Optimització

Ricard Torres Bargalló

Exercicis a càrrec de
Margarida Corominas Bosch
Anna Espinal Berenguer

PID_00186429

Universitat Oberta de Catalunya

Índex

Introducció

A la part d'aquest curs dedicada al càlcul univariant, ens hem trobat amb els conceptes bàsics de la teoria de l'optimització. Aquí en farem un tractament més sistemàtic. La teoria de l'optimització s'ocupa de tot el que té relació amb la cerca de valors màxims i mínims de funcions. Nosaltres classificarem aquí l'optimització en tres grans apartats:

1. Optimització sense restriccions.

2. Amb restriccions d'igualtat.

3. Amb restriccions de desigualtat.

El criteri de classificació depèn de les **restriccions** que té un determinat problema, que són condicions, en forma d'igualtat o de desigualtat, que han de complir les variables. Establim aquesta classificació perquè les tècniques de resolució que usarem canvien segons el tipus de problema. El nostre enfocament tracta més aviat de fer ressaltar la lògica comuna que hi ha en els diferents mètodes emprats.

Objectius

Al llarg d'aquest mòdul podreu assolir els objectius següents:

1. Aprendre a trobar analíticament la solució de problemes d'optimització, tant amb restriccions com sense.

2. Conèixer la regla dels multiplicadors de Lagrange per tal de resoldre problemes amb restriccions d'igualtat, i saber com es formulen les condicions de Kuhn i Tucker, que descriuen les solucions de problemes en què les restriccions són de desigualtat.

3. Conèixer les condicions de segon ordre per a determinar el caràcter local dels punts crítics, però sobretot les condicions de concavitat i/o convexitat que asseguren el caràcter global de les solucions.

4. Saber justificar de manera intuïtiva i gràfica els mètodes emprats, perquè, davant un problema que no s'emmotlla exactament a la teoria que heu après, sigueu capaços de tenir criteri per tal de raonar per què un cert mètode pot ser vàlid per a trobar la solució o per què no ho és.

1. Conceptes bàsics

En moltes circumstàncies de la nostra vida ens trobem que hem de triar, d'entre un nombre donat d'alternatives, aquella que ens interessa més. Suposem que, davant d'un problema d'aquest tipus, podem associar a cada una de les alternatives possibles un índex numèric que expressa la seva desitjabilitat relativa, de tal manera que, per saber quina de dues alternatives és millor, no hem de fer res més que comparar la magnitud dels seus índexs respectius. Ens trobem, en aquest cas, amb un problema d'optimització. El problema és de **maximització** si valors més grans de l'índex indiquen alternatives millors, i és de **minimització** quan el que preferim són valors més petits de l'índex. !

Exemple 1.1. Suposem que fem correspondre 0 a un suspès, 1 a un aprovat just, 2 a un aprovat, 3 a un notable i 4 a un excel·lent. Si volem trobar aquell estudiant que té el millor expedient de la UOC, només hem d'associar a cada estudiant la mitjana de les seves notes i resoldre el corresponent problema de maximització.

Per cert,...

... aquesta associació de nombres a notes no és pas arbitrària. Si algú de vosaltres s'anima, quan acabi els estudis, a fer un màster o un doctorat en un país com ara els Estats Units, veurà que a la sol·licitud d'admissió al programa li demanen un índex de notes com aquest.

Exemple 1.2. En Pepet és un gran afeccionat als cotxes potents però, fent honor al país, és més aviat del puny estret. Està planejant comprar-se un cotxe, i per això buscarà aquell pel qual sigui més baix el quocient entre el preu de compra i la potència (mesurada en CV). El que farà és resoldre un problema de minimització.

Exemple 1.3. Els serveis de meteorologia enregistren cada any la pluviositat total que hi ha hagut a tota una sèrie d'observatoris que tenen repartits pel territori. Trobar el lloc on ha plogut més és un problema de maximització, i el lloc on ha plogut menys un de minimització. El mateix s'aplica al registre d'altres magnituds com ara les temperatures.

Exemple 1.4. D'entre tots els rectangles que tenen un perímetre donat (diem-ne 1), volem trobar aquell que té l'àrea màxima. Això és un problema de maximització. Més endavant veurem com el podem resoldre. Observem, de moment, que hi ha una diferència essencial entre aquest exemple i els anteriors que hem vist: en els altres havíem de triar entre un nombre finit d'alternatives i, per tant, la seva resolució no requeria cap tècnica més sofisticada que la d'anar comparant alternatives, i armar-se de paciència si n'hi havia moltes; en aquest exemple, en canvi, aquesta tècnica no ens portaria gaire lluny, ja que hi ha un nombre infinit d'alternatives (la llargada de la base pot ser qualsevol nombre real entre 0 i 1/2).

Exemple 1.5. La Maria recull petxines a la platja i després les ven als turistes que hi passegen. Amb el temps, ha trobat que la quantitat de petxines que ven depèn del preu que en demana, d'acord amb la següent relació: $q = 100 - 2\,p$, en què q representa la quantitat i p el preu. La Maria voldria saber quin és el preu que hauria de posar per tal de tenir un ingrés més elevat. Aquí l'índex que ens permet comparar alternatives és l'ingrés, és a dir, el preu multiplicat per la quantitat:

$$\text{Ingrés} = I(p) = p\,q = p\,(100 - 2\,p) = 100\,p - 2\,p^2.$$

Si considerem que el preu pot ser qualsevol nombre real positiu, ens trobem de nou amb un problema en què hi ha un nombre infinit d'alternatives.

En la base de molts models econòmics que tracten de fer prediccions sobre el que passarà, hi ha el supòsit que els agents triaran, de totes aquelles alternatives que tenen a l'abast, aquella o aquelles que els donin resultats millors. Dit en altres paraules, aquests models econòmics suposen que els agents resolen algun tipus de problema d'optimització. Però l'optimització està present no sols en models abstractes, sinó en coses tan concretes com la planificació de la producció, de la política de finançament o de les estratègies de mercat d'una empresa.

Quan el nombre de possibles alternatives és finit, podem resoldre un problema d'optimització a base d'anar comparant alternatives i eliminant aquelles que són inferiors. El problema més difícil apareix qual el nombre d'alternatives és infinit. El que veurem aquí són una sèrie de tècniques, basades en la diferenciació de funcions, que serveixen per a aquest cas.

1.1. Definicions bàsiques. Existència de solucions

En un problema d'optimització distingirem tres elements:

1) El **conjunt factible** o **conjunt d'oportunitats**, que designarem aquí per la lletra X.

2) La **funció objectiu**, una funció definida a X i que pren valors numèrics, $f : X \to \mathbb{R}$, que és un índex que associem a cada element de X.

3) Un criteri de decisió: buscarem aquell element o elements de X que tenen associat un índex més gran (problemes de **maximització**) o bé un índex més petit (problemes de **minimització**).

Simbòlicament, representarem els problemes de maximització i de minimització com:

$$\begin{aligned} \max_{x} \;\; & f(x) \\ \text{s.a.} \;\; & x \in X \end{aligned} \qquad \text{i} \qquad \begin{aligned} \min_{x} \;\; & f(x) \\ \text{s.a.} \;\; & x \in X, \end{aligned}$$

en què les lletres *s.a.* volen dir "subjecte a". A sota del símbol **max** o **min** escrivim la variable x, que és la variable respecte a la qual estem optimitzant, i que anomenem **variable de decisió** o **variable de control**.

Exemple 1.6. Recordem l'exemple 1.4, en el qual ens plantejàvem trobar el rectangle d'àrea màxima de tots aquells que tenen perímetre 1. Definim x com la llargada de la base i y com l'alçària d'un rectangle. Llavors el conjunt factible del nostre problema d'optimització és

$$X = \left\{ (x, y) \in \mathbb{R}^2 \; : \; x \geq 0, \;\; y \geq 0, \;\; \text{i} \;\; 2x + 2y = 1 \right\}$$

i la funció objectiu (l'àrea) és $f(x, y) = x\,y$. En lloc de definir X separadament, solem escriure aquest problema d'optimització de la manera següent:

$$\begin{aligned} \max_{(x,y)} \;\; & x\,y \\ \text{s.a.} \;\; & 2\,x + 2\,y = 1 \\ & x \geq 0 \\ & y \geq 0. \end{aligned}$$

En aquesta formulació, el conjunt factible X està descrit per una sèrie de **restriccions** que imposem a les variables de decisió.

Exemple 1.7. Recordem ara l'exemple 1.5, en què volem trobar aquell preu p que fa més gran l'ingrés $I(p) = 100\,p - 2\,p^2$. La variable de decisió és el preu p, la funció objectiu és l'ingrés $I(p)$, i el conjunt factible, tots els preus possibles. Per exemple, com que la Maria vol cobrar per les petxines i no pas pagar-ne diners, podem suposar que el conjunt factible són tots els nombres reals no negatius.

Exercici

1.1. Una empresa usa dos factors de producció i produeix un únic bé. Si usa quantitats x_1 i x_2 dels dos factors, obté una quantitat $f(x_1, x_2)$ del bé produït. Els preus dels factors de producció són 2 i 5 respectivament, de tal manera que el cost d'usar unes quantitats (x_1, x_2) dels factors és $2\,x_1 + 5\,x_2$. L'empresa vol trobar aquella combinació (x_1, x_2) que fa més petit el cost de produir 10 unitats del producte. Identifiqueu les variables de decisió, la funció objectiu i el conjunt factible.

El primer que hem de tenir en compte és que un problema d'optimització, per senzill que sigui, no necessàriament té solució.

Exemple 1.8. Considerem el problema

$$\begin{aligned} \max_x \;& x \\ \text{s.a.}\;& x \in \mathbb{N}, \end{aligned}$$

en què $\mathbb{N}$ són els nombres naturals. En altres paraules, volem trobar el més gran de tots els nombres naturals (1, 2, 3, 4...), i aquest problema no té solució.

Diem que el problema

$$\begin{aligned} \max_x \;& f(x) \\ \text{s.a.}\;& x \in X \end{aligned}$$

té solució si hi ha un cert element x^* en el conjunt X tal que es compleix que $f(x^*) \geq f(x)$, per a tot x de X. En aquest cas diem que x^* és un **maximitzador** del problema, i que $f(x^*)$ és el **màxim** o **valor màxim** del problema.

Diem que el problema

$$\begin{aligned} \min_x \;& f(x) \\ \text{s.a.}\;& x \in X \end{aligned}$$

té solució si hi ha un cert element x^* en el conjunt X tal que es compleix que $f(x^*) \leq f(x)$, per a tot x de X. En aquest cas diem que x^* és un **minimitzador** del problema, i que $f(x^*)$ és el **mínim** o **valor mínim** del problema.

Un fet important que hem d'observar és que, quan un problema de maximització té solució, no necessàriament hi ha un únic maximitzador. El que sí que és necessàriament únic és el valor màxim: si hi hagués dos màxims diferents, sempre podríem quedar-nos amb el més gran dels dos, la qual cosa vol dir que l'altre no pot ser en realitat cap màxim. L'exemple següent ho il·lustra .

Exemple 1.9. Sigui $X = \{ -2, -1, 0, 1 \}$, i sigui $f(x) = x^2$, per a tot $x \in X$. Suposem que volem maximitzar f sobre X. Clarament, el problema té solució: el maximitzador és $x^* = -2$ i el valor màxim és $f(-2) = 4$.

Suposem ara que $X = \{ -2, -1, 0, 1, 2 \}$, i que $f(x) = x^2$, per a tot $x \in X$. Ara el problema té dos maximitzadors, $x^* = -2$ i $x^{**} = 2$, i el valor màxim continua sent 4. Observem que $f(2) = f(-2) = 4$.

El fet que un problema de maximització tingui solució o no en tingui, depèn de si el **recorregut** de la funció f té un element màxim o no. Recordem que, donada una funció $f : X \to \mathbb{R}$, el seu recorregut està definit com:

$$f(X) := \{ c \in \mathbb{R} : \text{ per algun } x \in X,\ f(x) = c \} .$$

Exemple 1.10. Uns quants recorreguts

- Si $X = \{ -2, -1, 0, 1, 2 \}$ i $f(x) = x^2$, llavors $f(X) = \{ 0, 1, 4 \}$.
- Si $X = [0, 2]$ i $f(x) = x^2$, llavors $f(X) = [0, 4]$.
- Si $X = [-2, 2]$ i $f(x) = x^2$, llavors $f(X) = [0, 4]$.
- Si $X = \mathbb{R}$ i $f(x) = e^x$, llavors $f(X) = (0, \infty)$.
- Si $X = (0, \infty)$ i $f(x) = \log(x)$, llavors $f(X) = \mathbb{R}$.

El següent lema resulta simplement d'interpretar el que significa la definició que hem donat de quan un problema de maximització (o de minimització) té solució.

Lema

El problema

$$\begin{aligned} &\max_x \ f(x) \\ &\text{s.a. } x \in X \end{aligned}$$

té solució si, i només si, el conjunt $f(X)$ (el recorregut de f) té un element màxim.

El problema

$$\begin{aligned} \min_x \;& f(x) \\ \text{s.a. } & x \in X \end{aligned}$$

té solució si, i només si, el conjunt $f(X)$ (el recorregut de f) té un element mínim.

Recordarem breument que, si A és un subconjunt de $\mathbb{R}$, diem que a^* és l'**element màxim** de A si $a^* \in A$ i, per a cada $a \in A$, es compleix $a^* \geq a$. Anàlogament, diem que a^* és l'**element mínim** de A si $a^* \in A$ i, per a cada $a \in A$, es compleix $a^* \leq a$. ❗

Una condició necessària, encara que no suficient, perquè un conjunt $A \subset \mathbb{R}$ tingui un màxim, és que A estigui fitat superiorment. Per veure que la condició no és suficient, considereu l'exemple següent: l'interval semiobert $[0, 1)$ és fitat superiorment però no té un màxim, ja que el nombre 1 no és dins l'interval, però qualsevol nombre positiu més petit que 1 sí que hi és (per exemple 0.9999999 amb tants nous com vulgueu).

Una condició suficient, però no necessària, perquè A tingui un element màxim i també un element mínim és que A sigui **compacte**, és a dir, fitat (superiorment i inferiorment) i tancat. Per veure que la condició no és necessària, considereu l'exemple següent: el conjunt $A = [0, 1) \cup 2$ (és a dir, el conjunt format per l'interval semiobert $[0, 1)$ i el nombre 2), té un màxim (el nombre 2), però no és compacte, ja que no és tancat.

Uns quants exemples mostraran diferents tipus d'inconvenients que es poden presentar en problemes d'optimització.

Exemple 1.11. El recorregut no és fitat

Considerem el problema

$$\begin{aligned} \max_x \;& x^2 \\ \text{s.a. } & x \in \mathbb{R}. \end{aligned}$$

Donat qualsevol nombre M arbitràriament gran, sempre podem trobar un $x \in \mathbb{R}$ tal que $x^2 > M$. Això implica que el problema no té solució. Fixeu-vos que aquí la pega és que el conjunt $f(X) = [0, \infty)$ no és fitat superiorment i, en conseqüència, no té un màxim. D'altra banda, observeu que el recorregut és en aquest cas fitat per sota, però això no té rellevància si volem maximitzar la funció.

El fet de restringir el conjunt factible a un interval fitat no necessàriament soluciona el problema.

Exemple 1.12. El recorregut és fitat, però no té màxim

Considerem

$$\begin{aligned} \max_{x} \;\; & x^2 \\ \text{s.a.} \;\; & x \in (0,1), \end{aligned}$$

en què el recorregut és $f(X) = (0,1)$, que no té màxim. Podem trobar punts $x \in X$ tals que $f(x)$ és arbitràriament proper a 1, però no hi ha cap punt que tingui 1 com a imatge.

El problema tampoc no se soluciona necessàriament tancant el domini X, a més de prendre'l fitat.

Exemple 1.13. El conjunt factible és fitat, però el recorregut no ho és

Considerem el problema

$$\begin{aligned} \max_{x} \;\; & f(x) \\ \text{s.a.} \;\; & x \in [0,1], \end{aligned}$$

en què la funció f és definida per

$$f(x) = \begin{cases} \frac{1}{x}\,, & \text{si } x > 0; \\ 0\,, & \text{si } x = 0. \end{cases}$$

En aquest cas, donat qualsevol nombre M arbitràriament gran, sempre podem trobar un x suficientment petit que compleixi $\frac{1}{x} > M$. Per tant, el recorregut no és (superiorment) fitat i el problema no té solució.

La pega en aquest exemple...

... resideix en la falta de continuïtat de la funció f: si aquesta hagués estat contínua, llavors el recorregut hauria estat fitat.

El problema de donar condicions que garanteixin l'existència d'una solució a un problema de maximització és similar, com hem vist, al problema d'assegurar l'existència d'element màxim a un conjunt de nombres reals: és fàcil descriure condicions *suficients* per tal que la solució existeixi, però hi ha molts casos en què una solució existeix encara que aquelles condicions no siguin satisfetes. Els exemples que acabem de discutir són il·lustratius del tipus de problemes que es poden presentar quan les condicions (suficients) del teorema següent són violades.

Teorema de Weierstrass

Si el conjunt factible X és fitat i tancat (compacte) i la funció objectiu f és contínua, llavors tant el problema de maximitzar f sobre X com el de minimitzar-la tenen solució.

Fixeu-vos que aquest teorema expressa condicions en termes del conjunt factible X i la funció objectiu f, que són en general molt més fàcils de verificar que requeriments sobre el recorregut $f(X)$, que sabem que és el que finalment importa. Aquest és un motiu addicional que fa que molts problemes tinguin solució encara que no compleixin les condicions del teorema de Weierstrass. Un altre motiu és que el teorema val tant per a problemes de maximització com per a problemes de minimització.

El motiu pel qual aquest teorema és cert és ben senzill: sempre que els dos requeriments siguin satisfets, el recorregut $f(X)$ també és un conjunt tancat i fitat (compacte) i, per tant, té un màxim i un mínim. La utilitat del teorema és relativa, ja que ens diu que existeix una solució del problema d'optimització, però no ens diu com la podem trobar.

Una aplicació del teorema de Weierstrass seria que, sempre que X sigui un conjunt finit, qualsevol problema d'optimització que tingui X com a conjunt factible té solució. Això és degut al fet que tot conjunt finit és compacte, i que tota funció definida sobre un conjunt finit és (trivialment) contínua sobre aquest conjunt de punts (no té per què ser contínua sobre un domini més gran, per això).

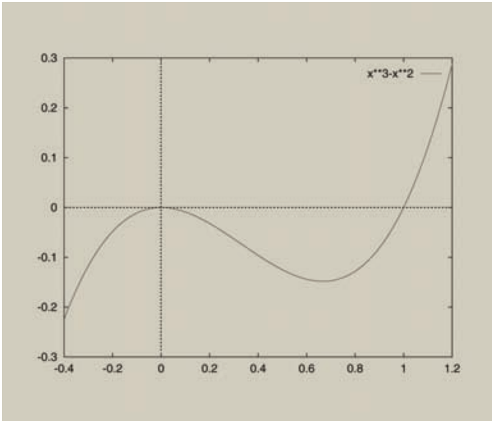
Un motiu important pel qual ens interessa tenir un resultat general que ens digui quan existeix una solució és que, si un problema no té solució però nosaltres apliquem tècniques de resolució com si en tingués, podem obtenir una resposta que ens sembli coherent i no adonar-nos de l'error que hem fet.

Exemple 1.14. Suposem que volem resoldre el programa

$$\begin{aligned} &\max_x \; x^3 - x^2 \\ &\text{s.a. } \; x \in \mathbb{R} \end{aligned}$$

i que sabem que, quan el problema té solució, la derivada de la funció objectiu en el maximitzador és igual a zero. Igualant la derivada de la funció objectiu a zero, obtenim l'equació $3\,x^2 - 2\,x = 0$, que té per solucions els punts $x^* = 0$ i $x^{**} = \frac{2}{3}$. Aplicant el criteri de la derivada segona, veurem que x^* és un maximitzador local i que x^{**} és un minimitzador local. Per tant, l'únic candidat a maximitzador és el punt $x^* = 0$. Si ara concloem que x^* és la solució del problema que havíem plantejat, estarem fent un error considerable, ja que en aquest cas el recorregut no és fitat superiorment ni inferiorment.

Una ullada a la gràfica de la funció ens pot ajudar a veure per què: la funció no és fitada superiorment ni inferiorment.



Gràfica de la funció $f(x) = x^3 - x^2$.

La funció $f(x) = x^3 - x^2$...

... té un màxim i un mínim locals, però no té cap màxim ni cap mínim globals.

Acabem aquest apartat amb un exemple que ens mostra que cap de les dues condicions que s'imposen en el teorema de Weierstrass no és necessària.

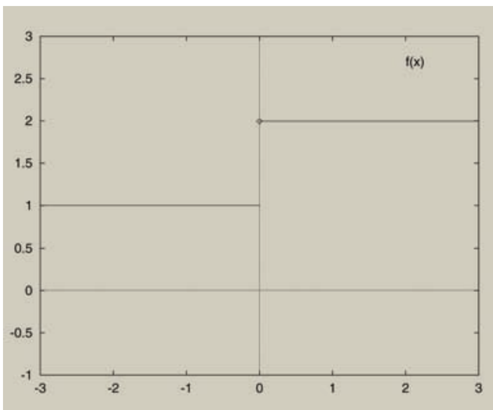
Exemple 1.15. Considereu el problema

$$\begin{array}{ll} \max\limits_{x} & f(x) \\ \text{s.a.} & x \in \mathbb{R}, \end{array}$$

en què la funció f és definida per

$$f(x) = \begin{cases} 1\,, & \text{si } x < 0; \\ 2\,, & \text{si } x \geq 0. \end{cases}$$

Aquí el valor màxim és 2 i qualsevol x en l'interval $[0, \infty)$ és un maximitzador, però f no és contínua i X no és fitat.



Una funció constant a trossos

Heus ací...

... una funció amb molts maximitzadors i molts minimitzadors, però que no compleix les condicions del teorema de Weierstrass.

1.2. Resolució de problemes. Algorismes

En problemes d'optimització no ens interessa només saber que hi ha una solució, sinó trobar-la i descriure'n les propietats. En aquest curs analitzarem fonamentalment una sèrie de tècniques que s'apliquen quan en els problemes d'optimització es donen unes certes condicions de diferenciabilitat que permeten descriure la solució mitjançant un sistema d'equacions (o de desigualtats).

En problemes pràctics d'optimització, és difícil que aquests requeriments siguin satisfets, així que probablement la resolució del problema es fa a través d'un ordinador. En aquest cas, haurem de programar en l'ordinador una seqüència de passos que esperem que ens acabin portant a la solució del problema o ens en donin una bona aproximació.

Un **algorisme** és un conjunt de regles que descriuen els passos que cal seguir per tal de buscar (o aproximar) la solució d'un problema determinat.

Exercici

1.2. Sigui $X = \{1, 2 \ldots n\}$ un conjunt finit, i sigui $f : X \to \mathbb{R}$ una funció qualsevol. Descriviu (de tal manera que poguéssim programar-ho en un ordinador) un algorisme que permeti trobar el màxim de la funció f sobre X.

Amb la (poca) informació que es dóna en el problema anterior, els algorismes que es poden definir passen tots per fer el que podem anomenar **enumeració** dels elements de X: bàsicament, hem d'anar agafant elements de X i comparar-los, i fins que no els haguem agafat tots no podrem saber amb seguretat quin és el màxim. Dit en altres paraules, quan X és finit, un problema d'optimització sempre té solució, però trobar-la pot ser molt costós: n pot ser molt gran. A tall d'exemple, suposem que hem de fer 2^{64} comparacions: si el nostre ordinador pot fer una comparació cada microsegon, tardaria uns 5.000 segles a obtenir el resultat final.

El nombre 2^{64}...

... està relacionat amb aquella llegenda d'un noble xinès que va demanar a un savi què volia en pagament d'un servei que aquest li havia fet, i el savi li va demanar que agafés un taulell d'escacs i posés un gra de blat a la primera casella i anés doblant el nombre de grans a cada casella successiva; el noble es va fregar les mans fins que va tractar de complir la promesa. El nombre total de grans de blat que hauria hagut de posar és $2^{64} - 1$. Podries dir per què?

En general, doncs, ens interessarà caracteritzar i explotar propietats que permetin simplificar els algorismes de solució de problemes. D'entre aquestes propietats veurem en aquest curs les de diferenciabilitat i concavitat/convexitat.

El gran èxit de la teoria de l'optimització ha estat el fet que es va descobrir un algorisme eficient per solucionar problemes de programació lineal, i que moltes situacions pràctiques que es presenten a diferents camps poden ser ben aproximades per programes lineals. Per exemple, la manera de determinar els horaris d'una companyia aèria.

El gran repte que hi ha en la branca algorítmica de la teoria de l'optimització avui en dia és el de desenvolupar algorismes eficients per a la resolució de problemes no lineals. En aquest curs no veurem aquesta vessant algorítmica de la teoria de l'optimització, que és més l'objecte d'un curs del que s'anomena Investigació Operativa.

Una anècdota...

... que il·lustra el que en el seu moment va significar el desenvolupament de la programació lineal és el fet que, durant la Segona Guerra Mundial, les tècniques de resolució de programes lineals als Estats Units eren un secret militar.

1.3. Diferenciació i optimització

A partir d'ara suposarem que la nostra variable de decisió és un vector n-dimensional, és a dir, que X és un subconjunt de $\mathbb{R}^n$. També suposarem en general que la funció objectiu f és diferenciable una vegada o més.

En aquest cas l'eina fonalmental per a trobar la solució del problema és la diferenciació. Com sabeu, la **diferenciació** és una manera d'aproximar localment entorn d'un cert punt una funció qualsevol mitjançant una funció lineal; això ens permet dir força coses sobre el comportament (a nivell local) de la funció original, en particular, si la funció tendeix a créixer o a decréixer en alguna direcció al voltant del punt de partida de l'aproximació.

La característica que fa que la diferenciació sigui tan important per resoldre problemes d'optimització és la relació que hi ha entre la derivada d'una funció (univariant) i el fet que aquesta sigui creixent o decreixent.

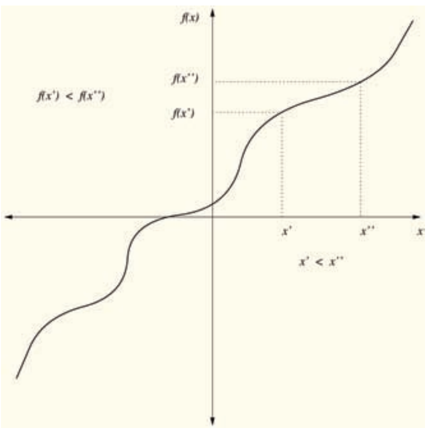
Dedicarem aquest apartat a fer un petit recordatori de la relació entre la derivada i el creixement d'una funció.

Sigui I un interval de nombres reals (on no descartem que $I = \mathbb{R}$), i sigui $f : I \to \mathbb{R}$ una funció.

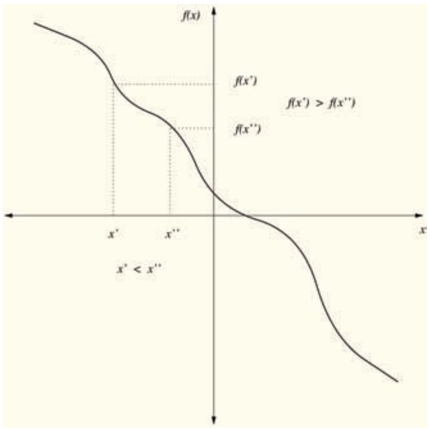
Diem que la funció f és **creixent** si $x' \leq x''$ implica que $f(x') \leq f(x'')$. Anàlogament, f és **decreixent** si $x' \leq x''$ implica que $f(x') \geq f(x'')$.

Diem que la funció f és **estrictament creixent** si $x' < x''$ implica que $f(x') < f(x'')$. Anàlogament, f és **estrictament decreixent** si $x' < x''$ implica que $f(x') > f(x'')$.

En altres paraules, una funció és estrictament creixent quan manté l'ordenació dels punts del domini, i estrictament decreixent quan la inverteix.

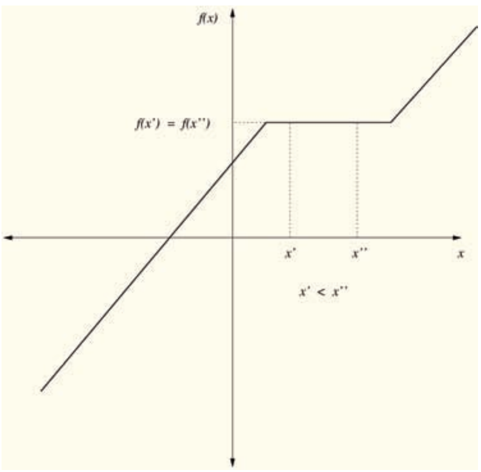


Una funció estrictament creixent



Una funció estrictament decreixent

La diferència entre una funció creixent en general i una de creixent *estrictament* és que la primera pot tenir trossos plans.



Creixent, però no estrictament

Observeu que...

... $x' < x''$, però $f(x') = f(x'')$. Això no passaria si la funció fos estrictament creixent.

Exemple 1.16. Tota funció constant és tant creixent com decreixent, però no ho és estrictament en cap sentit.

Recordem el teorema següent per a funcions diferenciables d'una variable.

Teorema del valor mitjà

Sigui $f : [a,b] \to \mathbb{R}$ una funció contínua i diferenciable a (a,b). Llavors hi ha un nombre c tal que $a < c < b$ i satisfà

$$f(b) - f(a) = f'(c)\,(b-a).$$

Els resultats que hi ha a continuació són conseqüència del teorema del valor mitjà.

Proposició 1. *Suposem que $f : I \to \mathbb{R}$ és diferenciable. Llavors tenim que:*

- *f és **creixent** a I si, i només si, $f'(x) \geq 0$ per a tot $x \in I$.*
- *f és **decreixent** a I si, i només si, $f'(x) \leq 0$ per a tot $x \in I$.*

Exemple 1.17. Si f és una funció constant, llavors la seva derivada és sempre 0. Per tant f és creixent. Els mateixos motius ens porten a afirmar que f és decreixent. De fet, les funcions constants són les úniques que poden ser creixents i decreixents alhora (no en sentit estricte, és clar).

La proposició anterior ens dóna una caracterització completa de funcions creixents o decreixents, en base a la seva derivada. A continuació veurem un altre resultat que, encara que no és sempre decisori, és extremament útil.

Proposició 2. *Suposem que $f : I \to \mathbb{R}$ és diferenciable. Llavors tenim que:*

- *Si $f'(x) > 0$ per a tot $x \in I$, llavors f és **estrictament creixent** a I.*
- *Si $f'(x) < 0$ per a tot x en I, llavors f és **estrictament decreixent** a I.*

Hi ha un cas força excepcional que és important destacar: una funció pot ser estrictament creixent o decreixent encara que la seva derivada sigui 0 en algun punt. Observem, però, que això només pot succeir en punts aïllats: si la derivada fos zero en un interval sencer, llavors la funció seria constant en aquest interval i, per tant, no podria ser creixent ni decreixent en sentit estricte. Un exemple n'és la funció $f(x) = x^3$, que és estrictament creixent, i la seva derivada $f'(x) = 3\,x^2$ és estrictament positiva a tot arreu, excepte en el punt $x = 0$, en què val 0. En general, però, és convenient associar intuïtivament una funció estrictament creixent amb una derivada positiva, i una d'estrictament decreixent amb una derivada negativa.

Exercici

1.3. En cada un dels casos que hi ha a continuació tenim que $f : I \to \mathbb{R}$. Digueu si la funció és creixent o decreixent sobre l'interval I en què està definida; si la vostra resposta és afirmativa, digueu si ho és en sentit estricte o no. Digueu també en cada cas quin és el recorregut J de la funció:

$$J = f(I) = \{\, y : \; y = f(x) \text{ per algun } x \in I \,\}\,.$$

Podeu recórrer a l'ajut de Gnuplot quan tingueu algun dubte.

a) $f(x) = x, \quad I = \mathbb{R}.$

b) $f(x) = -x, \quad I = \mathbb{R}.$

c) $f(x) = 2, \quad I = \mathbb{R}.$

d) $f(x) = 5 + 7\,x, \quad I = \mathbb{R}.$

e) $f(x) = 5 - 7\,x, \quad I = \mathbb{R}.$

f) $f(x) = x^2, \quad I = \mathbb{R}.$

g) $f(x) = x^2, \quad I = [0, \infty).$

h) $f(x) = -x^3, \quad I = \mathbb{R}.$

i) $f(x) = e^x, \quad I = \mathbb{R}.$

j) $f(x) = e^{-2\,x}, \quad I = \mathbb{R}.$

k) $f(x) = \log x, \quad I = (0, \infty).$

l) $f(x) = \log(1 + x^2), \quad I = \mathbb{R}.$

m) $f(x) = \log(1 + x^2), \quad I = [0, \infty).$

n) $f(x) = \frac{1}{x}, \quad I = (0, \infty).$

o) $f(x) = \sqrt{x}, \quad I = [0, \infty).$

p) $f(x) = \frac{1}{1+x^2}, \quad I = \mathbb{R}.$

q) $f(x) = \sin x, \quad I = \mathbb{R}.$

r) $f(x) = \sin x, \quad I = (0, \frac{\pi}{2}).$

s) $f(x) = \cos x, \quad I = \mathbb{R}.$

t) $f(x) = \cos x, \quad I = (0, \frac{\pi}{2}).$

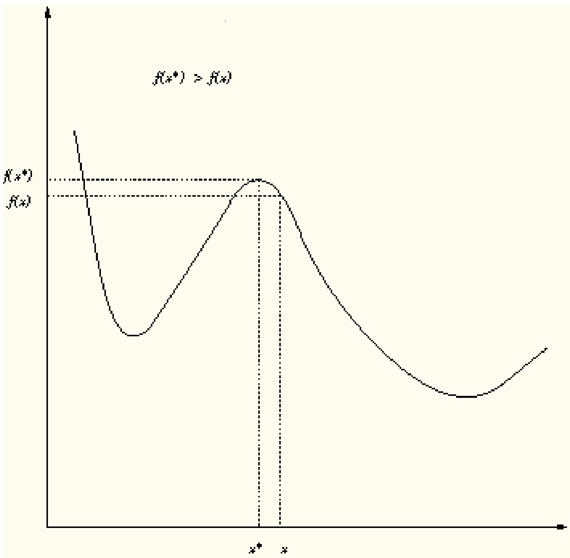
Fixeu-vos que, com que la diferenciació és una eina d'aproximació local d'una funció, tots els resultats que obtinguem a partir d'aquesta eina tindran només un caire local i, per tant, hem de tenir molta cura en la seva interpretació. En particular, en l'estudi de problemes d'optimització ens interessa introduir els conceptes següents:

Sigui $X \subset \mathbb{R}^n$ i $f : X \to \mathbb{R}$.

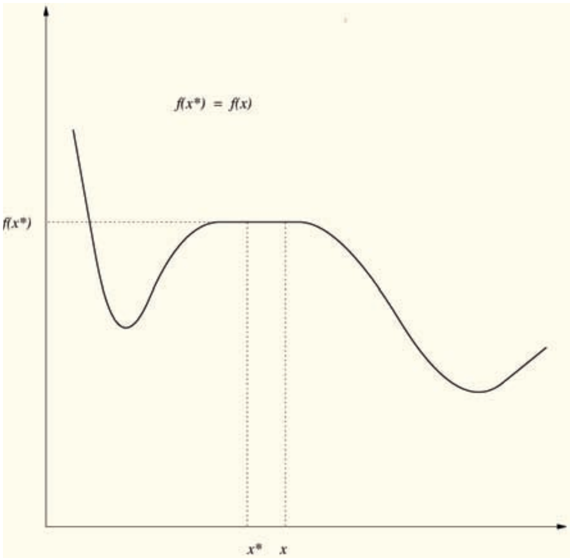
Diem que f ateny un **màxim local** a un punt $x^* \in X$ si hi ha un cert entorn de x^* tal que $f(x^*) \geq f(x)$, per a tot punt x d'aquell entorn. Si la desigualtat és estricta per a tot punt de l'entorn diferent de x^*, llavors diem que el màxim local és **estricte**.

Diem que f ateny un **mínim local** a un punt $x^* \in X$ si hi ha un cert entorn de x^* tal que $f(x^*) \leq f(x)$, per a tot punt x d'aquell entorn. Si la desigualtat és estricta per a tot punt de l'entorn diferent de x^*, llavors diem que el mínim local és **estricte**.

En un **màxim local estricte**, el valor de la funció en el maximitzador és estrictament més gran que en tots els punts del voltant. En canvi, en un **màxim local no estricte**, la funció pren un valor més gran o igual que en els punts del voltant.



Màxim local estricte



Màxim local no estricte

Per tal de fer la distinció entre els maximitzadors o minimitzadors locals i els maximitzadors o minimitzadors que solucionen un problema d'optimització, a vegades afegirem als darrers l'etiqueta de **globals**.

Alguns autors parlen d'extrems **relatius**, per designar els que nosaltres anomenem locals. Nosaltres preferim la terminologia de *locals versus globals*, més que la de *relatius versus absoluts*, perquè creiem que la primera és més descriptiva (a més, el seu ús s'està imposant cada vegada més).

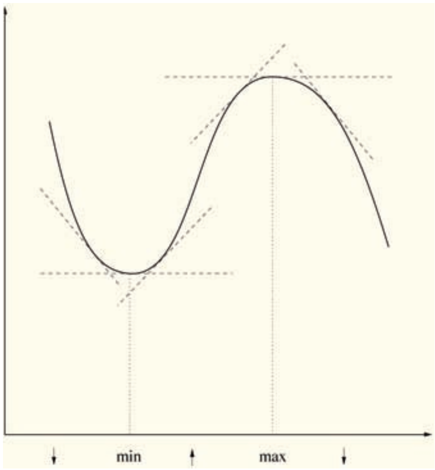
Usarem el terme *extrem*...

... per a referir-nos a un punt que pot ser tant un màxim com un mínim.

L'eina bàsica de detecció de màxims i mínims locals de funcions diferenciables d'una variable ja l'heu vista amb anterioritat, i consisteix en què en un extrem local la derivada de la funció ha de ser igual a zero. Acabem de veure que una funció creix quan la seva derivada és positiva i decreix quan aquesta és negativa. Com que en un mínim local la funció passa de decreixent a creixent, la derivada no pot ser ni negativa ni positiva, així que ha de ser zero. El mateix argument s'aplica als màxims locals, en què la funció passa de creixent a decreixent.

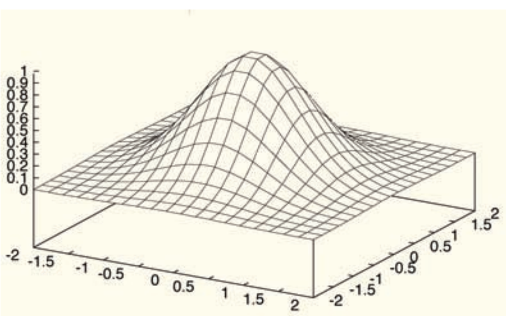
En un mínim local,...

... la funció passa de decreixent a creixent. En un màxim local, la funció passa de creixent a decreixent. Per tant, en ambdós punts la derivada ha de ser zero.

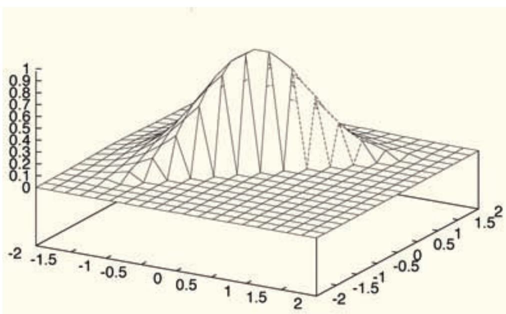


Màxims i mínims locals

En el cas de funcions multivariants, un resultat paral·lel és cert: en un extrem local totes les derivades parcials han de ser zero. El motiu és que, si la funció té un màxim local en un punt, llavors qualsevol secció vertical també ha de tenir un màxim local en el mateix punt, i les derivades parcials no són més que les derivades de les seccions verticals en les direccions dels eixos.



Màxim en dues variables



Secció vertical d'un màxim

Exercici

1.4. Feu la gràfica amb Gnuplot de la funció de dues variables

$$f(x, y) = e^{-x^2-y^2}$$

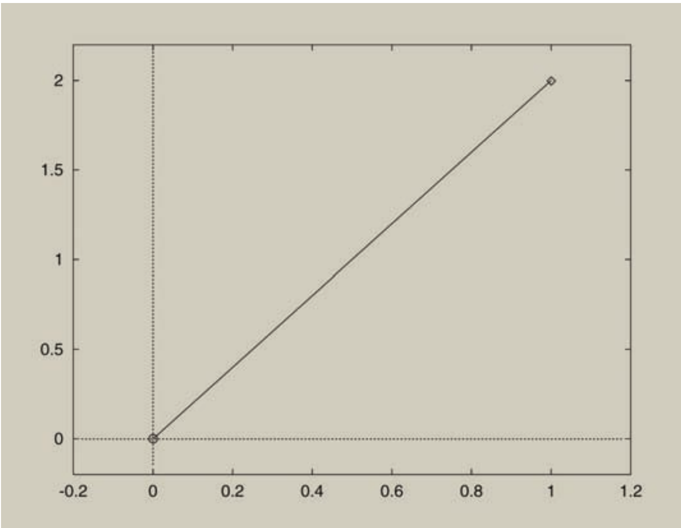
i comproveu que es correspon amb la que mostrem a dalt. En particular, la funció té un valor màxim al punt $(0, 0)$. A continuació feu seccions verticals que passin per l'origen i comproveu que totes tenen un màxim al mateix punt.

1.4. Tipus de programes d'optimització

El criteri que en un extrem les derivades parcials han de ser zero es basa, però, en un supòsit molt important: aquest extrem ha de ser un **punt interior**, és a dir, en qualsevol direcció a partir del punt hi ha d'haver altres que també siguin factibles. Quan aquest no sigui el cas, direm que estem tractant un problema **amb restriccions** i haurem de desenvolupar eines específiques per detectar els extrems locals.

Exemple 1.18. Un problema amb restriccions

Suposem que estem buscant els extrems locals de la funció $f(x) = 2\,x$, sobre el conjunt $X = [0, 1]$. Tenim que $f'(x) = 2$, per a tot x. Això semblaria indicar que la funció no té màxims ni mínims. Ara bé, una ullada a la gràfica hauria de bastar per a veure que f té un mínim (global) al punt $x = 0$, i un màxim (global) al punt $x = 1$.



Màxim i mínim restringits

Com que la funció...

... és estrictament creixent sobre $[0, 1]$, té un mínim a l'extrem inferior de l'interval i un màxim a l'extrem superior.

Observem que, si hi hagués d'altres punts a l'esquerra de $x = 0$, llavors aquest no seria un minimitzador; però com que estem imposant la restricció $x \geq 0$, llavors el fet que la derivada sigui positiva no impedeix que aquest punt sigui un minimitzador.

En particular, si el conjunt factible és un **conjunt obert**, tots els seus punts són interiors i, per tant, qualsevol maximitzador o minimitzador ha de satisfer la condició que les derivades parcials de la funció objectiu s'hi anul·len. Això motiva la definició següent.

Un problema d'**optimització sense restriccions** és aquell en què el conjunt factible és un conjunt obert.

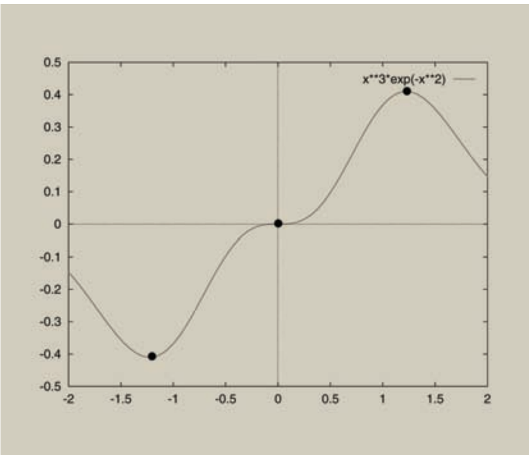
Ja hem vist anteriorment que, a més dels maximitzadors i minimitzadors locals del problema, hi pot haver altres punts en què les derivades parcials son iguals a zero.

Exemple 1.19. Extrems i punts d'inflexió

Sigui $f(x) = x^3\, e^{-x^2}$. En igualar la derivada a zero obtenim:

$$f'(x) = e^{-x^2}\,\left(3\,x^2 - 2\,x^4\right) = 0 \implies 3\,x^2 = 2\,x^4,$$

en què hi ha tres solucions: $x^* = \sqrt{\frac{3}{2}}$, $x^{**} = -\sqrt{\frac{3}{2}}$ i $x^{***} = 0$. Podeu comprovar ara que la funció és monòtona a les cues: si $x > x^*$ o $x < x^{**}$, llavors $3\,x^2 < 2\,x^4$ i, per tant, f és estrictament decreixent. A més, l'únic punt en què f talla l'eix de les x és quan $x = 0$. Tenint en compte el que acabem d'observar, la porció rellevant de la gràfica de la funció ens ajudarà a veure que la primera solució és el maximitzador (global), la segona el minimitzador (global), i la tercera no és ni maximitzador ni minimitzador, sinó un punt d'inflexió.



Gràfica de $x^3\, e^{-x^2}$

En la gràfica...

... podem visualitzar els tres punts crítics, que són, d'esquerra a dreta, un minimitzador, un punt d'inflexió i un maximitzador.

La tècnica que usarem en tots els problemes d'optimització que estudiarem (ja siguin sense restriccions o amb restriccions), té un tret comú: basant-nos en la diferenciació, formularem una sèrie de relacions que necessàriament ha de complir qualsevol punt que solucioni el problema d'optimització (si és que n'hi ha algun que ho faci). Anomenarem aquestes relacions **condicions de primer ordre**, ja que estan basades en les derivades de primer ordre. Els punts que solucionen les condicions de primer ordre són els **punts crítics**. Entre aquests punts crítics, n'hi pot haver alguns que solucionin el nostre problema d'optimització, i molts d'altres que no.

En els punts crítics que no són extrems, la funció necessàriament ha de créixer al llarg d'algunes direccions factibles i ha de decréixer al llarg d'altres. Quan no hi ha restriccions, i per tant totes les direccions són factibles, un punt crític que no és cap extrem és un **punt de sella** (la qual cosa no vol dir que necessàriament s'assembli a una sella de muntar).

Exercicis

1.5. Un punt de sella literal

Mostreu que $(2,3)$ és un punt crític de la funció $f(x,y) = -(x-2)^2 + (y-3)^2$. Useu Gnuplot per a generar la gràfica de la funció al voltant d'aquest punt, i comproveu a partir de la gràfica que es tracta d'un punt de sella.

1.6. Un punt de sella sense sella

També hi pot haver punts de sella amb els quals ens seria difícil muntar a cavall. Aquí en tenim un exemple. Mostreu, a partir de la seva derivada, que $(2,3)$ és un punt crític de la funció $f(x,y) = (x-2)^3 + (y-3)^3$. Useu Gnuplot per a generar la gràfica de la funció al voltant d'aquest punt. Comproveu en la gràfica que, a partir del punt, la funció creix en unes direccions i decreix en d'altres i que, per tant, es tracta d'un punt de sella, encara que la forma de la gràfica no s'assembli gaire a una sella de muntar.

Una qüestió que se'ns presenta és la de com es pot saber quan un punt crític és un maximitzador local, un minimitzador local o un punt de sella. En el cas d'una sola variable, el criteri que s'usa és el de mirar si els punts crítics satisfan les **condicions de segon ordre**, que són restriccions sobre la derivada segona: aquest criteri serà determinant sempre que la derivada segona tingui un signe estricte, i en cas contrari hem de continuar investigant. En el cas de més d'una variable, tenim criteris similars, en què en lloc de la derivada segona hem de considerar la matriu de derivades parcials segones. Més endavant ja veurem tot això amb detall.

Com hem vist, un problema d'optimització canvia radicalment quan hi ha punts des dels quals certes direccions són factibles i d'altres no ho són. En aquest cas, ja no és cert que un maximitzador o minimitzador local ha de tenir necessàriament totes les seves derivades parcials iguals a zero. Els problemes amb restriccions els classificarem en dues categories, segons si les restriccions estan formades per igualtats o per desigualtats.

Un problema d'**optimització amb restriccions d'igualtat** és aquell en què el conjunt factible està format per tots aquells punts que compleixen una sèrie de restriccions en forma d'igualtat (és a dir, que solucionen un sistema d'equacions).

Un problema d'**optimització amb restriccions de desigualtat** és aquell en què el conjunt factible està format per tots aquells punts que compleixen una sèrie de restriccions en forma de desigualtat (i potser també d'igualtat).

Els problemes...

... amb restriccions d'igualtat resultaran, en essència, molt semblants als problemes sense restriccions.

Exemple 1.20. Restriccions d'igualtat

El problema

$$\begin{aligned} \min_{(x,y)} \ & x^2 + y^2 \\ \text{s.a.} \ & x + y = 2 \end{aligned}$$

pot ser transformat en el problema (sense restriccions) amb una sola variable

$$\begin{aligned} \min_{x} \ & x^2 + (2-x)^2 \\ \text{s.a.} \ & x \in \mathbb{R}, \end{aligned}$$

però quan tractem aquest tipus de problemes veurem que hi ha una tècnica més poderosa que ens permet estalviar-nos aquesta substitució i trobar directament condicions que ha de satisfer qualsevol minimitzador local.

Quan tinguem problemes amb restriccions de desigualtat, la cosa serà una mica més complicada, ja que hem de ser capaços de determinar quines direccions són factibles i quines no ho són.

Exemple 1.21. Restriccions de desigualtat

El problema

$$\begin{aligned} \max_{x} \ & x^2 \\ \text{s.a.} \ & x \geq 0 \\ & x \leq 1 \end{aligned}$$

té un maximitzador global que és $x^* = 1$. Des de $x^* = 1$ és factible moure's cap al 0, però no cap a l'altra banda. En conseqüència, la condició que ha de satisfer la derivada de la funció objectiu a 1 per tal que f hi assoleixi un valor màxim no serà la igualtat a zero, sinó una desigualtat: sempre que la derivada sigui positiva, vol dir que, anant cap a 0, el valor de la funció objectiu disminuirà, així que ens trobarem sobre un maximitzador local.

En un problema amb restriccions de desigualtat, direm que una restricció és **efectiva** en un punt si aquest punt satisfà la restricció en forma d'igualtat. Per exemple, la restricció $x \leq 1$ és efectiva en el punt $x^* = 1$, però la restricció $x \geq 0$ no ho és. La tècnica que ens permetrà trobar la solució d'aquests problemes consisteix a escriure una sèrie de relacions que permeten investigar quines restriccions són efectives en l'òptim i quines no ho són.

En tots els tipus de problemes d'optimització que acabem de veure (amb restriccions o sense), la tècnica usada per a tractar de resoldre el problema és la diferenciació, perquè ens permet transformar el problema de trobar un punt òptim en el problema de resoldre un sistema d'equacions. Ara bé, essent la diferenciació una aproximació local, els resultats que obtenim de la seva aplicació també són de caire local, mentre que en els problemes d'optimització dels quals partim el que es busca és una solució global.

Les eines que ens permetran inferir el caire global de les solucions locals que hem trobat són dues:

a) el requeriment que el conjunt factible sigui convex,

b) El requeriment addicional que la funció objectiu tingui la curvatura (concavitat o convexitat) adequada segons que vulguem maximitzar o minimitzar.

En la part de càlcul univariant heu vist que una funció (diferenciable dues vegades) és convexa si, i només si, $f''(x) \geq 0$ a cada x, i que és còncava si la desigualtat oposada es compleix. Aquest resultat el generalitzarem al cas de diverses variables, fent servir matrius en lloc de nombres.

Exercicis

1.7. Considereu el problema

$$\begin{aligned} &\max_x \; 1 - x^2 \\ &\text{s.a. } x \in \mathbb{R}. \end{aligned}$$

a) Mostreu que l'únic punt crític és $x^* = 0$.

b) Usant Gnuplot, feu una gràfica de la funció al voltant de $x^* = 0$.

c) Mirant els signes de $f'(x)$, proveu que f és creixent a l'esquerra de 0 i decreixent a la seva dreta i que, per tant, 0 és un maximitzador global.

d) Una manera alternativa de veure que 0 és la solució del problema, és provar directament que, per a cada $x \in \mathbb{R}$, se satisfà $f(0) = 1 \geq 1 - x^2 = f(x)$. Feu-ho, i completeu l'argument.

1.8. Considereu el problema

$$\begin{aligned} &\min_x \; x \log x \\ &\text{s.a.} \;\; x \in (0,1), \end{aligned}$$

en què $x \in (0,1)$ vol dir que $0 < x < 1$; per tant, el conjunt factible és obert.

a) Usant Gnuplot, feu una gràfica de la funció.

b) Mostreu que l'únic punt crític és $x^* = \frac{1}{e}$.

c) Proveu que f és decreixent a l'esquerra de x^* i creixent a la seva dreta i que, per tant, x^* és un minimitzador global.

1.9. Considereu el problema

$$\begin{aligned} &\max_x \; \frac{1}{1+x^2} \\ &\text{s.a.} \;\; x \in \mathbb{R}. \end{aligned}$$

a) Mostreu que l'únic punt crític és $x^* = 0$.

b) Proveu que f és creixent a l'esquerra de 1 i decreixent a la seva dreta i que, per tant, 1 és un maximitzador global.

c) Proveu directament que, per a cada $x \in \mathbb{R}$, $\frac{1}{1+x^2} \le 1$.

d) Usant Gnuplot, feu una gràfica de la funció al voltant de $x^* = 1$.

1.5. Solucionari

1.1. Variables de decisió: quantitats dels factors de producció (x_1 i x_2).

Funció objectiu: cost de producció ($2\,x_1 + 5\,x_2$).

Conjunt factible: totes aquelles combinacions de factors, (x_1, x_2), que permeten produir 10 unitats:

$$\{\,(x_1, x_2) : \; f(x_1, x_2) \ge 10\,\}$$

1.2. L'algorisme consisteix a anar comparant els nombres successivament. Tenim un candidat a màxim que anem canviant a mesura que trobem nombres més grans. Comencem prenent $f(1)$ com a candidat a màxim. Llavors el comparem amb $f(2)$, i el més gran dels dos és el nou candidat a màxim, que comparem amb $f(3)$ i així successivament. El resultat de comparar el darrer candidat amb $f(n)$ ens donarà el màxim de la funció.

1.3.
a) $f(x) = x$, $\quad I = \mathbb{R}$: estrictament creixent, ja que, per a tot x, es compleix $f'(x) = 1 > 0$.

b) $f(x) = -x$, $\quad I = \mathbb{R}$: estrictament decreixent, ja que, per a tot x, es compleix $f'(x) = -1 < 0$.

c) $f(x) = 2$, $\quad I = \mathbb{R}$: constant (és a dir, és alhora feblement creixent i feblement decreixent).

d) $f(x) = 5 + 7\,x$, $\quad I = \mathbb{R}$: estrictament creixent, ja que, per a tot x, es compleix $f'(x) = 7 > 0$.

e) $f(x) = 5 - 7\,x$, $\quad I = \mathbb{R}$: estrictament decreixent, ja que, per a tot x, es compleix $f'(x) = -7 < 0$.

f) $f(x) = x^2$, $\quad I = \mathbb{R}$: no és creixent ni decreixent, ja que, per exemple, tenim $f'(-1) = -2 < 0 < 2 = f'(1)$.

g) $f(x) = x^2$, $\quad I = [0, \infty)$: estrictament creixent, ja que $f'(x) = 2\,x > 0$, per a tot $x > 0$.

h) $f(x) = -x^3$, $\quad I = \mathbb{R}$: estrictament decreixent, ja que $f'(x) = -3\,x^2 < 0$, per a tot $x \neq 0$.

i) $f(x) = e^x$, $\quad I = \mathbb{R}$: estrictament creixent, ja que, per a tot x, es compleix $f'(x) = e^x > 0$.

j) $f(x) = e^{-2\,x}$, $I = \mathbb{R}$: estrictament decreixent, ja que, per a tot x, es compleix $f'(x) = -2\,e^{-2\,x} < 0$.

k) $f(x) = \log x$, $I = (0, \infty)$: estrictament creixent, ja que, per a tot x, es compleix $f'(x) = 1/x > 0$.

l) $f(x) = \log(1 + x^2)$, $I = \mathbb{R}$: no és creixent ni decreixent, ja que, per exemple, tenim $f'(-1) = -1 < 0 < 1 = f'(1)$.

m) $f(x) = \log(1+x^2)$, $I = [0, \infty)$: estrictament creixent, perquè $f'(x) = \frac{2x}{1+x^2} > 0$, per a tot $x > 0$.

n) $f(x) = \frac{1}{x}$, $I = (0, \infty)$: estrictament decreixent, ja que, per a tot x, es compleix $f'(x) = -x^{-2} < 0$.

o) $f(x) = \sqrt{x}$, $I = [0, \infty)$: estrictament creixent, perquè $f'(x) = \frac{1}{2\sqrt{x}} > 0$, per a tot $x > 0$.

p) $f(x) = \frac{1}{1+x^2}$, $I = \mathbb{R}$: no és creixent ni decreixent, ja que, per exemple, tenim $f'(1) = -1/2 < 0 < 1/2 = f'(-1)$.

q) $f(x) = \sin x$, $I = \mathbb{R}$: no és creixent ni decreixent, ja que, per exemple, tenim $f'(\pi) = -1 < 0 < 1 = f'(0)$.

r) $f(x) = \sin x$, $I = (0, \frac{\pi}{2})$: estrictament creixent, perquè $f'(x) = \cos x > 0$, per a tot $0 \leq x < \pi/2$.

s) $f(x) = \cos x$, $I = \mathbb{R}$: no és creixent ni decreixent, ja que, per exemple, tenim $f'(\pi/2) = -1 < 0 < 1 = f'(3\pi/2)$.

t) $f(x) = \cos x$, $I = (0, \frac{\pi}{2})$: estrictament decreixent, perquè $f'(x) = -\sin x < 0$, per a tot $0 < x \leq \pi/2$.

1.4. La gràfica de la funció l'obtenim fent:

```
gnuplot> f(x,y) = exp(-x**2-y**2)

gnuplot> set isosamples 20

gnuplot> set hidden3d

gnuplot> splot [-2:2] [-2:2] f(x,y)
```

Per a obtenir el gràfic del tall vertical, fem:

```
gnuplot> g(x,y) = (x < y) ?  f(x,y) :  0

gnuplot> splot [-2:2] [-2:2] g(x,y)
```

1.5. Donada $f(x, y) = -(x - 2)^2 + (y - 3)^2$, les seves derivades parcials són

$$\frac{\partial f}{\partial x} = -2\,(x-2)$$
$$\frac{\partial f}{\partial y} = 2\,(y-3)$$

I és obvi que aquestes derivades parcials són zero al punt $(2, 3)$. Quan fixem $y = 3$, el tall vertical de la funció f que obtenim és $g(x) = -(x - 2)^2$, que té un màxim al punt $x = 2$. D'altra banda, si fixem $x = 2$, el tall vertical de la funció f que obtenim és $h(y) = (y-3)^2$, que té un mínim al punt $y = 3$. Per tant, es tracta d'un punt de sella.

Amb Gnuplot farem:

```
gnuplot> f(x,y)= - (x-2)**2 + (y-3)**2

gnuplot> set hidden3d

gnuplot> splot [1:3] [2:4] f(x,y)
```

Si ho volem veure millor, podem anar canviant la perspectiva usant la instrucció `set view`.

1.6. Donada $f(x,y) = (x-2)^3 + (y-3)^3$, les seves derivades parcials són

$$\frac{\partial f}{\partial x} = 3\,(x-2)^2$$
$$\frac{\partial f}{\partial y} = 3\,(y-3)^2$$

I és obvi que aquestes derivades parcials són zero al punt $(2,3)$. Quan fixem $y=3$, el tall vertical de la funció f que obtenim és $g(x) = (x-2)^3$, que és una funció estrictament creixent en x, per la qual cosa al punt $(2,3)$ la funció f no pot tenir-hi ni un màxim ni un mínim. Per definició, $(2,3)$ és un punt de sella.

Amb Gnuplot farem:

```
gnuplot> f(x,y)= (x-2)**3 + (y-3)**3

gnuplot> set hidden3d

gnuplot> splot [1:3] [2:4] f(x,y)
```

Podem observar que no hi ha cap semblança amb una sella de muntar.

1.7. Volem resoldre

$$\max_x \; 1 - x^2$$
$$\text{s.a. } x \in \mathbb{R}$$

Sigui $f(x) = 1 - x^2$, llavors els punts crítics solucionen l'equació

$$f'(x) = -2\,x = 0 \implies x = 0$$

A més, podem veure que f passa de ser creixent a decreixent a $x=0$, per la qual cosa es tracta d'un màxim (global):

$$x < 0 \implies f'(x) > 0, \qquad x > 0 \implies f'(x) < 0$$

Alternativament, comprovem que, per a cada x, $f(0) = 1 \geq 1 - x^2 = f(x)$, ja que $x^2 \geq 0$.

1.8. Tenim

$$\min_x \; x\,\log x$$
$$\text{s.a. } x \in (0,1)$$

Sigui $f(x) = x\,\log x$. Llavors els punts crítics han de complir:

$$f'(x) = 1 + \log(x) = 0 \implies x = e^{-1} \in (0,1)$$

També tenim

$$x < 1/e \implies f'(x) < 0, \qquad x > 1/e \implies f'(x) > 0$$

la qual cosa vol dir que $1/e$ és un minimitzador global.

1.9. Tenim

$$\max_x \; \frac{1}{1+x^2}$$
$$\text{s.a. } x \in \mathbb{R}$$

Sigui $f(x) = 1/(1+x^2)$. Per a obtenir els punts crítics fem:

$$f'(x) = -\frac{2\,x}{(1+x^2)^2} = 0 \implies x = 0$$

A més, podem veure que f passa de ser creixent a decreixent a $x = 0$, per la qual cosa es tracta d'un màxim (global):

$$x < 0 \implies f'(x) > 0, \qquad x > 0 \implies f'(x) < 0$$

Alternativament, tenim que, per a tot x, $f(0) = 1 \geq f(x)$, ja que $1 + x^2 \geq 1$ per a tot x real.

2. Optimització sense restriccions

2.1. Condicions de primer ordre

Hem definit un problema d'optimització sense restriccions com aquell en què la funció objectiu f és diferenciable una vegada o més i, addicionalment, el conjunt X és obert, és a dir, cada punt factible està rodejat d'altres punts que també ho són.

Exemple 2.1. Qualsevol problema en què el conjunt factible està format per tot l'espai n-dimensional (amb $n \geq 1$) és un problema sense restriccions. Per exemple:

$$\begin{aligned} \max_{(x,y)} \;& e^{-x^2} - y^4 \\ \text{s.a. } & (x,y) \in \mathbb{R}^2. \end{aligned}$$

Exemple 2.2. El problema

$$\begin{aligned} \max_{x} \;& x \log(x) \\ \text{s.a. } & x \in (0,1) \end{aligned}$$

és un problema sense restriccions, d'acord amb la definició que acabem de donar. El motiu és que si un punt satisfà $0 < x < 1$, llavors hi ha un cert nombre $\varepsilon > 0$ tal que l'interval $(x-\varepsilon, x+\varepsilon)$ està enterament contingut dins del conjunt factible. (Per exemple, podríem prendre ε com el més petit dels nombres $\frac{x}{2}$ i $\frac{1-x}{2}$.) Això vol dir que el punt x està rodejat d'altres punts factibles en qualsevol direcció.

Exemple 2.3. El problema

$$\begin{aligned} \min_{(x,y)} \;& 3\,x^2 + 2\,y^2 \\ \text{s.a. } & x^2 + y^2 < 1 \end{aligned}$$

és un altre problema sense restriccions, d'acord amb la definició que hem donat més amunt. El motiu és que, en ser la restricció de la forma $g(x,y) = x^2 + y^2 < 1$, en què $g(x,y)$ és una funció contínua, qualsevol punt que satisfà la desigualtat està necessàriament rodejat en qualsevol direcció d'altres punts que també la satisfan (això és cert sempre que tenim una funció contínua els valors de la qual satisfan una desigualtat estricta).

Exemple 2.4. El problema

$$\begin{aligned} &\min_x \; e^x \\ &\text{s.a.} \; x > 0 \end{aligned}$$

és un problema sense restriccions, d'acord amb la nostra definició, perquè el conjunt factible hi ve donat per una desigualtat estricta. En canvi, el problema

$$\begin{aligned} &\min_x \; e^x \\ &\text{s.a.} \; x \geq 0 \end{aligned}$$

és un problema amb una restricció de desigualtat. Si des del punt factible $x = 0$ ens movem cap a l'esquerra, hi trobem punts no factibles.

Abans hem vist el que de vegades s'anomena *regla de Fermat*, que diu que, en un problema univariant d'optimització sense restriccions, si la funció objectiu és diferenciable llavors la derivada ha de valdre zero en qualsevol maximitzador o minimitzador. Ho enunciem aquí formalment perquè en quedi constància.

Regla de Fermat

Sigui $f : X \to \mathbb{R}$ una funció diferenciable. Si x^* és un punt interior del conjunt X i f ateny un màxim o un mínim local a x^*, llavors totes les derivades parcials de f avaluades a x^* són iguals a zero. !

Pierre Simon de Ferrat...

... (Beaumont-de-Lomagne 1601 - Castus 1665) va idear el seu *Methodus ad Disquirendam Maximan et Miniman* l'any 1629.

Observeu que la regla de Fermat no proporciona un procediment per a calcular automàticament tots els maximitzadors o minimitzadors (locals) d'una funció, sinó que només ens diu que podem descartar com a tals tots aquells punts en què alguna derivada parcial sigui diferent de zero. En general, necessitarem criteris addicionals per a triar entre tots aquells punts que satisfan aquest primer sedàs, és a dir, entre tots els punts crítics de la funció.

Anomenem **condicions de primer ordre** en un problema d'optimització sense restriccions el sistema d'equacions que descriu els punts crítics de la funció objectiu, és a dir, aquell que resulta d'igualar a zero totes les derivades parcials de la funció.

El nostre objectiu a continuació és proporcionar eines d'anàlisi dels punts crítics. En primer lloc veurem com, a partir de les condicions de segon ordre, podrem en alguns casos saber si un punt crític és un maximitzador local, un minimitzador local o un punt de sella. Després veurem com l'anàlisi de la concavitat o convexitat de la funció objectiu ens serveix per a donar condicions que garanteixen que un punt crític és un maximitzador o un minimitzador global. !

2.2. Condicions de segon ordre

En molts casos, encara que no sempre, mitjançant l'estudi de les derivades parcials de segon ordre podem determinar si la funció creix o decreix al voltant d'un punt crític i, per tant, si es tracta d'un minimitzador o d'un maximitzador.

El primer que farem és un petit repàs de les propietats de les formes quadràtiques, ja que és estudiant formes quadràtiques formades a partir de les derivades parcials segones que formularem les condicions de segon ordre.

2.2.1. Formes quadràtiques

En el curs d'àlgebra lineal ja heu vist les definicions de formes quadràtiques i la seva classificació. Aquí repassem les definicions i fixem la terminologia que usarem en endavant.

Sigui

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix}$$

una matriu simètrica, és a dir, que per a cada $i \in \{1, 2 \dots n\}$ i per a cada $j \in \{1, 2, \dots, n\}$, es compleix $c_{ij} = c_{ji}$. Llavors podem definir una funció $Q : \mathbb{R}^n \to \mathbb{R}$ fent:

$$\forall v \in \mathbb{R}^n, \quad Q(v) = v'\, C\, v = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, v_i\, v_j, \tag{2.1}$$

en què v' denota el vector v transposat. Direm que Q és una **forma quadràtica**.

Exemple 2.5. Observeu la relació entre la següent matriu C i la corresponent forma quadràtica Q

$$C = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 7 \end{pmatrix}$$
$$Q(x,y,z) = x^2 + 4\,x\,y + 6\,x\,z + 5\,y^2 + 8\,y\,z + 7\,z^2.$$

Les files 1, 2 i 3 de la matriu corresponen a les variables x, y i z, respectivament. Anàlogament, les columnes 1, 2 i 3 també corresponen a les variables x, y i z, respectivament. Com que multipliquem cada terme dues vegades (primer multipliquem pel vector (x, y, x) i després premultipliquem pel mateix vector transposat), la suma dels exponents de les variables que apareixen en cada un dels sumands és 2. El coeficient de x^2 és el terme que correspon als índexs $(1,1)$ de la matriu, en aquest cas el nombre 1. El coeficient de y^2 és el terme que correspon als índexs $(2,2)$, en aquest cas el nombre 5. Finalment, el coeficient de z^2 és el terme que correspon als índexs $(3,3)$, en aquest cas el nombre 7.

El coeficient del sumand que conté el producte $x\,y$ és el resultat de sumar el terme $(2,1)$ amb el terme $(1,2)$, en aquest cas $2+2=4$. El coeficient del sumand que conté el producte $x\,z$ és el resultat de sumar el terme $(3,1)$ amb el terme $(1,3)$, en aquest cas $3+3=6$. Finalment, el coeficient del sumand que conté el producte $y\,z$ és el resultat de sumar el terme $(3,2)$ amb el terme $(2,3)$, en aquest cas $4+4=8$.

La classificació de formes quadràtiques que nosaltres usarem és la següent.

Sigui $Q : \mathbb{R}^n \to \mathbb{R}$ una forma quadràtica.

1) Diem que Q és **definida positiva** si, per a tot $v \in \mathbb{R}^n$ tal que $v \neq 0$, es compleix $Q(v) > 0$.

2) Diem que Q és **definida negativa** si, per a tot $v \in \mathbb{R}^n$ tal que $v \neq 0$, es compleix $Q(v) < 0$.

3) Diem que Q és **semidefinida positiva** si, per a tot $v \in \mathbb{R}^n$, es compleix $Q(v) \geq 0$.

4) Diem que Q és **semidefinida negativa** si, per a tot $v \in \mathbb{R}^n$, es compleix $Q(v) \leq 0$.

5) Diem que Q és **indefinida** si hi ha dos vectors v i w de $\mathbb{R}^n$ tals que: $Q(v) < 0 < Q(w)$.

Freqüentment, es defineix una forma quadràtica com a semidefinida si, a més de la condició que hem escrit a dalt, hi ha algun vector $v \neq 0$ tal que $Q(v) = 0$. Nosaltres aquí no ho fem així. D'acord amb la definició que acabem de donar, tota forma definida positiva també és semidefinida positiva, mentre que una forma indefinida és tota aquella que no és semidefinida positiva ni semidefinida negativa. Emprem aquesta definició perquè ens facilita molt més la presentació de tots els resultats basats en formes quadràtiques que veurem més endavant. Quan ens vulguem referir a una forma quadràtica que és semidefinida positiva (o negativa) però no definida positiva (negativa) ens hi referirem dient que la forma quadràtica és **només semidefinida**.

És important tenir en compte que l'expressió $v \neq 0$ que apareix en la definició anterior vol dir que si $v = (v_1, v_2 \dots v_n)$, llavors hi ha almenys un índex i tal que el nombre v_i és diferent de zero.

En la pràctica, tractarem més amb la matriu C que amb la corresponent forma quadràtica Q. Per abús de notació, direm també que *la matriu* C és definida positiva, indefinida, etc.

Les condicions que hi ha a continuació permeten verificar les definicions anteriors treballant només amb la matriu.

Criteri dels valors propis

Sigui $Q : \mathbb{R}^n \to \mathbb{R}$ una forma quadràtica, definida a partir de la matriu simètrica C com en l'equació 1. Siguin $\{\lambda_1, \lambda_2 \dots \lambda_n\}$ els valors propis de la matriu C. Llavors tenim:

1) Q és definida positiva si, i només si, per a tot i, $\lambda_i > 0$.

2) Q és definida negativa si, i només si, per a tot i, $\lambda_i < 0$.

3) Q és semidefinida positiva si, i només si, per a tot i, $\lambda_i \geq 0$.

4) Q és semidefinida negativa si, i només si, per a tot i, $\lambda_i \leq 0$.

5) Q és indefinida si, i només si, hi ha dos índexs i i j tals que $\lambda_i < 0 < \lambda_j$.

El teorema dóna una **caracterització exacta** de la categoria a la qual pertany la forma quadràtica Q a partir dels valors propis de la matriu C.

Si no tenim un ordinador a prop, ens podem estalviar feina sovint si usem el teorema següent, que proveeix només una caracterització parcial, basada en els menors preferents de la matriu C. Els **menors preferents** són els determinants de les successives submatrius quadrades que anem formant partint de l'element superior esquerre, fins arribar a la matriu sencera.

Criteri dels menors preferents

Sigui $Q : \mathbb{R}^n \to \mathbb{R}$ una forma quadràtica, definida a partir de la matriu C com en l'equació 1. Siguin $\{\Delta_1, \Delta_2 \ldots \Delta_n\}$ els menors preferents de la matriu C. Llavors tenim:

1) Q és definida positiva si, i només si, per a tot i, $\Delta_i > 0$.

2) Q és definida negativa si, i només si, per a tot i, $(-1)^i \Delta_i > 0$, és a dir, si els menors preferents d'ordre senar són negatius i els d'ordre parell són positius.

Dues observacions:

- El criteri dels menors preferents només és decisori quan la forma quadràtica és definida positiva o negativa. No permet distingir formes quadràtiques indefinides d'altres que només són semidefinides.

- La manera més senzilla de recordar les condicions perquè Q sigui definida negativa és considerar el cas en què C és una matriu diagonal (i, per tant, els elements de la diagonal principal són els seus valors propis); aleshores ens trobem: $\Delta_1 = \lambda_1 < 0$, $\Delta_2 = \lambda_1 \lambda_2 > 0$, etc.

Exemple 2.6. Sigui $Q(x, y) = 3\,x^2 + y^2$ i, per tant,

$$C = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

és una matriu diagonal, que té per valors propis $\lambda_1 = 3 > 0$ i $\lambda_2 = 1 > 0$. El criteri dels valors propis implica que Q és definida positiva.

Alternativament, en aquest cas és fàcil analitzar Q directament a partir de la definició de forma quadràtica definida positiva. Observem d'entrada que $\forall(x, y)$, $3\,x^2 + y^2 \geq 0$, perquè els quadrats de nombres reals sempre són no negatius. Això vol dir que, com a mínim, Q és semidefinida positiva. D'altra banda, com que el quadrat d'un nombre real és zero només quan aquest nombre és zero, és senzill veure que Q només pot ser zero quan tant x com y són iguals a zero. Això és precisament el que es requereix en la definició de **forma quadràtica definida positiva**.

Exemple 2.7. Sigui $Q(x,y) = x^2 + +x\,y + y^2$, d'on

$$C = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

Els menors preferents són aquí $\Delta_1 = 1 > 0$ i $\Delta_2 = \frac{3}{4} > 0$, de manera que el criteri dels menors preferents garanteix que Q també és definida positiva. Calcular els valors propis ens hauria portat força més feina.

Si volem analitzar Q directament, la podem escriure com a $Q(x,y) = \left(\frac{x}{2} + +y\right)^2 + \frac{3}{4}\,x^2$, d'on resulta senzill veure que és estrictament positiva sempre que x i y no siguin zero simultàniament.

Exemple 2.8. Sigui $Q(x,y) = 2\,x\,y$, d'on

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

El criteri dels menors preferents no és aplicable, ja que $\Delta_1 = 0$, així que hem de mirar els valors propis. Els valors propis són $\lambda_1 = 1$ i $\lambda_2 = -1$, la qual cosa indica que Q és indefinida.

A partir de la definició mateixa no és gens difícil provar que Q és indefinida; per exemple: $Q(-1,1) = -2 < 0 < 2 = Q(1,1)$. Així de senzilla és la cosa! Rellegiu la definició i veureu que, per tal de verificar que Q és indefinida, en tenim prou trobant dos elements qualssevol en els quals Q pren signes oposats.

A part dels criteris que acabem de mencionar, vosaltres heu vist una forma diferent de classificar formes quadràtiques, que es basa en els canvis de signe del polinomi característic. El criteri dels menors preferents té l'avantatge que no necessitem ni tan sols trobar el polinomi característic per tal d'aplicar-lo, i el desavantatge que no sempre ens diu com és la forma quadràtica.

Hi ha un altre criteri...

... el dels menors principals, que és una extensió del dels menors preferents i proporciona una caracterització completa del signe de la forma quadràtica; nosaltres no el donem aquí però el podeu trobar en molts llibres que parlen d'optimització o d'àlgebra lineal.

2.2.2. Formes quadràtiques i matrius hessianes

Donat $X \subset \mathbb{R}^n$, sigui $f : X \to \mathbb{R}$ una funció que té derivades parcials de segon ordre contínues, i sigui $\bar{v} = (\bar{v}_1, \bar{v}_2 \ldots \bar{v}_n)$ un punt donat de $\mathbb{R}^n$. $D^2 f(\bar{v})$ denota la matriu (hessiana) formada per les derivades parcials segones de f, avaluades en el punt $\bar{v}$.

L'extensió a $\mathbb{R}^n$ de les condicions de segon ordre que s'estudien en el càlcul univariant està basada en l'estudi de la matriu hessiana.

Condicions suficients de segon ordre

Sigui $X \subset \mathbb{R}^n$ un conjunt obert, i sigui $f : X \to \mathbb{R}$ una funció que té derivades parcials de segon ordre contínues. Suposem que $\bar{v} \in \mathbb{R}^n$ és un punt crític, és a dir, que per a cada i, $1 \leq i \leq n$, $D_i f(\bar{v}) = 0$. Tenim que:

1) Si la matriu $D^2 f(\bar{v})$ és definida positiva, llavors $\bar{v}$ és un minimitzador local (estricte) de la funció f.

2) Si la matriu $D^2 f(\bar{v})$ és definida negativa, llavors $\bar{v}$ és un maximitzador local (estricte) de la funció f.

3) Si la matriu $D^2 f(\bar{v})$ és indefinida, llavors $\bar{v}$ és un punt de sella de la funció f.

Recordem que...

... un punt x^* és un maximitzador local de la funció f, si hi ha un cert entorn de x^* en què es compleix que $f(x^*) \geq f(x)$, per a qualsevol punt x de l'entorn. Si la igualtat és estricta quan $x \neq x^*$, diem que x^* és un *maximitzador local estricte*. El concepte de minimitzador local estricte es defineix similarment.

Una observació important: cap de les implicacions del teorema anterior no és certa si la posem a l'inrevés. A continuació en veurem exemples.

Exemple 2.9. Sigui $f(x, y) = x^2 + y^2$. L'únic punt crític és $(0, 0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 2 > 0$ i $\lambda_2 = 2 > 0$ i, per tant, és definida positiva. El teorema anterior implica que $(0, 0)$ és un minimitzador local estricte.

Exemple 2.10. Sigui $f(x, y) = 1 - x^2 - y^2$. L'únic punt crític és $(0, 0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = -2 < 0$ i $\lambda_2 = -2 < 0$ i, per tant, és definida negativa. El teorema anterior implica que $(0, 0)$ és un maximitzador local estricte.

Exemple 2.11. Sigui $f(x, y) = x^2 - y^2$. L'únic punt crític és $(0, 0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 2 > 0$ i $\lambda_2 = -2 < 0$ i, per tant, és indefinida. El teorema anterior implica que $(0,0)$ és un punt de sella.

Exemple 2.12. Sigui $f(x,y) = x^4 + y^2$. L'únic punt crític és $(0,0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 0$ i $\lambda_2 = 2 > 0$ i, per tant, només és semidefinida positiva. El teorema anterior no és aplicable. Ara bé, és fàcil veure que, per a cada $(x,y) \in \mathbb{R}^2$, $f(x,y) = x^4 + y^2 \geq 0 = f(0,0)$, amb igualtat només quan $(x,y) = (0,0)$. Això vol dir que $(0,0)$ és un minimitzador global (i, per tant, local) estricte.

Exemple 2.13. Sigui $f(x,y) = 1 - x^4 - y^2$. L'únic punt crític és $(0,0)$, i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 0$ i $\lambda_2 = -2 < 0$ i, per tant, només és semidefinida negativa. El teorema anterior no és aplicable. Ara bé, és fàcil veure que, per a cada $(x,y) \in \mathbb{R}^2$, $f(x,y) = 1 - x^4 - y^2 \leq 1 = f(0,0)$, amb igualtat només quan $(x,y) = (0,0)$. Això vol dir que $(0,0)$ és un maximitzador global (i, per tant, local) estricte.

Exemple 2.14. Sigui $f(x,y) = y^2 - x^4$. L'únic punt crític és $(0,0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 0$ i $\lambda_2 = = 2 > 0$ i, per tant, només és semidefinida positiva. El teorema anterior no és aplicable. Ara bé, per a cada $x \neq 0$ i $y \neq 0$, tan a prop de 0 com vulguem, es compleix $f(x,0) = -x^4 < 0 < y^2 = f(0,y)$, mentre que $f(0,0) = 0$. Això vol dir que $(0,0)$ és un punt de sella, encara que la matriu no sigui indefinida.

Exemple 2.15. Sigui $f(x,y) = x^4 - y^2$. L'únic punt crític és $(0,0)$ i la matriu hessiana avaluada en aquest punt és

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}.$$

Com que aquesta és una matriu diagonal, els valors propis són $\lambda_1 = 0$ i $\lambda_2 = -2 < 0$ i, per tant, només és semidefinida negativa. El teorema anterior no és aplicable. Ara bé, per a cada $x \neq 0$ i $y \neq 0$, tan a prop de 0

com vulguem, es compleix $f(x,0) = x^4 > 0 > -y^2 = f(0,y)$, mentre que $f(0,0) = 0$. Això vol dir que $(0,0)$ és un punt de sella, encara que la matriu no sigui indefinida.

Per tant, les condicions de segon ordre poden no ser decisòries per a determinar la natura d'un punt crític. En alguns casos en què l'anterior teorema no és aplicable, podrem almenys descartar alguna possibilitat si tenim en compte el resultat següent.

Condicions necessàries de segon ordre

Sigui $X \subset \mathbb{R}^n$ un conjunt obert, i sigui $f : X \to \mathbb{R}$ una funció que té derivades parcials de segon ordre contínues. Llavors:

1) Si $\bar{v} \in \mathbb{R}^n$ és un minimitzador local (no necessàriament estricte), llavors $\bar{v}$ és un punt crític i la matriu hessiana $D^2f(\bar{v})$ és semidefinida positiva.

2) Si $\bar{v} \in \mathbb{R}^n$ és un maximitzador local (no necessàriament estricte), llavors $\bar{v}$ és un punt crític i la matriu hessiana $D^2f(\bar{v})$ és semidefinida negativa.

Observació important

Cap de les implicacions del teorema anterior no és certa si la posem a l'inrevés, com hem pogut veure en els exemples 2.14 i 2.15.

Com hem pogut apreciar en els exemples, les condicions de segon ordre no són sempre decisòries. Aleshores la pregunta és: què hem de fer quan ens trobem en aquest cas? Malauradament, no hi ha un mètode sistemàtic d'anàlisi; hi ha una sèrie de tècniques que funcionen millor en uns casos que en d'altres. El que es fa en general és procedir al que anomenem una **anàlisi local** del comportament de la funció al voltant del punt crític. !

Els exemples que hem presentat eren tan senzills que n'hem tingut prou amb mirar la funció directament; en general, haurem de fer una anàlisi considerant diferents trajectòries d'aproximació a la funció.

Exemple 2.16. Suposem que volem analitzar el punt crític $(0,0)$ de la funció

$$f(x,y) = x^4y^4 + x^3y^3.$$

Com que totes les variables tenen exponents elevats, és clar que la matriu hessiana tindrà zero en tots els components quan l'avaluem al punt $(0,0)$ i, per tant, les condicions de segon ordre no són decisòries. En aquest cas ens bastarà considerar el que passa al llarg de dues trajectòries lineals d'aproximació. Quan aproximem $(0,0)$ al llarg de la trajectòria lineal $y = x$, tenim que el valor de la funció és $f(x,x) = x^8 + x^6$, per la qual cosa aquesta funció presenta un mínim local al llarg de la trajectòria considerada. Quan aproximem $(0,0)$ al llarg de la trajectòria lineal $y = -x$, tenim que el valor de la funció és $f(x,-x) = x^8 - x^6 = -x^6\,(1 - x^2)$, per la qual cosa aquesta funció presenta un màxim local al llarg de la trajectòria considerada (la funció val 0 quan $x = 0$, i pren valors negatius quan $0 < |x| < 1$). Això prova que $(0,0)$ és un punt de sella de la funció.

Un darrer exemple ens servirà per veure que, de vegades, l'anàlisi local de la funció pot ser força complicada.

Exemple 2.17. Suposem que volem analitzar la funció $f(x,y) = x^4 - 3x^2y + y^2$. El seu únic punt crític és el $(0,0)$, però la matriu hessiana avaluada en aquest punt només és semidefinida, per la qual cosa les condicions de segon ordre no són decisòries. Per tal d'analitzar trajectòries d'aproximació lineal al punt, comencem observant que, si ens aproximem pels eixos (fem $x = 0$ o $y = 0$), llavors la funció presenta un mínim a $(0,0)$. Per a analitzar qualsevol trajectòria lineal, suposem que $y = tx$, on t és un cert nombre real (fix). Llavors tenim que $f(x,tx) = x^4 - 3tx^3 + t^2x^2$, i les eines habituals de càlcul mostren que f presenta un mínim local a $x = 0$, qualsevol que sigui t.

Tot això sembla suggerir que el punt crític $(0,0)$ és un minimitzador local, però si ens parem a considerar la trajectòria *quadràtica* d'aproximació $y = x^2$, podem comprovar que $f(x,x^2) = -x^4 < 0 = f(0,0)$, sempre que $x \neq 0$; és a dir, f presenta un màxim al nostre punt crític al llarg d'aquesta trajectòria quadràtica d'aproximació. Això mostra que $(0,0)$ és en realitat un punt de sella.

3. Concavitat i convexitat de funcions. Criteris de globalitat

3.1. Motivació. Cas univariant

Volem trobar un criteri que ens permeti afirmar que qualsevol maximitzador o minimitzador local ho és també en sentit global. És clar, doncs, que el tipus de condicions que hem d'imposar sobre el conjunt factible i la funció objectiu han de ser també de caire global; no ens podem limitar a dir com ha de ser la funció objectiu just al voltant del nostre candidat a maximitzador o minimitzador global, sinó que hem d'exigir una certa condició que se satisfaci en qualsevol punt de la funció objectiu.

Per a entendre els conceptes de concavitat i convexitat, el primer que hem de fer és analitzar el comportament de les funcions al voltant dels punts on tenen màxims i mínims (no importa si són locals o globals). !

Exemple 3.1. Màxim, mínim i funció derivada

Sigui $f(x) = x^3\, e^{-x^2}$. Abans hem vist, en l'exemple 1.19, que aquesta funció té un màxim i un mínim (locals i globals alhora) i un punt d'inflexió enmig dels dos. El que volem veure ara és com es comporta la derivada al voltant del màxim i del mínim. Deixarem que Gnuplot ens faci els càlculs. El que hem de fer és entrar al programa la funció i la seva derivada (que hem hagut de trobar nosaltres prèviament):

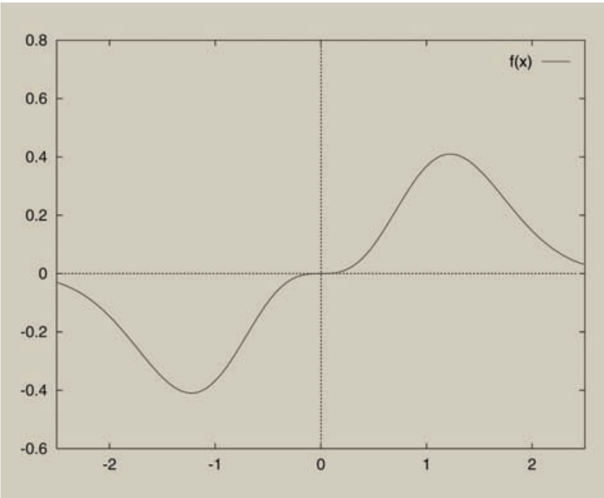
```
gnuplot> f(x) = x**3*exp(-x**2)
gnuplot> fp(x) = x**2*exp(-x**2)*(3-2*x**2)
```

Una mica d'experimentació amb els gràfics ens mostra quins són els valors de les variables més convenients per a la visualització:

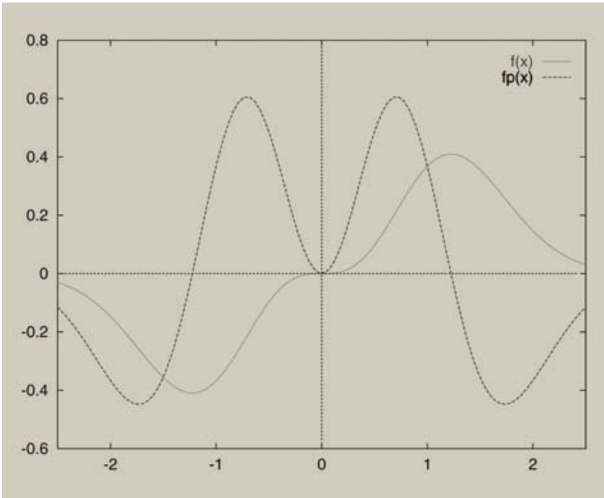
```
gnuplot> set xrange [-2.5:2.5]
gnuplot> set yrange [-0.6:0.8]
```

Fem primer una gràfica de la funció f sola, per tal de veure els punts crítics, i a continuació la presentem junt amb la seva derivada:

```
gnuplot> plot f(x)
gnuplot> plot f(x), fp(x)
```



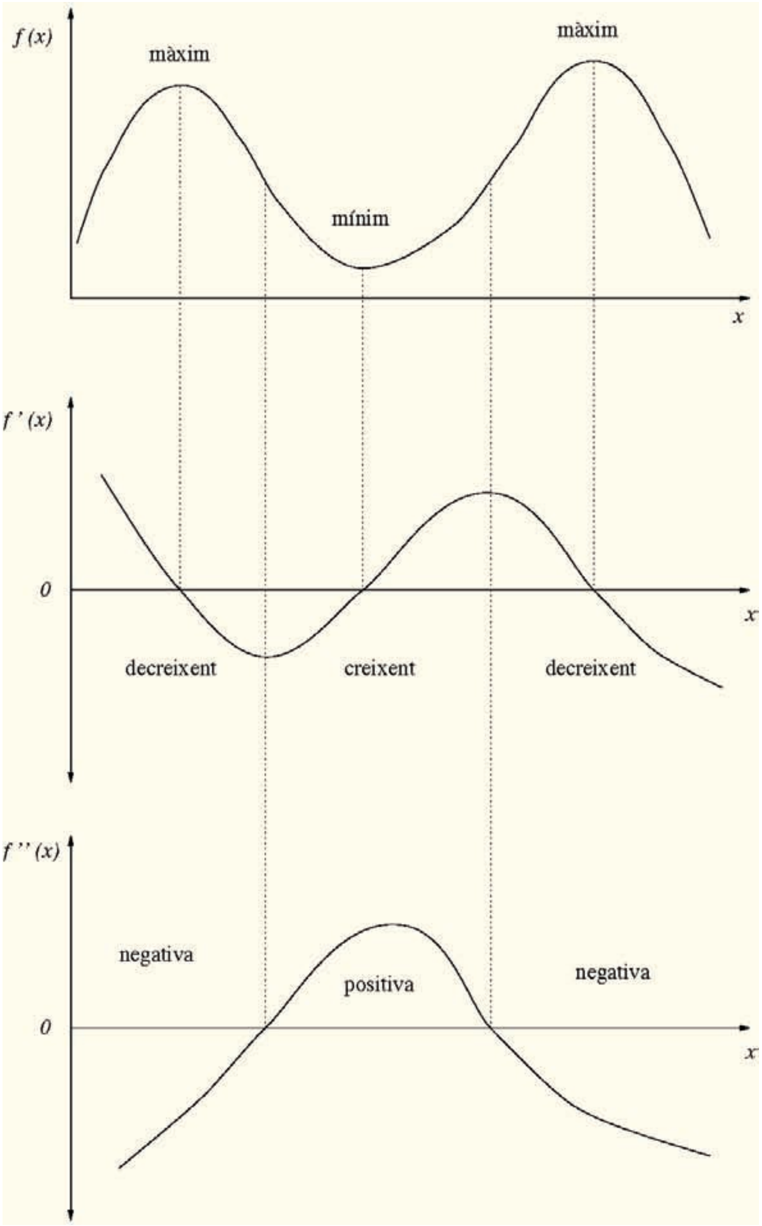
La funció $x^3\, e^{-x^2}$ sola



I amb la seva derivada

Al voltant del mínim la derivada és una funció creixent, i al voltant del màxim, una funció decreixent. Al voltant del punt d'inflexió que hi ha per a $x = 0$, no és ni una cosa ni l'altra.

En general, el comportament d'una funció i les seves derivades al voltant de punts màxims i mínims és semblant al que mostrem en la gràfica que tenim a continuació.



Una funció i les seves derivades

Un màxim local estricte ...

... és un punt en el qual la funció passa de tenir la derivada positiva a tenir-la negativa, quans ens movem d'esquerra a dreta. És a dir, al voltant del punt la derivada és una funció decreixent. De la mateixa manera, al voltant d'un mínim local la derivada és una funció creixent. El creixement o decreixement de la derivada es tradueix finalment en la positivitat o negativitat de la derivada segona.

Al voltant d'un màxim (local), la derivada és una funció decreixent i, per tant, la derivada segona és una funció negativa. Si exigim que la derivada sigui una funció decreixent a tot arreu, llavors la funció es comportarà globalment com al voltant del màxim del qual hem partit, és a dir, no n'hi podrà haver cap altre.

Una funció la derivada de la qual és decreixent globalment és una **funció còncava**. Més endavant veurem que els maximitzadors locals d'una funció còncava són sempre globals.

De la mateixa manera, una **funció convexa** és aquella la derivada de la qual és una funció creixent. Els minimitzadors locals d'una funció convexa són sempre globals.

Considerem una funció còncava. Fixem un cert punt a i triem un altre punt x qualsevol. Llavors sabem pel teorema del valor mitjà que hi ha un punt c entre a i x tal que:

$$f'(c) = \frac{f(x) - f(a)}{x - a}.$$

Si suposem que $a < x$, llavors tenim que $a < c$, i com que la funció és còncava:

$$a < c \implies f'(a) \geq f'(c) \implies f'(a) \geq \frac{f(x) - f(a)}{x - a}.$$

Multiplicant per $x - a$ (que és un nombre positiu), obtenim:

$$f(a) + f'(a)\,(x - a) \geq f(x).$$

D'altra banda, si $a > x$, llavors $a > c$ i tindríem:

$$a > c \implies f'(a) \leq f'(c) \implies f'(a) \leq \frac{f(x) - f(a)}{x - a}.$$

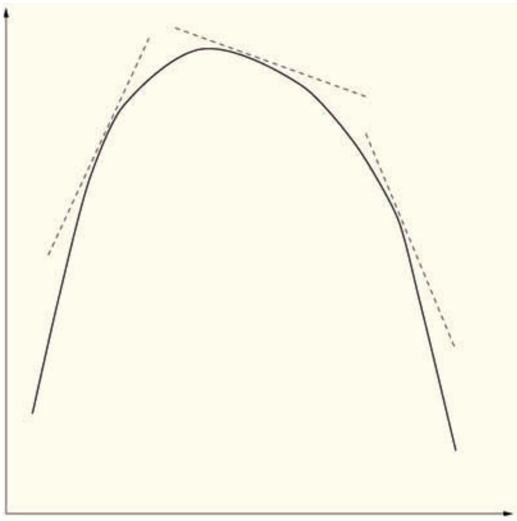
Però ara en multiplicar per $x - a$ hem d'invertir la desigualtat, ja que aquest nombre és negatiu. Amb això, ens torna a quedar

$$f(a) + f'(a)\,(x - a) \geq f(x).$$

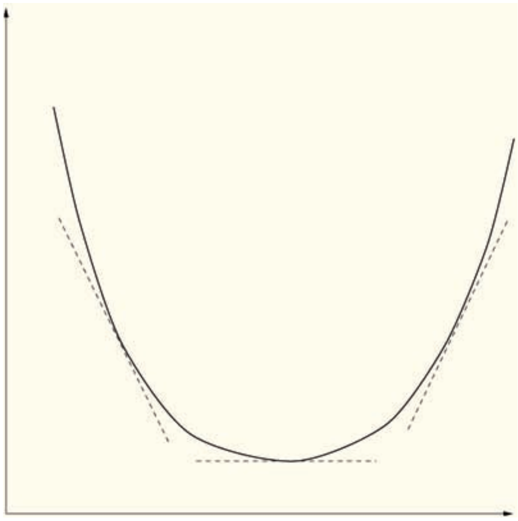
Resumint, qualssevol que siguin x i a es compleix la desigualtat anterior. Ara bé, el terme de l'esquerra de la desigualtat és l'expressió de la recta tangent a la funció f en el punt a. Per tant, la desigualtat ens està dient que, si f és una funció còncava, la recta tangent sempre té un valor superior a la funció. De fet, aquesta comparació entre la funció i les seves tangents no requereix ni tan sols que la funció sigui diferenciable, i nosaltres la prendrem com a definició provisional de concavitat.

Intuïtivament, una **funció còncava** és aquella per la qual les tangents sempre tenen un valor superior o igual al de la funció. Una **funció convexa** és aquella per la qual les tangents sempre tenen un valor inferior o igual al de la funció.

Els conceptes que apareixen en la definició són purament geomètrics: el graf de qualsevol tangent a una funció còncava sempre queda per damunt del graf de la funció. Anàlogament, el graf de qualsevol tangent a una funció convexa sempre queda per sota del graf de la funció.

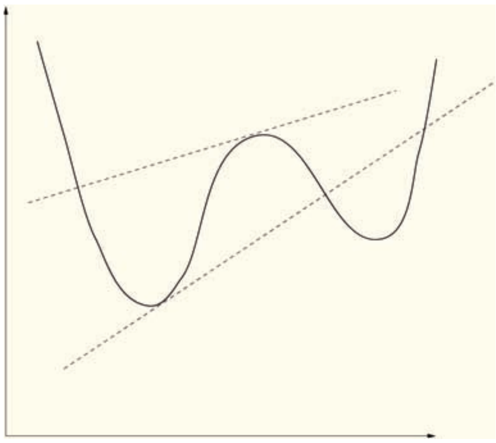


Una funció còncava



Una funció convexa

Una funció que no és còncava ni convexa, necessàriament ha de ser tallada per alguna de les seves tangents.



Ni concavitat ni convexitat

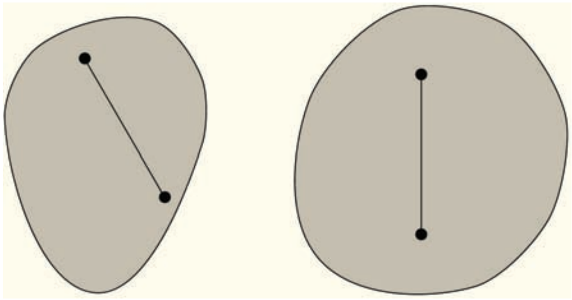
Alguna tangent...

... no és ni totalment a sota del gràfic ni totalment a sobre i, per tant, el talla.

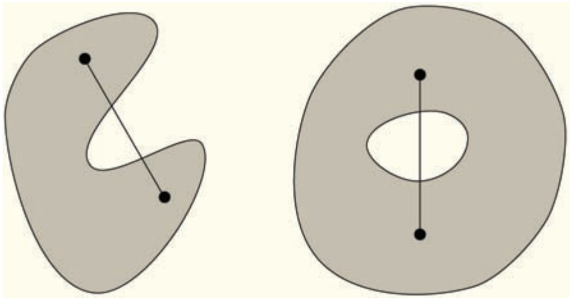
3.2. Extensió a múltiples variables

Com a primer requeriment en les definicions de funcions còncaves i convexes, imposarem que el domini sobre el qual les funcions estan definides (el conjunt factible) sigui un conjunt convex. Recordarem la seva definició.

Sigui X un subconjunt de $\mathbb{R}^n$. Diem que X és un **conjunt convex** si el segment que uneix dos punts qualssevol del conjunt hi és contingut enterament a dins.



Conjunts convexos



Conjunts no convexos

El requeriment que el domini de definició sigui un conjunt convex és molt important per tots els resultats que relacionen funcions còncaves i convexes amb optimització.

Més endavant...

... veurem un exemple de com les coses fallen quan el domini no és convex.

Hi ha una coincidència desafortunada de les denominacions que estem usant, però és molt important tenir present el següent:

Una qüestió terminològica molt important és que no s'han de confondre els **conjunts convexos** amb les **funcions convexes** o les **funcions còncaves**.

Si volem comprovar algèbricament si un conjunt és convex o no, haurem de ser capaços de descriure el segment que uneix dos punts donats. La manera de fer-ho és la següent: donats dos punts x' i x'', un punt x''' està sobre el segment que els uneix si, i només si, hi ha un nombre $\lambda \in (0,1)$, tal que $x''' = \lambda x' + (1-\lambda)x''$.

Nosaltres aquí en tenim prou amb la noció intuïtiva i no usarem aquest tipus de formalització, així que no insistirem sobre els aspectes algèbrics. Un resultat que sí que usarem és el següent.

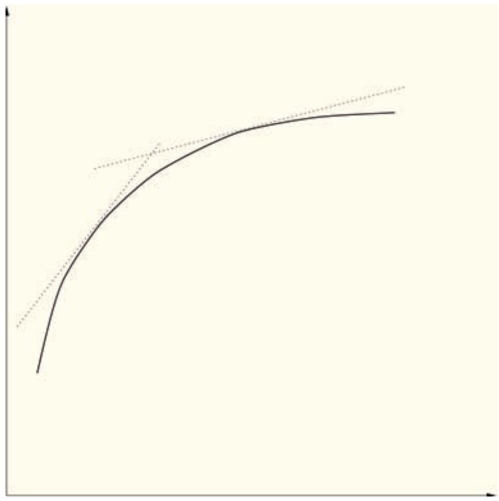
La intersecció de conjunts convexos produeix un conjunt que també és convex.

La definició formal de funció còncava i convexa és la mateixa que hem donat abans, però amb el requeriment que el domini de la funció sigui un conjunt convex.

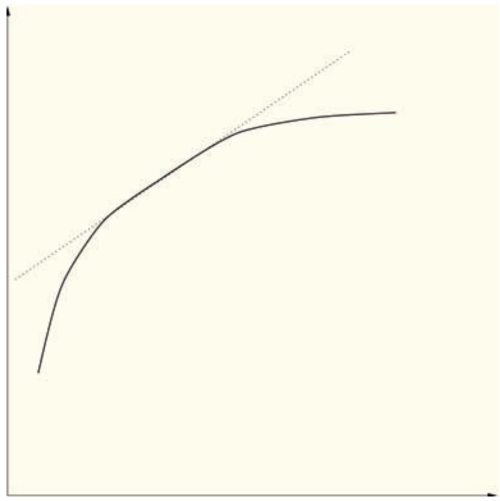
Sigui X un subconjunt convex de $\mathbb{R}^n$, i $f : X \to \mathbb{R}$ una funció.

1) Diem que f és una **funció còncava** sitotes les seves tangents queden per sobre del graf de la funció. La funció és **estrictament còncava** si cada tangent té un únic punt de contacte amb el graf.

2) Diem que f és una **funció convexa** si totes les seves tangents queden per sota del graf de la funció. La funció és **estrictament convexa** si cada tangent té un únic punt de contacte amb el graf.



Funció estrictament còncava



Funció còncava, no estrictaments

El graf d'una funció que és còncava, però no estrictament, té algun tram pla.

Ara suposem que f és una funció diferenciable. Anomenarem $\nabla f(\bar{x})$ el vector gradient de la funció f avaluat en el punt $\bar{x}$, és a dir, el vector compost per les derivades parcials de la funció f avaluades en el punt $\bar{x}$. Recordem que un punt '·' designa el producte escalar entre vectors. La tangent al graf de f en el punt $\big(\bar{x}, f(\bar{x})\big)$ és el graf de la funció:

$$g(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}).$$

Per tant, la condició geomètrica que hem dit que satisfà tota funció còncava és que, per a cada $\bar{x} \in X$ i cada $x \in X$, es compleix que

$$f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) \geq f(x).$$

El que hi ha a continuació no és més que una manera de reescriure les definicions per a funcions diferenciables. Observem que no es tracta més que de l'extensió a n dimensions del que ja havíem vist anteriorment.

Proposició 3. *Sigui $X \subset \mathbb{R}^n$ un conjunt convex i $f : X \to \mathbb{R}$ una funció diferenciable.*

- *f és una funció còncava si, i només si, donats dos punts qualssevol x' i x'', es compleix que*
$$f(x') + \nabla f(x') \cdot (x'' - x') \geq f(x'').$$

- *f és una funció estrictament còncava si, i només si, donats dos punts diferents x' i x'', es compleix que*
$$f(x') + \nabla f(x') \cdot (x'' - x') > f(x'').$$

- *f és una funció convexa si, i només si, donats dos punts qualssevol x' i x'', es compleix que*
$$f(x') + \nabla f(x') \cdot (x'' - x') \leq f(x'').$$

- *f és una funció estrictament convexa si, i només si, donats dos punts diferents x' i x'', es compleix que*
$$f(x') + \nabla f(x') \cdot (x'' - x') < f(x'').$$

Diem que una funció és **diferenciable amb continuïtat** si les seves derivades parcials són funcions contínues. Anàlogament, una funció és **diferenciable dues vegades amb continuïtat** si les seves derivades parcials segones són funcions contínues.

Un resultat lleugerament menys general, però més útil de cara a determinar quan una funció és còncava o convexa, és el següent, que requereix que la funció sigui diferenciable dues vegades. Aquest resultat es deriva de la desigualtat del teorema anterior quan es té en compte l'expansió de Taylor de segon ordre.

Teorema

Sigui $X \subset \mathbb{R}^n$ un conjunt obert convex i $f : X \to \mathbb{R}$ una funció diferenciable dues vegades amb continuïtat.

1) f és una funció còncava si, i només si, per a cada $x \in X$ la matriu hessiana $D^2 f(x)$ és semidefinida negativa.

2) f és una funció convexa si, i només si, per a cada $x \in X$ la matriu hessiana $D^2 f(x)$ és semidefinida positiva.

El que acabem de veure és una condició equivalent a concavitat o a convexitat. El que hi ha a continuació és un resultat útil, perquè garanteix que una funció sigui còncava o convexa en sentit estricte.

Teorema

Sigui $X \subset \mathbb{R}^n$ un conjunt obert convex i $f : X \to \mathbb{R}$ una funció diferenciable dues vegades amb continuïtat.

1) Si per a cada $x \in X$ la matriu hessiana $D^2 f(x)$ és definida negativa, llavors f és estrictament còncava.

2) Si per a cada $x \in X$ la matriu hessiana $D^2 f(x)$ és definida positiva, llavors f és estrictament convexa.

Hem de tenir en compte que hi ha funcions estrictament còncaves i funcions estrictament convexes tals que la seva matriu hessiana només és semidefinida en alguns punts aïllats.

Considerem, per exemple, la funció $f(x) = x^4$, que és estrictament convexa encara que $f''(0) = 0$. De fet, la relació entre funcions estrictament convexes i hessianes definides positives és la mateixa que la relació entre funcions estrictament creixents i derivades positives. Això no és cap casualitat: recordem que la convexitat per a funcions univariants vol dir que la derivada és una funció creixent.

Exercici

3.1. Apliqueu el test de la segona derivada per a determinar si les funcions següents són còncaves, convexes, o cap de les dues coses.

a) $f(x) = x^2$

b) $f(x) = x^3$

c) $f(x) = x^3$, quan $x > 0$

d) $f(x,y) = x + y$

e) $f(x,y) = x^2 - 2\,x\,y + y^2$

f) $f(x,y) = x^2 - x\,y + y^2$

g) $f(x,y) = x^4 + y^2$

h) $f(x,y) = 2\,x + 4\,y - 5\,x^2 - y^2$

3.3. Concavitat, convexitat i optimització

Dels teoremes següents se'n deriva la importància que tenen els conceptes de concavitat i convexitat en els problemes d'optimització.

Teorema local-global per a funcions còncaves

Sigui $X \subset \mathbb{R}^n$ un conjunt convex i $f : X \to \mathbb{R}$, una funció còncava, i suposem que el punt $x^* \in X$ és un maximitzador local de f. Llavors x^* també és un maximitzador global.

Teorema local-global per a funcions convexes

Sigui $X \subset \mathbb{R}^n$ un conjunt convex i $f : X \to \mathbb{R}$, una funció convexa, i suposem que el punt $x^* \in X$ és un minimitzador local de f. Llavors x^* també és un minimitzador global.

Observacions:

- En un problema de maximització, ens agradarà que la funció objectiu sigui còncava, ja que llavors sabrem que els màxims locals també ho són en sentit global.

- En un problema de minimització, ens agradarà que la funció objectiu sigui convexa, ja que llavors sabrem que els mínims locals també ho són en sentit global.

- Observeu que el maximitzador o minimitzador local de què es parla no té per què ser un punt interior. De fet, l'únic requeriment sobre el domini és que aquest sigui un conjunt convex. Això ens permetrà usar aquests teoremes per derivar resultats de globalitat en tot tipus de problemes d'optimització, encara que tinguin restriccions.

- Si una funció és estrictament còncava, el maximitzador de què parla el primer teorema és únic. Si una funció és estrictament convexa, el minimitzador de què parla el segon teorema és únic.

Exemple 3.2. Sigui $X = \mathbb{R}_{++} = (0, \infty)$ i $f(x) = x\ \log(x)$. Tenim que $f'(x) = 1 + \log(x)$ i $f''(x) = 1/x$. Com que el domini són els reals estrictament positius, tenim que, per a cada $x \in X$, es compleix $f''(x) = 1/x > 0$, per la qual cosa podem afirmar que f és una funció convexa.

Els punts crítics solucionen $f'(x) = 1+\log(x) = 0$, és a dir, l'únic punt crític és $x = 1/e$. Com que $f''(1/e) > 0$, sabem que $1/e$ és un minimitzador local i el teorema anterior ens assegura que també és un minimitzador global.

Exemple 3.3. Sigui $X = \mathbb{R}^2$ i $f(x, y) = 1 - x^2 - y^2$. Les derivades parcials són $D_1f(x, y) = -2x$ i $D_2f(x, y) = -2y$, i la matriu hessiana

$$D^2f(x, y) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Per tant, la matriu hessiana és definida negativa per a cada (x, y) i podem concloure que f és una funció còncava.

És senzill veure que l'únic punt crític és el $(0, 0)$ i que aquest és un maximitzador local, perquè $D^2f(0, 0)$ és definida negativa. El teorema que hem vist anteriorment ens assegura que $(0, 0)$ també és un maximitzador global.

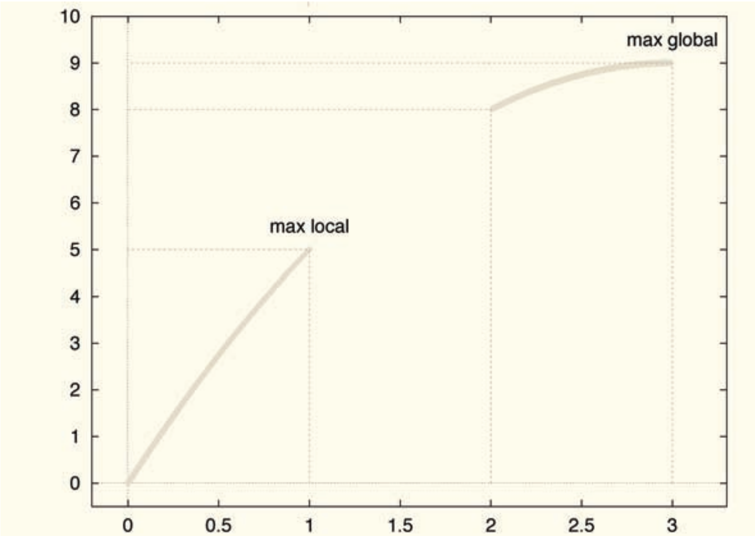
Hem de tenir present que és essencial que el domini de la funció sigui un conjunt convex.

Exemple 3.4. El domini no és convex

Sigui $f(x) = 6x - x^2$. És senzill veure, aplicant el criteri de la derivada segona, que aquesta funció és estrictament còncava. Suposem que volem resoldre el problema de maximitzar f sobre el conjunt $X = [0, 1] \cup [2, 3]$. Com que X és la unió de dos intervals que no es toquen, es tracta d'un conjunt **no convex**.

Dins d'aquest domini de definició, podem comprovar ràpidament que el punt $x^* = 1$ és un maximitzador local; això és degut al fet que els únics punts amb què el podem comparar localment són els de l'interval $[0, 1]$. També resulta senzill veure que $x^* = 1$ *no* és un maximitzador global; de fet, l'únic maximitzador global és $x^{**} = 3$.

Quin és el problema d'aquest exemple? Fixeu-vos que si mirem la definició formal de concavitat, veurem que un requeriment per poder afirmar que una funció és còncava és que el seu domini sigui un conjunt convex. És a dir, encara que la segona derivada de f sigui estrictament negativa, el fet que el seu domini no és convex implica que *no* podem dir que la funció sigui còncava, tal com hem fet alegrement en començar l'exemple.



Local però no global

El punt $x^* = 1$...

... és un maximitzador local, però el maximitzador global és $x^{**} = 3$.

Si observem la caracterització de funcions còncaves i convexes que són diferenciables, podem deduir immediatament un altre criteri que també és molt útil.

Teorema

Sigui $X \subset \mathbb{R}^n$ un conjunt convex i $f : X \to \mathbb{R}$ una funció diferenciable.

1) Si f és una funció còncava, llavors qualsevol punt en què totes les derivades parcials són zero és un maximitzador global.

2) Si f és una funció convexa, llavors qualsevol punt en què totes les derivades parcials són zero és un minimitzador global.

Observem que...

... aquest darrer resultat s'aplica fins i tot a problemes en els quals els criteris locals per problemes d'optimització sense restriccions no ens permeten decidir si un punt crític és, localment, un maximitzador, un minimitzador o un punt de sella.

Exemple 3.5. Sigui $f(x, y) = x^4 + y^4$. És fàcil veure que l'únic punt on s'anul·len les derivades parcials és el $(0, 0)$. La matriu hessiana és

$$D^2 f(x, y) = \begin{pmatrix} 12\,x^2 & 0 \\ 0 & 12\,y^2 \end{pmatrix}$$

i és semidefinida positiva *per a cada* $(x, y) \in \mathbb{R}^2$. Això vol dir que la funció és convexa i, per tant, l'únic punt crític $(0, 0)$ és un minimitzador global, encara que les condicions de segon ordre per a problemes d'optimització sense restriccions no ens permeten decidir ni tan sols si és un minimitzador local, perquè la matriu hessiana a $(0,0)$ és una matriu de zeros.

3.4. Solucionari

3.1.
a) $f(x) = x^2 \to f''(x) = 2 > 0$. Estrictament convexa.

b) $f(x) = x^3 \to f''(x) = 3\,x$. Positiva o negativa. Ni còncava ni convexa.

c) $f(x) = x^3 \to f''(x) = 3\,x > 0$. Estrictament convexa.

d) $f(x,y) = x + y$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

que és tant semidefinida positiva com negativa. Còncava i convexa.

e) $f(x,y) = x^2 - 2\,x\,y + y^2$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

és semidefinida positiva. Convexa.

f) $f(x,y) = x^2 - x\,y + y^2$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

és definida positiva. Estrictament convexa.

g) $f(x,y) = x^4 + y^2$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 2 \end{pmatrix}$$

que és definida positiva (excepte a l'origen). Estrictament convexa.

h) $f(x,y) = 2\,x + 4\,y - 5\,x^2 - y^2$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} -10 & 0 \\ 0 & -2 \end{pmatrix}$$

que és definida negativa. Estrictament còncava.

4. Optimització amb restriccions d'igualtat

4.1. Condicions de primer ordre: el mètode dels multiplicadors de Lagrange

Un problema d'optimització amb restriccions d'igualtat és aquell en el qual el conjunt factible està format per tots aquells punts que solucionen un determinat sistema d'equacions. Per exemple,

$$\max_{(x,y)} \; x\,y$$
$$\text{s.a.} \;\; x+y=1.$$

En aquest problema, en ser la restricció una equació lineal, podem expressar una de les variables en termes de l'altra i solucionar el problema (sense restriccions) que resulta després de substituir aquesta variable en la funció objectiu. Per exemple, com que qualsevol punt factible satisfà $y = 1 - x$, resulta que hem de solucionar el problema sense restriccions

$$\max_{x} \; x\,(1-x)$$
$$\text{s.a.} \;\; x \in \mathbb{R},$$

que té per solució única i global $x^* = 1/2$ i, per tant, $(x^*, y^*) = (1/2, 1/2)$ és el maximitzador del problema original.

Quan les restriccions estiguin definides per equacions no lineals, com per exemple en el problema

$$\max_{(x,y)} \; x\,y$$
$$\text{s.a.} \;\; x^2+y^2=2,$$

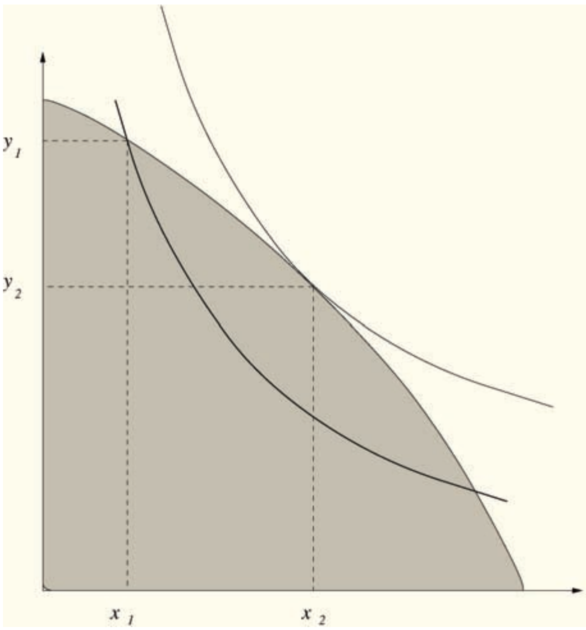
podem trobar les condicions necessàries de primer ordre sempre que el teorema de la funció implícita sigui aplicable als maximitzadors o minimitzadors. En aquestes notes no seguirem aquesta via, sinó que derivarem les condicions necessàries a partir de consideracions geomètriques.

La idea bàsica d'aquesta derivació és senzilla. Recordeu que, donada una funció $f : \mathbb{R}^n \to \mathbb{R}$ i donat qualsevol $v \in \mathbb{R}^n$, la corba de nivell que passa per v separa els punts on la funció pren valors superiors a $f(v)$ d'aquells punts on la funció pren valors inferiors. Considereu ara un problema d'optimització amb una restricció d'igualtat. El conjunt factible està format per tots aquells punts que satisfan la restricció; amb n variables, el conjunt factible consistirà, doncs, en una certa superfície $n-1$ dimensional. Si en un punt factible la corba de nivell de la funció objectiu *talla* la superfície factible, llavors aquest punt no pot ser maximitzador ni minimitzador, ja que hi ha punts factibles a totes dues bandes de la corba de nivell.

Per tant, una condició necessària per tal que un punt sigui un maximitzador o un minimitzador local serà que la corba de nivell que passa per aquest punt sigui **tangent** a la superfície definida pel conjunt factible. És per això que diem que aquesta condició necessària és la **condició de tangència**.

Fixeu-vos-hi

A (x_1, y_1) la corba de nivell talla el conjunt factible.
A (x_2, y_2) hi ha tangència.



Tangència

Tornem al problema que havíem formulat anteriorment, per tal de veure la translació en termes analítics de la condició de tangència.

Definim $f(x,y) = xy$, la funció objectiu, i $g(x,y) = x^2+y^2-2$, de tal manera que la restricció és $g(x,y) = 0$, la corba de nivell 0 de la funció g. Un punt factible (x^*, y^*) satisfà la condició de tangència si en aquest punt les corbes de nivell de f i de g són tangents. En el context que estem considerant, en què les corbes són definides per funcions diferenciables, dues corbes són tangents en un punt si hi ha una sola recta que és tangent a cada una de les corbes en aquest punt. Com que la tangent a una corba de nivell és perpendicular al vector gradient, la condició de tangència equival a dir que els vectors gradients de f i de g en el punt (x^*, y^*) són colineals, és a dir, hi ha un cert nombre real λ tal que

$$\nabla f(x^*, y^*) = \lambda\, \nabla g(x^*, y^*) \implies \begin{cases} \dfrac{\partial f}{\partial x} = \lambda\, \dfrac{\partial g}{\partial x} \\ \dfrac{\partial f}{\partial y} = \lambda\, \dfrac{\partial g}{\partial y} \end{cases}$$

Observeu que ara tenim una variable addicional, λ, que anomenarem **multiplicador de Lagrange** del problema. Si hi afegim la restricció, obtenim un sistema amb tres equacions i tres incògnites:

$$\begin{aligned} &y = \lambda\,(2\,x) \\ &x = \lambda\,(2\,y) \\ &x^2 + y^2 = 2. \end{aligned}$$

La solució d'aquest sistema ens dóna quatre candidats a maximitzadors o minimitzadors locals: $(1, 1)$, $(-1, -1)$, $(1, -1)$ i $(-1, 1)$. Una anàlisi gràfica ens mostra sense gaires dificultats que els dos primers són maximitzadors globals i els dos darrers, minimitzadors globals.

Considerem ara el cas general en què hi ha més d'una restricció. Observeu que el nombre de restriccions independents no pot excedir el nombre de variables; intuïtivament, cada restricció que afegim redueix el problema en una dimensió i, per tant, el problema no tindrà sentit si no queden dimensions lliures o "graus de llibertat". Ara bé, si el nombre de variables és igual que el de restriccions independents, llavors el problema és trivial, però encara té sentit formular-lo (això ens servirà per a quan tractem les restriccions de desigualtat).

Quan hi ha més d'una restricció, la condició de tangència encara és necessària: si la corba de nivell de la funció objectiu talla el conjunt factible, llavors hi haurà punts factibles en els quals la funció objectiu pren valors més grans i d'altres en què pren valors més petits. !

El que ara requereix una explicació addicional és l'expressió de la condició de tangència en termes dels vectors gradients de la funció objectiu i de les restriccions. Suposem que hi ha n variables i m restriccions, amb $m < n$. Anomenem f la funció objectiu i g_1, $g_2 \ldots g_m$, les funcions que defineixen les restriccions, tal com les hem definit abans (la i-èsima restricció és la corba de nivell 0 de g_i). Les corbes de nivell de f tenen $n - 1$ dimensions dins de $\mathbb{R}^n$, perquè estan definides per una equació. El conjunt factible, que està definit per m equacions dins de $\mathbb{R}^n$, té dimensió $n - m$. La condició de tangència requereix que el pla tangent (de dimensió $n - 1$) a la corba de nivell de f contingui el pla tangent (de dimensió $n - m$) al conjunt factible. Quan ho mirem en termes de les perpendiculars dels vectors gradients, les relacions s'inverteixen: si són linealment independents, els m gradients de les restriccions formen un pla m dimensional; si el gradient de f està contingut dins d'aquest pla, llavors el pla tangent a la corba de nivell de f contindrà el pla tangent al conjunt factible. La condició que busquem és, doncs, que el gradient de la funció objectiu sigui una combinació lineal dels gradients de les restriccions.

$$\nabla f(v) = \lambda_1 \nabla g_1(v) + \lambda_2 \nabla g_2(v) + \cdots + \lambda_m \nabla g_m(v).$$

Juntament amb les m restriccions, això ens proporciona un sistema amb $n+m$ equacions en què les $n+m$ incògnites són les n variables i els m **multiplicadors de Lagrange**. Observeu que hi ha un multiplicador de Lagrange associat a cada restricció.

Per tal de garantir la validesa general de la condició de tangència, necessitem una condició tècnica que ens asseguri que els plans tangents es poden representar a partir dels vectors gradients de les funcions respectives. Aquesta condició és que els gradients de les restriccions són linealment independents. Més endavant estudiarem amb més detall el significat d'aquesta condició. ❗

Regla dels multiplicadors de Lagrange

Siguin f, g_1, $g_2 \ldots g_m$ funcions definides en un subconjunt obert de $\mathbb{R}^n$ i prenent valors a $\mathbb{R}$, amb $m \leq n$. Suposem que totes les funcions són diferenciables amb continuïtat (les seves derivades parcials són funcions contínues).

Sigui v^* un maximitzador o un minimitzador local de f sobre el conjunt de punts que satisfan

$$\begin{aligned} g_1(v_1, v_2, \ldots, v_n) &= 0 \\ g_2(v_1, v_2, \ldots, v_n) &= 0 \\ \cdots\cdots\cdots\cdots \\ g_m(v_1, v_2, \ldots, v_n) &= 0 \end{aligned}$$

Si els gradients de g_1, $g_2 \ldots g_m$ avaluats en v^* són linealment independents, llavors hi ha nombres reals λ_1, $\lambda_2 \ldots \lambda_m$, que compleixen

$$\nabla f(v^*) \;=\; \lambda_1 \nabla g_1(v^*) + \lambda_2 \nabla g_2(v^*) + \cdots + \lambda_m \nabla g_m(v^*).$$ ❗

Observacions:

- Les condicions que hem derivat són necessàries, però no suficients. Això vol dir que en general qualsevol maximitzador o minimitzador les satisfarà, però també hi pot haver solucions que no són ni una cosa ni l'altra.

- Com que hem aproximat les funcions mitjançant la diferenciació, els maximitzadors o minimitzadors en principi només ho seran localment. Necessitarem criteris addicionals per a determinar-ne el caràcter global.

- Per tal que les condicions es compleixin, serà necessari que les tangents estiguin ben definides. Això serà cert sempre que els gradients de les restriccions siguin linealment independents quan els avaluem en els maximitzadors o minimitzadors.

Una manera convenient d'obtenir les condicions que esmenta el teorema anterior és definint una nova funció: el **lagrangià** o funció de Lagrange.

Donades funcions f, g_1, $g_2 \ldots g_m$, tal com en el teorema anterior, definim el **lagrangià** o funció de Lagrange com la funció de $n+m$ arguments:

$$L(v_1, v_2 \ldots v_n, \lambda_1, \lambda_2 \ldots \lambda_m) =$$
$$f(v_1, v_2 \ldots v_n) - \lambda_1\, g_1(v_1, v_2 \ldots v_n) - \ldots$$
$$\ldots - \lambda_2\, g_2(v_1, v_2 \ldots v_n) - \cdots - \lambda_m\, g_m(v_1, v_2 \ldots v_n).$$

La interpretació intuïtiva del lagrangià és que es tracta d'una funció objectiu substitutòria, que ens permet oblidar-nos del fet que les variables han d'obeir les restriccions.

Construïm el lagrangià...

... afegint a la funció objectiu original un terme de penalització per a cada restricció, de tal manera que els termes addicionals de penalització només tenen efecte quan les respectives restriccions no es compleixen.

Per a aplicar la regla dels multiplicadors de Lagrange, en primer lloc definim el lagrangià i després igualem a zero les derivades parcials d'aquest respecte a cada una de les variables originals. Finalment, afegim les restriccions per tal d'obtenir un sistema amb tantes equacions com incògnites.

Per exemple, en el cas de dues variables (x, y) i una restricció, tenim $L(x, y, \lambda) = f(x, y) - \lambda\, g(x, y)$. Igualant a zero les derivades parcials del lagrangià obtenim:

$$\begin{aligned}\frac{\partial L}{\partial x} &= \frac{\partial f}{\partial x} - \lambda\,\frac{\partial g}{\partial x} = 0 \ \text{ i}\\ \frac{\partial L}{\partial y} &= \frac{\partial f}{\partial y} - \lambda\,\frac{\partial g}{\partial y} = 0,\end{aligned}$$

que són exactament les condicions de tangència.

Exemple 4.1. Volem trobar els maximitzadors o minimitzadors de: $f(x, y, z) = x + y + 2z$ sobre la superfície definida per

$$\begin{aligned} x^2 + y^2 + z^2 &= 2\\ x + y + z &= 2.\end{aligned}$$

Definim $g_1(x, y, z) = x^2 + y^2 + z^2 - 2$ i $g_2(x, y, z) = x + y + z - 2$ i, per tant, el lagrangià és

$$L(x, y, z, \lambda_1, \lambda_2) = x + y + 2\,z - \lambda_1\,\left(x^2 + y^2 + z^2 - 2\right) - \lambda_2\,(x + y + z - 2)\,.$$

Per tal d'aplicar el mètode dels multiplicadors de Lagrange, igualem a zero les derivades parcials del lagrangià respecte a x, y i z, i després afegim les restriccions:

$$\begin{aligned}\frac{\partial L}{\partial x} &= 1 - 2\lambda_1 x - \lambda_2 = 0\\ \frac{\partial L}{\partial y} &= 1 - 2\lambda_1 y - \lambda_2 = 0\\ \frac{\partial L}{\partial z} &= 2 - 2\lambda_1 z - \lambda_2 = 0\\ x^2 + y^2 + z^2 &= 2\\ x + y + z &= 2.\end{aligned}$$

La primera equació i la segona impliquen:

$$\lambda_1\,(x - y) = 0.$$

Si fem $\lambda_1 = 0$ en la segona equació i en la tercera ens porta a una contradicció; per tant, podem concloure que $x = y$. Substituint-ho en les dues restriccions acabem per obtenir dues solucions:

$$\begin{aligned}(x, y, z, \lambda_1, \lambda_2) &= (1, 1, 0, -1/2, 2)\\ (x, y, z, \lambda_1, \lambda_2) &= (1/3, 1/3, 4/3, 1/2, 2/3).\end{aligned}$$

Una interpretació geomètrica (el conjunt factible resulta de tallar un casquet en diagonal a una esfera centrada a l'origen de $\mathbb{R}^3$) ens permet veure que el primer punt és un minimitzador global i el segon, un maximitzador global.

Exemple 4.2. Deveu estar familiaritzats amb funcions d'una sola variable, diguem-ne $h(x)$, en què les solucions de $h'(x) = 0$ no són maximitzadors ni minimitzadors, o bé només ho són localment. Per tal de veure que en els problemes amb restriccions d'igualtat ens podem trobar amb extrems locals que no ho són globalment, o amb punts crítics que no són maximitzadors ni minimitzadors, només cal que observem que els dos problemes següents són equivalents:

$$\begin{array}{ll} \max\limits_{x} & h(x) \\ \text{s.a.} & x \in \mathbb{R} \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \max\limits_{(x,y)} & y \\ \text{s.a.} & y = h(x). \end{array}$$

Nota

Us deixem la tasca de provar que, si apliquem el mètode de Lagrange al segon problema, les condicions de primer ordre es redueixen a $h'(x) = 0$.

Si en el problema anterior prenem $h(x) = x^3$, trobarem un punt crític que no és ni maximitzador ni minimitzador. D'altra banda, agafant $h(x) = 5x^2 - 3x^4 + \frac{1}{3}x^6$, trobarem maximitzadors i minimitzadors que ho són localment, però no globalment.

4.2. Regularitat

Hem remarcat abans que, per a aplicar el mètode de Lagrange, necessitem que els gradients de les restriccions avaluats en el maximitzador o minimitzador siguin linealment independents. Quan aquesta condició no es compleix, els gradients de les restriccions no descriuen bé el pla tangent al conjunt factible, i les coses fallen: hi pot haver solucions al problema d'optimització que no apareguin quan apliquem el mètode de Lagrange.

Exemple 4.3. Considerem el problema

$$\begin{array}{ll} \max\limits_{(x,y)} & -x^2 + y \\ \text{s.a.} & y^3 = 0. \end{array}$$

Observeu que la restricció és equivalent a $y = 0$, i substituint-ho la funció objectiu veiem que la solució és $x^* = y^* = 0$. Aplicant el mètode de Lagrange ens trobem amb les equacions

$$\begin{aligned} -2\,x &= 0, \\ 1 - 3\,\lambda\,y^2 &= 0, \\ y^3 &= 0. \end{aligned}$$

La tercera equació implica $y = 0$, i substituint-ho en la segona equació ens quedem amb $1 = 0$ i, per tant, no hi ha cap punt que satisfaci aquestes condicions. Recapitulant: ens hem trobat amb un problema que sabem que té solució, però el mètode de Lagrange ens diu que no n'hi ha cap.

Observeu que el gradient de la restricció és $\nabla g(x, y) = (0, 3y^2)$, i en el punt que sabem que és el maximitzador del problema tenim $\nabla g(0, 0) = (0, 0)$, que és linealment dependent (recordem que el vector de zeros mai no forma part d'un conjunt de vectors linealment independents).

Geomètricament, com que la restricció és l'eix de les x, el pla tangent a la restricció en qualsevol punt continua sent l'eix de les x. Ara bé, com que el gradient de g és el vector 0, el conjunt de vectors ortogonals al gradient de g no és el pla tangent, sinó tot l'espai $\mathbb{R}^2$.

Diem que un punt factible és **regular** si els vectors gradients de les restriccions són linealment independents quan els avaluem en aquest punt.

El teorema que hem vist anteriorment afirma que el mètode dels multiplicadors de Lagrange ens servirà per detectar qualsevol maximitzador o minimitzador local en un problema amb restriccions d'igualtat, sempre que aquests siguin punts regulars.

Observeu que la regularitat és una condició suficient, però no necessària, per a la validesa de la regla dels multiplicadors de Lagrange. Una anàlisi geomètrica ens pot ajudar a veure ens quins casos el mètode de Lagrange detectarà maximitzadors o minimitzadors en els quals falli la condició de regularitat, o deixarà de fer-ho. Ho il·lustrarem amb un parell d'exemples.

Exemple 4.4. Considerem un problema amb dues variables i la restricció $g(x,y) = x^2 + y^3 = 0$. Com que $\nabla g(x,y) = (2x, 3y^2)$, tots els punts factibles són regulars a excepció del punt $(0,0)$. Donada una funció f que té un maximitzador o minimitzador local sobre aquesta superfície a $(0,0)$, volem saber si el mètode de Lagrange tindrà aquest punt com a solució o no. Fixeu-vos que tot es redueix a verificar si es compleix la condició de tangència $\nabla f(0,0) = \lambda\, \nabla g(0,0)$. Com que $\nabla g(x,y) = (0,0)$, el mètode de Lagrange funcionarà si $\nabla f(0,0) = (0,0)$, és a dir, si $(0,0)$ és un punt crític (sense restriccions) de f. Per exemple, si $f(x,y) = y$, el mètode de Langrange no funcionarà, però sí que ho farà si $f(x,y) = x^2 + y^2$.

Així, en el cas d'una sola restricció, el mètode de Lagrange detectarà un maximitzador o minimitzador no regular si, i només si, totes les derivades parcials en aquest punt són nul·les. Quan hi ha més d'una restricció, també hi haurà d'altres casos.

Exemple 4.5. Considerem les dues restriccions

$$\begin{aligned} g_1(x,y,z) &= (x-1)^2 + y^2 = 1 \\ g_2(x,y,z) &= (x+1)^2 + y^2 = 1. \end{aligned}$$

Aquí tenim $\nabla g_1(x,y,z) = \big(2\,(x-1), 2\,y, 0\big)$ i $\nabla g_2(x,y,z) = \big(2\,(x+1), 2\,y, 0\big)$.

Com que el conjunt factible és

$$X = \left\{ (x,y,z) \in \mathbb{R}^3 : x = y = 0 \right\},$$

tots els punts factibles satisfan

$$\begin{aligned} \nabla g_1(x,y,z) &= \big(-2,0,0\big) \\ \nabla g_2(x,y,z) &= \big(2,0,0\big). \end{aligned}$$

És a dir, no hi ha cap punt factible que sigui regular. La funció $f(x,y,z) = x$ és constant sobre el conjunt factible i, per tant, té tant un màxim com un mínim global en cada punt factible; el mètode de Lagrange funcionarà en aquest cas, ja que $\nabla f(x,y,z) = (1,0,0)$ és linealment dependent de qualsevol dels dos gradients de les restriccions. En canvi, la funció $f(x,y,z)$ $= y$, que també és constant sobre el conjunt factible, té com a gradient $\nabla f(x,y,z) = (0,1,0)$, que no depèn linealment dels gradients de les restriccions.

Quina és la diferència en els dos casos? La condició de tangència exigeix que el pla tangent a la corba de nivell de la funció objectiu en el punt crític contingui l'eix de les z, que és el conjunt factible; ara bé, la dependència lineal dels gradients de les restriccions fa que analíticament només puguem detectar un dels molts plans tangents que satisfan aquesta propietat, que és el pla y–z. El mètode de Lagrange funcionarà si, i només si, el gradient de f en el punt crític és ortogonal al pla y–z.

4.3. Criteris de globalitat

En el cas amb restriccions d'igualtat, si volem aplicar directament el criteri de concavitat/convexitat, només ho podrem fer quan totes les restriccions siguin **equacions lineals**, perquè és essencialment l'únic cas (vegeu la nota al marge) en què el conjunt factible és convex.

Quan el conjunt factible sigui convex, el teorema local-global ens permet inferir el caràcter global dels maximitzadors o minimitzadors locals. Ara bé, hi ha una altra manera d'aplicar els criteris de concavitat/convexitat, que és a través d'una anàlisi del lagrangià. A continuació veurem com analitzant la curvatura del lagrangià obtindrem un criteri de globalitat.

Nota

Exceptuant restriccions que tinguin més d'una representació, per exemple, la restricció lineal $y = 0$, que també pot ser escrita en forma no lineal com a $y^3 = 0$. Per a veure que aquesta representació és anòmala, només cal repassar l'exemple en què això feia que el mètode dels multiplicadors de Lagrange no fos aplicable.

Teorema

Sigui (v^*, λ^*) un punt que soluciona les condicions de primer ordre del mètode de Lagrange, en què usem la notació vectorial $v = (v_1, v_2 \ldots v_n)$ i $\lambda = (\lambda_1, \lambda_2 \ldots \lambda_m)$. Definim la funció $h(v) = L(v, \lambda^*)$; és a dir, h és el lagrangià quan fixem els multiplicadors en els valors λ^*, i prenem el vector v com a variables. Llavors tenim:

1) Si h és una funció còncava, v^* és un maximitzador global del problema amb restriccions d'igualtat.

2) Si h és una funció convexa, v^* és un minimitzador global del problema amb restriccions d'igualtat.

Demostració del teorema:

Suposem que h és una funció còncava. Les condicions de tangència impliquen que per a cada $j = 1, 2 \ldots n$ es compleix

$$\frac{\partial L}{\partial v_j}(v^*, \lambda^*) = \frac{\partial h}{\partial v_j}(v^*) = 0,$$

és a dir, totes les derivades parcials de h són 0 quan les avaluem en el vector v^*. Com que h és una funció còncava, això implica que v^* és un maximitzador global de h, és a dir, $h(v^*) \geq h(v)$, per a tot v del domini (satisfaci o no les restriccions d'igualtat). Tenim que, per a tot v:

$$h(v^*) \geq h(v) \iff f(v^*) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v^*) \geq f(v) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v).$$

Com que v^* satisfà les restriccions d'igualtat, tenim que, per cada $i = 1, 2 \ldots m$, $g_i(v^*) = 0$, i per tant, qualsevol que sigui v:

$$f(v^*) \geq f(v) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v).$$

En particular, si v és un punt factible del problema amb restriccions d'igualtat, també es compleix que, per a cada $i = 1, 2 \ldots m$, $g_i(v) = 0$, i això implica que

$$f(v^*) \geq f(v).$$

És a dir, v^* és un maximitzador global del problema amb restriccions d'igualtat. Una prova paral·lela s'usa en el cas de minimització.

Exemple 4.6. Sigui $f(x, y) = x\,y$ i $g(x, y) = x^2 + y^2 - 2$, i considerem el problema de trobar els màxims o mínims de f sobre la superfície definida per $g(x, y) = x^2 + y^2 - 2 = 0$. El lagrangià és $L(x, y, \lambda) = x\,y - \lambda\,(x^2 + y^2 - 2)$, i les condicions de primer ordre són

$$\begin{aligned} \frac{\partial L}{\partial x} = y - 2\,\lambda\,x \ &= 0 \\ \frac{\partial L}{\partial y} = x - 2\,\lambda\,y \ &= 0 \\ x^2 + y^2 \ &= 2. \end{aligned}$$

Aquest sistema d'equacions té quatre solucions, en què (x, y, λ) prenen els valors

$$(1, 1, 1/2), \quad (-1, -1, 1/2), \quad (-1, 1, -1/2) \quad \text{i} \quad (1, -1, -1/2).$$

Per a analitzar el caràcter global de les solucions, comencem per la primera: $(x, y, \lambda) = (1, 1, 1/2)$. Per a definir la funció h de què parla el teorema, fixem $\lambda = 1/2$ i deixem (x, y) com a variables al lagrangià

$$h(x, y) = L(x, y, \frac{1}{2}) = x\,y - \frac{1}{2}\,\left(x^2 + y^2 - 2\right).$$

Les derivades de primer ordre de h són

$$\frac{\partial h}{\partial x} = y - x \qquad \text{i} \qquad \frac{\partial h}{\partial y} = x - y.$$

En particular, com ja sabíem per les condicions de tangència, es compleix

$$\frac{\partial h}{\partial x}(1, 1) = 0 \qquad \text{i} \qquad \frac{\partial h}{\partial y}(1, 1) = 0.$$

La matriu hessiana de h és

$$D^2 h(x, y) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix},$$

que podeu comprovar amb facilitat que és semidefinida negativa. Com que això és cert per a cada $(x, y) \in \mathbb{R}^2$, sabem que h és una funció còncava sobre $\mathbb{R}^2$. Com que h és còncava i les seves derivades parcials són zero en el punt $(1, 1)$, sabem que aquest punt és un maximitzador global de h. El teorema que acabem de veure ens assegura que $(1, 1)$ també és un maximitzador global de la funció $f(x, y) = x\,y$ sobre la superfície definida per $g(x, y) = x^2 + y^2 - 2 = 0$.

Una anàlisi similar...

... per a cada un dels altres punts, ens porta a la conclusió que tant $(1, 1)$ com $(-1, -1)$ són maximitzadors globals, mentre que $(1, -1)$ i $(-1, 1)$ són minimitzadors globals.

Exemple 4.7. Recordem l'exemple que havíem vist anteriorment, en què $f(x, y) = -x^2 + y$ i $g(x, y) = x^2 + y^2 - 1 = 0$. Les condicions de primer ordre ens donen quatre punts crítics:

$$\begin{aligned} (x^*, y^*, \lambda^*) \ &= (0, 1, \tfrac{1}{2}) \\ (x^*, y^*, \lambda^*) \ &= (0, -1, -\tfrac{1}{2}) \\ (x^*, y^*, \lambda^*) \ &= (\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1) \\ (x^*, y^*, \lambda^*) \ &= (-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1), \end{aligned}$$

dels quals els dos primers són maximitzadors locals i els dos darrers són minimitzadors locals. El lagrangià és $L(x, y, \lambda) = -x^2 + y - \lambda\,\left(x^2 + y^2 - 1\right)$.

Considerem el primer punt $(x^*, y^*, \lambda^*) = (0, 1, \frac{1}{2})$. Substituïm $\lambda = \frac{1}{2}$ al lagrangià per formar la funció h que hi correspon: $h(x, y) = -x^2 + y - -\frac{1}{2}\left(x^2 + y^2 - 1\right)$. La matriu hessiana de h és

$$D^2h(x, y) = \begin{pmatrix} -3 & 0 \\ 0 & -1 \end{pmatrix}.$$

Com que $D^2h(x, y)$ és semidefinida negativa en cada $(x, y) \in \mathbb{R}^2$, h és una funció còncava sobre $\mathbb{R}^2$. El nostre teorema de globalitat implica que $(x^*, y^*) = (0, 1)$ és un maximitzador global.

D'altra banda, si prenem el punt $(x^*, y^*, \lambda^*) = (0, -1, -\frac{1}{2})$, i substituïm $\lambda = -\frac{1}{2}$ al lagrangià, obtenim $h(x, y) = -x^2 + y + \frac{1}{2}\left(x^2 + y^2 - 1\right)$. Ara, la matriu hessiana de h és

$$D^2h(x, y) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Com que $D^2h(x, y)$ és indefinida en tot punt, h no és còncava ni convexa i el teorema no és aplicable. De fet, mirant el gràfic del problema, podem observar que aquest punt és un maximitzador local, però no global.

Si finalment considerem el punt $(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1)$ o el $(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1)$, com que tenen la mateixa $\lambda = -1$, els correspon a tots dos la mateixa funció $h(x, y) = = -x^2 + y + \left(x^2 + y^2 - 1\right)$. La matriu hessiana de h és

$$D^2h(x, y) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Com que $D^2h(x, y)$ és semidefinida positiva en cada $(x, y) \in \mathbb{R}^2$, h és una funció convexa sobre $\mathbb{R}^2$. El nostre teorema de globalitat implica que els dos punts són uns minimitzadors globals.

Si a una funció hi afegim termes lineals, les seves derivades segones no canvien i, per tant, la seva concavitat o convexitat és la mateixa. Com que el lagrangià resulta d'afegir les funcions que defineixen les restriccions a la funció objectiu, una conseqüència del teorema anterior és:

Corol·lari

Suposem que, en un problema amb restriccions d'igualtat, totes les restriccions són equacions lineals. Llavors tenim:

1) Si f és una funció còncava, tota solució de les condicions de primer ordre és un maximitzador global.

2) Si f és una funció convexa, tota solució de les condicions de primer ordre és un minimitzador global.

De fet, però, hi ha casos d'interès en els quals podem aplicar els criteris de concavitat/convexitat, però en què el teorema anterior no és aplicable.

Observeu que els supòsits del teorema garanteixen que la funció h (és a dir, el lagrangià) és còncava o convexa sobre tot el domini de les funcions, es compleixin o no les restriccions. Però per al problema d'optimització que nosaltres considerem, en tenim prou amb què la funció objectiu sigui còncava o convexa sobre el conjunt factible, sempre que aquest darrer sigui un conjunt convex. En aquest cas, podem aplicar directament el teorema local-global, sempre que coneguem el caràcter local dels punts que solucionen les condicions de primer ordre. !

Exemple 4.8. En el problema

$$\begin{aligned} \max_{(x,y)} \quad & x\,y \\ \text{s.a.} \quad & x+y=1 \end{aligned}$$

que hem vist al principi, és fàcil veure que les condicions suficients de segon ordre per a un maximitzador local s'apliquen al punt $(x^*, y^*, \lambda) =$ $= (1/2, 1/2, 1/2)$. Com que el conjunt factible és convex, si provem que la funció objectiu és còncava sobre el conjunt factible, podrem aplicar el teorema local-global per concloure que el punt crític és un maximitzador global. Podem veure que la funció objectiu és còncava sobre el conjunt factible simplement expressant, a partir de la restricció, una de les variables en termes de l'altra, amb la qual cosa ens queda una funció d'una variable que és còncava.

Una altra manera de fer-ho seria considerant la forma quadràtica formada a partir de la matriu hessiana de f (avaluada en qualsevol punt factible), però només al llarg de direccions factibles. En qualsevol cas, resulta un exercici senzill que deixem a l'estudiant que hi estigui interessat.

Considerem ara la funció

$$h(x,y) = L(x,y,\tfrac{1}{2}) = x\,y - \tfrac{1}{2}\,(x+y-1).$$

La matriu hessiana de h és

$$D^2h(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Per tant, h no és globalment còncava ni convexa (encara que quan la restringim al conjunt factible sigui còncava), per la qual cosa no podem aplicar el teorema de globalitat ni el seu corol·lari.

Quan ens trobem amb restriccions no lineals i el teorema de globalitat no és aplicable, tenim un criteri una mica *ad-hoc* que pot ser útil. Aquest criteri consisteix en una aplicació del teorema de Weierstrass; la idea és molt senzilla.

Suposem que tenim un problema amb restriccions d'igualtat que compleix les condicions següents:

- El conjunt factible és compacte (fitat i tancat).

- Tots els punts factibles són regulars.

Llavors, com que en els problemes que considerem aquí la funció objectiu sempre és contínua (és diferenciable), sabem el següent:

1) El Teorema de Weierstrass garanteix l'existència d'almenys un maximitzador global i d'almenys un minimitzador global.

2) El fet que tots els punts factibles són regulars garanteix que el mètode dels multiplicadors de Lagrange donarà com a punts crítics qualsevol maximitzador local i qualsevol minimitzador local.

La conjunció dels dos darrers resultats ens assegura que necessàriament hi ha un maximitzador global, que es troba entre els punts crítics, i necessàriament hi ha un minimitzador global, que es troba entre els punts crítics. Quan no hi ha gaires punts crítics, una simple inspecció del valor de la funció objectiu en cada un d'ells ens permet determinar quins optimitzadors són globals i quins són només locals.

Exemple 4.9. Suposem que volem trobar els extrems de $f(x,y) = \frac{2}{3}\,y^3$ sobre la superfície definida per $g(x,y) = x^2 + y^2 - 1 = 0$. El lagrangià és $L(x,y,\lambda) = \frac{2}{3}\,y^3 - \lambda\,\left(x^2 + y^2 - 1\right)$, i les condicions de primer ordre:

$$\begin{aligned} \frac{\partial L}{\partial x} = -2\,\lambda\,x &= 0 \\ \frac{\partial L}{\partial y} = 2\,y^2 - 2\,\lambda\,y &= 0 \\ x^2 + y^2 &= 1. \end{aligned}$$

Hi ha quatre solucions:

$$\begin{aligned}(x^*, y^*, \lambda^*) &= (0, 1, 1)\\ (x^*, y^*, \lambda^*) &= (0, -1, -1)\\ (x^*, y^*, \lambda^*) &= (1, 0, 0)\\ (x^*, y^*, \lambda^*) &= (-1, 0, 0).\end{aligned}$$

Observem que el teorema de globalitat que hem donat al principi d'aquest apartat no és aplicable. Per veure-ho, considerem el primer punt, $(x^*, y^*, \lambda^*) = (0, 1, 1)$. Substituint $\lambda = 1$ al lagrangià, formem la funció h, que és $h(x, y) = \frac{2}{3}y^3 - x^2 - y^2 + 1$. La matriu hessiana de h és

$$D^2h(x, y) = \begin{pmatrix} -2 & 0 \\ 0 & 4y - 2 \end{pmatrix}$$

i, per tant, la funció h no és (globalment) ni còncava ni convexa, ja que la matriu és definida negativa en alguns punts i indefinida en alguns altres. Per tant, no podem aplicar el teorema. L'anàlisi dels altres punts dóna el mateix resultat negatiu.

D'altra banda, com que el conjunt factible és una circumferència, es tracta d'un conjunt compacte (fitat i tancat). El gradient de la restricció és $\nabla g(x, y) = (2x, 2y)$, i només pot ser zero si x i y són zero simultàniament, la qual cosa no compleix cap punt factible; per tant, tots els punts factibles són regulars. Podem concloure que, entre els punts que solucionen les condicions de primer ordre, hi ha necessàriament un maximitzador global i un minimitzador global. Avaluant la funció objectiu als punts crítics trobem:

$$\begin{aligned}f(0, 1) &= 2/3\\ f(0, -1) &= -2/3\\ f(1, 0) &= 0\\ f(-1, 0) &= 0.\end{aligned}$$

La qual cosa ens indica que $(0, 1)$ és el maximitzador global i $(0, -1)$, el minimitzador global. Per a fer una anàlisi completa del problema, observem que la funció objectiu és una transformació estrictament creixent de la funció $q(x, y) = y$ i, per tant, els maximitzadors o minimitzadors de f són els mateixos que els de q. Això ens porta a la conclusió que, dels quatre punts que hem trobat, el primer és un maximitzador global, el segon un minimitzador global, i els altres dos no són ni maximitzadors ni minimitzadors (de fet, són punts de sella de la funció f).

4.4. Condicions de segon ordre

De la mateixa manera que hem fet en problemes sense restriccions, l'anàlisi de la matriu hessiana de la funció objectiu en problemes amb restriccions d'igualtat ens permet de vegades determinar la naturalesa local dels punts crítics. En aquest cas també hi ha un altre element que hem de tenir en compte: la restricció.

Nota important

Aquest apartat inclou material d'un nivell més avançat que el d'aquest curs, per la qual cosa s'ha de considerar a tots els efectes com a material complementari destinat només a qui tingui un interès particular en el tema. Presentem aquest apartat aquí purament com a referència per a l'estudiant interessat. L'estudiant d'aquest curs farà bé de saltar-se'l en una primera lectura del material.

En problemes sense restriccions, la matriu hessiana avaluada en un punt crític ens permet saber, quan aquesta és definida positiva o negativa o és indefinida, si la funció creix o decreix en alguna direcció a partir del punt crític. Quan hi ha restriccions, no totes les direccions ens interessen, sinó només aquelles direccions en què hi ha punts factibles.

Exemple 4.10. Recordem el problema

$$\begin{aligned} \max_{(x,y)} \;& x\,y \\ \text{s.a. }\; & x+y=1, \end{aligned}$$

en què l'únic punt crític és $x^* = y^* = \lambda^* = 1/2$. Si fem la substitució $y = 1 - x$, la funció objectiu esdevé $h(x) = x(1 - x)$, que és una funció estrictament còncava ($h''(x) = -2 < 0$, per a cada x). Això vol dir que les condicions de segon ordre d'aquest problema s'haurien de complir. Ara bé, la matriu hessiana de la funció objectiu és

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

i aquesta matriu és indefinida, la qual cosa sembla indicar que hi ha direccions en què la funció creix i d'altres direccions en què la funció decreix. Però en aquest problema no hem de considerar qualsevol direcció, sinó únicament aquelles direccions en què hi ha punts factibles. Les direccions factibles vénen donades pel pla tangent a la restricció, és a dir, són aquelles direccions perpendiculars al vector gradient $\nabla g(x^*, y^*) = (1, 1)$. Amb aquesta restricció és senzill mostrar que la forma quadràtica generada és definida negativa.

En un problema sense restriccions, la matriu hessiana en un punt crític ens indica com és la curvatura de la funció objectiu al voltant d'aquest punt, i això ens permet conèixer el creixement o decreixement de la funció. Quan hi ha restriccions, per saber si la funció està creixent o decreixent hem de tenir en compte, no solament la curvatura de la funció objectiu, sinó també la curvatura de la restricció. És per això que la matriu hessiana que hem de considerar és la del lagrangià, que a més de la funció objectiu inclou les restriccions.

Exemple 4.11. Considerem el problema

$$\min_{(x,y)} \; -x^2+y$$
$$\text{s.a. } 2\,x^2-y=0.$$

La restricció és $y=2x^2$, que substituïda en la funció objectiu mostra que el problema és equivalent a minimitzar la funció x^2, que és una funció convexa i, per tant, té un mínim global a $x=0$. El lagrangià és $L(x,y,\lambda) = = -x^2+y-\lambda(2x^2-y)$, i les condicions de primer ordre tenen com a única solució: $x^*=0$, $y^*=0$ i $\lambda^*=-1$. Ara bé, la matriu hessiana de la funció objectiu és

$$D^2 f(x,y) = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}.$$

que és semidefinida negativa, i ens indica que la funció decreix o es manté constant en qualsevol direcció **lineal**: això sembla indicar que el punt trobat podria ser un maximitzador, però en cap cas no pot ser un minimitzador. Abans hem vist que el punt és efectivament un minimitzador, i el problema és que hem d'analitzar la curvatura de la funció objectiu al llarg de la direcció **quadràtica** que defineix el conjunt factible. La matriu hessiana del Lagrangià respecte a les variables (x,y) és

$$D^2_{(x,y)} L(x,y,\lambda) = \begin{pmatrix} -2-4\lambda & 0 \\ 0 & 0 \end{pmatrix}.$$

Tenint en compte que en la solució tenim $\lambda^*=-1$, l'element superior esquerre de la hessiana és igual a $+2$ en aquesta solució i, per tant, la possibilitat d'un mínim ja no queda exclosa. De fet, si analitzem la forma quadràtica generada només al llarg de les direccions factibles, veurem que aquesta és definida positiva.

En definitiva, les condicions de segon ordre quan hi ha restriccions d'igualtat vénen donades a partir de la matriu hessiana del lagrangià, avaluant la forma quadràtica corresponent només al llarg de les direccions factibles. Observem que el lagrangià és funció tant de les variables de decisió com dels multiplicadors, i que per comprovar les condicions de segon ordre no ens interessa tota la matriu hessiana del lagrangià, sinó només aquella part generada a partir de les derivades parcials respecte a les variables de decisió.

Denotem per a $D^2_v L(v,\lambda)$ la **submatriu hessiana** formada per les derivades parcials segones del lagrangià respecte a les variables v, i, donat qualsevol vector $h \in \mathbb{R}^n$, sigui $D^2_v L(v,\lambda)\big(h,h\big)$ la forma quadràtica generada per aquella submatriu hessiana aplicada al vector h (escrivim el vector h dues vegades perquè hem de premultiplicar i postmultiplicar la matriu pel vector). Donat un cert v factible i regular, el pla tangent al conjunt factible que passa pel punt v ve donat per[1]

$$T(v) = \{\, h \in \mathbb{R}^n : \; \nabla g_i(v)\cdot h = 0, \; \text{per } 1 \le i \le m \,\}.$$

El punt v

Si el punt no és regular, el pla tangent està contingut dins del conjunt $T(v)$, però pot

Condicions suficients de segon ordre

Suposem que $m \leq n$ i que f, g_1, g_2 ... g_m són diferenciables dues vegades amb continuïtat (les seves derivades parcials segones són funcions contínues). Sigui v^* un punt factible que satisfà les condicions de primer ordre del mètode de Lagrange, i sigui λ^* el corresponent multiplicador de Lagrange. Llavors tenim:

1) *(Condició suficient per maximització local)* Suposem que $D_v^2 L(v^*, \lambda^*)\big(h, h\big) < 0$ per a cada $h \in T(v^*)$ tal que $h \neq 0$. Llavors v^* és un maximitzador local estricte.

2) *(Condició suficient per minimització local)* Suposem que $D_v^2 L(v^*, \lambda^*)\big(h, h\big) > 0$ per a cada $h \in T(v^*)$ tal que $h \neq 0$. Llavors v^* és un minimitzador local estricte.

Exemple 4.12. Siguin $f(x, y) = -x^2 + y$ i $g(x, y) = x^2 + y^2 - 1 = 0$. Les condicions de primer ordre ens donen quatre solucions:

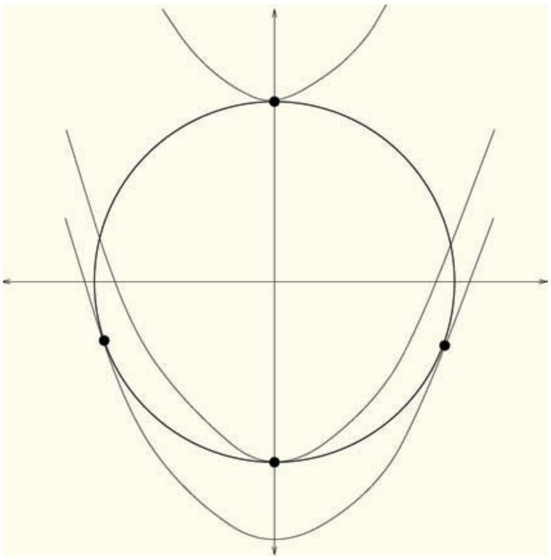
$$\begin{aligned}
(x, y, \lambda) &= (0, 1, \tfrac{1}{2})\\
(x, y, \lambda) &= (0, -1, -\tfrac{1}{2})\\
(x, y, \lambda) &= (\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1)\\
(x, y, \lambda) &= (-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1).
\end{aligned}$$

El vector gradient a la restricció és $\nabla g(x, y) = (2x, 2y)$, per la qual cosa el pla tangent al conjunt factible en el punt (x, y) està format per tots aquells vectors (h_1, h_2) tals que $x\,h_1 + y\,h_2 = 0$. D'altra banda, la matriu de derivades segones del lagrangià respecte a (x, y) i la forma quadràtica corresponent són

$$\begin{pmatrix} -2 - 2\,\lambda & 0 \\ 0 & -2\,\lambda \end{pmatrix} \quad \rightarrow \quad Q(h_1, h_2) = -(2 + 2\,\lambda)\, h_1^2 - 2\,\lambda\, h_2^2.$$

El que hem de fer per a cada punt crític (x, y, λ) és avaluar $Q(h_1, h_2)$ per a tots aquells $(h_1, h_2) \neq (0, 0)$ tals que $x\,h_1 + y\,h_2 = 0$.

$$\begin{array}{llrclcl}
(x, y, \lambda) = (0, 1, \tfrac{1}{2}), & h_2 = 0 & \rightarrow & Q(h_1, h_2) = -3\,h_1^2 < 0 & \rightarrow & \text{Max.}\\
(x, y, \lambda) = (0, -1, -\tfrac{1}{2}), & -h_2 = 0 & \rightarrow & Q(h_1, h_2) = -h_1^2 < 0 & \rightarrow & \text{Max.}\\
(x, y, \lambda) = (\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1), & \sqrt{3}\,h_1 - h_2 = 0 & \rightarrow & Q(h_1, h_2) = 6\,h_1^2 > 0 & \rightarrow & \text{Min.}\\
(x, y, \lambda) = (-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1), & -\sqrt{3}\,h_1 - h_2 = 0 & \rightarrow & Q(h_1, h_2) = 6\,h_1^2 > 0 & \rightarrow & \text{Min.}
\end{array}$$



Paràboles sobre cercle

Les condicions necessàries corresponents són similars. Recordeu que les condicions que usem en la resolució de problemes són les suficients, no les necessàries, ja que les primeres ens asseguren que un punt crític determinat és un maximitzador o minimitzador local.

Condicions necessàries de segon ordre

Suposem que $m \leq n$ i que f, g_1, g_2 … g_m són diferenciables dues vegades amb continuïtat (les seves derivades parcials segones són funcions contínues).

1) *(Condició necessària per maximització local)* Si v^* és un maximitzador local i és un punt regular, llavors existeix λ^* tals que (v^*, λ^*) satisfan les condicions de primer ordre del mètode de Lagrange, i $D_v^2 L(v^*, \lambda^*)\big(h, h\big) \leq 0$ per a cada $h \in T(v^*)$.

2) *(Condició necessària per minimització local)* Si v^* és un minimitzador local i és un punt regular, llavors existeix λ^* tals que (v^*, λ^*) satisfan les condicions de primer ordre del mètode de Lagrange, i $D_v^2 L(v^*, \lambda^*)\big(h, h\big) \geq 0$ per a cada $h \in T(v^*)$.

4.5. Derivació analítica dels resultats

Presentarem aquí de manera més aviat informal una possible derivació de la regla dels multiplicadors de Lagrange i de les condicions de segon ordre.

Comencem definint el marc en què treballarem. Hi ha un cert domini (obert) $D \subset \mathbb{R}^n$ i funcions diferenciables dues vegades amb continuïtat $f : D \to \mathbb{R}$ i $g_i : D \to \mathbb{R}$, $1 \leq i \leq m$, en què $m \leq n$. Definim $X = \{ x \in D : g_i(x) = 0,\ 1 \leq i \leq m \}$. L'argument que fem a continuació es basa en el concepte *geomètric* del pla tangent a la superfície X en un punt x^*. Per definir això formalment, primer necessitem un altre concepte.

Nota important

Aquest apartat inclou material d'un nivell més avançat que el d'aquest curs, per la qual cosa s'ha de considerar a tots els efectes com a material complementari destinat només a qui tingui un interès particular en el tema. Presentem aquest apartat aquí purament com a referència per a l'estudiant interessat. L'estudiant d'aquest curs farà bé de saltar-se'l en una primera lectura del material.

Donat un interval real obert I que conté el 0, una **trajectòria** sobre X és una funció $x : I \to X$ que té derivades segones contínues.

La noció intuïtiva del que seria una trajectòria sobre el conjunt X correspondria a **la imatge** d'una funció que complís les condicions anteriors.

Donat $x^* \in X$, definim el **pla tangent** a X en el punt x^* com el conjunt de totes les derivades $x'(0)$ de totes aquelles trajectòries $x(t)$ sobre X tals que $x(0) = x^*$.

Mitjançant l'ús del teorema de la funció implícita, es pot provar que sempre que x^* sigui un punt regular, el pla tangent està format per tots aquells vectors que són perpendiculars als vectors gradients

$$\{ \nabla g_1(x^*), \nabla g_2(x^*), \ldots, \nabla g_m(x^*) \}$$

(que pel supòsit de regularitat són linealment independents). És a dir, el pla tangent està format per tots aquells vectors u que satisfan

$$\nabla g_i(x^*) \cdot u = 0, \qquad 1 \leq i \leq m.$$

En aquest cas, el subespai vectorial engendrat pels gradients de les restriccions forma el pla perpendicular o **pla normal** al pla tangent, i els dos subespais són complements ortogonals (és a dir, engendren tot $\mathbb{R}^n$).

Per fixar les coses, a partir d'ara suposarem que la funció f té un màxim local en un punt x^* sobre la superfície X.

Nota

El cas d'un mínim local s'analitza de manera similar.

També suposarem que la condició de regularitat és satisfeta a x^*.

Imaginem una certa trajectòria $x(t)$ sobre X que passa pel punt x^*. Llavors la funció d'una variable $h(t) = f[x(t)]$ té un maximitzador local en el punt interior $t = 0$ i, per tant, sabem que $h'(0) = 0$. Aplicant la regla de la cadena:

$$h'(0) = \nabla f[x(0)] \cdot x'(0) = 0.$$

Aquesta condició de primer ordre implica que el vector gradient $\nabla f[x(0)]$ és perpendicular al pla tangent, i (donada la condició de regularitat) això vol dir que aquest vector gradient pertany al pla normal i, per tant, és combinació lineal dels vectors gradients de les restriccions, és a dir:

$$\nabla f[x(0)] = \lambda_1 \nabla g_1[x(0)] + \lambda_2 \nabla g_2[x(0)] + \cdots + \lambda_m \nabla g_m[x(0)].$$

Això és el que nosaltres hem anomenat abans **condició de tangència** de la regla dels multiplicadors de Lagrange.

Per a estudiar la condició suficient de segon ordre, el que hem de fer és analitzar la derivada segona $h''(0)$ de la funció d'una variable h. Aplicant de nou la regla de la cadena, obtenim

$$h''(0) = D^2 f[x(0)]\,(x'(0), x'(0)) + \nabla f[x(0)] \cdot x''(0),$$

en què, com abans, denotem per a $D^2 f[x(0)]\big(x'(0), x'(0)\big)$ la forma quadràtica generada a partir de la matriu hessiana de f i el vector $x'(0)$.

És interessant aturar-nos aquí i interpretar els termes que apareixen en l'expressió de la funció h'', que descriu la curvatura de h. El primer terme és la matriu hessiana de f avaluada al llarg de la direcció tangent donada per x' i, per tant, recull la influència de la curvatura de f. En el segon terme ens trobem amb la derivada segona de la trajectòria x'', que captura la influència de la curvatura de la superfície factible sobre h.

D'altra banda, com que per a tot $t \in I$ es compleix que:

$$\sum_{i=1}^{m} \lambda_i\, g_i[x(t)] = 0,$$

diferenciem dues vegades aquesta funció constant i, avaluant-la al punt 0, obtenim:

$$\sum_{i=1}^{m} \lambda_i \nabla g_i[x(0)] \cdot x''(0) + \sum_{i=1}^{m} \lambda_i\, D^2 g_i[x(0)]\,(x'(0), x'(0)) = 0.$$

Observem que la condició de tangència implica que

$$\nabla f[x(0)] \cdot x''(0) = \sum_{i=1}^{m} \lambda_i \nabla g_i[x(0)] \cdot x''(0)$$

i, per tant,

$$\nabla f[x(0)] \cdot x''(0) = -\sum_{i=1}^{m} \lambda_i \, D^2 g_i[x(0)] \, (x'(0), x'(0)) \,.$$

Aquesta darrera igualtat és interessant, ja que ens ensenya com podem expressar la curvatura de la superfície factible, recollida per la derivada segona de la trajectòria x'', en termes de les matrius hessianes de les restriccions aplicades a la direcció tangent x'. Substituint tot això en l'expressió de la derivada segona de h que hem trobat abans, obtenim finalment

$$h''(0) = D^2 f[x(0)] \, (x'(0), x'(0)) - \sum_{i=1}^{m} \lambda_i \, D^2 g_i[x(0)] \, (x'(0), x'(0)) \,.$$

Així, veiem que la curvatura de h ve donada per la matriu hessiana del lagrangià, aplicada a la direcció tangent x'. Recordem que el lagrangià d'aquest problema és $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i \, g_i(x)$, per la qual cosa, si designem $D_x^2 L(x, \lambda)$ la submatriu hessiana del lagrangià formada per les derivades parcials de segon ordre respecte a les variables x, ens queda

$$h''(0) = D_x^2 L[x(0), \lambda] \, (x'(0), x'(0)) \,.$$

A partir d'aquesta expressió podem derivar tant la condició necessària com la condició suficient de segon ordre. Per exemple, la condició suficient exigirà que la forma quadràtica definida a partir del lagrangià sigui negativa a $x'(0)$, per a qualsevol trajectòria que satisfaci $x(0) = x^*$ i $x'(0) \neq 0$. Dit d'una altra manera, que la forma quadràtica definida per $D_x^2 L(x^*, \lambda)$ prengui valors negatius per a tots els vectors (diferents del 0) que pertanyin al pla tangent.

5. Optimització amb restriccions de desigualtat

5.1. Una primera aproximació a problemes amb desigualtats

Considerem el següent problema de minimització:

$$\begin{aligned}&\min_{(x,y)} \ (x-1)^2+(y-1)^2\\ &\text{s.a. } \ x^2+y^2 \leq 8.\end{aligned}$$

Suposem que un cert punt, diem-ne (a,b), és un minimitzador local del problema (per cert, el teorema de Weierstrass ens assegura que hi ha almenys un minimitzador global). Llavors poden passar dues coses: o bé $a^2+b^2=8$ o bé $a^2+b^2<8$. En el primer cas, direm que la restricció és efectiva en el punt (a,b). Si la restricció és efectiva a (a,b), això vol dir que (a,b) també és solució del problema amb una restricció d'igualtat

$$\begin{aligned}&\min_{(x,y)} \ (x-1)^2+(y-1)^2\\ &\text{s.a. } \ x^2+y^2 = 8\end{aligned}$$

i, per tant, sabem que hi ha un cert λ^* tal que, si $f(x,y)=(x-1)^2+(y-1)^2$ i $g(x,y)=x^2+y^2-8$, llavors el triple (a,b,λ^*) soluciona el sistema d'equacions

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= \lambda\,\frac{\partial g}{\partial x}(x,y)\\ \frac{\partial f}{\partial y}(x,y) &= \lambda\,\frac{\partial g}{\partial y}(x,y)\\ g(x,y) &= 0.\end{aligned} \tag{5.1}$$

Suposem ara que la restricció *no* és efectiva a (a,b), és a dir, que $a^2+b^2<8$. Llavors (a,b) és solució del problema d'optimització

$$\begin{aligned}&\min_{(x,y)} \ (x-1)^2+(y-1)^2\\ &\text{s.a. } \ x^2+y^2 < 8.\end{aligned}$$

Aquest problema no sembla gaire diferent del que teníem originalment, però hi ha una diferència molt important: en ser la restricció una desigualtat estricta, el conjunt de punts factibles és obert, és a dir, tot punt factible està rodejat en qualsevol direcció d'altres punts factibles.

Si repasseu el raonament que vam fer en problemes sense restriccions per justificar que en aquell cas totes les derivades parcials han de ser zero, veureu que també s'aplica al cas que estem considerant, ja que el que és essencial en aquest raonament és que, en qualsevol direcció a partir del punt crític, hi hagi d'altres punts factibles. En definitiva, si (a, b) soluciona el darrer problema que hem escrit, llavors ha de ser solució del sistema de relacions següent:

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= 0 \\ \frac{\partial f}{\partial y}(x,y) &= 0 \\ g(x,y) &< 0. \end{aligned} \tag{5.2}$$

Per tant, qualsevol minimitzador local del problema que hem plantejat originalment satisfà o bé les relacions (5.1), o bé les relacions (5.3). Una manera compacta d'expressar les condicions que han de satisfer tots aquests candidats és la següent: buscarem aquells triples (x, y, λ) que satisfan

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \lambda\,\frac{\partial g}{\partial x}(x,y) \\ \frac{\partial f}{\partial y}(x,y) &= \lambda\,\frac{\partial g}{\partial y}(x,y) \\ g(x,y) &\leq 0 \\ g(x,y) &< 0 \text{ implica que } \lambda = 0. \end{aligned}$$

Observeu que la darrera condició fa que, quan un punt satisfà la desigualtat en forma estricta, la λ que hi correspon sigui zero i que, per tant, les condicions de tangència consisteixin senzillament a igualar les derivades parcials de f a zero.

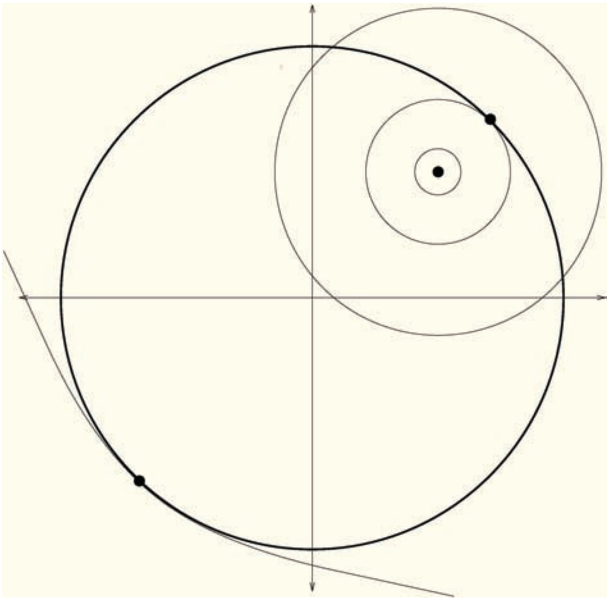
Les equacions que ens queden són

$$\begin{aligned} 2\,x - 2 &= 2\,\lambda\,x \\ 2\,y - 2 &= 2\,\lambda\,y \\ x^2 + y^2 &\leq 8 \\ x^2 + y^2 &< 8 \text{ implica que } \lambda = 0. \end{aligned}$$

Les solucions en termes de (x, y, λ) són

$$(2, 2, 1/2), \quad (-2, -2, 3/2) \quad \text{i} \quad (1, 1, 0).$$

En definitiva, tenim tres candidats en què les variables de decisió valen $(1, 1)$, $(2, 2)$ i $(-2, -2)$. Podem dir alguna cosa més sobre el problema? Fixeu-vos que el conjunt factible és convex, ja que és un cercle (sòlid), i la funció objectiu f és convexa; com que $(1, 1)$ és un punt crític de f (totes les derivades parcials hi són 0), tenim un teorema que ens assegura que $(1, 1)$ és un minimitzador global de f sobre el seu domini. És a dir, la solució del problema formulat originàriament és el punt $(1, 1)$. Per veure què són els altres punts, ens ajudarem amb un gràfic.



Restriccions de desigualtat

Punts crítics de...

...
$f(x,y) = (x-1)^2 + (y-1)^2$
sobre la superfície
$x^2 + y^2 \leq 8$.

Per exemple, considerem el punt $(2,2)$. Si des d'aquest punt ens movem cap a l'origen, f pren valors inferiors, mentre que si ens movem al llarg de la circumferència, f pren valors superiors. O sigui que $(2,2)$ no és ni maximitzador ni minimitzador local. Per què ens ha sortit com a punt crític? El motiu és que per trobar-lo hem aplicat el mètode de Lagrange prenent com a restricció només la circumferència, sobre la qual el punt és, efectivament, un minimitzador local. D'altra banda, a partir del gràfic podem veure que el punt $(-2,-2)$ és un maximitzador global.

En conclusió, combinant els mètodes d'optimització sense restriccions i d'optimització amb restriccions d'igualtat, podem solucionar els problemes amb desigualtats, però hi ha la dificultat que ens surten molts punts crítics espuris. Voldríem afinar una mica més la nostra anàlisi i ser capaços de descartar aquests punts addicionals. El resultat d'aquest afinament és el que anomenem **mètode de Kuhn i Tucker**.

5.2. El mètode de Kuhn-Tucker

En l'exemple anterior, els punts crítics espuris han aparegut quan hem imposat que la restricció es complís com una igualtat. El mètode de Kuhn i Tucker es deriva d'una anàlisi més detallada dels punts crítics que apareixen quan imposem que algunes de les restriccions siguin satisfetes amb igualtat.

El fonament d'aquest mètode consisteix a, donat un cert candidat, usar els gradients de les restriccions efectives en aquest punt per tal de saber cap a on es troba el conjunt factible, i el gradient de la funció objectiu per tal de saber en quina direcció hi ha punts que ens apropen més al nostre objectiu (de maximitzar o minimitzar); si hi ha coincidència entre punts factibles i punts en què ens apropem més al nostre objectiu, llavors el candidat no pot ser un extrem local.

Considerem de nou l'exemple de l'apartat 5.1. Recordem que hem definit $f(x,y) = (x-1)^2 + (y-1)^2$ i $g(x,y) = x^2 + y^2 - 8$. Considerem el punt crític $(x,y) = (2,2)$, en què la λ corresponent és $1/2$, ja que $\nabla f(2,2) = (2,2)$ i $\nabla g(2,2) = (4,4)$ (en els punts crítics amb una restricció d'igualtat, se satisfà $\nabla f = \lambda \nabla g$). El fet que λ és positiva indica que els dos gradients, a més de ser col·lineals, apunten en el mateix sentit. Com que la restricció del problema original és $g(x,y) \leq 0$, sabem que, al voltant del punt $(2,2)$ (en el qual g pren el valor 0 exactament), trobarem punts factibles si ens movem en la direcció oposada a la que marca el gradient.

D'altra banda, el punt $(2,2)$ serà un mínim local de la funció f sempre que no hi hagi punts factibles al voltant, en els quals la funció prengui un valor inferior. El que veiem és, doncs, que $(2,2)$ nop ot ser un minimitzador local, ja que just en la direcció en què sabem que la funció pren un valor inferior (la direcció oposada a la que assenyala el gradient), hi ha punts factibles. Una anàlisi similar mostra que el punt $(-2,-2)$ no pot ser un minimitzador local del problema.

La clau de l'argument anterior consisteix a anar comprovant, per a cada candidat, si hi ha punts factibles en la direcció de disminució del valor de la funció objectiu; si aquest és el cas, podem descartar aquest candidat, ja que no es pot tractar d'un minimitzador local. A més , la comprovació que hem de fer es limita a veure cap a on apunten els gradients (avaluats en el punt en qüestió), tenint en compte la forma en què hem escrit la desigualtat.

En la pràctica,...

... el més senzill és escriure els problemes de maximització i de minimització de tal manera que la comprovació es redueixi a verificar que els gradients de la funció objectiu i de la restricció apuntin en la mateixa direcció, és a dir, que la λ corresponent sigui positiva.

Diem que un problema d'optimització amb dues variables i una restricció de desigualtat està escrit en forma **normalitzada**, si definim les funcions f i g de tal manera que el problema esdevingui, segons si es tracta de maximitzar o de minimitzar:

$$\begin{array}{llcll} \max\limits_{(x,y)} & f(x,y) & \qquad & \min\limits_{(x,y)} & f(x,y) \\ \text{s.a.} & g(x,y) \leq 0 & & \text{s.a.} & g(x,y) \geq 0. \end{array}$$

És a dir, en el cas de maximització hem d'escriure la desigualtat com a $\leq$ i en el de minimització, com a $\geq$.

Les condicions de Kuhn-Tucker també seran diferents segons si volem maximitzar o minimitzar.

Maximització	Minimització
$\nabla f(x,y) = \lambda \nabla g(x,y)$, $g(x,y) \leq 0$, $\lambda \geq 0$, $g(x,y) < 0$ implica $\lambda = 0$.	$\nabla f(x,y) = \lambda \nabla g(x,y)$, $g(x,y) \geq 0$, $\lambda \geq 0$, $g(x,y) > 0$ implica $\lambda = 0$.

El requeriment que $\lambda = 0$ quan la restricció no és efectiva s'anomena **condició d'exclusió o de complementaritat**.

Una manera útil d'escriure les restriccions de factibilitat, no-negativitat de la λ i d'exclusió és, en el cas de maximització:

$$\lambda \geq 0, \qquad g(x,y) \leq 0, \qquad \lambda\, g(x,y) = 0.$$

Això és més senzill que la implicació que escrivíem abans, encara que òbviament és el mateix (tenint en compte les altres condicions).

Quan hi ha moltes variables i restriccions, tot el que acabem de raonar es generalitza. Hem de tenir en compte però que, quan hi ha restriccions de desigualtat, un detall canvia en relació a quan totes les restriccions són d'igualtat. Amb restriccions d'igualtat, el nombre de restriccions independents no pot excedir mai el nombre de variables, però això no és necessàriament cert quan les restriccions són de desigualtat. De fet, és fàcil imaginar problemes ben definits amb dues variables i un nombre arbitrari de restriccions (no trivials) de desigualtat: per exemple, desigualtats lineals que donen lloc com a conjunt factible a un polígon amb un nombre de cares donat (tan gran com es vulgui).

Comencem la part més formal d'aquest apartat amb definicions formals dels conceptes que hem anat veient en els exemples anteriors. A partir d'aquí, suposem que tenim donat un cert problema d'optimització amb restriccions de desigualtat (incloent-hi també el cas en què hi ha tant igualtats com desigualtats).

Diem que una restricció és **efectiva** en un punt si aquest la satisfà amb igualtat.

Exemple 5.1. La restricció $x^2 + y^2 \leq 8$ és efectiva en el punt $(2,2)$, però no ho és en el punt $(1,1)$.

Diem que un punt factible és **regular** si els gradients de totes les restriccions efectives en el punt són linealment independents (quan els avaluem en el punt).

Exemple 5.2. En el conjunt factible definit per:

$$\begin{aligned} (x-1)^2 + (y-2)^2 &\leq 1 \\ (x-1)^2 - y &\leq -1, \end{aligned}$$

tots els punts són regulars excepte $(1,1)$.

Podeu comprovar...

... que tots els punts són regulars excepte (1,1). És fàcil verificar mitjançant una representació gràfica que les dues restriccions són tangents en el punt $(1,1)$.

Diem que un problema de maximització amb n variables i m restriccions de desigualtat està escrit en forma **normalitzada** quan el podem expressar com a:

$$\begin{aligned} \max_{v \in \mathbb{R}^n} \quad & f(v) \\ \text{s.a.} \quad & g_1(v) \leq 0 \\ & g_2(v) \leq 0 \\ & \cdots\cdots\cdots \\ & g_m(v) \leq 0. \end{aligned} \tag{5.3}$$

Diem que un problema de minimització amb n variables i m restriccions de desigualtat està escrit en forma **normalitzada** quan el podem expressar com a:

$$\begin{aligned} \min_{v \in \mathbb{R}^n} \quad & f(v) \\ \text{s.a.} \quad & g_1(v) \geq 0 \\ & g_2(v) \geq 0 \\ & \cdots\cdots\cdots \\ & g_m(v) \geq 0. \end{aligned} \tag{5.4}$$

Considerem el problema normalitzat de maximització 5.3. Suposem que les funcions f, g_1, $g_2 \ldots g_m$ són diferenciables. Les **condicions de Kuhn-Tucker** del problema són

$$\nabla f(v) = \lambda_1 \nabla g_1(v) + \lambda_2 \nabla g_2(v) + \ldots + \lambda_m \nabla g_m(v) \rightarrow \text{Tangència}$$

$$\left.\begin{array}{llllll} \lambda_1 & \geq 0; & g_1(v) & \leq 0; & \lambda_1 g_1(v) & = 0 \\ \lambda_2 & \geq 0; & g_2(v) & \leq 0; & \lambda_2 g_2(v) & = 0 \\ \ldots\ldots & & \ldots\ldots & & \ldots\ldots & \\ \lambda_m & \geq 0; & g_m(v) & \leq 0; & \lambda_m g_m(v) & = 0 \end{array}\right\} \rightarrow \begin{array}{l}\text{Factibilitat} \\ \text{i exclusió}\end{array}$$

Considerem el problema normalitzat de minimització 5.4. Suposem que les funcions f, g_1, $g_2 \ldots g_m$ són diferenciables. Les **condicions de Kuhn–Tucker** del problema són

$$\nabla f(v) = \lambda_1 \nabla g_1(v) + \lambda_2 \nabla g_2(v) + \ldots + \lambda_m \nabla g_m(v) \rightarrow \text{Tangència}$$

$$\left.\begin{array}{llllll} \lambda_1 & \geq 0; & g_1(v) & \geq 0; & \lambda_1 g_1(v) & = 0 \\ \lambda_2 & \geq 0; & g_2(v) & \geq 0; & \lambda_2 g_2(v) & = 0 \\ \ldots\ldots & & \ldots\ldots & & \ldots\ldots & \\ \lambda_m & \geq 0; & g_m(v) & \geq 0; & \lambda_m g_m(v) & = 0 \end{array}\right\} \rightarrow \begin{array}{l}\text{Factibilitat} \\ \text{i exclusió}\end{array}$$

Tal com passa en el cas de restriccions d'igualtat, una manera convenient d'expressar les condicions de Kuhn i Tucker és fer-ho a través del lagrangià del problema d'optimització.

El **lagrangià**, o funció de Lagrange, del problema normalitzat de maximització 5.3, o del problema normalitzat de minimització 5.4, és la funció de $n + m$ arguments:

$$\begin{aligned} & L(v_1, v_2 \ldots v_n, \lambda_1, \lambda_2 \ldots \lambda_m) = \\ & \qquad f(v_1, v_2 \ldots v_n) - \lambda_1 g_1(v_1, v_2 \ldots v_n) - \ldots \\ & \qquad\qquad \ldots - \lambda_2 g_2(v_1, v_2 \ldots v_n) - \cdots - \lambda_m g_m(v_1, v_2 \ldots v_n). \end{aligned}$$

Tant si un problema és de maximització com de minimització, les condicions de tangència són aquelles que resulten d'igualar a zero les derivades parcials del lagrangià respecte a les variables de decisió.

Exemple 5.3. Considerem el problema de minimització

$$\begin{aligned} &\min_{(x,y)} \; x\,y \\ &\text{s.a. } x^2 + y^2 \leq 2. \end{aligned}$$

Per trobar les condicions de Kuhn i Tucker, el primer que fem és escriure'l de forma normalitzada:

$$\begin{aligned} &\min_{(x,y)} \; x\,y \\ &\text{s.a. } -x^2 - y^2 + 2 \geq 0. \end{aligned}$$

Per tant: $f(x,y) = x\,y$ i $g(x,y) = -x^2 - y^2 + 2$. A continuació definim el lagrangià:

$$L(x,y,\lambda) = f(x,y) - \lambda\, g(x,y) = x\,y - \lambda\left(-x^2 - y^2 + 2\right).$$

Les condicions de Kuhn i Tucker són:

$$\begin{aligned} \frac{\partial L}{\partial x} = y + 2\,\lambda\, x &= 0 && (a) \\ \frac{\partial L}{\partial y} = x + 2\,\lambda\, y &= 0 && (b) \\ \lambda &\geq 0 && (c) \\ x^2 + y^2 &\leq 2 && (d) \\ \lambda\left(-x^2 - y^2 + 2\right) &= 0 && (e). \end{aligned}$$

Per a resoldre aquestes condicions es típic començar per les condicions de tangència. Igualant (a) i (b) obtenim:

$$(x-y)\,(1 - 2\,\lambda) = 0.$$

Ara tenim dues possibilitats: o bé $x = y$, o bé $\lambda = 1/2$. Si explorem cada una d'aquestes possibilitats, acabarem per arribar a una solució o bé a una contradicció. Comencem per la primera:

$$x = y \xrightarrow{(a)} x\,(1 + 2\,\lambda) = 0 \xrightarrow{(c)} x = 0 = y \xrightarrow{(e)} \lambda = 0.$$

Per tant, hem arribat a la solució $(x,y,\lambda) = (0,0,0)$. D'altra banda:

$$\lambda = \frac{1}{2} \quad \to \quad \left\{ \begin{aligned} &\xrightarrow{(a)} x + y = 0 \\ &\xrightarrow{(e)} x^2 + y^2 = 2 \end{aligned} \right\} \quad \to \quad \left\{ \begin{aligned} &x = 1, \quad y = -1 \\ &x = -1, \; y = 1. \end{aligned} \right.$$

Per tant, els punts $(1,-1,1/2)$ i $(-1,1,1/2)$ també solucionen les condicions de Kuhn i Tucker.

Les condicions d'exclusió diuen que o bé una restricció és efectiva o bé el multiplicador és zero. En general, només una de les dues coses és certa, encara que no té per què ser així: és fàcil trobar exemples en què les dues coses poden ser certes simultàniament. De fet, aquests casos a vegades poden donar problemes, com quan volem fer una anàlisi de sensibilitat.

Anomenarem **solucions degenerades** aquelles en què es dóna aquesta circumstància.

Exemple 5.4. Considerem el programa de maximització:

$$\max_{(x,y)} \; x\,y$$
$$\text{s.a.} \; x+y \leq 2$$
$$y \geq 1.$$

Normalitzem:

$$\max_{(x,y)} \; x\,y$$
$$\text{s.a.} \; x+y-2 \leq 0$$
$$-y+1 \leq 0.$$

El lagrangià és $L(x,y,\lambda_1,\lambda_2) = x\,y - \lambda_1\,(x+y-2) - \lambda_2\,(-y+1)$, i les condicions de Kuhn-Tucker:

$$\frac{\partial L}{\partial x} = \quad y - \lambda_1 \quad = 0 \qquad (a)$$
$$\frac{\partial L}{\partial y} = x - \lambda_1 + \lambda_2 = 0 \qquad (b)$$
$$\lambda_1 \geq 0, \qquad x+y \leq 2, \qquad \lambda_1\,(x+y-2) = 0 \qquad (c)$$
$$\lambda_2 \geq 0, \qquad y \geq 1, \qquad \lambda_2\,(y-1) = 0 \qquad (d)$$

Per a resoldre aquestes condicions, comencem amb:

$$(a) \;\longrightarrow\; \lambda_1 \geq 1 > 0 \;\xrightarrow{(c)}\; x+y = 2.$$

D'altra banda, (a) i (b) impliquen que $-x+y = \lambda_2$. Sumant aquesta igualtat i la que acabàvem d'obtenir, tenim $y = 1 + \frac{\lambda_2}{2}$. Observem ara que aquesta igualtat, i la condició d'exclusió de (d), només poden ser vàlides si $y = 1$ i $\lambda_2 = 0$. Finalment, a partir d'això obtenim que $x = 1$ i $\lambda_1 = 1$.

Observem que...

... la restricció $y \geq 1$ és efectiva a la solució, però el multiplicador corresponent és zero. Es tracta d'una solució degenerada, tot i que una representació gràfica ajuda a veure que, de fet, el punt trobat és un maximitzador global.

5.3. Justificació alternativa de la no-negativitat dels multiplicadors

Suposem que estem considerant un problema de maximització escrit en la forma

$$\max_{v} \; f(v)$$
$$\text{s.a.} \; g(v) \leq 0.$$

I considerem un punt $\bar{v}$ que satisfà les condicions de Lagrange quan imposem igualtat a la restricció, és a dir $\nabla f(\bar{v}) = \lambda\,\nabla g(\bar{v})$.

El que ara volem fer és considerar quines direccions són factibles al voltant de $\bar{v}$. En els problemes amb restriccions d'igualtat, ja hem vist que les direccions al llarg del pla tangent a la restricció són factibles, és a dir, aquelles direccions h que satisfan

$$\nabla g(\bar{v}) \cdot h = 0.$$

Ara bé, la restricció és $g(v) \leq 0$ i, per tant, també seran factibles aquelles direccions que apuntin cap a on la funció g disminueix de valor (sabem que $g(\bar{v}) = 0$), a més de les direccions en què g es manté constant (que són les que hi ha sobre el pla tangent). L'aproximació lineal (és a dir, l'aproximació de Taylor de primer ordre) a la funció g al voltant de $\bar{v}$ ens mostra que les direccions factibles vindran donades per tots aquells vectors h que satisfan

$$\nabla g(\bar{v}) \cdot h \leq 0.$$

Ara bé, si hi ha alguna direcció factible h tal que $\nabla f(\bar{v}) \cdot h > 0$, llavors el punt $\bar{v}$ no pot ser un maximitzador local, ja que la funció f augmenta de valor si ens movem al llarg d'aquella direcció. Per tant, una condició necessària per tal que $\bar{v}$ sigui un maximitzador local és que

$$\nabla f(\bar{v}) \cdot h \leq 0$$

per a tota direcció factible h. Com que $\bar{v}$ satisfà les condicions de Lagrange d'igualtat, sabem que

$$\nabla f(\bar{v}) \cdot h = \lambda \, \nabla g(\bar{v}) \cdot h$$

i com que totes les h factibles satisfan $\nabla g(\bar{v}) \cdot h \leq 0$, llavors la condició necessària de maximització estarà garantida sempre que exigim que $\lambda \geq 0$.

5.4. Condicions de primer ordre

Tal com passa en el cas de restriccions d'igualtat, i exactament pels mateixos motius, les condicions de Kuhn i Tucker seran necessàries sempre que els maximitzadors o minimitzadors locals siguin punts regulars.

Condicions de primer ordre

Suposem que v^* és un maximitzador local del problema 5.3 o un minimitzador local del problema 5.4, i suposem addicionalment que les funcions f, g_1, $g_2 \ldots g_m$ són diferenciables amb continuïtat. Llavors, si v^* és un punt regular, aquest punt satisfà les condicions de Kuhn i Tucker del problema d'optimització respectiu.

És a dir, quan tots els punts factibles són regulars, podem eliminar com a candidats a maximitzadors o minimitzadors tots aquells punts que no satisfan les condicions de Kuhn i Tucker.

Exemple 5.5. En el problema

$$\begin{aligned} \min_{(x,y)} \; & x\,y \\ \text{s.a.} \; & x^2 + y^2 \leq 2 \end{aligned}$$

hem trobat que les solucions de les condicions de Kuhn i Tucker són els punts en què (x, y, λ) valen $(0, 0, 0)$, $(1, -1, 1/2)$ i $(-1, 1, 1/2)$. En l'apartat sobre restriccions en forma d'igualtat vam veure que tots els punts que hi ha sobre la circumferència són regulars. D'altra banda, qualsevol punt en què la restricció no és efectiva és regular, ja que la regularitat és una condició que s'ha de satisfer només en les restriccions efectives. En conclusió, tots els punts factibles són regulars. El teorema anterior implica que, si hi ha un minimitzador local, aquest és un dels tres punts que acabem d'escriure.

El teorema que hem vist no diu que, si les condicions de Kuhn i Tucker tenen solucions, totes siguin la solució del problema d'optimització, ni tan sols que aquest problema en tingui cap. Només afirma que les condicions de Kuhn i Tucker detectaran qualsevol maximitzador o minimitzador local.

Exemple 5.6. Considerem el problema (normalitzat) de minimització

$$\begin{aligned} \min_{(x,y)} \; & y \\ \text{s.a.} \; & -x^3 + y \geq 0. \end{aligned}$$

El lagrangià és $L(x, y, \lambda) = y - \lambda\,(-x^3 + y)$, i les condicions de Kuhn i Tucker

$$\frac{\partial L}{\partial x} = 3\,\lambda\,x^2 = 0$$
$$\frac{\partial L}{\partial y} = 1 - \lambda = 0$$
$$\lambda \geq 0, \qquad y \geq x^3, \qquad \lambda\,(-x^3 + y) = 0.$$

És fàcil veure que l'única solució és $(x, y, \lambda) = (0, 0, 1)$, i que tots els punts factibles són regulars. Això vol dir que hem trobat la solució del problema de minimització? Observeu que, prenent valors negatius molt grans de la variable x, podem fer que la variable y (la funció objectiu) sigui tan petita com vulguemi que, per tant, satisfaci la restricció. Això vol dir que la funció objectiu no és afitada inferiorment i, per tant, que no té un valor mínim.

5.5. Condicions de segon ordre

Les condicions locals de segon ordre no són més que la generalització natural del que passa en el cas de restriccions d'igualtat, amb una petita qualificació per al cas de suficiència. En el cas de restriccions de desigualtat, les restriccions que no siguin efectives no compten i les podem ignorar pel que fa a condicions locals de segon ordre. Pel que fa a les restriccions efectives, hem de considerar les direccions factibles a què donen lloc; hi ha dos tipus de direccions factibles: les que estan sobre el pla tangent i les que van cap a l'interior del conjunt factible. Però de les direccions factibles que van cap a l'interior del conjunt factible ja ens n'ocupem en exigir que els multiplicadors siguin no negatius. Per tant, la condició local de segon ordre és anàloga al cas amb restriccions d'igualtat, prenent només les restriccions efectives.

Nota important

Aquest apartat inclou material d'un nivell més avançat que el d'aquest curs, per la qual cosa s'ha de considerar a tots els efectes com a material complementari destinat només a qui tingui un interès particular en el tema. Presentem aquest apartat aquí purament com a referència per a l'estudiant interessat. L'estudiant d'aquest curs farà bé de saltar-se'l en una primera lectura del material.

Condicions Necessàries de segon ordre

Suposem que les funcions f, g_1, $g_2 \ldots g_m$ són diferenciables dues vegades amb continuïtat. Aleshores:

1) Suposem que v^* és un maximitzador local del problema normalitzat, i que és un punt regular de les restriccions. Llavors existeix λ^* tal que (v^*, λ^*) satisfan les condicions de Kuhn i Tucker, i a més la matriu hessiana $D_v^2 L(v^*, \lambda^*)$ és semidefinida negativa sobre el pla tangent a les restriccions efectives.

2) Suposem que v^* és un minimitzador local del problema normalitzat, i que és un punt regular de les restriccions. Llavors existeix λ^* tal que (v^*, λ^*) satisfan les condicions de Kuhn i Tucker, i a més la matriu hessiana $D_v^2 L(v^*, \lambda^*)$ és semidefinida positiva sobre el pla tangent a les restriccions efectives.

Condicions suficients de segon ordre

Suposem que les funcions f, g_1, $g_2 \ldots g_m$ són diferenciables dues vegades amb continuïtat, i que existeix λ^* tal que (v^*, λ^*) satisfan les condicions de Kuhn i Tucker. Aleshores:

1) Si la matriu hessiana $D_v^2 L(v^*, \lambda^*)$ és definida negativa sobre el pla tangent a totes les restriccions i tals que $\lambda_i^* > 0$, llavors v^* és un maximitzador local estricte del problema normalitzat.

2) Si la matriu hessiana $D_v^2 L(v^*, \lambda^*)$ és definida positiva sobre el pla tangent a totes les restriccions i tals que $\lambda_i^* > 0$, llavors v^* és un minimitzador local estricte del problema normalitzat.

Observeu que,...

... per a les condicions suficients, necessitem considerar no les restriccions efectives, sinó aquelles que tenen associat un multiplicador estrictament positiu, és a dir, les restriccions efectives no degenerades.

5.6. Criteris de globalitat

Una qüestió més restrictiva és saber sota quines condicions tot punt que compleixi les condicions de Kuhn i Tucker és necessàriament una solució del problema d'optimització. Essencialment, requerirem que el conjunt factible sigui convex i que la funció objectiu sigui còncava o convexa (o generalitzacions), segons si volem maximitzar o minimitzar. Un resultat molt més restrictiu del que és necessari, però extremament senzill de demostrar, i de comprovar a la pràctica, és el següent.

Teorema

1) Considerem el problema normalitzat de maximització 5.3. Suposem que les funcions f, g_1, $g_2 \dots g_m$ són diferenciables. Addicionalment, suposem que f és una funció còncava i que g_1, $g_2 \dots g_m$ són funcions convexes. Llavors, tot punt que solucioni les condicions de Kuhn i Tucker és un maximitzador global del problema.

2) Considerem el problema normalitzat de minimització 5.4. Suposem que les funcions f, g_1, $g_2 \dots g_m$ són diferenciables. Addicionalment, suposem que f és una funció convexa i que g_1, $g_2 \dots g_m$ són funcions còncaves. Llavors, tot punt que solucioni les condicions de Kuhn i Tucker és un minimitzador global del problema.

Exemple 5.7. Considerem el problema

$$\begin{aligned} \max_{(x,y)} \ & x - y + x\,y - \frac{1}{4}\,x^4 - \frac{1}{2}\,x^2 - y^2 \\ \text{s.a. } & x^2 + y^4 \leq 4 \\ & x \geq 0 \\ & y \geq 0. \end{aligned}$$

Normalitzem

$$\begin{aligned} \max_{(x,y)} \ & x - y + x\,y - \frac{1}{4}\,x^4 - \frac{1}{2}\,x^2 - y^2 \\ \text{s.a. } & x^2 + y^4 - 4 \leq 0 \\ & -x \leq 0 \\ & -y \leq 0. \end{aligned}$$

Definim

$$\begin{aligned} f(x,y) &= x - y + x\,y - \frac{1}{4}\,x^4 - \frac{1}{2}\,x^2 - y^2 \\ g_1(x,y) &= x^2 + y^4 - 4 \end{aligned}$$

$$\begin{aligned} g_2(x,y) &= -x \\ g_3(x,y) &= -y. \end{aligned}$$

És una qüestió rutinària comprovar, usant les respectives matrius hessianes, que f és una funció còncava i g_1, g_2 i g_3 són funcions convexes. Per tant, el teorema anterior implica que qualsevol solució de les condicions de Kuhn i Tucker és automàticament un maximitzador global.

Exemple 5.8. Considerem el problema

$$\begin{aligned} \min_{(x,y,z)} \quad & x+y+z+e^x+y^2+z^4 \\ \text{s.a.} \quad & x+y+z \geq -2 \\ & x^2+y^2+z^2 \leq 8. \end{aligned}$$

Normalitzem

$$\begin{aligned} \min_{(x,y,z)} \quad & x+y+z+e^x+y^2+z^4 \\ \text{s.a.} \quad & x+y+z+2 \geq 0 \\ & -x^2-y^2-z^2+8 \geq 0. \end{aligned}$$

Definim

$$\begin{aligned} f(x,y,z) &= x+y+z+e^x+y^2+z^4 \\ g_1(x,y,z) &= x+y+z+2 \\ g_2(x,y,z) &= -x^2-y^2-z^2+8. \end{aligned}$$

És una qüestió rutinària comprovar, usant les respectives matrius hessianes, que f és una funció convexa, i g_1 i g_2 són funcions còncaves. Per tant, el teorema anterior implica que qualsevol solució de les condicions de Kuhn i Tucker és automàticament un minimitzador global.

El resultat del teorema anterior, que ens servirà per a comprovar en molts casos la validesa dels punts obtinguts mitjançant les condicions de Kuhn i Tucker, es deriva de manera força immediata dels resultats que caracteritzen les funcions còncaves i convexes. A continuació en presentem una demostració.

Demostració del teorema

Provarem només el cas de maximització. Suposem que (v^*, λ^*) satisfan les condicions de Kuhn i Tucker. Un parell d'observacions elementals:

1) el negatiu d'una funció convexa és una funció còncava; i

2) la suma de funcions còncaves és una funció còncava. En conseqüència, la concavitat de f i la convexitat de les restriccions implica que el lagrangià

$L(v, \lambda^*) = f(v) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v)$ és una funció còncava en la variable v, ja que $\lambda^* \geq 0$. Per tant, donat qualsevol punt factible $\bar{v}$ tenim:

$$L(\bar{v}, \lambda^*) - L(v^*, \lambda^*) \leq \nabla_v L(v^*, \lambda^*) \, (\bar{v} - v^*).$$

Però la condició de tangència implica que $\nabla_v L(v^*, \lambda^*) = 0$, per la qual cosa

$$L(\bar{v}, \lambda^*) \leq L(v^*, \lambda^*).$$

Substituint en les definicions dels lagrangians respectius, tenim:

$$f(\bar{v}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{v}) \leq f(v^*) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v^*)$$

Ara bé, la condició d'exclusió implica que, per a cada i, $\lambda_i^* \, g_i(v^*) = 0$ i, per tant,

$$f(\bar{v}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{v}) \leq f(v^*).$$

També sabem que, per a cada i, $\lambda_i^* \geq 0$ i $g_i(\bar{v}) \leq 0$, o sigui que $\lambda_i^* \, g_i(\bar{v}) \leq 0$ (recordem que $\bar{v}$ no té per què satisfer la condició de KT i que, per tant, el darrer producte no té per què ser zero), o sigui que:

$$f(\bar{v}) \leq f(\bar{v}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{v}) \leq f(v^*).$$

Així, doncs, la funció objectiu pren un valor més gran en v^* que en $\bar{v}$. Com que $\bar{v}$ era un punt factible qualsevol, podem concloure que v^* és un maximitzador global.

El fet que les funcions que defineixen les restriccions siguin còncaves o convexes, segons si minimitzem o maximitzem, garanteix que el conjunt factible sempre és un conjunt convex.

Per a poder generalitzar aquest resultat, necessitarem generalitzar els conceptes de concavitat i de convexitat.

Sigui X un subconjunt convex de $\mathbb{R}^n$.

1) Diem que la funció $f : X \to \mathbb{R}$ és **quasicòncava** si, per a cada $v \in X$, el contorn superior $\{ \, w \in X : f(w) \geq f(v) \, \}$ és un conjunt convex.

2) Diem que la funció $f : X \to \mathbb{R}$ és **quasiconvexa** si, per a cada $v \in X$, el contorn inferior $\{ \, w \in X : f(w) \leq f(v) \, \}$ és un conjunt convex.

Quan f és una funció contínua, que és l'únic cas que nosaltres considerem, la definició és equivalent si la fem en termes dels contorns definits amb desigualtats febles o estrictes.

En general, comprovar si una funció és quasicòncava o quasiconvexa pot arribar a ser una tasca algèbricament força complexa. La manera més senzilla de fer-ho és a través de la següent proposició, que també ens serveix per a entendre l'origen de la terminologia.

Podeu...

... comprovar la validesa d'aquesta proposició a partir de les definicions algebraiques de funcions còncaves i convexes.

Proposició 4. *Tota funció còncava és quasicòncava, i tota funció convexa és quasiconvexa.*

Les equivalències següents són bàsicament reescriptures de les definicions respectives.

Proposició 5. *Sigui X un subconjunt convex de $\mathbb{R}^n$.*

1) *La funció $f : X \to \mathbb{R}$ és quasicòncava si, i només si, per a cada v i w de X, i per a cada $0 \leq t \leq 1$, es compleix que $f[tv + (1-t)w] \geq \min\{\, f(v), f(w)\,\}$.*

2) *La funció $f : X \to \mathbb{R}$ és quasiconvexa si, i només si, per a cada v i w de X, i per a cada $0 \leq t \leq 1$, es compleix que $f[tv + (1-t)w] \leq \max\{\, f(v), f(w)\,\}$.*

Recordem que dos problemes d'optimització són equivalents sempre que tinguin el mateix conjunt factible i les respectives funcions objectiu siguin transformacions estrictament creixents l'una de l'altra. El resultat següent ens diu que, si volem estudiar problemes que satisfacin la condició de globalitat que hem exposat al principi d'aquest apartat, no podem evitar considerar també problemes en els quals la funció objectiu sigui quasicòncava o quasiconvexa.

Proposició 6. *Tota transformació estrictament creixent d'una funció quasicòncava resulta en una nova funció quasicòncava. En particular, una transformació estrictament creixent d'una funció còncava és una funció quasicòncava (no necessàriament còncava).*

Tota transformació estrictament creixent d'una funció quasiconvexa resulta en una nova funció quasiconvexa. En particular, una transformació estrictament creixent d'una funció convexa és una funció quasiconvexa (no necessàriament convexa).

Exemple 5.9. La funció $f(x) = x^2$, definida sobre $X = [0, \infty)$, és quasicòncava, però no és còncava. La comprovació de la quasiconcavitat és senzilla a partir de la mateixa definició, ja que per a cada nombre real $t \geq 0$, el contorn superior de nivell t és l'interval $[\sqrt{t}, \infty)$. Com que, per a tot $x \in X$, $f''(x) = 2 > 0$, la funció no és còncava.

Exemple 5.10. La funció $f(x) = \log(x)$, definida sobre $X = (0, \infty)$, és quasiconvexa, però no és convexa. Per a cada nombre real t, el contorn inferior de nivell t és l'interval $(0, e^t]$, i, per tant, la funció és quasiconvexa. D'altra banda, per a cada $x \in X$, $f''(x) = -x^{-2} < 0$, per la qual cosa la funció no és convexa.

Exemple 5.11. La funció $f(x, y) = xy$, definida sobre

$$X = \mathbb{R}^2_{++} = \left\{ (x, y) \in \mathbb{R}^2 : x > 0,\ y > 0 \right\},$$

és quasicòncava, però no és còncava. Per a veure que no és còncava, només cal considerar la seva matriu hessiana

$$D^2 f(x, y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

que és indefinida en cada punt del domini. Per a veure que la funció és quasicòncava, considerem la funció $h(x, y) = \log[f(x, y)] = \log(x) + \log(y)$, que és una transformació estrictament creixent de f i que té per matriu hessiana

$$D^2 h(x, y) = \begin{pmatrix} -x^{-2} & 0 \\ 0 & -y^{-2} \end{pmatrix},$$

que és definida negativa en cada punt del domini, per la qual cosa h és (estrictament) còncava. Recordem de passada que la relació "ser transformació estrictament creixent de" és simètrica, és a dir, el fet que h ho és de f implica que f també ho és de h; en el nostre cas, és senzill comprovar que $f(x, y) = e^{h(x,y)}$.

Exemple 5.12. La funció $f(x, y) = \log(1+x^2+y^2)$, definida sobre $X = \mathbb{R}^2$, és quasiconvexa, però no és convexa. Amb la indicació que la funció $h(x, y) = 1 + x^2 + y^2$ és convexa, us deixem que comproveu la resta d'afirmacions tal com hem fet en el darrer exemple.

Deixem a l'estudiant...

... la tasca de comprovar la validesa de les afirmacions que hi ha a continuació.

Exemple 5.13. Sigui $X \subset \mathbb{R}^n$ i $f : X \to \mathbb{R}$ una funció quasicòncava. Definim una nova funció h sobre el mateix domini a través de $h(v) = -f(v)$, per a cada $v \in X$. Llavors h és una funció quasiconvexa.

Exemple 5.14. Sigui $X \subset \mathbb{R}^n$ i $f : X \to \mathbb{R}$ una funció quasicòncava. Sigui $q : \mathbb{R} \to \mathbb{R}$ una funció estrictament decreixent. Definim una nova funció $h : X \to \mathbb{R}$ a través de $h(v) = q[f(v)]$ (és a dir, h és una transformació estrictament *decreixent* de f). Llavors h és una funció quasiconvexa.

Quan totes les restriccions d'un problema normalitzat de maximització estiguin definides per funcions quasiconvexes, el conjunt factible sempre serà un conjunt convex. El mateix passarà en un problema de minimització quan les restriccions estiguin definides per funcions quasicòncaves. En aquest cas podem generalitzar el teorema de globalitat que hem donat a l'inici de l'apartat.

Teorema

Suposem que el problema normalitzat de maximització 5.3 satisfà les hipòtesis següents:

1) La funció objectiu f és còncava, i g_i és una funció quasiconvexa, per a $i = 1, 2 \ldots m$.

2) Totes les funcions són diferenciables amb continuïtat.

Llavors tot punt que satisfà les condicions de Kuhn i Tucker és un maximitzador global.

Teorema

Suposem que el problema normalitzat de minimització 5.4 satisfà les hipòtesis següents:

1) La funció objectiu f és convexa, i g_i és una funció quasicòncava, per a $i = 1, 2 \ldots m$.

2) Totes les funcions són diferenciables amb continuïtat.

Llavors tot punt que satisfà les condicions de Kuhn i Tucker és un minimitzador global.

Quan volem que la funció objectiu sigui quasicòncava o quasiconvexa, els resultats es generalitzen llevat d'un cas excepcional.

Teorema

Suposem que el problema normalitzat de maximització 5.3 satisfà les hipòtesis següents:

1) La funció objectiu f és quasicòncava, i g_i és una funció quasiconvexa, per a $i = 1, 2 \dots m$.

2) Totes les funcions són diferenciables amb continuïtat.

Llavors tot punt que satisfà les condicions de Kuhn i Tucker és un maximitzador global, sempre que no tingui totes les derivades parcials iguals a zero.

Teorema

Suposem que el problema normalitzat de minimització 5.4 satisfà les hipòtesis següents:

1) La funció objectiu f és quasiconvexa, i g_i és una funció quasicòncava, per a $i = 1, 2 \dots m$.

2) Totes les funcions són diferenciables amb continuïtat.

Llavors tot punt que satisfà les condicions de Kuhn i Tucker és un minimitzador global, sempre que no tingui totes les derivades parcials iguals a zero.

Exemple 5.15. Considerem el problema

$$\begin{aligned} \max_{(x,y)} \;& x\,y \\ \text{s.a.} \;& x^2 + y^2 \leq 2 \\ & x \geq 0 \\ & y \geq 0. \end{aligned}$$

Normalitzant, tenim:

$$\begin{aligned} \max_{(x,y)} \;& x\,y \\ \text{s.a.} \;& x^2 + y^2 - 2 \leq 0 \\ & -x \leq 0 \\ & -y \leq 0. \end{aligned}$$

Abans hem vist que la funció objectiu $f(x, y) = xy$ és quasicòncava sobre el conjunt factible. D'altra banda, les funcions que defineixen les restriccions (després de normalitzar) són totes convexes i, per tant, quasiconvexes. Les condicions de Kuhn-Tucker tenen per solució els dos punts en què

(x, y, λ) valen $(1, 1, 1/2)$ i $(0, 0, 0$. Ara bé, $\nabla f(1, 1) = (1, 1) \neq (0, 0)$, però $\nabla f(0, 0) = (0, 0)$. El teorema anterior ens assegura que el primer punt és un maximitzador global, però no es pronuncia sobre el segon punt. De fet, $(0, 0)$ *no* és un maximitzador, sinó que és un punt de sella de f, que apareix aquí ja que tot punt crític (sense restriccions) d'una funció sempre és solució de les condicions de KT, si està dins del conjunt factible. Si la funció fos còncava això no seria un problema, ja que tot punt en què totes les derivades parcials són zero és automàticament un maximitzador global, però això no és necessàriament cert quan una funció és quasicòncava, tal com il·lustra aquest exemple.

5.7. Problemes amb restriccions d'igualtat i de desigualtat

Quan tenim restriccions tant d'igualtat com de desigualtat, les condicions de Kuhn i Tucker són les mateixes, amb l'excepció que els multiplicadors associats a restriccions d'igualtat no tenen per què ser no negatius. Amb aquesta modificació, tots els resultats locals que hem derivat encara són vàlids per a aquest cas.

Pel que fa a les condicions de globalitat en termes de concavitat i convexitat, els teoremes que hem donat anteriorment continuen sent aplicables sempre que les restriccions d'igualtat siguin lineals (ho podem veure escrivint una igualtat com dues desigualtats amb signes oposats: $a = b \iff a \leq b$ i $a \geq b$).

Si tenim restriccions d'igualtat que no són lineals, podem aplicar un argument de globalitat com el que hem fet servir en el cas en què només hi ha restriccions d'igualtat. És a dir, considerem per a cada solució de les condicions de KT el lagrangià després de substituir el valor òptim dels multiplicadors, i mirem si la funció resultant és còncava (si maximitzem) o convexa (si minimitzem); si la resposta és afirmativa, llavors la solució corresponent és global.

5.8. Resum

En aquest mòdul, hem après unes tècniques d'optimització que ens permeten resoldre problemes no lineals sense restriccions o amb restriccions en forma d'igualtat o de desigualtat.

També hem vist condicions sota les quals podem afirmar que les solucions que trobem amb les nostres tècniques són de caire global. No hem vist demostracions matemàtiques, però en canvi sí que hem vist una il·lustració del per què les tècniques funcionen i exemples de casos en què no ho fan.

L'èmfasi de la presentació ha estat dotar-nos de tècniques per a resoldre problemes, i de suficient intuïció per a entendre quan i per què poden fallar aquestes tècniques.

Activitats

1. Analitzeu el punt crític $(0,0,0)$ de la funció $f(x,y,z) = x^3 + 3xy + 3xz + y^3 + 3yz + z^3$ mitjançant el criteri dels menors principals.

2. Donats els paràmetres a i b, sigui $f(x,y) = \frac{1}{2}\,a\,x^2 - a\,x\,y + \frac{1}{2}\,b\,y^2$.
a) Proveu que $(0,0)$ és un punt crític de f.
b) Trobeu tres valors del parell (a,b) que fan que $(0,0)$ sigui un maximitzador, minimitzador i punt de sella, respectivament. (*Nota*: la vostra resposta ha de ser de la forma: si $a = 12345$ i $b = 8888776$, llavors $(0,0)$ és un maximitzador, perquè ..., on ... conté la vostra justificació.)

3. Sigui $f\colon \mathbb{R} \to \mathbb{R}$ definida per $f(x) = 3\,x^4 - 4\,x^3$. Useu el fet que f és diferenciable per a respondre les qüestions següents:
a) En quins intervals f és creixent i en quins és decreixent?
b) En quins intervals f és còncava i en quins és convexa?

4. Considereu el problema

$$\min_{(x,y)} \; 2\,x^2 - 12\,x + y^2$$
$$\text{s.a. } \; y = 2\,x.$$

a) Solucioneu-lo substituint el valor de y en la funció objectiu.
b) Solucioneu-lo mitjançant el mètode de Lagrange, i comproveu que els resultats són els mateixos.
c) Discutiu el caràcter global de les solucions.

5. Trobeu els quatre punts crítics de la funció $f(x,y) = y - x^2$ sobre la superfície definida per $g(x,y) = x^2 + y^2 = 1$. El conjunt factible és una circumferència i, per tant, és compacte, però no és un conjunt convex. Discutiu el caràcter global dels punts crítics.

6. Trobeu els punts crítics de la funció $f(x,y) = x^2 + (y+1)^2$ sobre la superfície definida per $x^2 + y = 1$.

7. Considereu el problema següent:

$$\min_{(x,y)} \; x^2 + y^2$$
$$\text{s.a. } \; (x+1)^3 + y^2 = 0.$$

a) Mostreu que el mètode de Lagrange no dóna cap solució.
b) Substituïu el valor de y^2 que dóna la restricció en la funció objectiu i mostreu que el problema té solució.
c) Sabríeu dir per què el mètode de Lagrange ha fallat?

8. Considereu el problema

$$\begin{aligned} \min_{(x,y)} \;& x - y \\ \text{s.a. } \;& x^2 + y \le 1 \\ & x \le 0. \end{aligned}$$

Escriviu les condicions de Kuhn i Tucker, trobeu tots els punts crítics i després discutiu-ne el caràcter global.

9. Considereu el problema

$$\begin{aligned} \max_{(x,y)} \;& -x^2 + y \\ \text{s.a. } \;& x^2 + y^2 \le 1 \\ & y \ge 0. \end{aligned}$$

Escriviu les condicions de Kuhn i Tucker, trobeu tots els punts crítics i després discutiu-ne el caràcter global.

10. Considereu el problema

$$\max_{(x,y)} \; 2\,x + y$$
$$\text{s.a. } \; 2\,x - x^2 \ge y.$$

Escriviu les condicions de Kuhn i Tucker, trobeu tots els punts crítics i després discuteix el seu caràcter global.

Exercicis d'autoavaluació

1. Trieu l'opció correcta. La funció:

		x			
		0	1	2	3
y	1	0	1	2	3
	2	0	2	4	6
	3	0	3	6	9

a) té un màxim a $(3,3)$;
b) té un punt d'inflexió al $(2,1)$;
c) té un mínim a $(1,2)$.

2. El rectorat de la Universitat Opulenta de Catalunya ha decidit atorgar una beca d'estudis a aquell estudiant o estudiants que tinguin una mitjana més alta en el seu expedient acadèmic (fent suspès=0, aprovat=1, etc.). Formuleu això com un problema d'optimització, identificant clarament el conjunt factible i la funció objectiu. Què són en aquest problema el màxim i el(s) maximitzador(s)?

3. D'entre tots els triangles isòsceles que tenen perímetre 2, volem saber quin és el que té una àrea més gran. Formuleu un programa d'optimització que ens doni la resposta, identificant clarament la funció objectiu, el conjunt factible, les variables de decisió i les restriccions.

4. Trobeu els punts crítics de la funció $f(x,y) = \frac{1}{3}\,x^3 + x^2 + \frac{1}{2}\,y^2 - x\,y$, i determineu si són maximitzadors locals, minimitzadors locals o punts de sella.

5. Trobeu els punts crítics de la funció $f(x,y) = \frac{1}{3}\,x^3 - \frac{1}{2}\,x^2 - \frac{1}{2}\,y^2 - x\,y - x$, i determineu si són maximitzadors locals, minimitzadors locals o punts de sella.

6. Trobeu i analitzeu els punts crítics de la funció $f(x,y) = 2\,x\,y - \frac{1}{2}\,x^4 - y^2$.

7. Trobeu i analitzeu els punts crítics de la funció $f(x,y) = 2\,x\,y + \frac{1}{2}\,x^4 + y^2$.

8. Apliqueu el test de la segona derivada per a determinar si les següents funcions són còncaves, convexes o cap de les dues coses.
a) $f(x) = x^6$
b) $f(x) = \log x, \qquad (x > 0)$
c) $f(x) = e^x$
d) $f(x,y) = 2\,x - 3\,y$
e) $f(x,y) = x\,y$
f) $f(x) = 1 + x + e^{-x}$
g) $f(x,y) = x^2 + y^2$
h) $f(x,y) = x^3 - y^3$
i) $f(x,y) = x^{1/3}\,y^{1/3}, \qquad (x > 0,\;\; y > 0)$
j) $f(x,y) = \sqrt{x} + \sqrt{y}, \qquad (x \geq 0,\;\; y \geq 0)$.

9. Sigui $f(x,y) = \frac{1}{4}\,x^4 + \frac{1}{4}\,y^4 + \frac{1}{2}\,x^2\,y^2$. Trobeu i analitzeu els punts crítics. Discutiu-ne el caràcter global.

10. Solucioneu el problema

$$\begin{aligned} \min_{(x,y)} \;\; & x^2 + y^2 \\ \text{s.a.} \;\; & x\,y \;=\; 1. \end{aligned}$$

11. Resoleu el problema de maximització amb restriccions

$$\begin{aligned} \max_{(x,y)} \;\; & x\,y \\ \text{s.a.} \;\; & 0 \leq x \leq 1 \\ & 0 \leq y \leq 1. \end{aligned}$$

12. Mostreu que el problema

$$\begin{aligned} \min_{x} \;\; & e^x \\ \text{s.a.} \;\; & x > 0 \end{aligned}$$

no té cap solució i que, en canvi, el problema

$$\begin{aligned} \min_{x} \;\; & e^x \\ \text{s.a.} \;\; & x \geq 0 \end{aligned}$$

sí que en té.

Solucionari

1. a)

2. Per a cada estudiant tindrem un vector que té per components totes les seves notes. El conjunt factible és el conjunt de tots aquests vectors. La funció objectiu associa a cada vector de notes la seva mitjana aritmètica. El màxim és la nota mitjana més alta que hi hagi, i els maximitzadors són tots aquells estudiants que tinguin aquesta nota de mitjana.

3. Sigui x la meitat de la llargada de la base; y, l'alçària del triangle i z, la llargada del costat repetit. Llavors el problema que hem de solucionar és

$$\begin{aligned} \max_{(x,y,z)} \quad & x\,y \\ \text{s.a.} \quad & x^2+y^2-z^2=0 \\ & x+z= \end{aligned}$$

Observeu que la darrera igualtat és equivalent a $2x+2z=2$. Per a solucionar el problema hem d'escriure el lagrangià:

$$L(x,y,z,\lambda_1,\lambda_2) = x\,y - \lambda_1\left(x^2+y^2-z^2\right) - \lambda_2\,(x+z-1)\,.$$

Les condicions de primer ordre són

$$\begin{aligned} \frac{\partial L}{\partial x} = y - 2\lambda_1 x - \mu &= 0 \\ \frac{\partial L}{\partial y} = x - 2\lambda_1 y &= 0 \\ \frac{\partial L}{\partial z} = 2\lambda_1 z - \mu &= 0 \\ x^2+y^2-z^2 &= 0 \\ x+z &= 1 \end{aligned}$$

Operant algebraicament, trobem la solució

$$(x,y,z,\lambda_1,\lambda_2) = (1/3, 1/\sqrt{3}, 2/3, 1/2\sqrt{3}, 2/3\sqrt{3}).$$

Observeu que la solució és un triangle equilàter, ja que $z = 2\,x$.

4. Trobem els punts crítics igualant les derivades parcials a zero:

$$\begin{aligned} \tfrac{\partial f}{\partial x} &= x^2+2x-y &= 0, \\ \tfrac{\partial f}{\partial y} &= y-x &= 0. \end{aligned}$$

A partir de la segona equació tenim que $y = x$, i substituint en la primera, trobem que les solucions són $(x,y) = (0,0)$ i $(x,y) = (-1,-1)$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 2x+2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Substituint $(x,y) = (0,0)$, tenim que:

$$D^2 f(0,0) = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Com que aquesta matriu és definida positiva, deduïm que $(0,0)$ és un minimitzador local de f.

Si substituïm $(x,y) = (-1,-1)$, tenim que:

$$D^2 f(-1,-1) = \begin{pmatrix} 0 & -1 \\ -1 & 1 \end{pmatrix}.$$

Com que aquesta matriu és indefinida, deduïm que $(-1,-1)$ és un punt de sella de f.

5. Trobem els punts crítics igualant les derivades parcials a zero:

$$\begin{aligned} \tfrac{\partial f}{\partial x} &= x^2-x-y-1 &= 0, \\ \tfrac{\partial f}{\partial y} &= -y-x &= 0. \end{aligned}$$

A partir de la segona equació tenim que $y = -x$, i substituint en la primera trobem que les solucions són $(x,y) = (1,-1)$ i $(x,y) = (-1,1)$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 2x-1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Substituint $(x,y) = (1,-1)$, tenim que:

$$D^2 f(1,-1) = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Com que aquesta matriu és indefinida positiva, deduïm que $(1,-1)$ és un punt de sella de f.

Si substituïm $(x,y) = (-1,1)$, tenim que:

$$D^2 f(-1,1) = \begin{pmatrix} -3 & -1 \\ -1 & -1 \end{pmatrix}.$$

Com que aquesta matriu és definida negativa, deduïm que $(-1,1)$ és un maximitzador local de f.

6. Trobem els punts crítics igualant les derivades parcials a zero:

$$\begin{array}{rcccl} \frac{\partial f}{\partial x} & = & 2y - 2x^3 & = & 0, \\ \frac{\partial f}{\partial y} & = & 2x - 2y & = & 0. \end{array}$$

A partir de la segona equació tenim que $y = x$, i substituint en la primera trobem que les solucions són $(x,y) = (0,0)$, $(x,y) = (1,1)$ i $(x,y) = (-1,-1)$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} -6x^2 & 2 \\ 2 & -2 \end{pmatrix}.$$

Substituint $(x,y) = (0,0)$, tenim que:

$$D^2 f(0,0) = \begin{pmatrix} 0 & 2 \\ 2 & -2 \end{pmatrix}.$$

Com que aquesta matriu és indefinida, deduïm que $(0,0)$ és un punt de sella de f.

Si substituïm els altres valors de (x,y), tenim que:

$$D^2 f(1,1) = D^2 f(-1,-1) = \begin{pmatrix} -6 & 2 \\ 2 & -2 \end{pmatrix}.$$

Com que aquesta matriu és indefinida, deduïm que $(1,1)$ i $(-1,-1)$ són punts de sella de f.

7. Trobem els punts crítics igualant les derivades parcials a zero:

$$\begin{array}{rcccl} \frac{\partial f}{\partial x} & = & 2y + 2x^3 & = & 0, \\ \frac{\partial f}{\partial y} & = & 2x + 2y & = & 0. \end{array}$$

A partir de la segona equació tenim que $y = -x$, i substituint en la primera trobem que les solucions són $(x,y) = (0,0)$, $(x,y) = (1,-1)$ i $(x,y) = (-1,1)$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 6x^2 & 2 \\ 2 & 2 \end{pmatrix}.$$

Substituint $(x,y) = (0,0)$, tenim que:

$$D^2 f(0,0) = \begin{pmatrix} 0 & 2 \\ 2 & 2 \end{pmatrix}.$$

Com que aquesta matriu és indefinida, deduïm que $(0,0)$ és un punt de sella de f.

Si substituïm els altres valors de (x,y), tenim que:

$$D^2 f(1,-1) = D^2 f(-1,1) = \begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}$$

Com que aquesta matriu és definida positiva, deduïm que $(1,-1)$ i $(-1,1)$ són minimitzadors locals de f.

8.

a) $f''(x) = 12\,x^4 \implies$ convexa

b) $f''(x) = -\frac{1}{x^2} \implies$ còncava

c) $f''(x) = e^x \implies$ convexa

d) $D^2 f(x,y) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \implies$ còncava i convexa

e) $D^2 f(x,y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \implies$ ni còncava ni convexa

f) $f''(x) = e^{-x} \implies$ convexa

g) $D^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \implies$ convexa

h) $D^2 f(x,y) = \begin{pmatrix} 6x & 0 \\ 0 & -6y \end{pmatrix} \implies$ ni còncava ni convexa

i) $D^2 f(x,y) = \frac{2}{9} x^{-2/3} y^{-2/3} \begin{pmatrix} -2\frac{y}{x} & 1 \\ 1 & -2\frac{x}{y} \end{pmatrix} \implies$ còncava

j) $D^2 f(x,y) = \begin{pmatrix} -\frac{1}{4}x^{-3/4} & 0 \\ 0 & -\frac{1}{4}y^{-3/4} \end{pmatrix} \implies$ còncava

9. Trobem els punts crítics igualant les derivades parcials a zero:

$$\begin{aligned} \frac{\partial f}{\partial x} &= x^3 + xy^2 &= 0, \\ \frac{\partial f}{\partial y} &= y^3 + x^2 y &= 0. \end{aligned}$$

L'única solució és $(x,y) = (0,0)$. La matriu hessiana és

$$D^2 f(x,y) = \begin{pmatrix} 3x^2 + y^2 & 2xy \\ 2xy & 3y^2 + x^2 \end{pmatrix}.$$

Substituint $(x,y) = (0,0)$, tenim que:

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

El criteri local no és decisori.

Ara bé, qualsevol que sigui (x,y), podem veure que $D^2 f(x,y)$ és una matriu semidefinida positiva, i això implica que la funció és convexa. Això comporta que $(0,0)$, en ser un punt crític d'una funció convexa, sigui un minimitzador global.

10. El lagrangià és $L(x,y,\lambda) = x^2 + y^2 - \lambda(xy - 1)$, i les condicions de primer ordre són:

$$\begin{aligned} \frac{\partial L}{\partial x} &= 2x - \lambda y &= 0, \\ \frac{\partial L}{\partial y} &= 2y - \lambda x &= 0, \\ & \quad xy - 1 &= 0. \end{aligned}$$

Les solucions són $(x,y,\lambda) = (1,1,2)$ i $(x,y,\lambda) = (-1,-1,2)$.

11. Escrivim el problema en forma normalitzada:

$$\begin{aligned} \max_{(x,y)} \;& x\,y \\ \text{s.a.} \quad & -x \le 0, \\ & -y \le 0, \\ & x - 1 \le 0, \\ & y - 1 \le 0. \end{aligned}$$

El lagrangià és

$$L(x,y,\lambda_1,\lambda_2,\lambda_3,\lambda_4) = x\,y + \lambda_1 x + \lambda_2 y - \lambda_3(x-1) - \lambda_4(y-1).$$

Les condicions de Kuhn i Tucker són

$$\begin{aligned}
\frac{\partial L}{\partial x} &= y + \lambda_1 - \lambda_3 = 0,\\
\frac{\partial L}{\partial y} &= x + \lambda_2 - \lambda_4 = 0,\\
\lambda_1 \geq 0, &\quad x \geq 0, \quad \lambda_1 x = 0,\\
\lambda_2 \geq 0, &\quad y \geq 0, \quad \lambda_2 y = 0,\\
\lambda_3 \geq 0, &\quad x \leq 1, \quad \lambda_3(x-1) = 0,\\
\lambda_4 \geq 0, &\quad y \leq 1, \quad \lambda_4(y-1) = 0.
\end{aligned}$$

Les solucions són

$$\begin{aligned}
(x, y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) &= (0, 0, 0, 0, 0, 0),\\
(x, y, \lambda_1, \lambda_2, \lambda_3, \lambda_4) &= (1, 1, 0, 0, 1, 1).
\end{aligned}$$

La primera solució correspon a un punt de sella de la funció objectiu, i la podem descartar. La segona solució és el maximitzador del problema.

12. El problema

$$\begin{aligned}
&\min_x \ e^x\\
&\text{s.a. } x > 0
\end{aligned}$$

és un problema d'optimització sense restriccions, ja que el conjunt factible està definit per una desigualtat estricta i, per tant, és obert. Si hi ha alguna solució, llavors la derivada de la funció objectiu en aquest punt ha de ser 0. Però, si $f(x) = e^x$, llavors $f'(x) = e^x > 0$, per a tot x. Per tant, no hi pot haver cap solució.

El problema

$$\begin{aligned}
&\min_x \ e^x\\
&\text{s.a. } x \geq 0
\end{aligned}$$

és un problema d'optimització amb restriccions de desigualtat. El lagrangià és

$$L(x, \lambda) = e^x - \lambda\, x.$$

Les condicions de Kuhn-Tucker són:

$$\begin{aligned}
\frac{\partial L}{\partial x} &= e^x - \lambda = 0\\
\lambda \geq 0, &\quad x \geq 0, \quad \lambda\, x = 0.
\end{aligned}$$

Tenim

$$\lambda = e^x > 0 \implies x = 0$$

i aquesta és l'única solució. La restricció és lineal i la funció objectiu és convexa, així que la solució és global.

Glossari

Condicions d'exclusió: quan hi ha una restricció de desigualtat feble, si és satisfeta com a desigualtat estricta, el multiplicador corresponent és zero.

Condicions de Kuhn i Tucker: condicions necessàries per a l'optimalitat en un problema amb restriccions de desigualtat.

Condicions de primer ordre: condicions per a l'optimalitat en termes de les derivades de primer ordre.

Condicions de segon ordre: condicions per a l'optimalitat en termes de les derivades de segon ordre.

Condicions de tangència: en problemes amb restriccions, condicions que expressen que la corba de nivell corresponent al valor òptim de la funció objectiu no pot tallar el conjunt factible.

Funció objectiu: funció que volem maximitzar o minimitzar.

Lagrangià: funció objectiu modificada, en la qual introduïm un terme per a cada restricció, ponderat pel modificador corresponent.

Màxim: el valor més gran que assoleix la funció objectiu.

Màxim global: valor màxim de la funció objectiu sobre tot el domini en consideració (tenint en compte les possibles restriccions).

Màxim local: valor màxim de la funció objectiu, però només en un cert entorn.

Maximitzador: punt del domini en què la funció assoleix el valor màxim. Hi pot haver molts maximitzadors, però només hi ha un màxim.

Mínim: el valor més petit que assoleix la funció objectiu.

Mínim global: valor mínim de la funció objectiu sobre tot el domini en consideració (tenint en compte les possibles restriccions).

Mínim local: valor mínim de la funció objectiu, però només en un cert entorn.

Minimitzador: punt del domini en què la funció assoleix el valor mínim. Hi pot haver molts minimitzadors, però només hi ha un mínim.

Punt crític: candidat a maximitzador o minimitzador. Es tracta d'un punt que satisfà les condicions de primer ordre d'un problema d'optimització.

Punt de sella: punt en el qual totes les derivades parcials són zero, i a partir del qual la funció creix en alguna direcció i decreix en alguna altra.

Regularitat: independència lineal dels gradients de les restriccions que se satisfan amb igualtat en un punt.

Restriccions: condicions, en forma d'igualtat o de desigualtat, que han de complir les variables en un problema d'optimització.

Bibliografia

Trobareu una presentació clara i concisa del que hem tractat en aquest mòdul en el llibre següent:

Borrell, Josep (1990). *Métodos matemáticos para la economía. Programación matemática.* Madrid: Ediciones Pirámide.

Per a una exposició més completa i a un nivell més avançat, podeu consultar:

Chiang, Alpha (1987). *Métodos fundamentales de la economía matemática.* McGraw-Hill (Mèxic,1990).

Luenberger, David (1992). *Programación lineal y no lineal.* Wilmington: Addison-Wesley Iberoamericana.

Un llibre també més avançat, però amb bones explicacions intuïtives i gràfiques, i adreçat a economistes, és

Dixit, Avinash (). *Optimization in Economic Theory.* Oxford.

Finalment, el que fem aquí i molt més està molt ben explicat en

Sydsaeter, Knut; Hammond, Peter (1995). *Mathematics for Economic Analysis.* Englewood Cliffs, N.J.: Prentice-Hall International.