per aquesta secció són [4] i [8].

Originalment, el mètode de Newton serveix per trobar els zeros d'una funció. Tot i això, es pot usar per trobar mínims i màxims d'una funció tenint en compte la relació entre el mínim d'una funció i el seu gradient. Per tant, considerem el mètode de Newton aplicat al gradient d'una funció $g : \mathbb{R}^n \to \mathbb{R}$ que consisteix en, partint d'un punt inicial x_0, generar la següent successió:

$$x_{k+1} = x_k - \alpha_k(\nabla^2 g(x_k))^{-1}\nabla g(x_k),$$

on $\nabla g(x)$ és el gradient de la funció i $\nabla^2 g(x)$ la matriu Hessiana.

El mètode està ben definit en aquells conjunts oberts on la funció és de classe C^2 i la matriu Hessiana és invertible. Per poder aplicar el mètode de Newton a la nostra funció objectiu $f(x)$, haurem de considerar un conjunt que no contingui els vèrtexs a_i i que verifiqui que si x_k es troba en el conjunt, aleshores x_{k+1} també.

De la mateixa manera que amb l'aplicació del mètode de Weiszfeld, usarem la segona condició del teorema 3.1 que caracteritza el mínims en els vèrtexs. Primer de tot, prenem un vèrtex a_j que satisfaci que $f(a_j) \leqslant f(a_i)$, $\forall i \neq j$ (notem que pot haver més d'un vèrtex que satisfaci la desigualtat). Si a_j verifica la condició necessària i suficient definida en 3.1b, aleshores haurem trobat el mínim. En cas contrari, calcularem una direcció de descens d_j de f en a_j i un pas idoni $t_j > 0$ de manera que $x_0 := a_j + t_j d_j < a_j$ i prendrem x_0 com a punt inicial pel mètode de Newton.

Per escollir la direcció de descens d_j farem servir els mateixos arguments que en la modificació del mètode de Weiszfeld a l'apartat Subsecció 3.3 i obtenim que a l'igual que en (3.10):

$$d_j = -\frac{R_j}{\|R_j\|}$$

satisfà que $a_j + t_j d_j < a_j$ per t_j definit com en (3.11), per exemple.

Ara podem definir el següent conjunt no buit:

$$B(x_0) := \{x \in \mathbb{R}^n : f(x) \leqslant f(x_0)\}.$$

És evident que aquest conjunt no conté cap vèrtex i, per tant, la nostra funció objectiu és de classe C^2 en $B(x_0)$.

La matriu Hessiana $\nabla^2 f(x)$ d'una funció $f : \mathbb{R}^n \to R$ és la matriu que té la derivada parcial $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ a la columna i i la fila j. La matriu Hessiana és simètrica ja que $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$. Anem a veure com és la matriu Hessiana de la funció objectiu f. Per evitar confusions de notació, denotarem com x^i la component i-èssima del punt x i el mateix amb un vèrtex a_k. Fem les derivades parcials corresponents i obtenim que a la diagonal de la matriu tindrem la següent derivada parcial, on i serà el nombre de la fila i la columna,

$$\frac{\partial^2 f}{\partial^2 x^i}(x) = \sum_{k=1}^{m} \frac{w_k}{\|x - a_k\|^3}(\|x - a_k\|^2 - (x^i - a_k^i)^2).$$

I la resta de la matriu l'omplirem amb la següent expressió quan la fila i sigui diferent de la columna j,

$$\frac{\partial^2 f}{\partial x^i \partial x^j}(x) = \sum_{k=1}^{m} \frac{w_k}{\|x - a_k\|^3}(-(x^i - a_k^i)(x^j - a_k^j)).$$

Per veure que el mètode està ben definit ens falta demostrar la següent condició de la matriu Hessiana.

Proposició 4.1. *Sigui $x \in B(x_0)$. La matriu Hessiana $\nabla^2 f(x)$ és definida positiva, on f és la funció objectiu del problema de Weber 2.9 quan els vèrtexs no són col·lineals.*

Demostració. Sigui $x \in B(x_0)$. Del fet que f és una funció convexa, tenim que la matriu Hessiana $\nabla^2 f(x)$ és semidefinida positiva. Per veure que és definida positiva ens cal veure que la desigualtat és estricta, és a dir, que $\forall y \in \mathbb{R}^n$, $y \neq 0$, $y^T \nabla^2 f(x) d > 0$. De manera fàcil, es pot veure que la matriu Hessiana descrita en aquest apartat també es pot escriure com

$$\nabla^2 f(x) = \sum_{i=1}^{m} \frac{w_i}{\|x - a_i\|^3}(\|x - a_i\|^2 I - (x - a_i)(x - a_i)^T).$$

Aquí podem usar la desigualtat de Cauchy-Schwartz per veure que $\forall y \in \mathbb{R}^n$, $y \neq 0$,

$$\begin{aligned} y^T \nabla^2 f(x) d &= \sum_{i=1}^{m} \frac{w_i}{\|x - a_i\|^3}(\|x - a_i\|^2 \|d\|^2 - (d^T (x - a_i))^2) \leqslant \\ &\leqslant \sum_{i=1}^{m} \frac{w_i}{\|x - a_i\|^3}(\|x - a_i\|^2 \|d\|^2 - \|d\|^2 \|x - a_i\|^2) = 0 \end{aligned}$$

La igualtat és dóna en el cas en què els vectors y i $x - a_i$ són linealment independents per tot $i \in \{1, \ldots, m\}$. Així doncs la desigualtat implica que els vèrtexs són col·lineals, demostrant que si no són col·lineals la Hessiana és definida positiva. □

El mètode de Newton és convergent de manera local. Per poder garantir la convergència global del mètode, haurem d'aplicar la regla d'Armijo com en [8].

Bàsicament, a la pràctica, la regla d'Armijo ens dóna una manera de triar un pas que es redueix en cada iteració i que ens permet demostrar la convergència. Hi ha altres maneres d'implementar una cerca del pas més eficients, com les condicions de Wolfe, però la regla d'Armijo és més simple que d'altres. Per aplicar aquesta regla, primer fixarem uns valors $\rho \in (0,1)$ i $\sigma \in (0, 1/2)$. En una iteració k, triarem el pas t_k de la següent manera. Incialitzant $t_k = 1$, mentre no se satisfaci la desigualtat

$$f(x_k + t_k d_k) \leqslant f(x_k) + \sigma t_k \nabla f(x_k)^T d_k,$$

actualitzem $t_k \leftarrow t_k \rho$. Observem que com $\rho \in (0,1)$, al multiplicar t_k per ρ, anem reduint t_k i, al final, ens quedarem amb el pas més gran que satisfaci la desigualtat.

El principal inconvenient del mètode de Newton és que requereix resoldre

$$\nabla^2 f(x_k) d = -\nabla f(x_k)$$

és molt costós, sobretot en dimensions grans. Com que el problema de Weber normalment es pren en $\mathbb{R}^2$ i en $\mathbb{R}^3$, el mètode de Newton serà força bo, però al generalitzar a grans dimensions caldria esperar una velocitat de convergència baixa.

Finalment, tenim el següent teorema de convergència del mètode i la seva demostració es pot consultar en [8].

Teorema 4.2. *Si l'algorisme de Newton prèviament definit no acaba al primer pas, aleshores la successió $\{x_k\}$ generada per aquest mètode convergeix a la solució del problema de Weber 2.9, X^*. A més a més, el local rate of convergence és quadràtic, és a dir, existeix una constant $c > 0$ tal que $\|x_{k+1} - x^*\| \leqslant c\|x_k - x^*\|^2$, per tot $k \in \mathbb{N}$ prou gran.*

Hem exposat el mètode de Newton i els principals factors a tenir en compte a l'hora d'aplicar-ho al problema de Weber.

4.2 Implementació

En aquest apartat comentarem la implementació dels mètodes de Weiszfeld i de Newton que s'han fet. Una còpia del codi en llenguatge $C++$ es pot trobar a l'annex.

A l'hora de programar el mètode de Weiszfeld sense modificar s'ha de tenir en compte que l'operador T no està definit en A. Per tant, en cada iterat de l'algorisme cal verificar que no és un vèrtex abans de poder continuar. Per solucionar aquest inconvenient sense fer la modificació es podria fer comprovació prèvia dels vèrtexs. Pel teorema 3.1, el vèrtex a_i, $i \in \{1, 2, \ldots, m\}$ és mínim de f si i només si

$$\left\| \sum_{j=1, j\neq i}^{m} w_j \frac{a_i - a_j}{\|a_i - a_j\|} \right\| \leqslant w_i.$$

Així doncs, abans de començar a aplicar el mètode es podria calcular l'anterior valor per cada vèrtex i mirar si el mínim és en un dels vèrtexs. En cas positiu, s'hauria acabat i, en cas negatiu, es començaria a aplicar el mètode.

Nogensmenys, tot i que hem comentat que a la pràctica aquest cas pràcticament no es dóna, encara ens podríem trobar un vèrtex a l'iterar que ens impedís assolir el mínim. Per això mateix s'ha programat el mètode de Weiszfeld modificat tal i com es descriu en l'apartat Subsecció 3.3. La implementació que s'ha fet és per a qualsevol dimensió de l'espai $n \geqslant 1$ i qualsevol nombre de vèrtexs $m \geqslant 1$. A més, s'ha suposat no s'ha tingut en compte cap comprovació del cas on els vèrtexs són col·lineals.

Pel que fa a la tria de punt inicial, tal i com es fa en [5], s'ha calculat el centre de gravetat dels vèrtexs ja que és fàcil de calcular i té sentit que sigui una bona aproximació de la solució.

$$x_0 = \frac{\sum_{i=1}^m w_i a_i}{\sum_{j=1}^m w_j}$$

La condició de parada del mètode s'ha de satisfer quan ens trobem un mínim, així que, per la condició per 3.1 a, té sentit utilitzar que $\nabla f(x_k) = 0$ quan $x_k \notin A$ i $\|R_i\| \leqslant w_i$ quan $x_k = a_i$. Per poder comparar $\nabla f(x_k) = 0$, s'ha fixat una tolerància $tol = 10^{-5}$ i la condició de parada en aquest cas serà $\|\nabla f(x)\| < tol$.

Algorisme de Weiszfeld modificat

Pas 1: Inicialització. Fixar paràmetre $tol = 10^{-5}$. Calcular el punt inicial x_0 com:

$$x_0 = \frac{\sum_{i=1}^m w_i a_i}{\sum_{j=1}^m w_j}$$

Pas 2: Criteri de parada. Si $x_k \notin A$ i $\|\nabla f(x_k)\| < tol$, STOP i, si $x_k = a_i \in A$ i $\|R_i\| \leqslant w_i$, STOP.

Pas 3: Iteració $k+1$. Aplicar funció d'iteració:

$$x_{k+1} = \bar{T}(x_k) = \begin{cases} T(x_k) & si \quad x_k \notin A \\ \\ a_i & si \quad x_k = a_i \quad i \quad \|R_i\| \leqslant w_i \\ \\ S(a_i) & si \quad x_k = a_i \quad i \quad \|R_i\| > w_i \end{cases}$$

Posar $k \leftarrow k+1$ i anar al Pas 2: Criteri de parada.

Pel que fa al mètode de Newton, implementarem l'algorisme seguint els comentaris de l'apartat anterior. Primer, farem la comprovació prèvia en els vèrtexs usant la primera condició del teorema 3.1 tal i com hem descrit anteriorment. En el cas de trobar el mínim haurem acabat. Si no, el punt inicial s'escull a partir d'un vèrtex a_j tal que $f(a_j) \leqslant f(a_i)$, $\forall i \neq j$ fent un pas en la direcció de descens prenent la direcció d_j definida en (3.10) i el pas t_j definit en (3.11). La condició de parada serà la mateixa que amb el mètode de Weiszfeld: $\nabla f(x_k) = 0$ que, al implementar, esdevé $\|\nabla f(x_k)\| < tol$, on hem pres la tolerància $tol = 10^{-5}$.

Algorisme de Newton

Pas 1: Incialització. Fixar els paràmetres $tol = 10^{-5}$, $\rho = 0,5$, $\sigma = 1e-4$. Trobar a_j tal que $f(a_j) \leqslant f(a_i)$, $\forall i \neq j$. Si a_j verifica la condició de ser mínim (3.1b), parem. Si no, calcular d_j i t_j com en (3.10) i Equació 3.11 respectivament i posar $x_0 = a_j + t_j d_j$, $k = 0$

Pas 2: Criteri de parada. Si $\|\nabla f(x_k)\| \leqslant tol$, aleshores parem.

Pas 3: Direcció descens. Càlcul de la direcció d_k resolent $\nabla^2 f(x_k) d = -\nabla f(x_k)$.

Pas 4: Pas t_k. Escollim el pas $t_k \in \{1, \rho, \rho^2, \dots\}$ més gran que satisfà

$$f(x_k + t_k d_k) \leqslant f(x_k) + \sigma t_k \nabla f(x_k)^T d_k.$$

Pas 5: Iteració $k+1$. Posar $x_{k+1} = x_k + t_k d_k$, $k \leftarrow k+1$ i tornar al Pas 2: Cirteri de parada.

El problema principal que ens hem trobat al implementar l'algorisme de Newton és el càlcul de la direcció de descens. S'ha programat el càlcul de la Hessiana i la seva inversa en el cas $n = 2$, però no s'han obtingut els resultats desitjats.

4.3 Resultats

Primer de tot, s'han generat alguns casos senzills per verificar que la programació dels mètodes era la correcta. Pel que fa al mètode de Weiszfeld s'ha vist que sí que es trobava el mínim en tots els casos, tant si era un vèrtex com si no. Pel que fa a la implementació del mètode de Newton s'ha detectat un error: el mètode no convergeix. Això deu ser degut a algun error en el programa que no s'ha pogut detectar. S'adjunta als annexos igualment el codi.

Per veure els resultats de la implementació del mètode de Weiszfeld s'ha usat un programa en llenguatge $C++$ per generar casos aleatoris diferents. S'han fixat la dimensió de l'espai $n = 2, 3, 4, 5$ i el nombre de vèrtexs $m = 10, 100, 1000$ i s'han generat 100 casos en cada cas. Els vèrtexs s'han escollitd e manera que cadascuna de les seves components es troba en l'interval $[0, 100]$ i el mateix amb els pesos. El nombre d'iteracions del mètode de Weiszfeld en cadascun dels casos esmentats es presenten a la taula següent.

	$m = 10$	$m = 100$	$m = 10^3$
$n = 2$	57	31	27
$n = 3$	36	19	17
$n = 4$	26	14	13
$n = 5$	20	12	11

Taula 1: Nombre d'iteracions de l'implementació del mètode de Weiszfeld modificat en funció de la dimensió de l'espai i el nombre de vèrtexs.

A primer cop d'ull, observem que tant a l'augmentar la dimensió de l'espai com a l'augmentar el nombre de vèrtexs, disminueix el nombre d'iteracions. Tot i això, comentar que encara de manera subtil, en augmentar el nombre de vèrtexs i d'iteracions el programa té un temps d'execució major. Això és degut al fet que les operacions triguen més quan hi ha més vèrtexs a calcular o aquests tenen més components. Així doncs, per poder fer una millor comparativa seria adient enregistrar els temps de computació, a més de les iteracions. Aquest temps s'hauria de calcular, també, sense tenir en compte el temps de lectura o escriptura de fitxers i directament el temps per iteració.

Malauradament, no s'ha pogut fer una bona implementació del mètode de Newton i comparar els resultats amb el mètode de Weber. De les referències [8] observem que el resultat esperat hauria sigut obtenir millor resultat en dimensions petites amb el mètode de Newton. El mètode de Newton convergeix ràpid però té un cost molt elevat de càlculs. Per això, s'encareix molt ràpid en dimensions més grans i sortiria a compte aplicar el mètode de Weiszfeld en aquests casos.

5 Conclusions

Al llarg d'aquest treball hem pogut donar una perspectiva històrica i una introducció matemàtica del problema de Weber, així com la seva rellevància avui en dia. Tot i que breu, hem vist una base d'optimització que ens ha sigut útil a l'estudiar el problema. Per una altra banda, hem explicat com s'ha desenvolupat el mètode de Weiszfeld i hem pogut donar una demostració rigorosa de les seves validesa i convergència. Finalment, hem pogut ampliar el mètode de Weiszfeld donant una modificació de millora, adaptar el mètode de Newton al problema de Weber i implementar l'algorisme de Weiszfeld.

No s'ha assolit l'objectiu d'implementar el mètode de Newton i, per tantm tampoc s'ha pogut fer una bona reflexió dels resultats del mètode. Del que sí que podem comentar és de la trajectòria que s'ha seguit en exposar el problema de Weber i el mètode de Weiszfeld, que ha sigut majoritàriament rigorosa i amb bons resultats.

Tot el treball s'ha fet pensant en la distància més habitual, la euclidiana, però a vegades un problema pot quedar millor representat amb una altra distància. Així que una perspectiva interessant a tenir en compte per continuar aquest treball seria estudiar la validesa dels resultats obtinguts canviant la norma escollida.

Una altra branca per la qual es podria tirar seria considerar generalitzacions del problema de Weber. Com s'ha comentat, aquest és un problema molt estudiat que ha donat peu a moltes variacions. Es podria, per exemple, veure com afecten pesos negatius al mètode de Weiszfeld.

A Annexos

A.1 Mètode de Weiszfeld modificat

```
#include<bits/stdc++.h>
using namespace std;


long double normaxy(vector<long double> x,
                    vector<long double> y,
                    int n){
    // Retorna un real: la norma euclideana de x-y.
    long double norma = 0;
    for(int i = 0; i < n; i++){
        norma = norma + (x[i] - y[i])*(x[i] - y[i]);
    }
    return sqrt(norma);
}

long double f(vector<long double> x,
              vector<vector<long double>> a,
              vector<long double> w, int n, int m){
    // Retorna f(x)
    long double fx = 0;
    for(int i = 0; i < m; i++){
        fx += w[i]*normaxy(x, a[i], n);
    }
    return fx;
}

vector<long double> Ri(int i, vector<vector<long double>> a,
                           vector<long double> w, int n, int m){
    // Retorna R_i del vertex a[i]
    vector<long double> zero(n, 0);
    vector<long double> R(n, 0);
    long double coef = 0;
    for(int j = 0; j < m; j++){
        if(j != i){
            coef = w[j]/normaxy(a[i], a[j], n);
            for(int k = 0; k < n; k++){
                R[k] += coef*(a[i][k]-a[j][k]);
            }
        }
    }
    return R;
}
```

```
vector<long double>
descens_vertex(int i,
               vector<vector<long double>> a,
               vector<long double> w, int n, int m){
    // Prenem un pas en la direccio d des del vertex a[i]
    // Calcul de R i la seva norma del vertex amb menor valor
    vector<long double> R = Ri(i, a, w, n, m);
    vector<long double> zero(n, 0);
    long double normaR = normaxy(R, zero, n);
    // Calcul direccio de descens d
    vector<long double> d(n, 0);
    for(int j = 0; j < n; j++) d[j] = -R[j]/normaR;
    // Calcul de L(a)
    long double L = 0;
    for(int j = 0; j < m; j++){
        if(j != i) L += w[j]/normaxy(a[i], a[j], n);
    }
    // Calcul de t usant L
    long double t = (normaR - w[i])/L;
    // Prenem un pas en la direccio d des del vertex
    vector<long double> x(n, 0);
    for(int j = 0; j < n; j++) x[j] = a[i][j] + t*d[j];

    return x;
}

vector<long double>
calcul_punt_inicial_weiszfeld(vector<vector<long double>> a,
                              vector<long double> w,
                              int n, int m){
    // Calcula una mitjana per trobar un bon punt inicial
    vector<long double> numerador(n, 0);
    long double denominador = 0;
    for(int i = 0; i < m; i++){
        denominador += w[i];
        for(int j = 0; j < n; j++){
            numerador[j] += a[i][j]*w[i];
        }
    }
    for(int i = 0; i < n; i++) numerador[i]/=denominador;
    return numerador;
}

vector<long double>
funcio_iteracio_T(vector<long double> x,
```

```
            vector<vector<long double>> a,
            vector<long double> w,
            int n, int m){
    // Retorna T(x)
    vector<long double> numerador(n, 0);
    long double denominador = 0;
    long double norma;
    for(int i = 0; i < m; i++){
        norma = normaxy(x, a[i], n);
        denominador += (w[i]/norma);
        for(int j = 0; j < n; j++){
            numerador[j] += a[i][j]*w[i]/norma;
        }
    }
    for(int i = 0; i < n; i++) numerador[i]/=denominador;
    return numerador;
}

vector<long double>
funcio_iteracio_S(int i,
            vector<vector<long double>> a,
            vector<long double> w, int n, int m){
    // S(ai) es fer un pas en la direcci  de descens
    cout << "Usem l'operador S." << endl;
    return descens_vertex(i, a, w, n, m);
}

vector<long double> gradf(vector<long double> x,
                    vector<vector<long double>> a,
                    vector<long double> w, int n, int m){
    // Retorna el grad f(x)
    vector<long double> grad(n, 0);
    long double coef = 0;
    for(int i = 0; i < m; i++){
        coef = w[i]/normaxy(x, a[i], n);
        for(int j = 0; j < n; j++){
            grad[j] += coef*(x[j]-a[i][j]);
        }
    }
    return grad;
}

int esVertex(vector<long double> x,
        vector<vector<long double>> a,
        int n, int m, long double tol){
    // Retorna l'index del vertex que coincideix amb x
```

```
    // si es un dels v rtexs
    for(int i = 0; i < m; i++){
        if(normaxy(x, a[i], n) < tol) return i;
    }
    return -1;
}

vector<long double>
metodeWeiszfeld(vector<long double> x,
                vector<vector<long double>> a,
                vector<long double> w, int n,
                int m, long double tol){
    // Aplica el m tode de Weiszfeld modificat
    // Condicions de parada: trobem un m nim.
    // Retorna el m nim
    int vertex;
    bool minim, minimvertex = false;
    int it = 0;
    vector<long double> zero(n,0);
    vertex = esVertex(x, a, n, m, tol);
    if(vertex < 0 and
        normaxy(gradf(x,a, w, n, m), zero, n) < tol){
            minim = true;
        }
    if(vertex >= 0 and
        normaxy(Ri(vertex,a,w,n,m), zero, n) <= w[vertex]){
            minimvertex = true;
    }
    while(!minim and !minimvertex){
        it += 1;
        if(vertex >= 0) x = funcio_iteracio_S(vertex,a,w,n,m);
        else x = funcio_iteracio_T(x,a,w,n,m);
        vertex = esVertex(x, a, n, m, tol);
        if(vertex < 0 and
            normaxy(gradf(x,a, w, n, m), zero, n) < tol){
                minim = true;
        }
        if(vertex >= 0 and
            normaxy(Ri(vertex,a,w,n,m), zero, n) <= w[vertex]){
                minimvertex = true;
        }
    }
    cout << "El metode de Weiszfeld ha acabat." << endl;
    cout << "Nombre d'iteracions: " << it << endl;
    if(minimvertex){
        cout << "El minim en un vertex!" << endl;
```

```
    }

    return x;
}

int metodeWeiszfeldIteracions(vector<long double> x,
                              vector<vector<long double>> a,
                              vector<long double> w, int n,
                              int m, long double tol){
    // Aplica el m tode de Weiszfeld modificat
    // Condicions de parada: trobem un m nim.
    // Retorna el nombre d'iteracions
    int vertex;
    bool minim, minimvertex = false;
    int it = 0;
    vector<long double> zero(n,0);
    vertex = esVertex(x, a, n, m, tol);
    if(vertex < 0 and
        normaxy(gradf(x,a, w, n, m), zero, n) < tol){
            minim = true;
    }
    if(vertex >= 0 and
         normaxy(Ri(vertex,a,w,n,m), zero, n) <= w[vertex]){
            minimvertex = true;
    }
    while(!minim and !minimvertex){
        it += 1;
        if(vertex >= 0) x = funcio_iteracio_S(vertex,a,w,n,m);
        else x = funcio_iteracio_T(x,a,w,n,m);
        vertex = esVertex(x, a, n, m, tol);
        if(vertex < 0 and
             normaxy(gradf(x,a, w, n, m), zero, n) < tol){
            minim = true;
        }
        if(vertex >= 0 and
            normaxy(Ri(vertex,a,w,n,m), zero, n) <= w[vertex]){
            minimvertex = true;
        }
    }
    return it;
}
```

A.2 Mètode de Newton

```
#include<bits/stdc++.h>
using namespace std;


long double normaxy(vector<long double> x,
                    vector<long double> y,
                    int n){
    // Retorna un real: la norma euclideana de x-y.
    long double norma = 0;
    for(int i = 0; i < n; i++){
        norma = norma + (x[i] - y[i])*(x[i] - y[i]);
    }
    return sqrt(norma);
}

long double f(vector<long double> x,
            vector<vector<long double>> a,
            vector<long double> w, int n, int m){
    // Retorna f(x)
    long double fx = 0;
    for(int i = 0; i < m; i++){
        fx += w[i]*normaxy(x, a[i], n);
    }
    return fx;
}

vector<long double> Ri(int i, vector<vector<long double>> a,
                        vector<long double> w, int n, int m){
    // Retorna R_i del vertex a[i]
    vector<long double> zero(n, 0);
    vector<long double> R(n, 0);
    long double coef = 0;
    for(int j = 0; j < m; j++){
        if(j != i){
            coef = w[j]/normaxy(a[i], a[j], n);
            for(int k = 0; k < n; k++){
                R[k] += coef*(a[i][k]-a[j][k]);
            }
        }
    }
    return R;
}

vector<long double>
```

```
descens_vertex(int i,
               vector<vector<long double>> a,
               vector<long double> w, int n, int m){
    // Prenem un pas en la direccio d des del vertex a[i]
    // Calcul de R i la seva norma del vertex amb menor valor
    vector<long double> R = Ri(i, a, w, n, m);
    vector<long double> zero(n, 0);
    long double normaR = normaxy(R, zero, n);
    // Calcul direccio de descens d
    vector<long double> d(n, 0);
    for(int j = 0; j < n; j++) d[j] = -R[j]/normaR;
    // Calcul de L(a)
    long double L = 0;
    for(int j = 0; j < m; j++){
        if(j != i) L += w[j]/normaxy(a[i], a[j], n);
    }
    // Calcul de t usant L
    long double t = (normaR - w[i])/L;
    // Prenem un pas en la direccio d des del vertex
    vector<long double> x(n, 0);
    for(int j = 0; j < n; j++) x[j] = a[i][j] + t*d[j];

    return x;
}

vector<long double> gradf(vector<long double> x,
                          vector<vector<long double>> a,
                          vector<long double> w, int n, int m){
    // Retorna el grad f(x)
    vector<long double> grad(n, 0);
    long double coef = 0;
    for(int i = 0; i < m; i++){
        coef = w[i]/normaxy(x, a[i], n);
        for(int j = 0; j < n; j++){
            grad[j] += coef*(x[j]-a[i][j]);
        }
    }
    return grad;
}

int esVertex(vector<long double> x,
             vector<vector<long double>> a,
             int n, int m, long double tol){
    // Retorna l'index del vertex que coincideix amb x
    // si es un dels v rtexs
    for(int i = 0; i < m; i++){
```

```
        if(normaxy(x, a[i], n) < tol) return i;
    }
    return -1;
}

int menor_vertex(vector<vector<long double>> a,
                 vector<long double> w, int n, int m){
    // Retorna l'index i del v rtex que satisf
    // f(ai) <= f(aj) per tot j
    long double fj;
    long double min = f(a[0], a, w, n, m);
    int minIndex = 0;
    for(int j = 1; j < m; j++){
        fj = f(a[j], a, w, n, m);
        if(fj < min){
            min = fj;
            minIndex = j;
        }
    }
    return minIndex;
}

vector<long double>
calcul_punt_inicial_newton(vector<vector<long double>> a,
                           vector<long double> w, int n, int m){
    // Calcula el punt inicial tal i com s'ha vist a la memoria
    // Index i del vertex que satisf  f(ai) <= f(aj) per tot j
    int vertexIndex = menor_vertex(a,w,n,m);
    // Prenem el pas
    return descens_vertex(vertexIndex, a, w, n, m);
}

vector<vector<long double>>
Hessian(vector<long double> x,
        vector<vector<long double>> a,
        vector<long double> w, int n, int m){
    vector<vector<long double>> H(n, vector<long double>(n, 0));
    long double coef = 0;
    for(int i = 0; i < m; i++){
        long double norm = normaxy(x, a[i], n);
        coef = w[i]/(norm*norm*norm);
        for(int k = 0; k < n; k++){
            for(int j = 0; j <= k; j++){
                H[k][j] -= coef*(x[k] - a[i][k])*(x[j] - a[i][j]);
            }
            H[k][k] += coef*norm*norm;
```

```
        }
    }
    for(int k = 0; k < n; k++){
        for(int j = 0; j < k; j++) H[j][k] = H[k][j];
    }
    return H;
}

long double determinant(vector<vector<long double>> H, int n){
    if(n == 2) return H[0][0] * H[1][1] - H[0][1]*H[1][0];
    return 0;
}

vector<vector<long double>>
    inversaHessian(vector<vector<long double>> H, int n){
    long double det = determinant(H, n);
    cout << "DET" << det << endl;
    vector<vector<long double>> IH(n, vector<long double>(n, 0));
    if(n == 2){
        IH[0][0] = H[1][1]/det;
        IH[0][1] = IH[1][0] = -H[0][1]/det;
        IH[1][1] = H[0][0]/det;
    }
    return IH;

}

vector<long double>
calcul_direccio_newton(vector<long double> x,
                        vector<vector<long double>> a,
                        vector<long double> w, int n, int m){
    vector<long double> d(n,0);
    vector<long double> gradx = gradf(x,a,w,n,m);
    vector<vector<long double>> H(n, vector<long double>(n));
    H = Hessian(x,a,w,n,m);
    H = inversaHessian(H, n);
    for(int i = 0; i < n; i++){
        for(int j = 0; j < n; j++){
            d[i] += H[i][j]*gradx[j];
        }
    }

    return d;
}

long double
```

```
calcul_pas_newton(vector<long double> x,
                  vector<long double> d,
                  vector<vector<long double>> a,
                  vector<long double> w, int n, int m){
    long double ro = 0.5;
    long double sigma = 1e-4;
    long double t = 1;

    long double fx = f(x,a,w,n,m);

    vector<long double> gradx = gradf(x,a,w,n,m);

    long double gradd = 0;
    for(int j = 0; j < n; j++) gradd += gradx[j]*d[j];

    vector<long double> xk(n,0);
    for(int j = 0; j < n; j++) xk[j] = x[j] + t*d[j];

    double fxk = f(xk,a,w,n,m);
    int it = 0;
    while(fxk <= fx + sigma * t * gradd and it < 100){
        it++;
        t*=ro;
        fx = fxk;
        gradd = 0;
        gradx = gradf(xk,a,w,n,m);
        for(int j = 0; j < n; j++) gradd += gradx[j]*d[j];
        for(int j = 0; j < n; j++) xk[j] = xk[j] + t*d[j];
        fxk = f(xk,a,w,n,m);
    }
    return t;
}

vector<long double>
metodeNewton(vector<long double> x,
             vector<vector<long double>> a,
             vector<long double> w, int n, int m,
             long double tol){
    int it = 0;
    bool minim = false;
    vector<long double> zero(n,0);
    vector<long double> d(n,0);
    vector<long double> sol(2,0);
    sol[0] = 0.5;
    sol[1] = 1;
    long double t;
```

```
    if(normaxy(gradf(x,a, w, n, m), zero, n) < tol){
        minim = true;
    }
    while(!minim and it < 2){
        cout << it << endl;
        cout << "DIST" << normaxy(x,sol,n ) << endl;
        it += 1;
        d = calcul_direccio_newton(x,a,w,n,m);
        t = calcul_pas_newton(x,d,a,w,n,m);
        for(int j = 0; j < n; j++) x[j] = x[j] + t*d[j];
        if(normaxy(gradf(x,a, w, n, m), zero, n) < tol){
            minim = true;
        }
    }
    cout << "El metode de Newton ha acabat." << endl;
    cout << "Nombre d'iteracions: " << it << endl;
    return x;
}
```

Bibliografia

[1] Stephen Abbott. *Understanding analysis*, chapter 2: Sequences and Series. Springer, second edition, 2015.

[2] Amir Beck and Shoham Sabach. Weiszfeld's method: Old and new results. *Journal of Optimization Theory and Applications*, 164(1):1–40, 2015.

[3] Dimitri Bertsekas. *Convex optimization theory*, volume 1, chapter 1: Basic Concepts of Convex Analysis. Athena Scientific, 2009.

[4] Dimitri Bertsekas. *Nonlinear Programming*. Athena Scientific, third edition, 2016.

[5] Roberto J. Cañavate Bernal, María Belén Cobacho Tornel, and José Miguel Rodríguez Gómez. El algoritmo de weiszfeld para la resolución del problema económico de weber. 2002.

[6] Richard Courant, Herbert Robbins, and Ian Stewart. *What is Mathematics?: an elementary approach to ideas and methods*. Oxford University Press, USA, second edition, 1996.

[7] Horst A Eiselt and Vladimir Marianov. Pioneering developments in location analysis. In *Foundations of location analysis*, pages 3–22. Springer, 2011.

[8] Simone Görner and Christian Kanzow. On newton's method for the fermat–weber location problem. *Journal of Optimization Theory and Applications*, 170(1):107–118, 2016.

[9] Harold W Kuhn. A note on fermat's problem. *Mathematical programming*, 4(1):98–107, 1973.

[10] Harold W Kuhn and Robert E Kuenne. An efficient algorithm for the numerical solution of the generalized weber problem in spatial economics. *Journal of Regional Science*, 4(2):21–33, 1962.

[11] Yaakov S Kupitz, Horst Martini, and Margarita Spirova. The fermat–torricelli problem, part i: A discrete gradient-method approach. *Journal of Optimization Theory and Applications*, 158(2):305–327, 2013.

[12] Frank Plastria. The weiszfeld algorithm: proof, amendments, and extensions. In *Foundations of location analysis*, pages 357–389. Springer, 2011.

[13] Endre Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, First Series*, 43:355–386, 1937.

[14] Endre Weiszfeld and Frank Plastria. On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, 167(1):7–41, 2009.