DIPLOMATURA D'ESTADÍSTICA

APLICACIONS DE LA PROGRAMACIÓ NO LINEAL

Algorismes i exercicis

Narcís Nabona
F.Javier Heredia

UPC FACULTAT DE MATEMÀTIQUES I ESTADÍSTICA

Aplicacions de la Programació No Lineal.

Algorismes i exercicis.

Narcís Nabona
F. Javier Heredia
Departament d'Estadística i Investigació Operativa
Secció d'Informàtica
UPC

Índex

1 Condicions de mínim.

1.1 Procediment

Consideri's el problema d'optimització:

$$(\mathbf{P}) \min_{x \in \Omega \subseteq \Re^n} f(x)$$

on Ω indica una certa regió factible. Definirem els conceptes de *direcció factible* i de *direcció de descens*:

Definició 1.1 Direcció factible : Sigui el problema $(\mathbf{P})$ i el punt factible $x \in \Omega$. Es diu que $d \in \Re^n$ és una direcció factible sobre x si:

$$\exists \bar{\alpha} > 0 \mid \forall \alpha \in [0, \bar{\alpha}] \ : \ x + \alpha d \in \Omega$$

Definició 1.2 Direcció de descens : Sigui el problema $(\mathbf{P})$ i el punt factible $x \in \Omega$. Es diu que $d \in \Re^n$ és una direcció de descens sobre x si:

$$\exists \bar{\alpha} > 0 \mid \forall \alpha \in [0, \bar{\alpha}] \ : \ f(x + \alpha d) < f(x)$$

Les direccions que s'on alhora factibles i de descens representen direccions al llarg de les quals la funció objectiu pot millora sense deixar la regió factible Ω. Això ens permet formular una condició necessària de mínim local:

Teorema 1.1 : Donat un problema $(\mathbf{P})$, si $x^* \in \Omega$ és mínim local, llavors no existeixen dirección factibles de descens sobre x^*

Si el conjunt Ω és un polídre (és a dir, Ω està definit a partir d'un conjunt de constriccions lineals) llavors l'anterior condició és necessària i suficient.

1.2 Exercicis

Exercici 1.1. : pcomin1/condicions d'òptim

Donada la funció $f(X) = x_1^2 + 2x_2^2 + 3x_3^2 + x_1x_2 - x_2x_3 - 5x_2 - 10x_3$:

- Determineu i justifiqueu si aquesta funció és estrictament convexa.
- A través de la propietat de l'òptim X^* d'una funció $f(X)$ en un domini, per la qual $\nabla f(X^*)D \geq 0$ per a tota direcció D, a partir de X^*, factible respecte al domini, trobeu

(si existeix/teixen) el/els punt/s del domini definit per:

$$x_1 \geq 0 \quad , \quad x_2 \geq 0 \quad , \quad x_3 \geq 0$$

que minimitzin la funció.

SOLUCIÓ

VECTOR GRADIENT I MATRIU HESSIANA.

$$\nabla f(X)' = \begin{bmatrix} 2x_1 + x_2 \\ 4x_2 + x_1 - x_3 - 5 \\ 6x_3 - x_2 - 10 \end{bmatrix} \quad ; \quad \nabla^2 f(X) = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 6 \end{bmatrix}$$

CONDICIÓ DE $f(X)$ ESTRICTAMENT CONVEXA.

$\nabla^2 f(X)$ és definit positiu perquè els menors principals són tots positius:

$$2 > 0 \quad , \quad \det \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} = 7 > 0 \, , \text{ i } \det \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 6 \end{bmatrix} = 40 > 0$$

Així doncs, $\nabla^2 f(X)$ és definit positiu per a $\forall X$ (perquè $\nabla^2 f(X)$ és constant), i això és una condició suficient perquè $f(X)$ sigui estrictament convexa.

MÍNIM DE $f(X)$ A $\mathbb{R}^n$

$f(X)$ és una funció quadràtica amb $Q = \nabla^2 f(X)$ definida positiva. Primer comprovarem si el mínim d'aquesta funció quadràtica està dins del domini definit per $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$.

Per a una funció quadràtica $\frac{1}{2}X'QX - B'X$ el gradient és $G(X) = QX - B$. Per a $X = \underline{0}$ es té:

$$G(\underline{0}) = -B \quad , \quad G(\underline{0}) = \nabla f(\underline{0})' = \begin{bmatrix} 0 \\ -5 \\ -10 \end{bmatrix} \Rightarrow B = \begin{bmatrix} 0 \\ 5 \\ 10 \end{bmatrix}$$

d'on $f(X) = \frac{1}{2}X' \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 6 \end{bmatrix} X - [0 \quad 5 \quad 10]X$. X^* és tal que $QX^* = B$. Resolent el sistema $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 6 \end{bmatrix} X = \begin{bmatrix} 0 \\ 5 \\ 10 \end{bmatrix}$ amb $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 6 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1/2 & 1 & \\ 0 & -2/7 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ & 7/2 & -1 \\ & & 40/7 \end{bmatrix}$ s'obté:

$$\begin{bmatrix} 1 & & \\ 1/2 & 1 & \\ 0 & -2/7 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ & 7/2 & -1 \\ & & 40/7 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1/2 & 1 & \\ 0 & -2/7 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 10 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 80/7 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 \\ & 7/2 & -1 \\ & & 40/7 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 80/7 \end{bmatrix} \Rightarrow \boxed{\begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}}$$

que es troba fora de $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$.

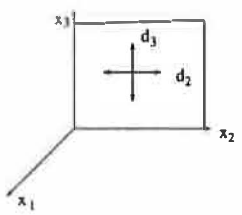
MÍNIM DE f(X) AL DOMINI DEFINIT PER $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$.

El mínim al domini $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ haurà de satisfer $\nabla(X^*)D \geq 0 \, \forall D$ factible, i estarà sobre la frontera. S'haurà doncs de satisfer:

$$\begin{bmatrix} 2x_1 + x_2 & 4x_2 + x_1 - x_3 - 5 & 6x_3 - x_2 - 10 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \geq 0$$

Segons a quina faceta ens trobem de la frontera definida per $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, es poden donar els casos següents:

a) $x_1 = 0$, d_2 i d_3 qualsevol i $d_1 > 0$:



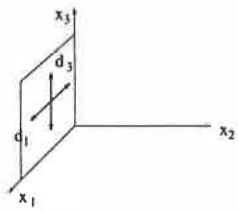
$$\begin{bmatrix} x_2 & 4x_2 - x_3 - 5 & 6x_3 - x_2 - 10 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \geq 0 \Rightarrow \begin{cases} x_2 \geq 0 & \text{(perquè } d_1 \geq 0\text{)} \\ 4x_2 - x_3 - 5 = 0 & \text{(perquè } d_2 \text{ pot ser molt negatiu)} \\ 6x_3 - x_2 - 10 = 0 & \text{(perquè } d_2 \text{ pot ser molt negatiu)} \end{cases}$$

d'on el sistema

$$\left\{ \begin{array}{l} 4x_2 - x_3 \\ -x_2 + 6x_3 = 10 \end{array} \right. \Rightarrow \begin{array}{l} x_2 = 40/23 \geq 0 \\ x_3 = 45/23 \geq 0 \end{array} \Rightarrow \boxed{\tilde{X}^* = \begin{bmatrix} 0 \\ 40/23 \\ 45/23 \end{bmatrix}}$$

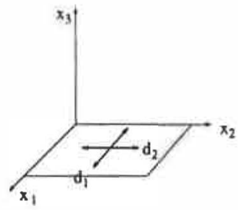
on $\tilde{X}^*$ és un mínim a $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$.

b) $x_2 = 0$, d_1 i d_3 qualsevol i $d_2 > 0$:



$$[2x_1 \quad x_1 - x_3 - 5 \quad 6x_3 - 10] \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \geq 0 \Rightarrow \begin{cases} x_2 = 0 & d_1 \text{ i } d_3 \text{ qualssevol, i } d_2 \geq 0 \\ x_1 = 0 & \text{(perquè } d_1 \text{ pot ser molt negatiu)} \\ -x_3 - 5 \geq 0 & \text{(perquè } d_2 \geq 0) \Rightarrow \boxed{x_3 \leq -5 \text{ IMPOSSIBLE}} \end{cases}$$

c) $x_3 = 0$, d_1 i d_2 qualsevol i $d_3 > 0$:



$$[2x_1 + x_2 \quad 4x_2 + x_1 - 5 \quad -x_2 - 10] \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \geq 0 \Rightarrow \begin{cases} 2x_1 + x_2 = 0 \text{ (perquè } d_1 \text{ pot ser molt negatiu). Llavors, si } x_1 > 0 \Rightarrow x_2 < 0 \text{ i viceversa} \Rightarrow \boxed{\text{IMPOSSIBLE}} \end{cases}$$

L'únic mínim és doncs $\boxed{\bar{X}^* = [0 \quad 40/23 \quad 45/23]'}$.

Exercici 1.2. : pcomin2/Condicions de mínim

Donada la funció $f(x) = x_1^2 + x_2^2 + x_1 x_2 - 6x_1$, trobeu (si existeix) el mínim de la funció subjecte a $x_1 \geq 2$, $x_2 \geq 0$.

(Empreu la propietat de l'òptim x^* $\nabla f(x^*) D \geq 0$ per a tota direcció factible D respecte al domini de definició).

2 Optimització sense constriccions.

2.1 Mètode de Nelder i Mead

2.1.1 Procediment

Sigui el problema:

$$\underset{X}{\text{minimitzeu}} \quad f(X) \tag{1}$$

amb $X \in \mathbb{R}^n$ i $f(X)$: $\mathbb{R}^n \rightarrow \mathbb{R}$. el qual ens proposem de resoldre amb només avaluacions de la funció i sense utlilitzar derivades (ni aproximar–les per diferencies finites).

L'algorisme de Nelder–Mead es basa en considerar un poliedre amb $n+1$ vèrtexs (en un espai de dimensió n). De cada vèrtex cal emmagatzemar les seves coordenades i el valor de la seva funció, i a cada iteració les coordenades de com a mínim un dels vèrtexs canvien, de forma que el poliedre va canviant la seva forma i posició a $\mathbb{R}^n$. És doncs necessari emmagatzemar una matriu $n \times n$ densa amb les coordenades dels vèrtexs, el qual fa que el mètode no sigui apropiat per a n gran.

Es comença amb un poliedre regular amb un vèrtex a l'origen i situat de forma simètrica a l'ortant positiu. Les coordenades dels vèrtexs són aleshores els de la matriu $D \in \mathbb{R}^{n \times (n+1)}$:

$$D = \begin{bmatrix} 0 & a & b & b & \dots & b \\ 0 & b & a & b & \dots & b \\ 0 & b & b & a & \dots & b \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & b & b & b & \dots & a \end{bmatrix} \quad \text{amb} \quad \begin{cases} a = \dfrac{t}{n\sqrt{2}}(\sqrt{n+1} + n - 1) \\ b = \dfrac{t}{n\sqrt{2}}(\sqrt{n+1} - 1) \\ t\text{: distància entre vèrtexs} \end{cases} \tag{2}$$

Cal avaluar $f(X)$ a cada vèrtex formant un vector d'avaluacions $F \in \mathbb{R}^{n+1}$ amb components $F(1)=f(X_1)$, $F(2)=f(X_2)$, ..., $F(n+1)=f(X_{n+1})$, sent X_1, X_2, ..., X_{n+1} les coordenades dels vèrtexs. En particular convé determinar el vèrtex amb millor (més baix) valor de la funció, que designarem per X_m, i el pitjor (amb més alt valor de la funció), que designarem per X_M.

L'operació bàsica, que cal repetir a cada iteració, consisteix en determinar X_M de entre tots els vèrtexs, i a calcular el punt X_G o centre de gravetat de tots els vèrtexs excepte el pitjor X_M, el qual tindrà components:

$$x_{G\,j} = \frac{1}{n}\left[\left(\sum_{i=1}^{n+1} x_{i\,j}\right) - x_{M\,j}\right] \quad j = 1, \dots, n$$

A partir de X_G es determina un nou punt X_R que és el *reflex* del punt pitjor X_M respecte X_G:

$$X_R = X_G + \alpha(X_G - X_M) \quad \text{amb } \alpha > 0 \quad (\text{normalment } \alpha = 1)$$

i avaluem la funció a X_R trobant $f(X_R)$.

L'algorisme compara $f(X_R)$ amb $f(X_m)$ i, si escau, amb $f(X_M)$ i pren les decisions següents:

⋆ Si $f(X_R) \leq f(X_m)$ fa una *expansió*
  calculem $X_E = X_G + \gamma(X_G - X_M)$ amb $\gamma > 1$ (normalment $\gamma = 2$)
  ▷ si $f(X_E) < f(X_R)$ canviem X_M per X_E i tornem a trobar el nou X_M i a calcular X_G
  ▷ si $f(X_E) \not< f(X_R)$ canviem X_M per X_R i tornem a trobar el nou X_M i a calcular X_G

⋆ Si $f(X_R) > f(X_i)$ $\forall i \neq M$ fem una *contracció*
  calculem $X_C = X_G + \beta(X_G - X_M)$ amb $0 < \beta < 1$ (normalment $\beta = 0.5$)
  ▷ si $f(X_C) < f(X_M)$ canviem X_M per X_C i tornem a trobar el nou X_M i a calcular X_G
  ▷ si $f(X_C) \not< f(X_M)$ anem a l'etapa de *reducció* que es descriu a continuació

⋆ Si $f(X_R) > f(X_M)$ fem una *reducció* (fem tot el poliedre més petit només conservant el vèrtex millor)
  calculem $X_i = X_i + \delta(X_i - X_m)$ $\forall i \neq m$ amb $0 < \delta < 1$ (normalment $\delta = 0.5$) i avaluem $f(X_i)$ als nous punts i tornem a determinar els nous X_M, X_m i a clacular X_G.

L'algorisme s'acaba quan el poliedre es fa molt petit o quan els valors de la funció dels distints vèrtexs es indistingible.

Pot trobar–se l'algorisme de Nelder–Mead descrit a [BERT95] pp. 144–148.

2.1.2 Exercicis

Exercici 2.1. : pnelm1/mètode de Nelder-Mead

S'aplica la metodologia de Nelder-Mead a minimitzar una funció en un espai de dimensió 2, i tenim un poliedre amb tres vèrtexs X_a, X_b i X_c on la funció objectiu pren respectivament els valors $f(X_a)=3$, $f(X_b)=4$ i $f(X_c)=5$. Un cop realitzada l'operació de reflexió obtenim un nou punt X_r on la funció val $f(X_r)$. Indiqueu els valors que hauria de prendre $f(X_r)$ per tal que el que seguís en l'aplicació de l'algorisme fos cadascuna de les operacions:

- una expansió
- una contracció
- una reducció a partir de X_a
- que prengessim X_r i descartessim X_c.

SOLUCIÓ

EXPANSIÓ

Quan $f(X_r)$ té valor inferior al millor vèrtex ($f(X_a)$) i a més si determinant un vèrtex expandit X_e al llarg de la direcció de X_c a X_r, el vèrtex expandit és encara millor. D'on hi haurà Expansió si:

$$f(X_r) \leq 3 = f(X_a) \quad \text{i} \quad f(X_e) < f(X_r)$$

CONTRACCIÓ

Quan $f(X_r)$ té valor pitjor que tots els actuals menys el pitjor:

$$f(X_b) = 4 < f(X_r) \leq f(X_c) = 5$$

REDUCCIÓ

Quan $f(X_r)$ té valor pitjor que tots els actuals

$$f(X_r) > 5 = f(X_c)$$

SUBSTITUCIÓ DE X_c PER X_r

Quan $f(X_r)$ té valor entre el millor i el pitjor actuals i no és només millor que el pitjor actual (en el qual cas fóra contracció):

$$f(X_a) = 3 < f(X_r) \leq 4 = f(X_b)$$

També en el cas d'expansió infructuosa:

$$f(X_r) < 3 = f(X_a) \quad \text{i} \quad f(X_e) \geq f(X_r)$$

Exercici 2.2. : pnelm2/mètode de Nelder-Mead

Sigui la funció objectiu $f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2$. En la resolució del problema $\min_{x \in \Re^3} f(x)$ pel mètode de Nelder-Mead s'ha arribat a la iteració k-éssima amb el següent poliedre:

$$\begin{array}{r c} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{array}{r} \\ P = \\ \\ f(x_j) = \end{array} & \left[\begin{array}{cccc} 0 & -1.495 & 0.546 & -0.747 \\ 0 & 1.157 & 2.314 & 1.462 \\ 0 & 1.157 & 0.546 & -1.189 \\ 2 & 122.463 & 406.791 & 9.563 \end{array}\right] \end{array}$$

Apliqueu una passa del mètode de Nelder-Mead a partir de P amb $\alpha = 1$, $\beta = 0.5$ i $\gamma = 2$

2.2 Exploració lineal

2.2.1 Procediment

Consideri's que en la resolució del problema $\min_{X \in \Re^n} f(x)$ s'ha arribat a l'iterat X_k amb la direcció de descens D_k. L'exploració lineal aproximada per ajustos de corbes consisteix en trobar l'escalar λ_k, una aproximació de la longitud de pas òptima , definida com $\lambda_k = \text{argmin}\{f(X_k + \lambda D_k) | \lambda > 0\}$ El procediment iteratiu que s'aplica consisteix en trobar, a cada iteració, un nou valor de $\tilde{\lambda}$ a partir de certa informació sobre valors de la funció objectiu i les seves derivades sobre diferents valors de λ. El procés es repeteix fins que el valor $\tilde{\lambda}$ es considera *acceptable*, segons les *condicions d'Armijo-Goldstein*, es pren $\lambda_k := \tilde{\lambda}$.

A continuació es descriurà, en primer lloc, el procés de càlcul del valor $\tilde{\lambda}$ per *ajustos de corbes* i, seguidament, els criteris d'acceptabilitat (*Condicions d'Armijo-Goldstein*).

Ajustos quadràtics:

Aquests mètodes calculen $\tilde{\lambda}$ a partir de l'òptim d'un model quadràtic

$$m_q(\lambda) = a\lambda^2 + b\lambda + c$$

de la representació de la funció objectiu a partir de X_k al llarg de D_k:

$$g(\lambda) = f(X_k + \lambda D_k)$$

Algorisme A2.1 : Ajust quadràtic I

$\boxed{0}$ Siguin conegudes les següents quantitats:
$g(0) := f(X_k)$; $g'(0) := \nabla f(X_k)D_k$; $g(\lambda_1) := f(X_k + \lambda_1 D_k)$

$\boxed{1}$ Càlcul dels coeficients del model $m_q(\lambda)$

$$c := g(0);\ d := g'(0);\ a := \frac{1}{\lambda_1^2}(g(\lambda_1) - g'(0)\lambda_1 - g(0))$$

$\boxed{2}$ Càlcul del mínim de $m_q(\lambda)$

<u>Si</u> $a > 0$ <u>Llavors</u>

$$\tilde{\lambda} := -\frac{b}{2a} = -\frac{g'(0)}{\frac{2}{\lambda_1^2}(g(\lambda_1) - g'(0)\lambda_1 - g(0))}$$

<u>Altrament</u>

$m_q(X)$ és còncava

<u>Fi Si</u>

Algorisme A2.2 : Ajust quadràtic II

$\boxed{0}$ Siguin conegudes les següents quantitats:
$g(0) := f(X_k)$; $g(\lambda_1) := f(X_k + \lambda_1 D_k)$; $g(\lambda_2) := f(X_k + \lambda_2 D_k)$

$\boxed{1}$ Càlcul dels coeficients del model $m_q(\lambda)$

$$c := g(0);\ \begin{bmatrix} a \\ b \end{bmatrix} := \frac{1}{(\lambda_1 - \lambda_2)} \begin{bmatrix} 1/\lambda_2 & -1/\lambda_2 \\ -\lambda_2/\lambda_1 & \lambda_1/\lambda_2 \end{bmatrix} \begin{bmatrix} g(0) - g(\lambda_1) \\ g(0) - g(\lambda_2) \end{bmatrix}$$

$\boxed{2}$ Càlcul del mínim de $m_q(\lambda)$

<u>Si</u> $a > 0$ <u>Llavors</u>

$$\tilde{\lambda} := -\frac{b}{2a}$$

<u>Altrament</u>

$m_q(X)$ és còncava

<u>Fi Si</u>

Ajustos cúbics:

Aquests mètodes calculen $\tilde{\lambda}$ a partir de l'òptim d'un model cúbic

$$m_c(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d$$

de la representació de la funció objectiu a partir de X_k al llarg de D_k:

$$g(\lambda) = f(X_k + \lambda D_k)$$

Algorisme A2.3 : Ajust cúbic I

$\boxed{0}$ Siguin conegudes les següents quantitats:
$g(0) := f(X_k)$; $g'(0) := \nabla f(X_k)D_k$; $g(\lambda_1) := f(X_k + \lambda_1 D_k)$; $g(\lambda_2) := f(X_k + \lambda_2 D_k)$

$\boxed{1}$ Càlcul dels coeficients del model $m_c(\lambda)$

$$d := g(0);\ c := g'(0);\ \begin{bmatrix} a \\ b \end{bmatrix} := \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} 1/\lambda & -1/\lambda_2^2 \\ -\lambda_2/\lambda_1^2 & \lambda_1/\lambda_2^2 \end{bmatrix} \begin{bmatrix} g(\lambda_1) - g(0) - g'(0)\lambda_1 \\ g(\lambda_2) - g(0) - g'(0)\lambda_2 \end{bmatrix}$$

$\boxed{2}$ Càlcul del mínim de $m_c(\lambda)$

<u>Si</u> $(b^2 - 3ag'(0)) > 0$ Llavors

$$\tilde{\lambda} := \frac{-b + \sqrt{b^2 - 3ag'(0)}}{3a}$$

Altrament

$m_c(\lambda)$ no té mínim

Fi Si

Algorisme A2.4 : Ajust cúbic II

$\boxed{0}$ Siguin conegudes les següents quantitats:

$g(0) := f(X_k)$; $g'(0) := \nabla f(X_k)D_k$

$g(\lambda_1) := f(X_k + \lambda_1 D_k)$; $g'(\lambda_1) := \nabla f(X_k + \lambda_1 D_k)D_k$

$\boxed{1}$ Càlcul dels coeficients del model $m_c(\lambda)$

$$d := g(0);\ c := g'(0);\ \begin{bmatrix} a \\ b \end{bmatrix} := \begin{bmatrix} -2/\lambda_1^3 & -1/\lambda_1^2 \\ 3/\lambda_1^2 & -1/\lambda_1 \end{bmatrix} \begin{bmatrix} g(\lambda_1) - g(0) - g'(0)\lambda_1 \\ g'(\lambda_1) - g'(0) \end{bmatrix}$$

$\boxed{2}$ Càlcul del mínim de $m_c(\lambda)$

Si $(b^2 - 3ag'(0)) > 0$ Llavors

$$\tilde{\lambda} := \frac{-b + \sqrt{b^2 - 3ag'(0)}}{3a}$$

Altrament

$m_c(\lambda)$ no té mínim

Fi Si

Condicions d'Armijo-Goldstein.

AG1 Condició primera d'Armijo-Goldstein: sigui el punt X_k i la direcció D_k i l'escalar $\alpha \in (0,1)$. Direm que la longitud de pas λ satisfà la primera condició d'Armijo-Goldstein associada a α si es verifica

$$f(X_k + \lambda D_k) \leq f(X_k) + \alpha \nabla f(X_k)D_k$$

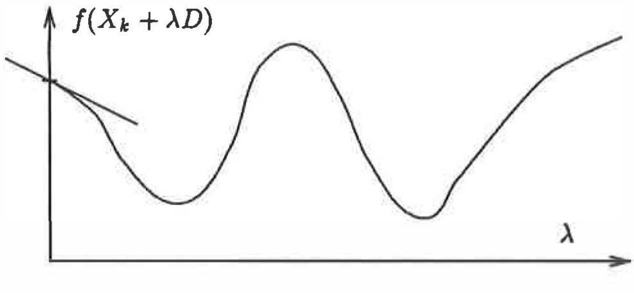
AG2 Condició segona d'Armijo-Goldstein: sigui el punt X_k i la direcció D_k i l'escalar $\beta \in (\alpha, 1)$. Direm que la longitud de pas λ satisfà la segona condició d'Armijo-Goldstein associada a β si es verifica

$$\nabla f(X_k + \lambda D_k)D_k \geq \beta \nabla f(X_k)D_k$$

2.2.2 Exercicis

Exercici 2.3. : pargol/criteris d'Armijo-Goldstein

Donada la funció de la figura adjunta, la qual expressa la variació del valor de $f(X_k + \lambda D)$ en funció de λ, sent $X_k, D \in \mathbb{R}^n$ i $f(X)$: $\mathbb{R}^n \to \mathbb{R}$, copieu la figura i indiqueu les zones de l'eix de les abscisses on es satisfà la 1ª condició d'Armijo-Goldstein i, de forma diferenciada, les zones on es satisfà la 2ª condició d'Armijo-Goldstein. Indiqueu els valors numèrics que utilitzeu per definir els paràmetres que caracteritzen les esmentades condicions.

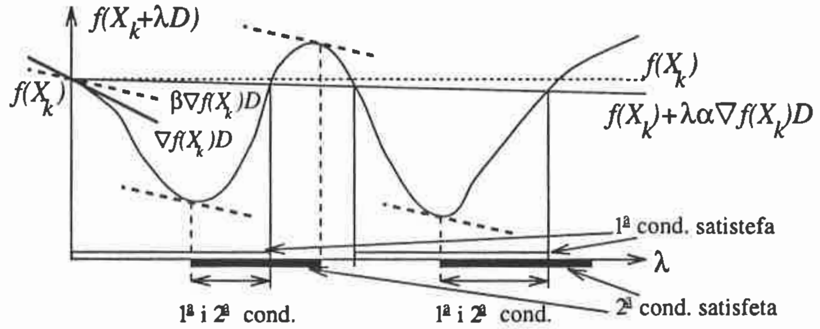


SOLUCIÓ

1^er^ CRITERI D'ARMIJO-GOLDSTEIN

El 1^er^ criteri d'acceptabilitat d'una passa λ és que:

$$f(X_k + \lambda D) \leq f(X_k) + \lambda\alpha\nabla f(X_k)D \qquad 0 < \alpha < 1 \quad \text{generalment } \alpha=0,1$$



2^on^ CRITERI D'ARMIJO-GOLDSTEIN

Cal que:

$$\nabla f(X_k + \lambda D)D \geq \beta\nabla f(X_k)D \qquad \alpha < \beta < 1 \quad \text{generalment } \beta=0,5$$

La figura mostra les zones on es satisfà el 1^er^ i el 2^on^ criteri d'Armijo-Goldstein d'acceptabilitat de la passa. En el dibuix s'han utilitzat (aproximadament) els valors α=0,1 i β=0,5

Exercici 2.4. : pargo2/criteris d'Armijo-Goldstein

Donada la funció objeciu $f(x) = x_1^2 + x_1x_2 + \frac{2x_1}{x_2}$, el punt iterat $X_0 = [1 \quad -2]'$ i la direcció de cerca $D = [2 \quad 1]'$:

2.4.1.– Feu exploració lineal a partir de X_0 al llarg de D amb ajust quadràtic emprant el valor de la derivada de la funció objectiu a X_0. Comproveu si el punt obtingut satisfà la 1ª condició d'Armijo-Goldstein.

2.4.2.– Feu exploració lineal a partir de X_0 al llarg de D amb ajust cúbic emprant el valor de la derivada de la funció objectiu a X_0. Comproveu si el punt obtingut satisfà la 2ª condició d'Armijo-Goldstein. En cas de no fer l'ajust cúbic, comproveu aquesta condició sobre el punt associat a $\lambda = 1/8$.

SOLUCIÓ

CÀLCULS PRELIMINARS :

$$f(x) = x_1^2 + x_1x_2 + \frac{2x_1}{x_2} \quad ; \quad \nabla f(x)' = \begin{bmatrix} 2x_1 + x_2 + \frac{2}{x_2} \\ x_1 - \frac{2x_1}{x_2^2} \end{bmatrix}$$

$$X_0 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad ; \quad f(X_0) = -2 \quad ; \quad \nabla f(X_0)' = \begin{bmatrix} -1 \\ \frac{1}{2} \end{bmatrix}$$

$$\nabla f(X_0) D = [-1 \quad \tfrac{1}{2}] \begin{bmatrix} 2 \\ 1 \end{bmatrix} = -\frac{3}{2} < 0 \Rightarrow \text{D de descens}$$

AJUST QUADRÀTIC :

Es pren el punt $X_0 + D$ com a punt addicional :

$$X_0 + D = \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad ; \quad f(X_0 + D) = 0$$

Ajust quadràtic : $q(\lambda) = a\lambda^2 + b\lambda + c$

$$\left.\begin{array}{lll} \lambda = 0 & : & q(0) = c = f(X_0) = -2 \Rightarrow \boxed{c = -2} \\ & & q(\lambda)' = 2a\lambda + b \quad ; \quad q(0)' = b = \nabla f(X_0) D = -\frac{3}{2} \Rightarrow \boxed{b = -\frac{3}{2}} \\ \lambda = 1 & : & q(1) = a + b + c = f(X_0 + D) = 0 \Rightarrow a + b + c = 0 \end{array}\right\} \Rightarrow \begin{array}{c} a - \frac{3}{2} - 2 = 0 \\ \boxed{a = \frac{7}{2}} \end{array}$$

Càlcul de λ_q : $q(\lambda_q)' = 2a\lambda_q + b = 0 \quad ; \quad \lambda_q = -\frac{b}{2a} = -\frac{-3/2}{7} \Rightarrow \boxed{\lambda_q = \frac{3}{14}}$ mínim, doncs $a > 0$.

COMPROVACIÓ DE LA 1^a COND. D'ARMIJO-GOLDSTEIN A $X_0 + \lambda_q D$:

Condició AG1 : $f(X_0 + \lambda D) \leq f(X_0) + \alpha \nabla f(X_0) D \lambda \quad ; \quad \alpha \in [0, 1]$. S'acostuma a prendre un escalar α molt petit. Usarem $\alpha = 0.05$.

$$\left.\begin{array}{l} X_0 + \lambda_q D = \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \frac{3}{14} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{10}{7} \\ -\frac{25}{14} \end{bmatrix} \\ f(X_0 + \lambda_q D) = \underline{-2.1102} < \\ < f(X_0) + \alpha \nabla f(X_0) D \lambda_q = -2 + 0.05(-\frac{3}{2})\frac{3}{14} = \underline{-2.0161} \end{array}\right\} \Rightarrow \boxed{\text{Es satisfà AG1}}$$

AJUST CÚBIC :

Es prenen els punts $X_0 + D$ i $X_0 + \frac{5}{9} D$ com a punts addicionals :

$$X_0 + D = \begin{bmatrix} 3 \\ -1 \end{bmatrix} ; f(X_0 + D) = 0 ; X_0 + \frac{5}{9} D = \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \frac{5}{9} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{19}{9} \\ \frac{-13}{9} \end{bmatrix} ; f(X_0 + \frac{5}{9} D) = -1.5156$$

Ajust cúbic : $u(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d$

$$\left.\begin{array}{l}
\lambda = 0 \ : u(0) = d = f(X_0) = -2 \Rightarrow \boxed{d=-2} \\
\qquad u(\lambda)' = 3a\lambda^2 + 2b\lambda + c \\
\qquad u(0)' = c = \nabla f(X_0)D = -\frac{3}{2} \Rightarrow \boxed{c=-\frac{3}{2}} \\
\lambda = 1 \ : u(1) = a+b+c+d = f(X_0+D) = 0 \Rightarrow \underline{a+b=\frac{7}{2}} \\
\lambda = \frac{5}{9} \ : u(\frac{5}{9}) = a\frac{125}{729} + b\frac{25}{81} + c\frac{5}{9} - 2 = f(X_0+\frac{5}{9}D) = -1.5156
\end{array}\right\} \left.\begin{array}{l} a+b=\frac{7}{2} \\ \frac{125}{729}a + \frac{25}{81}b = 1.3176 \end{array}\right\}(*)$$

$$(*) \Rightarrow \boxed{\begin{array}{l} a = -1.7307 \\ b = 5.2307 \end{array}}$$

Càlcul de λ_c :

$$u(\lambda_u)' = 3a\lambda_u^2 + 2b\lambda_u + c = 0 \quad ; \quad -5.1921\lambda_u^2 + 10.4614\lambda_u - 1.5 = 0$$

$$\lambda_u = \frac{-10.4614 + \sqrt{109.4408 - 31.1526}}{-10.3842} = 0.1553 \Rightarrow \boxed{\lambda_u = 0.1553}$$

COMPROVACIÓ DE LA 2^A COND.D'ARMIJO-GOLDSTEIN A $X_0 + \lambda_u D$:

Condició AG2 : $\nabla f(X_0+\lambda D)D \geq \beta\nabla f(X_0)D$; $\beta \in [\alpha, \frac{1}{2}]$. Usarem $\beta = 0.3$.

$$X_0 + \lambda_u D = \begin{bmatrix} 1 \\ -2 \end{bmatrix} + 0.1553\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.3107 \\ -1.8446 \end{bmatrix}$$

$$\nabla f(X_0+\lambda_u D)D = [-0.3074 \quad 0.5403]\begin{bmatrix} 2 \\ 1 \end{bmatrix} = -0.07448$$

$$\nabla f(X_0+\lambda_u D)D = \underline{-0.07448} > \beta\nabla f(X_0)D = 0.3(-\frac{3}{2}) = \underline{-0.45} \Rightarrow \boxed{\text{Es satisfà AG2}}$$

COMPROVACIÓ DE LA 2^A COND. D'ARMIJO-GOLDSTEIN A $X_0 + \frac{1}{8}D$:

$$X_0 + \lambda_u D = \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \frac{1}{8}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.25 \\ -1.875 \end{bmatrix}$$

$$\nabla f(X_0+\lambda_u D)D = [-0.4416 \quad 0.5388]\begin{bmatrix} 2 \\ 1 \end{bmatrix} = -0.3444$$

$$\nabla f(X_0+\frac{1}{8}D)D = \underline{-0.3444} > \beta\nabla f(X_0)D = 0.3(-\frac{3}{2}) = \underline{-0.45} \Rightarrow \boxed{\text{Es satisfà AG2}}$$

Exercici 2.5. : pargo3/criteris d'Armijo-Goldstein

Per a la funció objectiu $f(x) = x_1^3 + x_1x_2 + 2x_1/x_2$:

2.5.1.– Verifiqueu si la direcció $D = [0 \quad -1]'$ és de descens per a la funció $f(x)$ a partir del punt inicial $X_0 = [1 \quad 1]'$. Si no ho fos, trobeu-ne una que ho sigui.

2.5.2.– Partint del punt X_0 donat, efectueu exploració lineal per ajust quadràtic en la direcció de descens que tingueu. Efectueu a continuació exploració lineal per ajust cúbic del tipus $y = a\lambda^3 + b\lambda^2 + c\lambda + d$ emprant els valors de la funció a X_0 i al punt trobat a l'ajust quadràtic i el valor de la derivada a X_0.

2.5.3.– Efectueu per als punts trobats en les exploracions lineals per ajust quadràtic X_q i ajust cúbic X_c, una comprovació d'acceptabilitat que no sigui la simple verificació de que $f(X_q)$ i $f(X_c)$ valen menys que $f(X_0)$.

2.3 Mètode del gradient

2.3.1 Procediment

El mètode del gradient és un algorisme de programació no lineal que permet resoldre problemes de minimització sense constriccions:

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

on $f : \mathbb{R}^n \rightarrow \mathbb{R} \quad , \quad f \in C^1$. Amb les següent notació:

$$\begin{aligned} k &: \text{índex de la iteració actual.} \\ x^k \in \mathbb{R}^n &: \text{vector de variables a la iteració k-éssima.} \\ d^k \in \mathbb{R}^n &: \text{direcció de descens a la iteració k-éssima.} \\ \alpha^k \in \mathbb{R} &: \text{longitud de pas a la iteració k-éssima.} \end{aligned}$$

l'algorisme del mètode del gradient o del máxim descens ("steepest descent") consisteix en les següents passes :

Mètode del gradiente :

0.-	**Inicialització**	:	x^0. $k = 0$.
1.-	**Detecció d'òptim**	:	Si $\Vert \nabla f(x^k) \Vert = 0$: $x^* = x^k$ STOP.
2.-	**Direcció de descens**	:	$d^k = -\nabla f(x^k)$
3.-	**Longitud de pas**	:	$\alpha^k \quad \leftarrow \quad \min_{\alpha > 0} \{ f(x^k + \alpha d^k) \}$
4.-	**Actualizació**	:	$x^{k+1} := x^k + \alpha^k d^k$ $k := k + 1$ Ir a **1**.

2.3.2 Exercicis

Exercici 2.6. : pgraq/mètode del gradient amb funcions quadràtiques

Obtenció d'un punt des d'on el mètode del gradient arriba en una sola passa al mínim d'una funció objectiu quadràtica.

$$\frac{1}{2}X'\begin{bmatrix} 7/2 & 1 & 0 & -1/2 \\ 1 & 11/2 & -1/2 & -2 \\ 0 & -1/2 & 11/2 & -1 \\ -1/2 & -2 & -1 & 15/2 \end{bmatrix} X - [21/2 \quad 47 \quad -58 \quad 41/2]X \quad ; \quad X^* = \begin{bmatrix} 1 \\ 9 \\ -9 \\ 4 \end{bmatrix}$$

Valors propis de Q : $\lambda_i = \{9, 6, 4, 3\}$

SOLUCIÓ

Prenem $\lambda_4 = 3$ i determinem U_4 : $Q - \lambda_4 \mathbf{I} = \begin{bmatrix} 1/2 & 1 & 0 & -1/2 \\ 1 & 5/2 & -1/2 & -2 \\ 0 & -1/2 & 5/2 & -1 \\ -1/2 & -2 & -1 & 9/2 \end{bmatrix}$

Per a resoldre el sistema homogeni $(Q - \lambda_4 \mathbf{I})U_4 = \mathbf{0}$ prenem $u_{44} = 1$ i considerem les tres primeres equacions :

$$\begin{bmatrix} 1/2 & 1 & 0 \\ 1 & 5/2 & -1/2 \\ 0 & -1/2 & 5/2 \end{bmatrix} \begin{bmatrix} u_{41} \\ u_{42} \\ u_{43} \end{bmatrix} = \begin{bmatrix} 1/2 \\ 2 \\ 1 \end{bmatrix}$$

triangularitzant i resolent s'obté:

$$\begin{bmatrix} 1/2 & 1 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 5/2 \end{bmatrix} \begin{bmatrix} u_{41} \\ u_{42} \\ u_{43} \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ 1 \end{bmatrix} ; \begin{bmatrix} 1/2 & 1 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} u_{41} \\ u_{42} \\ u_{43} \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ 2 \end{bmatrix} \Rightarrow \underline{U_4 = \begin{bmatrix} -5 \\ 3 \\ 1 \\ 1 \end{bmatrix}}$$

Qualsevol punt $X_0 = X^* + \alpha U_4$ satisfà la condició demanada. Podem prendre $X_0 = X^* + U_4$:

$$X_0 = X^* + U_4 = \begin{bmatrix} 1 \\ 9 \\ -9 \\ 4 \end{bmatrix} + \begin{bmatrix} -5 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 12 \\ -8 \\ 5 \end{bmatrix} \quad ; \quad \boxed{X_0 = \begin{bmatrix} -4 \\ 12 \\ -8 \\ 5 \end{bmatrix}}$$

Exercici 2.7. : pgrad1/mètode del gradient

Feu una passa de minimització de la funció quadràtica:

$$q(X) = \frac{1}{2}X' \begin{bmatrix} 19/3 & -8\sqrt{2}/3 \\ -8\sqrt{2}/3 & 11/3 \end{bmatrix} X - [19 \quad -8\sqrt{2}]'X$$

des del punt $X_0 = [0 \quad 1]'$ obtenint el nou punt X_1. Comproveu que la funció d'error $\mathcal{E}(X) = \frac{1}{2}(X - X^*)'Q(X - X^*)$ (sent X^* el mínim de $q(X)$) als punts X_0 i X_1 satisfà la taxa de convergència local pròpia de les funcions quadràtiques, i comproveu que el gradient a X_1 és ortogonal a la passa $(X_1 - X_0)$.

SOLUCIÓ

PASSA PEL MÈTODE DEL GRADIENT A PARTIR DE X_0

Donada una funció quadràtica: $q(X) = \frac{1}{2}X'QX - B'X$ amb $X, B \in \mathbb{R}^n$, i $Q=Q' \in \mathbb{R}^{n \times n}$, l'expressió de l'algorisme del gradient és:

$$X_{k+1} = X_k - \alpha_k G_k \quad \text{amb } G_k = QX_k - B \text{ el gradient a } X_k \text{ i } \alpha_k = \frac{G_k'G_k}{G_k'QG_k}$$

$$G_0 = QX_0 - B = \begin{bmatrix} 19/3 & -8\sqrt{2}/3 \\ -8\sqrt{2}/3 & 11/3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 19 \\ -8\sqrt{2} \end{bmatrix} = \begin{bmatrix} -\frac{8\sqrt{2}+57}{3} \\ \frac{11+24\sqrt{2}}{3} \end{bmatrix}$$

$$G_0'G_0 = 742,94082 \quad G_0'QG_0 = 6680,2367 \quad \alpha_0 = 0,1112147$$

$$X_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 0,1112147 \begin{bmatrix} -\frac{8\sqrt{2}+57}{3} \\ \frac{11+24\sqrt{2}}{3} \end{bmatrix} = \begin{bmatrix} 2,5324972 \\ -0,6660379 \end{bmatrix}$$

FUNCIÓ D'ERROR A X_0 I A X_1

Per tal de calcular $\mathcal{E}(X)=\frac{1}{2}(X - X^*)'Q(X - X^*)$ necessitem coneixer X^* el qual és solució de $QX^* = B$. Resolent el sistema:

$$\begin{bmatrix} 19/3 & -8\sqrt{2}/3 \\ -8\sqrt{2}/3 & 11/3 \end{bmatrix} X^* = \begin{bmatrix} 19 \\ -8\sqrt{2} \end{bmatrix} \quad \text{obtenim } X^* = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

Fent els càlculs per a X_0 i per al punt trobat X_1 trobem que $\mathcal{E}(X_0)=\frac{1}{2}(X_0 - X^*)'Q(X_0 - X^*)$=41,647042 i que $\mathcal{E}(X_1)=\frac{1}{2}(X_1 - X^*)'Q(X_1 - X^*)$=0,33111455 .

FITA SUPERIOR A LA TAXA DE CONVERGÈNCIA

Sabem que:

$$\frac{\mathcal{E}(X_{k+1})}{\mathcal{E}(X_k)} \leq \left(\frac{A-a}{A+a}\right)^2 \quad \text{on } A \text{ i } a \text{ són el més gran i el més petit valor propi de } Q.$$

Hem de calcular els valors propis de Q i ho farem resolent el polinomi característic $\det(Q - \lambda\mathbf{1}) = 0$ (de grau 2 en aquest cas).

$$(\frac{19}{3} - \lambda)(\frac{11}{3} - \lambda) - \frac{128}{9} = 0 \implies \lambda^2 - 10\lambda + 9 = 0 \implies \begin{cases} \lambda_1 = A = 9 \\ \lambda_2 = a = 1 \end{cases}$$

COMPROVACIÓ DE LA TAXA DE CONVERGÈNCIA LOCAL

Efectivament:

$$\frac{\mathcal{E}(X_1)}{\mathcal{E}(X_0)} = \frac{0,33111455}{41,647042} = 0,00795 \leq \left(\frac{A-a}{A+a}\right)^2 = \left(\frac{9-1}{9+1}\right)^2 = 0,64$$

COMPROVACIÓ DE L'ORTOGONALITAT DE G_1 I $(X_1 - X_0)$

$$G_1 = QX_1 - B = \begin{bmatrix} 19/3 & -8\sqrt{2}/3 \\ -8\sqrt{2}/3 & 11/3 \end{bmatrix} \begin{bmatrix} 2,5324972 \\ -0,6660379 \end{bmatrix} - \begin{bmatrix} 19 \\ -8\sqrt{2} \end{bmatrix} = \begin{bmatrix} -0,4490648 \\ -0,6790754 \end{bmatrix}$$

$$G_1'(X_1 - X_0) = [-0,4490648 \quad -0,6790754] \begin{bmatrix} 2,5324972 \\ -1,6660379 \end{bmatrix} = 0 \quad \text{tal com diu la teoria}$$

Exercici 2.8. : pgrad2/mètode del gradient

Donada la funció $f(x_1, x_2) = 50x_1^2 + 3x_1x_3 + 4x_2^2 + 15x_3^2 - 6x_1 + 8x_2 - 60x_3$ es demana que:

2.8.1.– apliqueu una passa completa del mètode del gradient per a minimitzar aquesta funció a partir de $X_0 = [0 \quad -1 \quad 0]'$.

2.8.2.– comproveu, a través de la funció d'error $\epsilon(X) = (X - X^*)'Q(X - X^*)/2$, que la convergència local de l'algorisme entre X_0 i X_1 (que haureu trobat) està dins del límit teòric.

Exercici 2.9. : pgrad3/mètode del gradient

La taula que ve a continuació mostra el valor de la funció objectiu al llarg de tres iteracions, a prop de l'optim, del procés de resolució d'un problema d'optimització no lineal sense constriccions $\min_{x \in \Re^n} f(x)$ que té un minim local amb valor $f(x^*) = 1$:

$f(x^k) =$	0.9991
$f(x^{k+1}) =$	0.999325
$f(x^{k+2}) =$	0.99949375

2.9.1.– Raoneu, a partir dels valors de la taula, si el mètode d'optimització aplicat pot ser el mètode de Newton.

2.9.2.– Indiqueu, raonadament, si els valors que s'indiquen a la taula poden correspondre a l'aplicació del métode del Gradient sobre la funció objectiu $f(x) = x_1^3 + x_2^2 - 3x_1 - 2x_2 + 4$

Exercici 2.10. : pgrad4/Métode del gradient.

Considereu el problema **(P)** consistent en la minimització de la funció $f(x) = 2x_1^2 - x_2^2 + x_1x_3 - 3x_2 + x_3$.

2.10.1.– Efectueu una passa del métode del gradient a partir de $x^0 = (1 \quad 1 \quad 1)'$ amb exploració lineal exacta.

2.10.2.– Considereu que l'aplicació del métode del gradient en la resolució de **(P)** a partir de x^0 convergeix a un punt extrem x^*. És x^* la solució del problema **(P)**? Per què? Quína és la solució del problema **(P)**?

2.10.3.– Calculeu la fita superior a la taxa de convergència β del métode del gradient aplicat sobre **(P)**.

SOLUCIÓ

2.10.1.– $x^1 = [-2.8571 \quad 4.8571 \quad -0.5928]$

2.10.2.– Si es calcula la definició de la matriu Hessiana (constant, doncs $f(x)$ és quadràtica), s'obté que és indefinida. Així doncs, el problema **(PNL)** no té solució, doncs no n'hi ha cap punt de $\mathbb{R}^3$ que satisfaci les condicions necessàries de segon ordre ($\nabla f(x) = [0]$ i $\nabla^2 f(x)$ semidef +).

2.10.3.– Atés que el problema no té solució, no té sentit calcular la taxa de convergència. Si es calculés s'obtindria $\beta = 7.77 > 1$, que no té sentit.

2.4 Mètode del gradient conjugat

2.4.1 Procediment

2.4.1.1 Procediment de direccions conjugades per a funcions quadràtiques

Sigui la minimització:

$$\underset{X}{\text{minimitzeu}} \quad \frac{1}{2}X'QX - B'X \tag{3}$$

amb $B, X \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n\times n}$ simètrica i definida positiva.

Entenem per direccions Q–conjugades els vectors $D_j \in \mathbb{R}^n$ $(j = 0, ..., n-1)$ tals que:

$$D_i'QD_j = 0 \quad \forall j \neq i$$

El teorema de les direccions conjugades diu que, donades n direccions Q–conjugades D_j $(j = 0, ..., n-1)$ a $\mathbb{R}^n$, partint d'un punt X_0 qualsevol, la seqüència $\{X_k\}$ generada amb $X_{k+1} = X_k + \alpha_k D_k$, sent α_k la passa d'exploració lineal exacta al llarg de la direcció Q–conjugada D_k, arribarem al mínim X^* de la funció quadràtica.

Per passa analítica exacta des d'un punt X_k en la direcció D_k entenem la que minimitza l'escalar α de l'expressió:

$$\underset{\alpha}{\text{minimitzeu}} \quad \frac{1}{2}(X_k + \alpha D_k)'Q(X_k + \alpha D_k) - B'(X_k + \alpha D_k) \tag{4}$$

Pot comprovar-se fàcilment que la passa òptima és:

$$\alpha = \frac{-G_k' D_k}{D_k' Q D_k} \quad \text{sent: } G_k = QX_k - B \text{ gradient de la funció quadràtica a } X_k$$

2.4.1.2 Mètode del gradient conjugat

Sigui la minimització (3). L'algorisme següent, del *gradient conjugat*, troba el mínim del problema.

0) Per a un punt inicial qualsevol X_0 calculem $G_0 = QX_0 - B$ i fem $D_0 = -G_0$.
1) Fem $X_{k+1} = X_k + \alpha D_k$ amb $\alpha_k = \dfrac{-G_k' D_k}{D_k' Q D_k}$.
2) Calculem $G_{k+1} = G_k + QD_k$ i obtenim $G_{k+1}' G_{k+1}$
   - ▷ si $G_{k+1}' G_{k+1} > \epsilon$ anem a 3)
   - ▷ si $G_{k+1}' G_{k+1} \leq \epsilon$ $\boxed{\text{FI}}$: ja som a l'òptim.
3) Calculem $D_{k+1} = -G_{k+1} + \beta_k D_k$, amb $\beta_k = \dfrac{G_{k+1}' G_{k+1}}{G_k' G_k}$.
4) Fem $k := k+1$ i anem a 1).

2.4.1.3 Bibliografia

El mètode de les direccions conjugades està descrit a [LUEN84] pp. 238–243. El mètode del gradient conjugat està a la mateixa referència [LUEN84] pp. 243–254.

2.4.2 Exercicis

Exercici 2.11. : pdircl/direccions conjugades.

Donada la funció quadràtica a $\mathbb{R}^2$:

$$\frac{1}{2}X'\begin{bmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 5 \end{bmatrix}X - [15-\sqrt{3} \quad 5-5\sqrt{3}]X$$

i les direccions Q-conjugades D_0 i D_1, calculeu, a partir del resultat del teorema de les direccions conjugades, quina ha de ser la direcció D_1 suposant que $X_0=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ i que $D_0=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Trobeu també D_1 si $X_0=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ i $D_0=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Representeu gràficament la solució del problema.

(El teorema de les direccions conjugades diu que, donades n direccions Q–conjugades D_j ($j = 0, ..., n-1$) a $\mathbb{R}^n$, partint d'un punt X_0 qualsevol, la seqüència $\{X_k\}$ generada amb $X_{k+1}=X_k+\alpha_k D_k$, sent α_k la passa d'exploració lineal exacta al llarg de la direcció Q–conjugada D_k, arribarem al mínim X^* de la funció quadràtica).

Exercici 2.12. : pdircl/direccions conjugades.

Ja que estem a $\mathbb{R}^2$, si deduïm X_1, sabem que D_2 ha de ser tal que:

$$X^* - X_1 = \alpha_1 D_1 \quad \text{i} \quad QX^* = B$$

OBTENCIÓ DE X^*

Del sistema $QX^*=B$

$$\begin{bmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 5 \end{bmatrix}X^* = \begin{bmatrix} 15-\sqrt{3} \\ 5-5\sqrt{3} \end{bmatrix} \quad \rightarrow \quad X^* = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

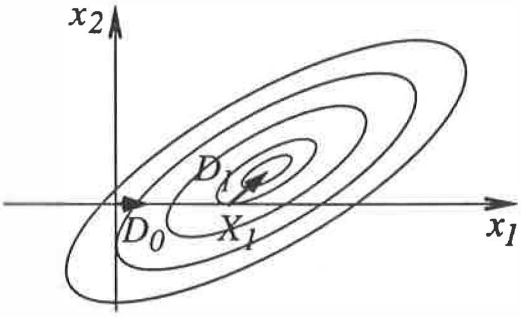
OBTENCIÓ DE X_1 QUAN $D_0=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ Calculem $\alpha_0=-\frac{G_0'D_0}{D_0'QD_0}$

$$G_0 = QX_0 - B \quad G_0 = -B \text{ si } X_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad G_0 = \begin{bmatrix} \sqrt{3}-15 \\ 5\sqrt{3}-5 \end{bmatrix} \quad \alpha_0 = \frac{15-\sqrt{3}}{3}$$

$$X_1 = X_0 + \alpha_0 D_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{15-\sqrt{3}}{3}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5-\frac{\sqrt{3}}{3} \\ 0 \end{bmatrix}$$

OBTENCIÓ DE D_1 QUAN $D_0=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$D_1 = X^* - X_1 \text{ (o qualsevol vector multiple d'aquest)} \quad D_1 = \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 5-\frac{\sqrt{3}}{3} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{3} \\ 1 \end{bmatrix}$$



Efectivament tenim que $D_0'QD_1=0$

OBTENCIÓ DE X_1 QUAN $D_0=\begin{bmatrix}0\\1\end{bmatrix}$

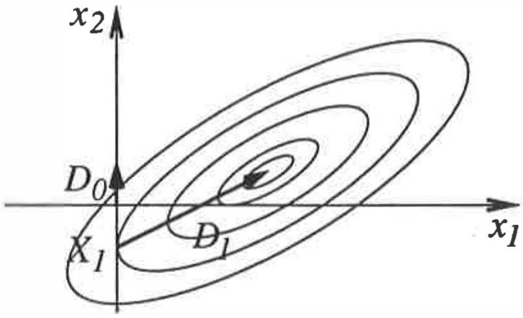
Calculem $\alpha_0=-\frac{G_0'D_0}{D_0'QD_0}$

$$G_0 = QX_0 - B = \begin{bmatrix}\sqrt{3}-15\\5\sqrt{3}-5\end{bmatrix} \quad \alpha_0 = \frac{5-5\sqrt{3}}{5} = 1-\sqrt{3}$$

$$X_1 = X_0 + \alpha_0 D_0 = \begin{bmatrix}0\\0\end{bmatrix} + (1-\sqrt{3})\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\1-\sqrt{3}\end{bmatrix}$$

OBTENCIÓ DE D_1 QUAN $D_0=\begin{bmatrix}0\\1\end{bmatrix}$

$D_1 = X^* - X_1$ (o qualsevol vector multiple d'aquest) $D_1 = \begin{bmatrix}5\\1\end{bmatrix} - \begin{bmatrix}0\\1-\sqrt{3}\end{bmatrix} = \begin{bmatrix}5\\\sqrt{3}\end{bmatrix}$



També tenim que $D_0'QD_1=0$

Exercici 2.13. : pgrcol/gradient conjugat

Determineu una fita superior a la funció d'error $\mathcal{E}(X)$ d'una funció quadràtica $q(X) = \frac{1}{2}X'QX - B'X$, amb $X, B \in \mathbb{R}^n$, i $Q=Q' \in \mathbb{R}^{n\times n}$, en el punt X_3 obtingut després de tres iteracions del mètode del gradient conjugat a partir d'un punt X_0 per a un cas amb n=8 , sabent que els valors propis de Q són:

$$\lambda_i = \{\, 126,81 \quad 126,8 \quad 126,75 \quad 30,04 \quad 30,03 \quad 30 \quad 29,98 \quad 2,7 \,\}$$

, i que $\mathcal{E}(X_0)$=100.

L'expressió de la funció d'error és $\mathcal{E}(X)=\frac{1}{2}(X-X^*)'Q(X-X^*)$ (sent X^* el mínim de $q(X)$), i la propietat de l'algorisme del gradient conjugat aplicat a funcions quadràtiques que lliga $\mathcal{E}(X)$ als valors propis de Q és:

$$\mathcal{E}(X_{k+1}) \leq \max_{\forall \lambda_i} \left[1+\lambda_i P_k(\lambda_i)\right]^2 \mathcal{E}(X_0)$$

sent $P_k(\lambda_i)$ un polinomi qualsevol de grau k en λ_i i sent X_k els successius punts iterats del gradient conjugat.

SOLUCIÓ

DEFINICIÓ DEL POLINOMI DE GRAU $k+1$ ($k+1=3$)

Després de tres passes a partir de X_0 tindrem X_3. Es tracta doncs de definir $\mathcal{E}(X_{k+1})$=$\mathcal{E}(X_3)$.

Com a polinomi de grau 3 (amb terme independent unitat) corresponent a $1+\lambda_i P_2(\lambda_i)$ podem prendre:

$$1+\lambda_i P_2(\lambda_i) = \frac{(126,8-\lambda_i)(30-\lambda_i)(2,7-\lambda_i)}{126,8\cdot 30\cdot 2,7}$$

MAXIMITZACIÓ DEL POLINOMI DE GRAU $k+1$ PER ALS VALORS PROPIS DE Q

De tots els valors propis $\lambda_i = \{\, 126,81 \quad 126,8 \quad 126,75 \quad 30,04 \quad 30,03 \quad 30 \quad 29,98 \quad 2,7 \,\}$, aquell que pot maximitzar:

$$\max_{\forall \lambda_i} \frac{(126,8-\lambda_i)(30-\lambda_i)(2,7-\lambda_i)}{126,8\cdot 30\cdot 2,7}$$

és el valor propi 126,75 ja que és el dels més grans que difereix més de 126,8 (constant del multiplicand del numerador que té el més gran terme independent). Així:

$$\max_{\forall \lambda_i} 1+\lambda_i P_2(\lambda_i) = \frac{(126,8-126,75)(30-126,75)(2,7-126,75)}{126,8\cdot 30\cdot 2,7}$$
$$= \frac{0,05(-96,75)(-124,05)}{126,8\cdot 30\cdot 2,7} = 0,058427$$

FITA A LA FUNCIÓ D'ERROR A X_3

$$\mathcal{E}(X_3) \leq (0,0584)^2 \mathcal{E}(X_0) = 0,0034137\cdot 100 = 0,34137$$

Exercici 2.14. : pgrco2/gradient conjugat

Considereu la funció quadràtica $f(X) = \frac{1}{2}X' \begin{pmatrix} 2 & 1 & 2 \\ 1 & 3 & 0 \\ 2 & 0 & 4 \end{pmatrix} X - (5 \quad 4 \quad 6)X$ i el punt $X_0 = (1 \quad 1 \quad 1)'$.

2.14.1.– Apliqueu una passa de la minimització de $f(X)$ amb el mètode del gradient conjugat a partir de X_0. Calculeu el nou punt iterat X_1 així com la direcció D_1 del mètode del gradient conjugat a partir de X_1.

2.14.2.– Sabent que els valors propis de la matriu Q de $f(X)$ són $\lambda \approx \{0.47108, 3.16745, 5.36147\}$, calculeu una fita superior de $\mathcal{E}(X_1)$, la funció d'error sobre X_1.

Exercici 2.15. : pgrco3/gradient conjugat

Considereu el problema $\min_{X\in\Re^3} f(X) = (x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2$. Fent una primera iteració del Mètode del Gradient Conjugat (MGC) a partir del punt inicial $X_0 = [0 \quad 0 \quad 0]'$, s'arriba al punt $X_1 = [1/3 \quad 0 \quad 1/3]'$.

2.15.1.– Obtingueu l'iterat X_2 aplicant una nova iteració del MGC a partir de X_1, prenent com a longitud de pas el resultat d'un ajust quadràtic entre $\alpha = 0$ i $\alpha = 1$.

2.15.2.– Comproveu la segona condició d'Armijo-Goldstein pel punt trobat a l'ajust quadràtic de l'apartat anterior, prenent $\beta = 0.5$. A la vista del resultat obtingut indiqueu, sense fer el producte escalar $\nabla f(X_2)D_2$, si la direcció D_2 que prendria el MGC a partir de X_2 seria de descens.

2.5 Mètode de Newton

2.5.1 Procediment

Sigui la minimització d'una funció qualsevol:

$$\underset{X}{\text{minimitzeu}} \quad f(X) \tag{5}$$

amb $X\in\mathbb{R}^n$, $f(X) : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(X)\in\mathcal{C}^2$ (contínua i fins a segones derivades contínues).

L'algorisme <u>pur</u> de Newton consistiria en obtenir punts

$$X_{k+1} = X_k - \alpha_k \nabla^2 f(X_k)^{-1} \nabla f(X_k)' \tag{6}$$

on α_k és una passa adequada d'exploració lineal. Es demostra que, localment, aquest algorisme presenta convergència d'ordre 2.

Ara bé, sabem d'una banda per les condicions de mínim de la solució X^* que l'hessiana $\nabla^2 f(X^*)$ és definida positiva, i que per continuitat $\nabla^2 f(X_k)$ també serà definida positiva en entorn (petit) de X^* ($\|X_k - X^*\| < \epsilon$). Fora d'aquest entorn no tenim cap garantia que l'hessiana $\nabla^2 f(X_k)$ sigui definida positiva.

D'altra banda sabem que la direccó $P_k = -M\nabla f(X_k)'$ de la família d'algorismes:

$$X_{k+1} = X_k - \alpha_k M \nabla f(X_k)' \tag{7}$$

és de descens si la matriu $M \in \mathbb{R}^{n\times n}$ és definida positiva, i no podem garantir que ho sigui si M no és definida positiva

Com a conseqüència deduïm que fora d'un entorn de la solució no podem estar segurs que l'algorisme de Newton pur (6) ens doni direccions $-\nabla^2 f(X_k)^{-1}\nabla f(X_k)$ que siguin de descens, violant així una condició necessària per garantir la convergència global de l'algorisme.

Cal doncs modificar el mètode de Newton per tal de garantir la convergència global. Les modificacions habituals consisteixen en prendre $H_k \in \mathbb{R}^{n\times n}$ definida positiva en comptes de $\nabla^2 f(X_k)$:

$$X_{k+1} = X_k - \alpha_k H_k^{-1}\nabla f(X_k)' \tag{8}$$

2.5.1.1 Factorització de Cholesky d'una matriu

Donada una matriu simètrica H (tal com ho és l'hessiana $\nabla^2 f(X_k)$), ens interessa trobar la descomposició d'H en producte de matrius subtriangular amb uns a la diagonal S, diagonal D, i sobretriangular S' (transposada de la subtriangular S): $H{=}SDS'$:

$$\begin{bmatrix} h_{11} & h_{21} & h_{31} & & h_{n1} \\ h_{21} & h_{22} & h_{32} & \cdots & h_{n2} \\ h_{31} & h_{32} & h_{33} & & h_{n3} \\ & \cdots & & \ddots & \vdots \\ h_{n1} & h_{n2} & h_{n3} & \cdots & h_{nn} \end{bmatrix} =$$
$$\begin{bmatrix} 1 & & & & \\ s_{21} & 1 & & \mathbf{0} & \\ s_{31} & s_{32} & 1 & & \\ & \cdots & & \ddots & \\ s_{n1} & s_{n2} & s_{n3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} d_1 & & & \\ & d_2 & \mathbf{0} & \\ & & d_3 & \\ & \mathbf{0} & & \ddots & \\ & & & & d_n \end{bmatrix} \begin{bmatrix} 1 & s_{21} & s_{31} & & s_{n1} \\ & 1 & s_{32} & \cdots & s_{n2} \\ & & 1 & & s_{n3} \\ & \mathbf{0} & & \ddots & \vdots \\ & & & & 1 \end{bmatrix} \tag{9}$$

Les expressions recursives que ens permeten obtenir els elements de la factorització són:

$$d_j = h_{jj} - \sum_{i=1}^{j-1} d_i s_{ji}^2 \quad j = 1,\ldots,n \tag{10}$$

$$s_{kj} = \frac{1}{d_j}\Big(h_{kj} - \sum_{i=1}^{j-1} d_i s_{ji} s_{ki}\Big) \quad k = j+1,\ldots,n \tag{11}$$

2.5.1.2 Obtenció de la direcció P_k

$P_k{=}-H_k^{-1}\nabla f(X_k)'$ l'obtindrem per resolució del sistema $H_k P_k{=}-\nabla f(X_k)'$.

Si fessim la factorització de Cholesky de l'hessiana $\nabla^2 f(X_k){=}S_k D_k S_k'$, sent S_k una matriu triangular inferior amb uns a la diagonal i D_k una matriu diagonal, podrien haver-hi a D_k elements diagonals negatius. També podria ser que aquesta factorització no existís o fos pràcticament impossible d'obtenir, si $\nabla^2 f(X_k)$ fos singular o quasi, perquè en fer els càlculs (11) tindriem una divisió per zero (o quasi) que ens faria abortar la factorització.

La modificació computacionalment més económica de $\nabla^2 f(X_k)$ es prendre $H_k{=}\ S_k\tilde{D}_k S_k'$ sent $\tilde{D}_k$ una matriu diagonal amb tots els elements diagonals positius. Amb aquesta H_k tenim garantit que obtindrem direccions de descens. Una forma, proposada per Luenberger, d'obtenir $\tilde{D}_k$ a partir de D_k, quan aquesta té elements negatius, és prendre $\tilde{D}_k{=}D_k{+}\mu\mathbf{1}$ amb $\mu{=}\delta{+}\left|\min\{d_i|\forall d_i < 0\}\right|$ sent δ una petita constant positiva (p.e. $\delta{=}0.01$). D'aquesta forma asegurem l'obtenció de direccions de descens mentres que H_k no està excessivament allunyada de $\nabla^2 f(X_k)$ i així tractem de conservar la propietat de convergència quadràtica del métode de Newton.

2.5.1.3 Algorisme de Newton variant de Luenberger

L'algorisme de minimització d'una funció qualsevol per la variant de Luenberger del mètode de Newton és el següent:

0) Fixem:
   — una constant petita $\delta>0$ (p.e. $\delta=0.01$),
   — una constant, dita de desplaçament, ρ (p.e. $\rho=0.1$),
   — una tolerància petita ϵ (p.e. $\epsilon=0.00001$), i
   — prenem un punt inicial qualsevol X_0 i en calculem el gradient $\nabla f(X_0)$ i l'hessiana $\nabla^2 f(X_0)$.
1) Fem la factorització de Cholesky de l'hessiana $\nabla^2 f(X_k)=S_k D_k S_k'$. (Noti's que en fer la factorització el programa que implementi aquest algorisme pot abortar per una divisió per zero si $\nabla^2 f(X_k)$ és semidefinida o quasi).
2) Calculem $\|\nabla f(X_k)\|^2$ i el comparem amb una tolerància petita ϵ
   ▷ si $\|\nabla f(X_k)'\|^2>\epsilon$ anem a 3)
   ▷ si $\|\nabla f(X_k)'\|^2\leq\epsilon$ i les diagonals D_k de la factorització son totes positives $\boxed{\text{FI}}$: ja som a l'òptim.
   ▷ si $\|\nabla f(X_k)'\|^2\leq\epsilon$ i les diagonals D_k tenen un o més elements negatius, som a prop d'un punt de sella i cal desplaçar-se del punt actual (p.e. fent $X_{k+1} = X_k \pm \rho\mathbf{1}$) i anem a 5)
3) Calculem la direcció P_k:
   ▷ si tots els elements diagonals de D_k són superiors a δ calculem P_k resolent $S_k D_k S_k' P_k=-\nabla f(X_k)'$ i anem a 4)
   ▷ si no tots els elements diagonals de D_k són superiors a δ fem $\tilde{D}_k=D_k+\mu\mathbf{1}$ amb $\mu=\delta+\left|\min\left\{d_i|\forall d_i<0\right\}\right|$, resolem $S_k\tilde{D}_k S_k' P_k=-\nabla f(X_k)'$ i anem a 4)
4) Fem exploració lineal a partir de X_k en la direcció P_k de $f(X)$ obtenint la passa α_k i actualitzem el punt fent $X_{k+1}=X_k+\alpha_k P_k$
5) Calculem els nous gradient $\nabla f(X_{k+1})$ i hessiana $\nabla^2 f(X_{k+1})$
6) Fem $k:=k+1$ i anem a 1).

2.5.1.4 Triangularització de Gill-Murray

Qan la matriu H de (9) és indefinida o és definida positiva per molt poc, d_j obtingut amb (10) pot ser molt petit, i aleshores els valors de s_{kj} obtinguts amb (11) poden ser molt grans en valor absolut. Aleshores tant els posteriors d_{j+1} com els $s_{k\,j+1}$ poden arrossegar un important error numèric que desvirtua els càlculs que es puguin efectuar amb S i D. A més, ens pot interessar quan $H=\nabla^2 f(X_k)$ sigui indefinida, d'obtenir uns factors $\tilde{S}$ i $\tilde{D}$ que corresponguin a una matriu definida positiva el més propera possible (en norma) a H.

Principis

Una forma d'obtenir això és imposar que, per a una certa $\delta>0$, tinguem:

$$d_j > \delta \tag{12}$$

$$|s_{kj}\sqrt{d_j}| \leq \beta \quad k=j+1,\ldots,n \tag{13}$$

amb $\beta>0$ determinada tal com es veurà més endavant.

La condició (12) assegura que la matriu a la que corresponen els factors és definida positiva, i la condició (13) assegura que $|s_{kj}|$ queda limitat en mida.

Per tal d'assegurar que es satisfagin les condicions (12) i (13), podem procedir a triangularitzar H de la forma següent:

Suposant que ja haguem obtingut fins a la diagonal d_{j-1} i obtingut els $s_{k\,j-1}$ $(k{=}j,\ldots,n)$, fem:

0) $\xi_j{=}h_{jj}$ i $e_{jj}{=}0$ (e_{jj} serà un terme a afegir a l'element diagonal h_{jj} de H)

1) $\gamma_j{=}\left|\xi_j{-}\sum_{i=1}^{j-1} d_i s_{ji}^2\right|$

2) $\overline{d}{=}\max\{\gamma_j, \delta\}$ (amb el qual assegurem el compliment de (12))

3) Calculem per a $k{=}j{+}1,\ldots,n$

$$s_{kj} = \frac{1}{\overline{d}}\left(h_{kj} - \sum_{i=1}^{j-1} d_i s_{ji} s_{ki}\right) \quad \text{i comprovem si } \left|s_{kj}\sqrt{\overline{d}}\right| \leq \beta$$

quan aquesta darrera condició es satisfà per a tota $k{=}j{+}1,\ldots,n$, queda cumplida la condició (13), però això no té perquè passar així per a qualsevol β (per a β gran (13) sempre es satisfà, per a β petita (13) pot no satisfer-se mai). Si no es satisfés per a una β donada, la política a seguir és la d'afegir un escalar positiu e_{jj} a h_{jj} de forma que amb $\xi_j{=}h_{jj}{+}e_{jj}$ i tornant a 1 tinguessim que, en el pitjor dels casos $\max_{\forall k}\left|s_{kj}\sqrt{d_j}\right|{=}\beta$

Aquest algorisme descrit, el qual pot variar elements diagonals (a la passa 2) i afegir–hi escalars positius (a la passa 3), equival a fer una factorització normal d'una matriu modificada: $H{+}E{=}\widetilde{S}\widetilde{D}\widetilde{S}'$, on E és la matriu diagonal d'elements e_{jj}.

D'altra banda l'algorisme descrit deixa mal resolta la passa 3 perquòbliga a un procés de prova–i–error ajustant la quantitat a afegir e_{jj} (si cal) per satisfer la condició (13). A més no s'especifica com fixar β. Per tot això és necessari establir un procediment directe que no tingui aquests inconvenients.

Satisfacció de la condició (13)

Si en el pitjor dels casos: $\max_{\forall k}|s_{kj}\sqrt{d_j}|{=}\beta$ o $\max_{\forall k} s_{kj}^2 d_j{=}\beta^2$, tenint en compte (11) obtenim:

$$\max_{\forall k} s_{kj}^2 d_j = \max_{\forall k}\left\{\frac{1}{d_j}\left(h_{kj} - \sum_{i=1}^{j-1} d_i s_{ji} s_{ki}\right)\right\}^2 d_j = \frac{1}{d_j}\left(\max_{\forall k}\left\{\left(h_{kj} - \sum_{i=1}^{j-1} d_i s_{ji} s_{ki}\right)^2\right\}\right) = \beta^2$$

d'on, posant $\theta_j{=}\max_{\theta_j \forall k}\left|h_{kj} - \sum_{i=1}^{j-1} d_i s_{ji} s_{ki}\right|$, queda: $\frac{\theta_j^2}{d_j}{=}\beta^2$ o bé $d_j{=}\frac{\theta_j^2}{\beta^2}$, és a dir,una expressió de d_j en funció de β i que satisfà (13), per tant podem posar:

$$\widetilde{d}_j = \max\left\{\gamma_j, \delta, \frac{\theta_j^2}{\beta^2}\right\} \quad \text{i} \quad e_{jj} = \widetilde{d}_j - \left(h_{jj} - \sum_{i=1}^{j-1} \widetilde{d}_i \widetilde{s}_{ji}^2\right) \tag{14}$$

Determinació de β

Tenim d'una banda que $\beta^2{\geq}s_{kj}^2 d_j$ $\forall j$, $\forall k{>}j$, i que quan H sigui suficientment definida positiva, la matriu diagonal E de correccions e_{jj} hauria de ser zero, i naturalment, de (10) tindrem que $d_k{>}0$, és a dir, $h_{kk}{>}s_{kj}^2 d_j$ $\forall k{>}j$, d'on el màxim valor de h_{kk}, que designarem per $\overline{h}_{jj}$, és una bona fita superior de β^2.

D'altra banda volem evitar correccions e_{jj} excessivament grans. Podem tenir en compte que es pot deduir (però aquí no ho fem) que una fita superior a $\|E(\beta)\|_\infty$ és:

$$\|E(\beta)\|_\infty \leq \left(\frac{\overline{h}_{kj}}{\beta} + (n-1)\beta\right)^2 + 2(n-1)\beta^2 + 2\overline{h}_{jj} + \delta$$

sent $\overline{h}_{kj}$ el més gran element no diagonal d'H , i que la fita superior de l'expressió anterior té un mínim a $\beta^2 = \dfrac{\overline{h}_{kj}}{\sqrt{n^2-1}}$.

Finalment també cal tenir en compte que la intervenció de $\overline{h}_{jj}$ i de $\overline{h}_{kj}$ a les fites trobades i la possibilitat que els elements d'H siguin molt petits aconsella incloure la precisió de la màquina, anomenada aquí p_m , com a fita de β^2 , així:

$$\beta^2 = \max\left\{\overline{h}_{jj}, \frac{\overline{h}_{kj}}{\sqrt{n^2-1}}, p_m\right\} \tag{15}$$

Algorisme

Tenint en compte (14) i (15), la forma final de la triangularització de Gill–Murray queda com:

0) $\beta^2 = \max\left\{\overline{h}_{jj}, \dfrac{\overline{h}_{kj}}{\sqrt{n^2-1}}, p_m\right\}$, $j=1$

1) $\gamma_j = \left|h_{jj} - \sum_{i=1}^{j-1} \tilde{d}_i \tilde{s}_{ji}^2\right| \quad \theta_j = \max\limits_{j+1 \leq k \leq n} \left|h_{kj} - \sum_{i=1}^{j-1} \tilde{d}_i \tilde{s}_{ji} \tilde{s}_{ki}\right|$

$\quad \tilde{d}_j = \max\left\{\gamma_j, \delta, \dfrac{\theta_j^2}{\beta^2}\right\}$ i $e_{jj} = \tilde{d}_j - \left(h_{jj} - \sum_{i=1}^{j-1} \tilde{d}_i \tilde{s}_{ji}^2\right)$

2) $\tilde{s}_{kj} = \dfrac{1}{\tilde{d}_j}\left(h_{kj} - \sum_{i=1}^{j-1} \tilde{d}_i \tilde{s}_{ji} \tilde{s}_{ki}\right) \quad k = j+1, \ldots, n$

3) $j := j+1$, i anem a 1)

Així obtenim la triangularització $\tilde{S}\tilde{D}\tilde{S}' = \tilde{H} = H + E$, sent E una matriu diagonal amb elements e_{jj} (correccions a les diagonals) calculades a la passa 1). La norma de Frobenius de la diferència entre $\tilde{H}$ i H és $\|\tilde{H} - H\|_F = \|E\|_F = \sqrt{\sum_{j=1}^n e_{jj}^2}$.

Direcció de corbatura negativa

Una interessant propietat addicional de la triangularització de Gill–Murray és que permet utilitzar un procediment eficient d'obtenir una direcció de corbatura negativa P_- (tal que $P_-' H P_- < 0$). Aquesta direcció P_- és especialment útil quan $H = \nabla^2 f(X_k)$ és indefinida i el gradient $\nabla f(X_k)$ és nul o gairebé nul.

La direcció de corbatura negativa s'obté resolent el sistema $\tilde{S}' P_- = \mathcal{E}_m$ sent $\mathcal{E}_m$ la columna m de la matriu unitat, i sent m l'índex de la més gran e_{jj} obtinguda a la passa 1):

$$e_{mm} = \max\left\{e_{jj}\ ,\ \forall j \left| h_{jj} - \sum_{i=1}^{j-1} \tilde{d}_i \tilde{s}_{ji}^2 < 0\right.\right\} .$$

En efecte:

$$P'_{-}HP_{-} = P'_{-}(\tilde{H} - E)P_{-} = P_{-}\tilde{S}\tilde{D}\tilde{S}'P_{-} - P'_{-}EP_{-} = \mathcal{E}'_m\tilde{D}\mathcal{E}_m - \sum_{j=1}^{n} e_{jj}p_{-j}^2 = \tilde{d}_m - \sum_{j=1}^{n} e_{jj}p_{-j}^2$$

Ara bé, donat que la solució P_{-} del sistema:

$$\begin{bmatrix} 1 & \tilde{s}_{21} & \dots & \tilde{s}_{m1} & & \tilde{s}_{n1} \\ & 1 & \dots & \tilde{s}_{m2} & \dots & \tilde{s}_{n2} \\ & & \ddots & & & \\ & \mathbf{0} & & 1 & \dots & \tilde{s}_{nm} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{bmatrix} \begin{bmatrix} p_{-1} \\ p_{-2} \\ \vdots \\ p_{-m} \\ \vdots \\ p_{-n} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

mostra que $p_{-m}=1$ i que $p_{-j}=0$, $j=m+1,\dots,n$, el sumatori de l'expressió anterior queda:

$$P'_{-}HP_{-} = \tilde{d}_m - \sum_{j=1}^{n} e_{jj}p_{-j}^2 = \tilde{d}_m - e_{mm} - \sum_{j=1}^{m-1} e_{jj}p_{-j}^2 = h_{mm} - \sum_{j=1}^{m-1} \tilde{d}_i\tilde{s}_{mi}^2 - \sum_{j=1}^{m-1} e_{jj}p_{-j}^2 < 0$$

La derrera desigualtat és certa perquè el derrer sumatori és tot de termes positius i perquè el terme $h_{mm} - \sum_{j=1}^{m-1} \tilde{d}_i\tilde{s}_{mi}^2$ és negatiu (de fet és el més negatiu de tots, ja que ha requerit la més gran correcció e_{mm}). Així $P'_{-}HP_{-}<0$ □.

2.5.1.5 Mètode de Newton amb triangularització de Gill-Murray

Substituint la factorització de Cholesky per la de Gill-Murray a l'algorisme de Newton obtenim sempre direccions de descens ja que $\tilde{H}$ és definida positiva.

A més, en cas de gradient petit prop d'un punt de sella, en podrem sortir prenent una direcció de corbatura negativa generada a partir de la triangularització de Gill-Murray.

2.5.1.6 Mètode de Newton segons la variant de Dennis–Schnabel

La variant de Dennis–Schnabel proposa una correcció uniforme de l'hessiana:

$$H = \nabla^2 f(X_k) + \mu\mathbf{1}$$

amb la constant positiva μ calculada a partir de la factorització de Gill-Murray i de l'aplicació de la fita inferior de Gerxgorin als valors propis de $\nabla^2 f(X_k)\nabla^2 f(X_k)$.

Sigui ν la fita inferior als valors propis de $\nabla^2 f(X_k)$, (la qual serà negativa sempre que l'hessiana sigui indefinida) i sigui e_{mm} la més gran correcció realitzada a la digonal en la factorització de Gill–Murray, aleshores:

$$\mu = \min\{|\nu|\ ,\ e_{mm}\}$$

és la constant de la correcció uniforme de l'hessiana.

Aquesta variant requereix dues factoritzacions per iteració quan $\nabla^2 f(X_k)$ és indefinida: la de Gill–Murray de $\nabla^2 f(X_k)$ i la de Cholesky de $\nabla^2 f(X_k)+\mu\mathbf{1}$.

2.5.1.7 Bibliografia.

La convergència del mètode de Newton està analitzada a [LUEN84] pp. 225–226. La modificació de Luenberger està descrita en l'edició de 1973 del mateix text , pp. 157–158. La triangularització de Gill-Murray està a [GILL81] pp. 107–112. El procediment d'obtenció d'una direcció de corbatura negativa es troba a [GILL74]. La variant de Dennis–Schnabel està descrita a [DESCH83] pp. 99–103, i pot trobar–se un enunciat del teorema de Gerxgorin a [DESCH83] p. 60.

2.5.2 Exercicis

Exercici 2.16. : pluenl/mètode de Newton, variant de Luenberger

La funció

$$f(X) = x_1^3 + x_2^2 + \frac{11}{2}x_3^2 + x_4^3 + \frac{5}{2}x_5^2 + x_1x_2^2 + 6x_1x_3 + 3x_1x_5 + 2x_2x_4 + x_2^2x_5 + 6x_3x_5 + x_4x_5$$

té l'hessiana indefinida al punt $X_0 = [1/2 \quad 0 \quad -1 \quad 1/2 \quad 0]'$

Trobeu la direcció de minimització a partir de X_0 pel mètode de Newton en la variant de Luenberger i comproveu que la direcció P obtinguda és de descens. Efectueu exploració lineal per ajust quadràtic en la direcció P i en la $[2 \quad 0 \quad -1 \quad 0 \quad 0]'$.

SOLUCIÓ

VECTOR GRADIENT, MATRIU HESSIANA I LA SEVA TRIANGULARITZACIÓ.

$$\nabla f(X_0)' = \begin{bmatrix} 3x_1^2 + x_2^2 + 6x_3 + 3x_5 \\ 2x_2 + 2x_1x_2 + 2x_4 + 2x_2x_5 \\ 11x_3 + 6x_1 + 6x_5 \\ 3x_4^2 + 2x_2 + x_5 \\ 5x_5 + 3x_1 + x_2^2 + 6x_3 + x_4 \end{bmatrix} \quad ; \quad \nabla f(X_0)' = \begin{bmatrix} -21/4 \\ 1 \\ -8 \\ 3/4 \\ -4 \end{bmatrix}$$

$$\nabla^2 f(X) = \begin{bmatrix} 6x_1 & 2x_2 & 6 & 0 & 3 \\ 2x_2 & 2x_1 + 2x_5 + 2 & 0 & 2 & 2x_2 \\ 6 & 0 & 11 & 0 & 6 \\ 0 & 2 & 0 & 6x_4 & 1 \\ 3 & 2x_2 & 6 & 1 & 5 \end{bmatrix} \quad ; \quad \nabla^2 f(X_0) = \begin{bmatrix} 3 & 0 & 6 & 0 & 3 \\ 0 & 3 & 0 & 2 & 0 \\ 6 & 0 & 11 & 0 & 6 \\ 0 & 2 & 0 & 3 & 1 \\ 3 & 0 & 6 & 1 & 5 \end{bmatrix}$$

$$\nabla^2 f(X_0) = TDT' = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & \mathbf{0} & \\ 2 & 0 & 1 & & \\ 0 & 2/3 & 0 & 1 & \\ 1 & 0 & 0 & 3/5 & 1 \end{bmatrix} \begin{bmatrix} 3 & & & & \\ & 3 & & \mathbf{0} & \\ & & -1 & & \\ & \mathbf{0} & & 5/3 & \\ & & & & 7/5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ & 1 & 0 & 2/3 & 0 \\ & & 1 & 0 & 0 \\ & \mathbf{0} & & 1 & 3/5 \\ & & & & 1 \end{bmatrix}$$

CORRECCIÓ UNIFORME δ DE LUENBERGER A LA MATRIU DIAGONAL D PER TAL QUE EL MÍNIM VALOR DIAGONAL RESULTANT SIGUI $\epsilon = 0.1 > 0$.

La correció necessària és $\delta = 1.1$:

$$\tilde{D} = D + \delta I = \begin{bmatrix} 3 & & & & \\ & 3 & & \mathbf{0} & \\ & & -1 & & \\ & \mathbf{0} & & 5/3 & \\ & & & & 7/5 \end{bmatrix} + 1.1 \begin{bmatrix} 1 & & & & \\ & 1 & & \mathbf{0} & \\ & & 1 & & \\ & \mathbf{0} & & 1 & \\ & & & & 1 \end{bmatrix} = \begin{bmatrix} 4.1 & & & & \\ & 4.1 & & \mathbf{0} & \\ & & 0.1 & & \\ & \mathbf{0} & & 2.7\hat{6} & \\ & & & & 2.5 \end{bmatrix}$$

RESOLUCIÓ DE $T\tilde{D}T'P = -\nabla f(X_0)$.

$$\begin{bmatrix} 1 & & & & \\ 0 & 1 & & \mathbf{0} & \\ 2 & 0 & 1 & & \\ 0 & 2/3 & 0 & 1 & \\ 1 & 0 & 0 & 3/5 & 1 \end{bmatrix} \begin{bmatrix} 4.1 & & & & \\ & 4.1 & & \mathbf{0} & \\ & & 0.1 & & \\ & \mathbf{0} & & 2.7\hat{6} & \\ & & & & 2.5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ & 1 & 0 & 2/3 & 0 \\ & & 1 & 0 & 0 \\ & \mathbf{0} & & 1 & 3/5 \\ & & & & 1 \end{bmatrix} P = \begin{bmatrix} 21/4 \\ -1 \\ 8 \\ -3/4 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & & & & \\ 0 & 1 & & \mathbf{0} & \\ 2 & 0 & 1 & & \\ 0 & 2/3 & 0 & 1 & \\ 1 & 0 & 0 & 3/5 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 21/4 \\ -1 \\ 8 \\ -3/4 \\ 4 \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 21/4 \\ -1 \\ -5/2 \\ -1/12 \\ -6/5 \end{bmatrix}$$

$$\begin{bmatrix} 4.1 & 0 & 8.2 & 0 & 4.1 \\ & 4.1 & 0 & 2.7\hat{3} & 0 \\ & & 0.1 & 0 & 0 \\ & \mathbf{0} & & 2.7\hat{6} & 1.66 \\ & & & & 2.5 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} 21/4 \\ -1 \\ -5/2 \\ -1/12 \\ -6/5 \end{bmatrix} \Rightarrow \boxed{\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} 10.11951 \\ -0.31384 \\ -5.8 \\ 0.11337 \\ -0.48 \end{bmatrix}}$$

COMPROVACIÓ DE QUE P ÉS DE DESCENS: $\nabla f(X_0)P < 0$.

$$[-21/4 \quad 1 \quad -8 \quad 3/4 \quad -4] \begin{bmatrix} 10.11951 \\ -0.31384 \\ -5.8 \\ 0.11337 \\ -0.48 \end{bmatrix} = -5.01375 < 0$$

EXPLORACIÓ LINEAL PER AJUST QUADRÀTIC EN LA DIRECCIÓ P OBTINGUDA.

$$q(X_0 + \lambda P) = a\lambda^2 + b\lambda + c \quad , \quad q'(X_0 + \lambda P) = 2a\lambda + b \quad , \quad q''(X_0 + \lambda P) = 2a$$

$$\underline{\lambda = 0:} \quad \begin{cases} q(X_0) = f(X_0) = \boxed{c = -8.25} \\ q'(X_0) = \nabla f(X_0)P = \boxed{b = -5.01375} \end{cases}$$

$$\underline{\lambda = 1:} \quad \begin{cases} X_0 + P = \begin{bmatrix} 1/2 \\ 0 \\ -1 \\ 1/2 \\ 0 \end{bmatrix} + \begin{bmatrix} 10.11951 \\ -0.31384 \\ -5.8 \\ 0.11337 \\ -0.48 \end{bmatrix} = \begin{bmatrix} 10.61951 \\ -0.31384 \\ -6.8 \\ 0.61337 \\ -0.48 \end{bmatrix} \Rightarrow f(X_0 + P) = 1024.2122 \\ q(X_0 + P) = a + b + c = f(X_0 + P) = 1024.2122 \Rightarrow \boxed{a = 1037.4759} \end{cases}$$

$$\min q(X_0 + \lambda P) \Rightarrow \boxed{\lambda_q = -\frac{b}{2a} = -\frac{-5.01375}{2 \times 1037.4759} = 0.00241632} \quad (\ddot{q}(X_0 + \lambda_q P) = 2074.95 > 0 : \text{mínim})$$

EXPLORACIÓ LINEAL PER AJUST QUADRÀTIC EN LA DIRECCIÓ $\hat{P} = [2 \quad 0 \quad -1 \quad 0 \quad 0]'$.

$$q(X_0+\lambda\hat{P}) = \hat{a}\lambda^2 + \hat{b}\lambda + \hat{c} \quad , \quad q'(X_0+\lambda\hat{P}) = 2\hat{a}\lambda + \hat{b} \quad , \quad q''(X_0+\lambda\hat{P}) = 2\hat{a}$$

$$\nabla f(X_0)\hat{P} = [-21/4 \quad 1 \quad -8 \quad 3/4 \quad -4]\begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} = -2.5$$

$$\underline{\lambda = 0}: \quad \begin{cases} q(X_0) = f(X_0) = \boxed{\hat{c} = -8.25} \\ q'(X_0) = \nabla f(X_0)\hat{P} = \boxed{\hat{b} = -2.5} \end{cases}$$

$$\underline{\lambda = 1}: \quad \begin{cases} X_0 + \hat{P} = \begin{bmatrix} 1/2 \\ 0 \\ -1 \\ 1/2 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5/2 \\ 0 \\ -2 \\ 1/2 \\ 0 \end{bmatrix} \Rightarrow f(X_0+\hat{P}) = 7.75 \\ q(X_0+\hat{P}) = \hat{a} + \hat{b} + \hat{c} = f(X_0+\hat{P}) = 7.75 \Rightarrow \boxed{\hat{a} = 7.75} \end{cases}$$

$$\min q(X_0+\lambda\hat{P}) \Rightarrow \boxed{\hat{\lambda}_q = -\frac{\hat{b}}{2\hat{a}} = -\frac{-2.5}{2\times 7.5} = \frac{1}{6}} \quad (\ddot{q}(X_0+\hat{\lambda}_q\hat{P}) = 15 > 0 : \text{mínim})$$

Exercici 2.17. : pluen2/mètode de Newton, variant de Luenberger

Efectueu una passa del mètode de Newton en la variant de Luenberger aplicat sobre la funció $f(x) = x_1^2 + x_2^2 + x_1x_2x_3^2$ a partir del punt $X_0 = [0 \quad 1/2 \quad 1]'$

Exercici 2.18. : pluen3/mètode de Newton, variant de Luenberger

Considereu la funció $f(x) = -1/3(x_1-3)^3 + (x_2-4)^2 + (x_3-2)^2 + x_1x_2x_3$ i el punt $x_0 = (2,-3,0)'$.

2.18.1.– Indiqueu, raonadament, si seria correcte aplicar una passa del mètode de Newton sense modificar.

2.18.2.– Calculeu la direcció de descens de $f(x)$ a partir de x_0 pel mètode de Newton en la variant de Luenberger amb $\epsilon = 0.5$.

2.18.3.– Si es calcula la direcció de descens pel mètode de Newton sense modificar s'obté la direcció $p = (1.3, 6.4, 0.5)'$, que és de descens. Comenteu aquest fet i la seva relació amb l'aplicació d'un mètode de Newton modificat.

Exercici 2.19. : pluen4/mètode de Newton, variant de Luenberger

Considereu el problema d'optimització no lineal sense constriccions:

$$\textbf{(P)} \quad \min_{x\in\Re^3} f(x) = e^{2x_1-x_2} + e^{x_2-x_1} - x^3$$

Es vol resoldre aquest problema a partir del punt inicial $x^0 = [0 \quad 0 \quad 1]'$.

2.19.1.– Indiqueu, raonadament, si seria correcte aplicar una passa del mètode de Newton sense modificar a partir de x^0.

2.19.2.– Calculeu la direcció de descens de $f(x)$ a partir de x^0 pel mètode de Newton modificat en la variant de Luenberger amb $\epsilon = 0.5$.

2.19.3.– Considereu que heu aplicat el mètode de de Newton en la variant de Luenberger al problema **(P)** i n'heu obtingut com a resultat el punt x^*. Indiqueu raonadament si x^* pot ser una solució (mínim local) del problema **(P)**.

Exercici 2.20. : pgimul/mètode de Newton, variant de Gill-Murray

La funció $f(X) = x_1^3 - 3x_1x_2^2 + 2x_1^2x_4 + x_2x_3x_4 + 6x_2$ té gradient zero i hessiana indefinida en el punt $X_0 = [1 \quad 1 \quad -2 \quad 0]'$.

2.20.1.– Trobeu una direcció de descens D a partir de X_0 sabent que la triangularització de Gill-Murray de l'hessiana $\nabla^2 f(X_0)$ és:

$$\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & 0 & 1 & \\ 2/3 & 1/6 & 6 & 1 \end{bmatrix} \begin{bmatrix} 6 & & & \\ & 12 & & \\ & & 1/6 & \\ & & & 9 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 2/3 \\ & 1 & 0 & 1/6 \\ & & 1 & 6 \\ & & & 1 \end{bmatrix}$$

2.20.2.– Per al problema anterior, efectueu exploració lineal per ajust cúbic del tipus $y = a\lambda^3 + b\lambda^2 + c\lambda + d$ emprant la informació dels valors de la funció objectiu i el seu gradient als punts X_0 i $X_0 + D$. Indiqueu si observeu alguna circumstància atípica en l'ajust cúbic en la direcció D.

2.20.3.– Efectueu la mateixa exploració que l'apartat anterior prenent la direcció $P = [7/3 \quad 1 \quad 0 \quad 0]'$.

SOLUCIÓ

OBTENCIÓ DIRECCIÓ DE DESCENS AMB DIRECCIONS DE CORBATURA NEGATIVA

Sent el gradient zero i l'hessiana indefinida, cal utilitzar com a direcció de descens una direcció de corbatura negativa D_{CN}. Trobarem D_{CN} utilitzant el procediment que es deriva de la triangularització de Gill-Murray de l'hessiana: $\tilde{S}\tilde{D}\tilde{S}' D_{CN} = E_m$, sent E la matriu diagonal d'elements e_{jj} (correccions a les diagonals). L'obtenció de D_{CN} és per resolució del sistema $\tilde{S}' D_{CN} = E_m$, sent E_m la columna m de la matriu unitat i m l'índex de la més gran correcció e_{mm} a E.

Obtenció de E per diferència entre $\tilde{S}\tilde{D}\tilde{S}'$ i $\nabla^2 f(X_0)$ i determinació de e_{mm}.

$$\tilde{S}\tilde{D}\tilde{S}' = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & 0 & 1 & \\ 2/3 & 1/6 & 6 & 1 \end{bmatrix} \begin{bmatrix} 6 & & & \\ & 12 & & \\ & & 1/6 & \\ & & & 9 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 2/3 \\ & 1 & 0 & 1/6 \\ & & 1 & 6 \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 6 & -6 & 0 & 4 \\ -6 & 18 & 0 & -2 \\ 0 & 0 & 1/6 & 1 \\ 4 & -2 & 1 & 18 \end{bmatrix}$$

$$\nabla f(X)' = \begin{bmatrix} 3x_1^2 - 3x_2^2 + 4x_1x_4 \\ -6x_1x_2 + x_3x_4 + 6 \\ x_2x_4 \\ 2x_1^2 + x_2x_3 \end{bmatrix} \quad ; \quad \nabla^2 f(X) = \begin{bmatrix} 6x_1 + 4x_4 & -6x_2 & 0 & 4x_1 \\ -6x_2 & -6x_1 & x_4 & x_3 \\ 0 & x_4 & 0 & x_2 \\ 4x_1 & x_3 & x_2 & 0 \end{bmatrix}$$

$$\nabla f(X_0)' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad ; \quad \nabla^2 f(X_0) = \begin{bmatrix} 6 & -6 & 0 & 4 \\ -6 & -6 & 0 & -2 \\ 0 & 0 & 0 & 1 \\ 4 & -2 & 1 & 0 \end{bmatrix}$$

$$E = \tilde{S}\tilde{D}\tilde{S}' - \nabla^2 f(X_0) = \begin{bmatrix} 0 & & & \\ & 24 & & \\ & & 1/6 & \\ & & & 18 \end{bmatrix}$$

d'on $m = 2$, ja que $e_{mm} = e_{22} = 24$, i el sistema a resoldre és:

$$\tilde{S}' D_{CN} = \begin{bmatrix} 1 & -1 & 0 & 2/3 \\ & 1 & 0 & 1/6 \\ & & 1 & 6 \\ & & & 1 \end{bmatrix} \begin{bmatrix} d_{CN_1} \\ d_{CN_2} \\ d_{CN_3} \\ d_{CN_4} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow D_{CN} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

comprovació:

$$D_{CN}' \nabla^2 f(X_0) D_{CN} = [1 \quad 1 \quad 0 \quad 0] \begin{bmatrix} 6 & -6 & 0 & 4 \\ -6 & -6 & 0 & -2 \\ 0 & 0 & 0 & 1 \\ 4 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = -12 < 0$$

<u>Nota:</u> també $-D_{CN}$ és una direcció de corbatura negativa a partir de X_0 on $\nabla f(X_0)' = \mathbf{0}$.

EXPLORACIÓ LINEAL PER AJUST CÚBIC AMB $u(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d$.

$$u(\lambda) = a\lambda^3 + b\lambda^2 + c\lambda + d \quad , \quad u'(\lambda) = 3a\lambda^2 + 2b\lambda + c \quad , \quad u''(\lambda) = 6a\lambda + 2b$$

a) En la direcció D_{CN}:

$$\underline{\lambda = 0}: \begin{cases} u(0) = f(X_0) = \boxed{d = 4} \\ u'(0) = \nabla f(X_0) D_{CN} = \boxed{c = 0} \end{cases}$$

$$\underline{\lambda = 1}: \begin{cases} X_0 + D_{CN} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -2 \\ 0 \end{bmatrix} \Rightarrow f(X_0 + D_{CN}) = -4 \\ \nabla f(X_0 + D_{CN}) = [0 \quad -18 \quad 0 \quad 4] \quad ; \quad \nabla f(X_0 + D_{CN}) D_{CN} = -18 \\ \left.\begin{array}{l} u(1) = a + b + c + d = f(X_0 + D_{CN}) = -4 \\ u'(1) = 3a + 2b + c = \nabla f(X_0 + D_{CN}) D_{CN} = -18 \end{array}\right\} \left.\begin{array}{l} a + b = -8 \\ 3a + 2b = -18 \end{array}\right\} \Rightarrow \boxed{\begin{array}{l} a = -2 \\ b = -6 \end{array}} \end{cases}$$

Tenim doncs $u(\lambda) = -2\lambda^3 - 6\lambda^2 + 4$. Igualant a zero $u'(\lambda_u) = 3a\lambda_u^2 + 2b\lambda_u + c = 0$ arribem a

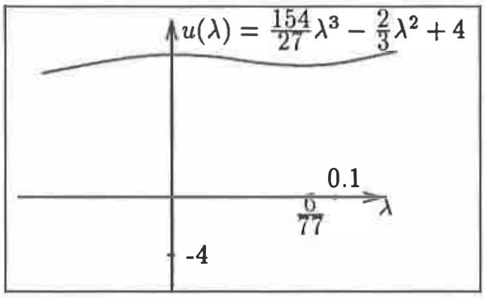


Figura 3

Exercici 2.21. : pden2/Mètode de Newton, variant de Dennis-Schnabel

La funció:

$$\begin{aligned} f(x) = & x_1^3 + x_2^2 + (11/2)x_3^2 + x_4^3 + (5/2)x_5^2 + x_1x_2^2 + 6x_1x_3 + 3x_1x_5 + \\ & + 2x_2x_4 + x_2^2x_5 + 6x_3x_5 + x_4x_5 \end{aligned}$$

té l'hessià indefinit al punt $X_0 = [1/2, 0, -1, 1/2, 0]'$. Trobeu la direcció de minimització a partir de X_0 pel mètode de Dennis-Schnabel, sabent que la triangularització de Gill-Murray de l'hessià a X_0 afegeix a l'hessià la matriu:

$$E = \begin{pmatrix} 0.272727 & & & & \\ & 0 & & & \\ & & 0.022727 & & \\ & & & 0 & \\ & & & & 18.7 \end{pmatrix}$$

Exercici 2.22. : pden3/Mètode de Newton, variant de Dennis-Schnabel

Considereu al següent problema d'optimització sense constriccions:

$$\min f(x) = x_1^2 + 2x_2^2 + 3x_1x_2 + 3x_2x_3^2 + x_1x_2x_3$$

La triangularització de Gill-Murray de $\nabla^2 f(x)$ sobre el punt $x^0 = [-1, 1, 0]'$ és:

$$\begin{pmatrix} 1. & 0 & 0 \\ 1.5 & 1. & 0 \\ 0.5 & -2.4 & 1. \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1.0417 & 0 \\ 0 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} 1. & 1.5 & 0.5 \\ 0 & 1. & -2.4 \\ 0 & 0 & 1. \end{pmatrix}$$

2.22.1.– Calculeu la direcció de descens de f(x) a partir de x^0 segons el mètode de Dennis-Schnabel.

2.22.2.– Considereu ara una nova funció objectiu $\tilde{f}(x)$ obtinguda afegint a $f(x)$ els termes $-x_1 - x_2 + x_3$, és a dir:

$$\tilde{f}(x) = x_1^2 + 2x_2^2 + 3x_1x_2 + 3x_2x_3^2 + x_1x_2x_3 - x_1 - x_2 + x_3$$

Calculeu una direcció de descens de $\tilde{f}(x)$ a partir de x^0. Comproveu que la direcció obtinguda és de descens.

Exercici 2.23. : pden4/Mètode de Newton, variant de Dennis-Schnabel

Considereu al següent problema d'optimització sense constriccions:

$$\min_{x\in\Re^3} f(x) = x_1^2 + 3x_2^2 + x_1x_2/x_3$$

La correcció diagonal de la triangularització de Gill-Murray de la Hessiana de $f(x)$ sobre el punt $X_k = [1/4, 1/5, -1/2]'$ és:

$$E = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 3.86 \end{pmatrix}$$

Efectueu una iteració del mètode de Newton modificat en la variant de Dennis-Schnabel. Feu exploració lineal per ajust quadràtic entre $\lambda = 0$ i $\lambda = 1$. Preneu com a criteri d'acceptabilitat de la longitud de pas la primera condició d'Armijo-Goldstein amb $\alpha = 0.1$.

2.6 Mètodes quasi Newton

2.6.1 Procediment

2.6.1.1 Procediments quasi-Newton per a funcions quadràtiques

Sigui o bé la minimització de:

$$\underset{X}{\text{minimitzeu}} \quad \frac{1}{2}X'QX - B'X \tag{16}$$

amb $B, X\in\mathrm{I\!R}^n$, $Q\in\mathrm{I\!R}^{n\times n}$ simètrica i definida positiva.

O bé la minimització d'una funció qualsevol:

$$\underset{X}{\text{minimitzeu}} \quad f(X) \tag{17}$$

amb $X\in\mathrm{I\!R}^n$, $f(X) : \mathrm{I\!R}^n \longrightarrow \mathrm{I\!R}$, $f(X)\in\mathcal{C}^2$ (contínua i fins a segones derivades contínues).

Hi ha dos tipus de procediments quasi-Newton (també anomenats *de la secant*). Els que utilitzen una aproximació H_k de Q (o de l'hessiana $\nabla f(X_k)$ per a una funció qualsevol), i els que utilitzen una aproximació S_k de la inversa de Q (o de $\nabla f(X_k)^{-1}$). L'algorisme per als dos tipus de procediment només difereix en la determinació de la direcció D_k a cada passa: en un cas cal resoldre el sistema $H_kD_k{=}{-}G_k$ i en l'altre cal efectuar el producte $D_k{=}{-}S_kG_k$. G_k és el gradient (de la funció quadràtica, o de la funció qualsevol).

2.6.1.2 Mètode quasi-Newton (o de la secant) per a una funció quadràtica

L'algorisme de minimització d'una funció quadràtica (16) per un mètode quasi-Newton és el següent:

0) Prenem un punt inicial qualsevol X_0 i calculem el gradient $G_0{=}QX_0{-}B$. Prenem una aproximació a Q H_0 simètrica i definida positiva, p.e. $H_0{=}\mathbf{1}$ (o bé a la inversa de Q $S_0{=}S_0'$ i def+) .
1) Calculem $G_k'G_k$
   - ▷ si $G_k'G_k{>}\epsilon$ anem a 2)
   - ▷ si $G_k'G_k{\le}\epsilon$ $\boxed{\text{FI}}$: ja som a l'òptim.
2) Calculem D_k resolent $H_kD_k{=}{-}G_k$ (o multiplicant $-S_kG_k{=}D_k$) .

3) Calculem $P_k{=}\alpha_k D_k$, amb $\alpha_k{=}\frac{G_k' D_k}{D_k' Q D_k}$ i fem $X_{k+1}{=}X_k{+}P_k$.
4) Obtenim $G_{k+1}{=}QX_{k+1}{-}B$ i $Y_k{=}\;G_{k+1}{-}G_k$.
5) Actualitzem l'aproximació $H_{k+1}{=}h(H_k,P_k,Y_k)$ (o bé $S_{k+1}{=}s(S_k,P_k,Y_k)$).
6) Fem $k:=k{+}1$ i anem a 1).

2.6.1.3 Mètode quasi-Newton (o de la secant) per a una funció qualsevol

L'algorisme de minimització d'una funció qualsevol (17) per un mètode quasi-Newton és el següent:

0) Prenem un punt inicial qualsevol X_0 i calculem el gradient $G_0{=}QX_0{-}B$. Prenem una aproximació a Q H_0 simètrica i definida positiva, p.e. $H_0{=}\mathbf{1}$ (o bé a la inversa de Q $S_0{=}S_0'$ i def+) .
1) Calculem $G_k' G_k$
   ▷ si $G_k' G_k{>}\epsilon$ anem a 2)
   ▷ si $G_k' G_k{\leq}\epsilon$ $\boxed{\text{FI}}$: ja som a l'òptim.
2) Calculem D_k resolent $H_k D_k{=}{-}G_k$ (o multiplicant $-S_k G_k{=}D_k$) .
3) Calculem $P_k{=}\alpha_k D_k$, amb $\alpha_k{=}\frac{G_k' D_k}{D_k' Q D_k}$ i fem $X_{k+1}{=}X_k{+}P_k$.
4) Obtenim $G_{k+1}{=}QX_{k+1}{-}B$ i $Y_k{=}\;G_{k+1}{-}G_k$.
5) Actualitzem l'aproximació $H_{k+1}{=}h(H_k,P_k,Y_k)$ (o bé $S_{k+1}{=}s(S_k,P_k,Y_k)$).
6) Fem $k:=k{+}1$ i anem a 1).

2.6.1.4 Aproximacions a l'hessiana o a la inversa de l'hessiana

Les funcions d'actualització de les aproximacions a l'hessiana $H_{k+1}{=}h(H_k,P_k,Y_k)$ o a la inversa de l'hessiana $S_{k+1}{=}s(S_k,P_k,Y_k)$ són l'única diferència entre els diversos mètodes quasi Newton existents. Podriem esmentar per exemple:

$$H_{k+1} = H_k + \frac{Y_k Y_k'}{P_k' Y_k} - \frac{H_k P_k P_k' Y_k}{P_k' H_k P_k} \quad \text{fòrmula Broyden-Fletcher-Goldfarb-Shanno} \tag{18}$$

$$S_{k+1} = S_k + \frac{(P_k - S_k Y_k)(P_k - S_k Y_k)'}{Y_k'(P_k - S_k Y_k)} \quad \text{correcció de rang u} \tag{19}$$

$$S_{k+1} = S_k + \frac{P_k P_k'}{P_k' Y_k} - \frac{S_k Y_k Y_k' S_k}{Y_k' S_k Y_k} \quad \text{fòrmula Davidon-Fletcher-Powell} \tag{20}$$

2.6.1.5 Obtenció de la direcció D_k

Les passes de l'algorisme que en l'aplicació computacional requereixen més temps de CPU són l'actualització de S_{k+1} o H_{k+1} , i l'obtenció de la direcció D_k.

— Si utilitzem l'aproximació de la inversa S_k perquè cal fer el producte $S_k G_k$ el qual requereix $O(n^2)$ operacions aritmètiques (recordi's que S_k és densa) i també l'actualització $S_k{\rightarrow}S_{k+1}$, ja que cal actualitzar la part simètrica de n^2 termes.

— Si utilitzem l'aproximació de l'hessiana H_k perquè cal resoldre el sistema $H_k D_k{=}{-}G_k$ el qual requereix $O(n^2)$ si H_k està factoritzada (la factorització existeix perquè H_k és definida positiva). L'actualització de la factorització $H_k{=}L_k L_k'{\rightarrow}L_{k+1}L_{k+1}{=}H_{k+1}$ es pot fer en $O(n^2)$ operacions aritmètiques perquè es poden utilitzar els algorismes d'actualització de factoritzacions quan s'afegeixen o sostreuen matrius de rang u, ja que podem rescriure (18) com:

$$L_{k+1}L_{k+1}' = L_k L_k' + \frac{Y_k Y_k'}{P_k' Y_k} - \frac{H_k P_k P_k' Y_k}{P_k' H_k P_k}$$

sent L_k i L_{k+1} els factors no sobrediagonals abans i després d'actualitzar la factorització de l'aproximació de l'hessiana.

Cal tenir present que fins i tot quan Q o $\nabla f(X_k)$ siguin matrius esparses, S_k o H_k seran matrius (simètriques) totalment denses.

2.6.1.6 Bibliografia

Els mètodes quasi–Newton actualitzant la inversa de l'hessiana estan descrit a [LUEN84] pp. 260–271. Els que actualitzen l'hessiana es poden trobar a [DESCH83] pp. 194–207. Per a una descripci'o succinta de l'actualització de la factoritzacó quan s'afegeix una matriu de rang u, pot veure's [GILL81] pp. 41–43.

2.6.2 Exercicis

Exercici 2.24. : pdafp1/mètode Davidon-Fletcher-Powell

Considerem la minimització a $X\in\mathbb{R}^3$ de la funció:

$$\min_X \frac{1}{x_1} + x_1x_2^2 + 2x_1(x_3-1)^2 + \frac{x_1}{4}$$

pel mètode de Davidon-Fletcher-Powell que aproxima la inversa de l'hessià S_k segons l'actualització:

$$S_{k+1} = S_k + \frac{P_kP_k'}{P_k'Y_k} - \frac{S_kY_kY_k'S_k}{Y_k'S_kY_k}$$

sent $P_k{=}X_{k+1}{-}X_k$ i $Y_k{=}\nabla f(X_{k+1})'{-}\nabla f(X_k)'$.

$$\text{Sigui} \quad X_k = \begin{bmatrix} 0,9005 \\ -0,01431 \\ 1.0854 \end{bmatrix} \quad \text{un punt iterat, i} \quad S_k = \begin{bmatrix} 0,00831 & & (\text{sim.}) \\ -0,0132 & 1,01 & \\ 0,0116 & 0,0911 & 0,318 \end{bmatrix}$$

la matriu obtinguda que aproxima la inversa de l'hessià a X_k.

A partir de X_k l'algorisme faria una passa α_k=1,225 per arribar a X_{k+1}. Comproveu i indiqueu si fent una passa d'exploració $\tilde{\alpha}_k$=0,5 la nova aproximació S_{k+1} podria esdevenir indefinida.

SOLUCIÓ

GRADIENT A X_k I DIRECCIÓ D_k

$$\nabla f(X)' = \begin{bmatrix} -\frac{1}{x_1^2} + x_2^2 + 2(x_3-1)^2 + \frac{1}{4} \\ 2x_1x_2 \\ 4x_1(x_3-1) \end{bmatrix} \qquad \nabla f(X_k)' = \begin{bmatrix} -0,9684 \\ -0,02577 \\ 0,3076 \end{bmatrix}$$

$$D_k = -S_k\nabla f(X_k)' = \begin{bmatrix} 0,004139 \\ -0,014776 \\ -0,084239 \end{bmatrix}$$

DETERMINACIÓ DE $\widetilde{X}_{k+1}$, $\nabla f(\widetilde{X}_{k+1})$, $\widetilde{P}_k$ and $\widetilde{Y}_k$,

$$\widetilde{X}_{k+1} = X_k + \widetilde{\alpha}_k D_k = \begin{bmatrix} 0,90256 \\ -0,02169 \\ 1,04328 \end{bmatrix} \quad \nabla f(\widetilde{X}_{k+1})' = \begin{bmatrix} -0,97333 \\ -0,03917 \\ 0.15625 \end{bmatrix} \quad \widetilde{P}_k = \begin{bmatrix} 0,00207 \\ -0,00739 \\ -0,04212 \end{bmatrix}$$

$$\widetilde{Y}_k = \begin{bmatrix} -0,00493 \\ -0,01340 \\ -0,15136 \end{bmatrix}$$

DETERMINACIÓ DE L'ESCALAR $\widetilde{P}_k'\widetilde{Y}_k$ I EL SEU SIGNE

Podem comprovar que $\widetilde{P}_k'\widetilde{Y}_k$=0,00646>0 . Això és suficient per assegurar que S_{k+1} és definida positiva, ja que donat un vector qualsevol $V \in \mathbb{R}^n$

$$V'S_{k+1}V = V'S_kV + \frac{V'P_kP_k'V}{P_k'Y_k} - \frac{V'S_kY_kY_k'S_kV}{Y_k'S_kY_k}$$

Donat que S_k és definida positiva, podem fer els canvis $T{=}S_k^{\frac{1}{2}}V$ i $U{=}S_k^{\frac{1}{2}}Y_k$:

$$V'S_{k+1}V = T'T - \frac{(T'U)^2}{U'U} + \frac{(P_k'V)^2}{P_k'Y_k} = \frac{(T'T)(U'U) - (T'U)^2}{U'U} + \frac{(P_k'V)^2}{P_k'Y_k}$$

El numerador del primer trencat és sempre més gran que zero (perquè el producte dels quadrats de les normes de dos vectors sempre és superior al quadrat del seu producte escalar) i el denominador és la norma d'un vector al quadrat. El numerador del segon trencat és l'escalar $P_k'V$ al quadrat, d'on basta que $P_k'Y_k$ sigui positiu per assegurar que $V'S_{k+1}V$ sigui positiu. (Cal notar que, quan $P_k'Y_k{>}0$, $(T'T)(U'U) - (T'U)^2$ i $P_k'V$ no es poden anul·lar simultàniament, ja que només si T i U fossin colineals, és a dir: $V{=}\beta Y_k$ —amb $\beta{>}0$ —, s'anul·laria $(T'T)(U'U) - (T'U)^2$, però aleshores $P_k'V{=}\beta P_k'Y_k{>}0$ tal com hem suposat).

Si $P_k'Y_k$ hagués estat negatiu hauria sigut necessari construir S_{k+1} i comprovar si era definida positiva.

Exercici 2.25. : pdafp2/mètode Davidon-Fletcher-Powell

Considerem la minimització d'una funció quadràtica:

$$\min_X \frac{1}{2}X'QX - B'X \quad \text{amb} \quad Q = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \quad \text{i} \quad B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

a partir del punt $X_0{=}\underline{0}$ pel mètode de Davidon-Fletcher-Powell que aproxima la inversa de l'hessiana S_k segons l'actualització:

$$S_{k+1} = S_k + \frac{P_kP_k'}{P_k'Y_k} - \frac{S_kY_kY_k'S_k}{Y_k'S_kY_k}$$

sent $P_k{=}X_{k+1}{-}X_k$ i $Y_k{=}G_{k+1}{-}G_k$ (G_k: gradient al punt X_k), .

Començant amb $S_0{=}\mathbf{1}$, comproveu que:

2.25.1.– P_0 és un vector propi amb valor propi unitat de $S_1 Q$.

2.25.2.– P_0 i P_1 són Q-conjugades.

SOLUCIÓ

PASSA A PARTIR DE X_0

$$G_0 = QX_0 - B = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

$$D_0 = -S_0 G_0 = -\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \alpha_0 = \frac{-G_0' D_0}{D_0' Q D_0} = \frac{1}{2}$$

$$X_1 = X_0 + \alpha_0 D_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} \qquad P_0 = X_1 - X_0 = \alpha_0 D_0 = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix}$$

ACTUALITZACIO DE S_k

$$G_1 = QX_1 - B = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$Y_0 = G_1 - G_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad P_0' Y_0 = 1$$

$$S_0 Y_0 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad Y_0' S_0 Y_0 = 3 \quad \text{i sent} \quad S_1 = S_0 + \frac{P_0 P_0'}{P_0' Y_0} - \frac{S_0 Y_0 Y_0' S_0}{Y_0' S_0 Y_0}$$

$$S_1 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} + \begin{bmatrix} 1/4 & 0 & 1/4 \\ 0 & 0 & 0 \\ 1/4 & 0 & 1/4 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 11/12 & -1/3 & -1/12 \\ -1/3 & 2/3 & -1/3 \\ -1/12 & -1/3 & 11/12 \end{bmatrix}$$

PASSA A PARTIR DE X_1

$$D_1 = -S_1 G_1 = -\begin{bmatrix} 11/12 & -1/3 & -1/12 \\ -1/3 & 2/3 & -1/3 \\ -1/12 & -1/3 & 11/12 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix} \qquad \alpha_1 = \frac{-G_1' D_1}{D_1' Q D_1} = \frac{3}{2}$$

$$X_2 = X_1 + \alpha_1 D_1 = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad P_1 = X_2 - X_1 = \alpha_1 D_1 = \begin{bmatrix} 1/2 \\ -1 \\ 1/2 \end{bmatrix}$$

COMPROVACIÓ DE QUE $S_1QP_0 = P_0$

(P_0 és un vector propi amb valor propi unitat de S_1Q)

$$S_1QP_0 = -\begin{bmatrix} 11/12 & -1/3 & -1/12 \\ -1/3 & 2/3 & -1/3 \\ -1/12 & -1/3 & 11/12 \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} = P_0$$

COMPROVACIÓ DE QUE $P_1'QP_0 = 0$

(P_0 i P_1 són Q-conjugades)

$$P_1'QP_0 = [1/2 \quad -1 \quad 1/2] \begin{bmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix} = 0$$

Exercici 2.26. : pdafp3/mètode Davidon-Fletcher-Powell

Considereu el problema $\min_{x \in \Re^2} f(x) = (x_1 - 2)^4 + (x_1 - 2)^2 x_2^2 - (x_2 + 1)^2$ i el punt inicial $x^0 = [1 \quad 1]'$:

2.26.1.– Apliqueu una passa completa del métode DFP que actualitza S^k a partir de la expressió:

$$S^{k+1} = S^k + \frac{p^k p^{k'}}{p^{k'} y^k} - \frac{S^k y^k y^{k'} S^k}{y^{k'} S^k y^k}$$

sent $p^k = x^{k+1} - x^k$ i $y^k = \nabla f(x^{k+1})' - \nabla f(x^k)'$ .Preneu $S^0 = \mathbf{I}$ i $\alpha^0 = 1$

2.26.2.– Calculeu la direcció de cerca d^1 obtinguda pel mètode DFP sobre el punt x^1 trobat a l'apartat anterior. Sense fer el producte escalar $\nabla f(x^1) d^1$ indiqueu si la direcció d^1 és de descens.

Exercici 2.27. : pbfgs1/mètodes quasi-Newton, actualització BFGS

Es vol resoldre el problema:

$$\min_{x \in \mathbf{R}^2} f(X) = -(x_1 - 2)^4 - (x_1 - 2)^2 x_2^2 - (x_2 + 1)^2$$

aplicant l'actualització secant simètrica definida positiva (BFGS) donada per l'expressió:

$$H_{k+1} = H_k + \frac{Y_k Y_k'}{Y_k' P_k} - \frac{H_k P_k P_k' H_k}{P_k' H_k P_k}$$

amb $Y_k = \nabla f(X_{k+1}) - \nabla f(X_k)$ i $P_k = X_{k+1} - X_k$. El punt inicial és $X_0 = [1 \quad 1]'$ i s'inicialitza l'aproximació quasi-Newton de l'hessiana fent $H_0 = I$.

2.27.1.– Apliqueu una passa completa del mètode indicat, incloent el càlcul de l'actualització BFGS sobre X_1, prenent com a longitud de pas $\alpha_0 = 1$.

2.27.2.– Calculeu la direcció de moviment D_1 a partir de X_1 i comproveu que no és de descens. Per quína raó D_1 no és de descens?

SOLUCIÓ

2.27.1.- PASSA DEL MÈTODE.

Càlculs preliminars :

$$X_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad , \quad \nabla f(X)' = \begin{bmatrix} -4(x_1-2)^3 - 2(x_1-2)x_2^2 \\ -2(x_1-2)^2 x_2 - 2(x_2+1) \end{bmatrix} \quad , \quad \nabla f(X_0)' = \begin{bmatrix} 6 \\ -6 \end{bmatrix} \quad , \quad H_0 = \mathbf{I}$$

Passa: $X_1 = X_0 - \alpha_0 H_0^{-1} \nabla f(X_0)' = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 \times \mathbf{I} \times \begin{bmatrix} 6 \\ -6 \end{bmatrix} = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$

Acualització BFGS:

$$\nabla f(X_1)' = \begin{bmatrix} 2058 \\ -702 \end{bmatrix} , \; Y_0 = \nabla f(X_1)' - \nabla f(X_0)' = \begin{bmatrix} 2058 \\ -702 \end{bmatrix} - \begin{bmatrix} 6 \\ -6 \end{bmatrix} = \begin{bmatrix} 2052 \\ -696 \end{bmatrix} , \; P_0 = \begin{bmatrix} -6 \\ 6 \end{bmatrix}$$

$$\boxed{H_1 = H_0 + \frac{Y_0 Y_0'}{Y_0' P_0} - \frac{H_0 P_0 P_0' H_0}{P_0' H_0 P_0} = \begin{bmatrix} -254.880 & 87.120 \\ 87.120 & -28.88 \end{bmatrix}}$$

2.27.2.- DIRECCIÓ DE MOVIMENT D_1

$$H_1 D_1 = -\nabla f(X_1)'$$

$$\begin{bmatrix} -254.880 & 87.120 \\ 87.120 & -28.88 \end{bmatrix} \begin{bmatrix} d_{1_1} \\ d_{1_2} \end{bmatrix} = \begin{bmatrix} -2058 \\ 702 \end{bmatrix} \quad ; \quad \begin{bmatrix} -254.879 & 87.120 \\ 0. & 0.898 \end{bmatrix} \begin{bmatrix} d_{1_1} \\ d_{1_2} \end{bmatrix} = \begin{bmatrix} -2058 \\ -1.440 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{cases} d_{1_1} = \frac{1}{-254.879}(-2058 - (87.120 \times (-1.603))) = 7.526 \\ d_{1_2} = -1.440/0.898 = -1.603 \end{cases} \quad ; \quad \boxed{D_1 = \begin{bmatrix} 7.526 \\ -1.603 \end{bmatrix}}$$

Comprovació direcció de descens :

$$\nabla f(X_1) D_1 = [2058 \quad -702] \begin{bmatrix} 7.526 \\ -1.603 \end{bmatrix} = 16613.814 > 0 \Rightarrow D_1 \text{ és d'ascens}$$

D_1 és d'ascens perquè la matriu H_1 no és definida positiva. L'actualització BFGS només serà definida positiva si el producte $Y_k' P_k$ és > 0. En el nostre cas es té:

$$Y_0' P_0 = [2052 \quad -696] \begin{bmatrix} -6 \\ 6 \end{bmatrix} = -16488 < 0$$

la qual cosa implica que H_1 no és definida positiva. Tanmateix cal tenir present que, donat que H_1 és indefinida ($\lambda_1 = 0.804$, $\lambda_2 = -284.564$), la direcció P_1 tant hauria pogut ser d'ascens com de descens.

3 Condicions de Kuhn-Tucker.

3.1 Procediment

3.1.1 Condicions d'optimalitat amb constriccions d'igualtat.

En aquest apartat s'exposaran les condicions d'optimalitat per a problemes del tipus

$$(\mathbf{PN}) \begin{cases} \min\limits_{x\in\Re^n} f(x) \\ \qquad h(x) = 0 \end{cases}$$

amb $f : \Re^n \longrightarrow \Re$, $h : \Re^n \longrightarrow \Re^m$. Prèviament a la presentació de les condicions d'optimalitat s'introdueix el concepte de *regularitat d'un punt x respecte de les constriccions $h(x) = 0$* i de *pla tangent de les constriccions $h(x) = 0$ sobre x^**:

Definició 3.1 Punt regular amb constricions d'igualtat : $x^* \in \Re^n$ és un punt regular de $h(x) = 0$ si es satisfà:

i) $h(x^*) = 0$

ii) El conjunt de vectors $\nabla h_i(x^*)$, $i = 1, \ldots, m$ són linealment independents

Definició 3.2 Pla tangent : sigui $x^* \in \Re^n$ punt regular de les constriccions $h(x) = 0$. El pla tangent de les constriccions $h(x) = 0$ sobre x^* es defineix com :

$$\mathrm{M} = \{y \in \Re^n \mid \nabla h(x^*)y = 0\}$$

S'enuncien ara les condicions necessàries i suficients de minim local del problema **(PN)**.

Teorema 3.2 : Condicions necessàries de primer ordre de mínim local. Sigui x^* punt extrem local de $f(x)$ subjecte a $h(x) = 0$. Suposem que x^* és punt regular de $h(x) = 0$. Aleshores, existeix el vector $\lambda \in \Re^m$, dit de *multiplicadirs de Lagrange* tal que:

$$\nabla f(x^*) + \lambda' \nabla h(x^*) = 0$$

Teorema 3.3 : Condicions necessàries de segon ordre de mínim local. Sigui x^* mínim local de f subjecte a $h(x) = 0$ i punt regular de $h(x) = 0$. Llavors es satisfà:

i) $\exists\lambda \in \Re^m$ tal que $\nabla f(x^*) + \lambda'\nabla h(x^*) = 0$

ii) Sigui M el pla tangent de les constriccions $h(x) = 0$ sobre x^*. La matriu

$$\nabla_x^2 \mathcal{L}(x^*) = \nabla^2 f(x^*) + \sum_{j=1}^{m} \lambda_j \nabla^2 h_j(x^*)$$

és semidefinida positiva sobre M

Teorema 3.4 : Condicions suficients de segon ordre de mínim local. Sigui $x^* \in \Re^n$ tal que $h(x^*) = 0$, i sigui $\lambda \in \Re^m$ el vector de multiplicadors de Lagrange que satisfan:

i) $\nabla f(x^*)\lambda'\lambda h(x^*) = 0$

ii) La matriu $\nabla_x^2 \mathcal{L}(x^*)$ és definida positiva sobre el subespai M, pla tangent de les constriccions $h(x) = 0$ sobre x^*.

Llavors x^* és un mínim local estricte de f subjecte a $h(x) = 0$.

3.1.2 Condicions d'optimalitat amb constriccions de desigualtat.

S'estudiaran ara les condicions d'optimalitat del següent problema d'optimització no lineal amb constriccions d'igualtat i desigualtat:

$$(\mathbf{PN}) \begin{cases} \min_{x \in \Re^n} f(x) \\ \quad h(x) = 0 \\ \quad g(x) \leq 0 \end{cases}$$

amb $f : \Re^n \to \Re$, $h : \Re^n \to \Re^m$, $g : \Re^n \to \Re^p$. Es dona a continuació la definició de punt regular en presència de constriccions de desigualtat.

Definició 3.3 Punt regular amb constricions de desigualtat : $x^* \in \Re^n$ és un punt regular de $h(x) = 0$ i $g(x) \leq 0$ si es satisfà:

i) $h(x^*) = 0$ i $g(x^*) \leq 0$.

ii) El conjunt de vectors $\nabla h_i(x^*)$, $i = 1, \ldots, m$ i $\nabla g_j(x^*)$, $j \in \mathrm{J} = \{j \mid g_j(x^*) = 0\}$ són linealment independents

Teorema 3.5 : Condicions necessàries de primer ordre de mínim local. Sigui x^* mínim local de **(PN)**, punt regular. Llavors existeixen els vectors $\lambda \in \Re^m$ i $\mu \in \Re^p$, dits *multiplicadors de Lagrange* que verifiquen les següents condicions (*Condicions de Karush-Kuhn-Tucker*):

i) $\nabla f(x^*) + \lambda'\nabla h(x^*) + \mu'\nabla g(x^*) = 0$

ii) $\mu' g(x^*) = 0$

iii) $\mu \geq 0$

Teorema 3.6 : Condicions necessàries de segon ordre de mínim local. Sigui $f, h, g \in \mathcal{C}^2$ i $x^* \in \Re^n$ punt regular de $h(x) = 0$ i $g(x) \leq 0$. Si x^* és mínim local de **(PN)** llavors es satisfan les següents condicions:

i) Existeixen els vectors $\lambda \in \Re^m$ i $\mu \in \Re^p$ que verifiquen les condicions de Karush-Khun-Tucker:

a) $\nabla f(x^*) + \lambda' \nabla h(x^*) + \mu' \nabla g(x^*) = 0$

b) $\mu' g(x^*) = 0$

c) $\mu \geq 0$

ii) La matriu $\nabla_x^2 \mathcal{L}(x^*) = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j=1}^m \mu_j \nabla^2 g_j(x^*)$ és semidefinida positiva sobre el subespai tangent a les constriccions actives $\mathrm{M} = \{y \in \Re^n \mid \nabla h(x^*) y = 0, \nabla g_j(x^*) y = 0 \, \forall j \in \mathrm{J}\}$

Teorema 3.7 : Condicions suficients de segon ordre de mínim local. Sigui $f, h, g \in \mathcal{C}^2$ i $x^* \in \Re^n$ punt factible del problema **(PN)**. Si el punt x^* satisfà les següents condicions:

i) Existeixen els vectors $\lambda \in \Re^m$ i $\mu \in \Re^p$ que verifiquen les condicions de Karush-Khun-Tucker:

a) $\nabla f(x^*) + \lambda' \nabla h(x^*) + \mu' \nabla g(x^*) = 0$

b) $\mu' g(x^*) = 0$

c) $\mu \geq 0$

ii) La matriu $\nabla_x^2 \mathcal{L}(x^*) = \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j=1}^m \mu_j \nabla^2 g_j(x^*)$ és definida positiva sobre el subespai M′ definit com $\mathrm{M}' = \{y \in \Re^n \mid \nabla h(x^*) y = 0, \nabla g_j(x^*) y = 0 \, \forall j \in \mathrm{J}'\}$ amb $\mathrm{J}' = \{j \mid g_j(x^*) = 0, \mu_j > 0\}$

Llavors x^* és mínim local estricte del problema **(PN)**.

3.2 Exercicis

Exercici 3.1. : pkutu1/condicions de Kuhn-Tucker

En la minimització:

$$\begin{aligned} \min \; & {x_1}^2 + {x_2}^2 + 2x_1x_3 \\ \text{subj.} \; & -{x_1}^2 + {x_2}^2 + 4x_3 = 0 \end{aligned}$$

3.1.1.– determineu si el punt $\widetilde{X} = [-\frac{4}{3} \quad 0 \quad \frac{4}{9}]'$, el qual satisfá les condicions de $1^{\underline{\text{er}}}$ ordre, és un mínim.

SOLUCIÓ

EXPRESSIÓ DE LES CONDICIONS DE $1^{\underline{\text{er}}}$ ORDRE (obtenció del multiplicador de Lagrange de la constricció)

Definim el Lagrangià: $\mathcal{L}(X, \lambda) = {x_1}^2 + {x_2}^2 + 2x_1x_3 + \lambda(-{x_1}^2 + {x_2}^2 + 4x_3)$

Derivant respecte a X i a λ,

$$\nabla_X \mathcal{L}(X,\lambda)' = \begin{bmatrix} 2x_1 + 2x_3 - 2\lambda x_1 \\ 2x_2 + 2\lambda x_2 \\ 2x_1 + 4\lambda \end{bmatrix} = \underline{0} \quad \nabla_X \mathcal{L}(\tilde{X},\lambda)' = \begin{bmatrix} -\frac{16}{9} + \frac{8}{3}\lambda \\ 0 \\ -\frac{8}{3} + 4\lambda \end{bmatrix} = \underline{0} \quad \Rightarrow \quad \tilde{\lambda} = \frac{2}{3}$$

$$\nabla_\lambda \mathcal{L}(X,\lambda)' = -{x_1}^2 + {x_2}^2 + 4x_3 = 0 \quad \text{es satisfà a } \tilde{X}.$$

EXPRESSIÓ DE L'HESSIÀ DEL LAGRANGIÀ

$$\nabla^2_X \mathcal{L}(X,\lambda) = \nabla^2_X f(X) + \sum_i \lambda_i \nabla^2_X h_i(X)$$

$$\nabla^2_X f(X) = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 0 \end{bmatrix} \quad \text{(constant)} \quad \nabla^2_X h_i(X) = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(constant)}$$

$$\nabla^2_X \mathcal{L}(\tilde{X},\tilde{\lambda}) = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} -2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & 0 & 2 \\ 0 & \frac{10}{3} & 0 \\ 2 & 0 & 0 \end{bmatrix} \quad \text{la qual és indefinida}$$

PLA TANGENT I BASE Z DEL PLA TANGENT A $\tilde{X}$

Els punts Y del pla tangent a $\tilde{X}$ han de satisfer: $\nabla h(\tilde{X})Y=0$, $\nabla h(X)=[-2x_1 \quad 2x_2 \quad 4]$ i $\nabla h(\tilde{X})$ és $[\frac{8}{3} \quad 0 \quad 4]$, així:

$$[\tfrac{8}{3} \quad 0 \quad 4] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0 \quad \Rightarrow \quad y_3 = -\frac{2}{3} y_1 \quad \text{condició de pla tangent a } \tilde{X}$$

$X \in \mathbb{R}^3$ i només hi ha una constricció: calen doncs dos vectors per definir una base al pla tangent. Prenen dos vectors ortonormals. Un pot ser $Z_1 = [0 \quad 1 \quad 0]'$ (el qual satisfà $y_3 = -\frac{2}{3}y_1$). L'altre, ortogonal a Z_1, podria ser: $[1 \quad 0 \quad -\frac{2}{3}]'$ normalitzat, és a dir: $Z_2 = [\frac{3}{\sqrt{13}} \quad 0 \quad -\frac{2}{\sqrt{13}}]'$. Una base al pla tangent a $\tilde{X}$ és doncs:

$$Z = \begin{bmatrix} 0 & \frac{3}{\sqrt{13}} \\ 1 & 0 \\ 0 & -\frac{2}{\sqrt{13}} \end{bmatrix}$$

PROJECCIÓ DE L'HESSIÀ DEL LAGRANGIÀ AL PLA TANGENT

Calculem $Z'\nabla^2_X \mathcal{L}(\tilde{X},\tilde{\lambda})Z$:

$$Z'\nabla^2_X \mathcal{L}(\tilde{X},\tilde{\lambda})Z = \begin{bmatrix} 0 & 1 & 0 \\ \frac{3}{\sqrt{13}} & 0 & -\frac{2}{\sqrt{13}} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & 0 & 2 \\ 0 & \frac{10}{3} & 0 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{3}{\sqrt{13}} \\ 1 & 0 \\ 0 & -\frac{2}{\sqrt{13}} \end{bmatrix} = \begin{bmatrix} -\frac{18}{13} & 0 \\ 0 & \frac{10}{3} \end{bmatrix} \text{(indefinida)}$$

Sent l'hessià del Lagrangià projectat al pla tangent indefinit, podem assegurar que $\tilde{X}$ **no és un mínim**.

Exercici 3.2. : pkutu2/condicions de Kuhn-Tucker

Considereu el problema d'optimització:

$$\min f(x) = x_1^2x_3 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 4x_1x_3 + 2x_2x_3 - 4x_1 - 7x_2 + 3x_3$$
$$\text{subj. a } h(x) = \frac{1}{2}x_1^2 + x_2 + x_3 - 4 \leq 0$$

amb un punt extrem $\tilde{X}$ tal que $\tilde{x}_1 = -2$. Calculeu la resta de components de $\tilde{X}$ i determineu la seva condició de màxim, mínim o punt de sella.

SOLUCIÓ

Si $\tilde{X}$ satisfà les condicions necessàries de primer ordre (i té $\tilde{x}_1 = -2$) hem de determinar el valor del multiplicador de Lagrange μ de la constricció de desigualtat per a saber si és activa o no.

CONDICIONS NECESSÀRIES DE PRIMER ORDRE :

El Lagrangià$\mathcal{L}(X, \mu)$ és :

$$\mathcal{L}(X,\mu) = x_1^2x_3+2x_2^2+3x_3^2+2x_1x_2+4x_1x_3+2x_2x_3-4x_1-7x_2+3x_3+\mu\left(\frac{1}{2}x_1^2+x_2+x_3-4\right)$$

$$\left.\begin{array}{l} \dfrac{\partial\mathcal{L}(X,\mu)}{\partial x_1} = 2x_1x_3 + 2x_2 + 4x_3 + \mu x_1 - 4 = 0 \\ \dfrac{\partial\mathcal{L}(X,\mu)}{\partial x_2} = 2x_1 + 4x_2 + 2x_3 + \mu - 7 = 0 \\ \dfrac{\partial\mathcal{L}(X,\mu)}{\partial x_3} = x_1^2 + 4x_1 + 2x_2 + 6x_3 + \mu + 3 = 0 \end{array}\right\} \rightarrow \tilde{x}_1 = -2 \rightarrow \left\{\begin{array}{r} 2\tilde{x}_2 - 2\tilde{\mu} = 4 \\ 4\tilde{x}_2 + 2\tilde{x}_3 + \tilde{\mu} = 11 \\ 2\tilde{x}_2 + 6\tilde{x}_3 + \tilde{\mu} = 1 \end{array}\right.$$

havent-se de complir que $\tilde{\mu}\left(\frac{1}{2}\tilde{x}_1^2 + \tilde{x}_2 + \tilde{x}_3 - 4\right) = 0$.

DETERMINACIÓ DE LA CONDICIÓ D'ACTIVA DE LA CONSTRICCIÓ DE DESIGUALTAT, I OBTENCIÓ DE $\tilde{X}$ I $\tilde{\mu}$:

a) Suposem que la constricció no sigui activa: llavors $\frac{1}{2}\tilde{x}_1^2 + \tilde{x}_2 + \tilde{x}_3 - 4 \neq 0$, el qual força a que $\tilde{\mu}$ sigui zero i el sistema que queda és :

$$\left.\begin{array}{r} 2\tilde{x}_2 = 4 \Rightarrow \tilde{x}_2 = 2 \\ 4\tilde{x}_2 + 2\tilde{x}_3 = 11 \Rightarrow \tilde{x}_3 = \dfrac{3}{2} \\ 2\tilde{x}_2 + 6\tilde{x}_3 \neq 1 \quad\quad\quad \end{array}\right\}$$ **NO ÉS COMPATIBLE.** així doncs el punt $\tilde{X}$ (amb $\tilde{x}_1 = -2$) no pot tenir la constricció no activa

b) Suposem que la constricció sigui activa: llavors $\frac{1}{2}\tilde{x}_1^2 + \tilde{x}_2 + \tilde{x}_3 - 4 = 0$, que, tenint en compte que $\tilde{x}_1 = -2$ queda $\tilde{x}_2 + \tilde{x}_3 = 2$, i pot ser $\tilde{\mu} \neq 0$. El sistema que queda té quatre equacions i només tres incògnites. Plantegem un subsistema i després comprovarem si l'equació resultant és compatible. Agafant, per exemple, la 3^a, la 4^a i la 1^a equació i ordenant les variables μ, x_3 i x_2 queda el sistema :

$$\begin{bmatrix} 1 & 6 & 2 \\ 0 & 1 & 1 \\ -2 & 0 & 2 \end{bmatrix}\begin{bmatrix} \tilde{\mu} \\ \tilde{x}_3 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 6 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -6 \end{bmatrix}\begin{bmatrix} \tilde{\mu} \\ \tilde{x}_3 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -18 \end{bmatrix} \Rightarrow \begin{array}{r} \tilde{x}_2 = 3 \\ \tilde{x}_3 = -1 \\ \tilde{\mu} = 1 \end{array}$$

L'equació 2^a $4 \times 3 + 2 \times (-1) + 1 = 11$ es satisfà, així doncs el punt $\tilde{X} = [-2 \quad 3 \quad -1]'$ amb $\tilde{\mu} = 1$ verifica les condicions necessàries de 1er ordre i la constricció és activa.

CONDICIÓ DE MÀXIM, MÍNIM O PUNT DE SELLA DE $\tilde{X}$:

Donat que $\tilde{\mu} = 1 > 0$, $\tilde{X}$ podria ser un mínim, perquè $\tilde{\mu} > 0$ és condició necessària de mínim, però no és condició suficient. Per a comprovar si $\tilde{X}$ és un mínim cal que es verifiqui a més la condició suficient de segon ordre (Hessià del Lagrangià definit positiu al pla tangent).

OBTENCIÓ DE L'HESSIÀ DEL LAGRANGIÀ :

$$\nabla f(X)' = \begin{bmatrix} 2x_1x_3 + 2x_2 + 4x_3 \\ 2x_1 + 4x_2 + 2x_3 \\ x_1^2 + 4x_1 + 2x_2 + 6x_3 \end{bmatrix} \; ; \; \nabla^2 f(X) = \begin{bmatrix} 2x_3 & 2 & 2x_1 + 4 \\ 2 & 4 & 2 \\ 2x_1 + 4 & 2 & 6 \end{bmatrix}$$

$$\nabla^2 f(\tilde{X}) = \begin{bmatrix} -2 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix}$$

$$\nabla h(X)' = \begin{bmatrix} x_1 \\ 1 \\ 1 \end{bmatrix} \quad ; \quad \nabla^2 h(X) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad ; \quad \nabla^2 h(\tilde{X}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\nabla^2 \mathcal{L}(\tilde{X}, \tilde{\mu}) = \nabla^2 f(\tilde{X}) + \tilde{\mu} \nabla^2 h(\tilde{X}) = \begin{bmatrix} -2 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} + 1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix}$$

OBTENCIÓ D'UNA BASE ORTONORMAL DEL PLA TANGENT A $\tilde{X}$:

El pla tangent té dimensió 2 perquè $X \in \Re^3$ i tenim una constricció activa. Els vectors del pla tangent Y han de satisfer que :

$$\nabla h(\tilde{X}) Y = 0 \quad ; \quad [-2 \quad 1 \quad 1] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0 \Rightarrow -2y_1 + y_2 + y_3 = 0$$

hem de cercar dos vectors ortonormals que satisfagin aquesta condició. Els vectors :

$$Z_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad ; \quad Z_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

satisfan aquesta condició i són mutuament ortogonals ($Z_1' Z_2 = 0$). La base del pla tangent associada a aquests vectors és :

$$Z = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

PROJECCIÓ DE L'HESSIÀ DEL LAGRANGIÀ SOBRE EL PLA TANGENT, I

COMPROVACIÓ DE LA CONDICIÓ SUFICIENT DE SEGON ORDRE :

$$Z'\nabla^2\mathcal{L}(\tilde{X},\tilde{\mu})Z = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -1 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{17}{3} & 0 \\ 0 & 3 \end{bmatrix}$$

$$Z'\nabla^2\mathcal{L}(\tilde{X},\tilde{\mu})Z = \begin{bmatrix} \frac{17}{3} & 0 \\ 0 & 3 \end{bmatrix} \text{def} + \Rightarrow \boxed{\tilde{X} \text{ és un mínim de } f(x) \text{ subjecte a } h(x) \leq 0}$$

Exercici 3.3. : pkutu3/condicions de Kuhn-Tucker

Donat el següent problema de minimització:

$$\begin{array}{rl} \text{opt.} & f(X) = x_1^2x_3 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 4x_1x_3 + 2x_2x_3 - 28x_1 - 10x_2 - 33x_3 \\ \text{subj.a :} & \frac{1}{2}x_1^2 + x_2 + x_3 = \frac{11}{2} \end{array}$$

3.3.1.– Determineu el caràcter de màxim, mínim o punt de sella d'un punt $\tilde{X}$ que satisfà les condicions de 1$^{\text{er}}$ ordre i tal que $\tilde{x}_1 = 1$.

3.3.2.– Si la constricció fos de desigualtat:

$$\text{subj. a :} \quad \frac{1}{2}x_1^2 + x_2 + x_3 \leq \frac{11}{2}$$

$\tilde{X}$ seria també un mínim?

Exercici 3.4. : pkutu4/condicions de Kuhn-Tucker

3.4.1.– Enuncieu les condicions necessàries de segon ordre de mínim local del problema:

$$(\mathbf{P}) \left\{ \begin{array}{rl} \min\limits_{x\in\Re^n} & f(x) \\ \text{subj. a :} & h(x) = 0 \\ & g(x) \leq 0 \end{array} \right.$$

3.4.2.– Considereu el problema **(P)** amb $f(x) = \frac{x_1x_2}{x_3+1}$ i $h(x) = x_1^2 + x_2^2 - x_3 - 1$, sense constriccions $g(x) \leq 0$. La solució $x^* = [1 \;\; 1 \;\; 1]'$ és un punt estacionari de **(P)** amb multiplicador de Lagrange associat $\lambda^* = [-1/4]$. Comproveu si x^* és màxim, mínim o punt de sella de **(P)**.

Exercici 3.5. : pkutu4a/condicions de Kuhn-Tucker

Considereu el següent problema de programació no lineal amb constriccions :

$$(\mathbf{PNL}) \left\{ \begin{array}{rl} \min & f(x_1,x_2) = (x_1-1)^2x_2 + (x_2-x_1)^2 \\ \text{s.a :} & (x_1+1)(x_2+1) + 2x_1^2 + 2x_2^2 \leq 4 \\ & x_1 \geq 0 \,,\; x_2 \geq 0 \end{array} \right.$$

3.5.1.– Comproveu, a partir de les condicions de Kuhn-Tucker si el punt $x^0 = [0 \quad 1]'$ és un mínim local del problema **(PNL)**.

3.5.2.– És **(PNL)** un problema de programació convexa? Justifiqueu la vostra resposta.

SOLUCIÓ

3.5.1.– El punt $x^0 = [0 \quad 1]'$ no és un mínim local de **(PNL)** perquè viola la condició de signe dels multiplicadors de Lagrange ($\lambda_1 = -3/5$).

3.5.2.– El conjunt factible és convex, doncs la primera constricció és de menor o igual i la funció que la defineix és convexa ($\nabla^2 g_1(x) = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$). Les constriccions de signe $x \geq 0$ també defineixen conjunts convexos, i la intersecció de conjunts convexos és un conjunt convex. La matriu Hessiana de la funció objectiu és:

$$\nabla^2 f(x) = \begin{bmatrix} 2x_2 + 2 & 2(x_1 - 1) - 2 \\ 2(x_1 - 1) - 2 & 2 \end{bmatrix}$$

Hem d'estudiar la definició d'aquesta matriu:

- ▷ $\Delta_1 = 2x_2 + 2 \geq 0 \Rightarrow x_2 \geq -1$. Això sempre es satisfà per a tota solució factible **(PNL)**.
- ▷ $\Delta_2 = -4x_1^2 + 16x_1 + 4x_2 - 12 \geq 0$: aquesta condició es pot satisfer en funció dels valors de x_1 i x_2. Si provem, però, el la solució factible de l'anterior apartat $x^0 = [0 \quad 1]'$ s'obté $\Delta_2 = -8 \not\geq 0$. Així doncs, $f(x)$ no és convexa ($\nabla^2 f(x)$ no és semidefinida positiva) sobre la regió factible, i, llavors, **(PNL)** no és un problema de programació convexa.

Exercici 3.6. : pkutu5/condicions de Kuhn-Tuck.

Considereu el problema d'optimització no lineal següent :

$$(\mathbf{P}) \begin{cases} \min & f(x) = e^{x_1(x_2+1))} \\ \text{s.a.} & x_1^2 - x_2 \leq 0 \\ & 2x_1^3 - 2x_1^2 - x_2 + 1 \leq 0 \end{cases}$$

3.6.1.– Formuleu les condidicions necessàries de primer ordre de **(P)**. Indiqueu el procediment que s'hauria de seguir si es volgués resoldre **(P)** a partir d'aquestes condicions. Quína seria la dificultat més rellevant d'aquest procés?.

Si es resol **(P)** amb el paquet GINO s'obté la següent solució, que indicarem per x^*:

```
SOLUTION STATUS:  OPTIMAL TO TOLERANCES.  DUAL CONDITIONS:  SATISFIED.

                    OBJECTIVE FUNCTION VALUE

              1)          .529462

        VARIABLE          VALUE          REDUCED COST
              X1        -.455418            .000000
              X2         .396285           -.000007

             ROW    SLACK OR SURPLUS          PRICE
              2)         .188880            .000000
              3)        -.000009           -.241119
```

3.6.2.– Comproveu si x^* és un mínim local de **(P)**.

3.6.3.– Trobeu una base del subespai tangent de les constriccions actives sobre x^*.

SOLUCIÓ

3.6.1.– Calulem les primeres derivades de la funció objectiu i de les constriccions:

$$\nabla f(x)' = \begin{bmatrix} (x_2+1)e^{x_1(x_2+1)} \\ x_1 e^{x_1(x_1+1)} \end{bmatrix} \quad ; \quad \nabla g_1(x) = \begin{bmatrix} 2x_1 \\ -1 \end{bmatrix} \quad ; \quad \nabla g_2(x) = \begin{bmatrix} 6x_1^2 - 4x_1 \\ -1 \end{bmatrix}$$

Condicions de Kuhn i Tucker: n'hi ha dos multiplicadors de Lagrange μ_1 i μ_2 associats al problema **(PNL)**

i) $\nabla f(x) + \sum_{j=1}^{2} \mu_j \nabla g_j(x) = 0$

$$(x_2+1)e^{x_1(x_2+1)} + 2\mu_1 x_1 + \mu_2 6x_1^2 - 4\mu_2 x_1 = 0 \tag{1}$$
$$x_1 e^{x_1(x_2+1)} - \mu_1 - \mu_2 = 0 \tag{2}$$

ii) $\mu_j g_j(x) = 0$, $j = 1, \ldots, 2$:

$$\mu_1(x_1^2 - x_2) = 0 \tag{3}$$
$$\mu_2(2x_1^3 - 2x_1^2 - x_2 + 1) = 0 \tag{4}$$

iii) $\mu_j \geq 0$, $j = 1, 2$

$$\mu_1 \geq 0 \tag{5}$$
$$\mu_2 \geq 0 \tag{6}$$

Condicions de factibilitat:

$$x_1^2 - x_2 \leq 0 \tag{7}$$
$$2x_1^3 - 2x_1^2 - x_2 + 1 \leq 0 \tag{8}$$

Per a resoldre **(PNL)** a partir de les equacions (1)-(8) caldria explorar els quatre casos següents:

1) Dues constriccions actives: hem de trobar un punt solució del sistema (1)-(8) sencer.
2) Cap constricció activa: de (3) i (4) s'obté $\mu_1 = \mu_2 = 0$ i el sistema es redueix a calcular $\nabla f(x) = 0$, és a dir, un punt estacionari de $f(x)$ a l'interior de la regió factible de **(PNL)**.
3) $g_1(x) = 0$, $g_2(x) < 0$: llavors $\mu_2 = 0$ i s'ha de resoldre (1), (2), (5), (7) i (8), el·liminant μ_2 de (1) i (2).
4) $g_1(x) < 0$, $g_2(x) = 0$: de (3) tenim $\mu_1 = 0$ i s'ha de resoldre (1), (2), (6) i (8), el·liminant μ_1 de (1) i (2).

La major dificultat consisteix en que els sistemes que s'obtenen en els quatre casos anteriors són sistemes d'equacions no lineals que, en general, no es poden resoldre de forma directa.

3.6.2.– A la sortida de GINO observem que el valor d'un dels multiplicadors de Lagrange, μ_2, és negatiu, violant la condició (6): el punt $\tilde{x} = \begin{bmatrix} -0.455418 \\ 0.396285 \end{bmatrix}$ no és mínim local de **(PNL)**.

3.6.3.– Comprovem primer que el punt $\tilde{x}$ és regular. Calculem la jacobiana de les constriccions actives $J = \{2\}$:

$$\nabla g_2(\tilde{x}) = [3.0661 \quad -1] \qquad \text{rang}(\nabla g_2(\tilde{x})) = 1 \Rightarrow \text{rang complet}$$

El subespai tangent sobre $\tilde{x}$ regular és:

$$M = \{y \mid \nabla g_j(\tilde{x})y = 0\,,\, j \in J\} = \{y \mid [3.0661 \quad -1]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0\} = \{y \mid y_2 = 3.0661 y_1\}$$

La dimensió del subespai M és dim(M)=1, llavors qualsevol vector de M serveix com a base d'aquest subespai:

$$\boxed{Z = \begin{bmatrix} 1 \\ 3.0661 \end{bmatrix}}$$

Exercici 3.7. : pkutu6/condicions de Kuhn-Tucker.

Considereu el problema d'optimització no lineal següent :

$$\textbf{(P)} \begin{cases} \min & f(x) = e^{x_1(x_2+1))} \\ \text{s.a.} & x_1^2 - x_2 \geq 0 \\ & 2x_1^3 - 2x_1^2 - x_2 + 1 \geq 0 \end{cases}$$

Comproveu si el punt $x^* = [-0.5 \quad .25]$ és solució del problema **(P)**

Exercici 3.8. : pkutu6a/condicions de Kuhn-Tucker.

Considereu el problema d'optimització no lineal següent :

$$
(\mathbf{PNL})\begin{cases} \min & f(x) = e^{x_1(x_2+1)} \\ \text{s.a.} & x_1^2 - x_2 \geq 0 \\ & 2x_1^3 - 2x_1^2 - x_2 + 1 \geq 0 \end{cases}
$$

Comproveu si el punt $x^* = [-1/2 \quad 1/4]$ és solució del problema **(PNL)**.

SOLUCIÓ

Per tal de comprovar si x^* és òptim de **(PNL)** hem de comprovar les condicions suficients de segon ordre (que inclouen, recordeu, a les de primer ordre). El primer pas és trobar els multiplicadors de Lagrange μ_1 i μ_2 associats a x^*, si es que aquests existeixen. Per tal de calcular aquests multiplicadors usarem les condicions necessàries de primer ordre (condicions de Khun i Tucker). Calculem el gradient i la matriu Jacobiana sobre el punt x^*:

$$
\nabla f(x)' = \begin{bmatrix} (x_2+1)e^{x_1(x_2+1)} \\ x_1 e^{x_1(x_2+1)} \end{bmatrix} \quad , \quad \nabla f(x^*)' = \begin{bmatrix} 0.6691 \\ -0.2675 \end{bmatrix}
$$

$$
\nabla g(x) = \begin{bmatrix} 2x_1 & -1 \\ 6x_1^2 - 4x_1 & -1 \end{bmatrix} \quad , \quad \nabla g(x^*) = \begin{bmatrix} -1 & -1 \\ 3.5 & -1 \end{bmatrix}
$$

S'observa que la matriu Jacobiana és de rang complet, sent doncs x^* regular. Les dues constriccions són actives sobre x^*. Si plantejem la primera condició de Khun i Tucker tenim:

$$
\nabla f(x^*) + [\mu_1^* \quad \mu_2^*]\nabla g(x^*) = 0
$$

$$
\left.\begin{array}{r} 0.6691 - \mu_1^* + 3.5\mu_2^* = 0 \\ -0.2676 - \mu_1^* - \mu_2^* = 0 \end{array}\right\} \boxed{\mu_1^* = -0.05947 \, , \, \mu_2^* = -0.20816}
$$

Les constriccions del problema **(PNL)** estan plantejades com de ≥ 0. Això implica que la condició de signe sobre els multiplicador (tercera condició de Kuhn i Tucker) és $\mu^* \leq 0$, que és satisfeta pel vector trobat. Així doncs, el parell $x^* = [-1/2 \quad 1/4]'$, $\mu^* = [-0.05947 \quad -0.20816]$ satisfan les condicions necessàries de primer ordre. Comprovem ara les de segon ordre, calculant la definició de la matriu Hessiana de la funció Lagrangiana sobre el subespai tangent M. Calculem primer $\nabla_x^2 \mathcal{L}(x^*, \mu^*)$:

$$
\nabla^2 f(x^*) = \begin{bmatrix} 0.8363 & 0.2007 \\ 0.2007 & 0.1338 \end{bmatrix} , \nabla^2 g_1(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} , \nabla^2 g_2(x^*) = \begin{bmatrix} -10 & 0 \\ 0 & 0 \end{bmatrix}
$$

$$
\nabla_x^2 \mathcal{L}(x^*, \mu^*) = \nabla^2 f(x^*) + \sum_{j=1}^{2} \mu_j \nabla^2 g_j(x^*) =
$$

$$
= \begin{bmatrix} 0.8363 & 0.2007 \\ 0.2007 & 0.1338 \end{bmatrix} - 0.05947 \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} - 0.20816 \begin{bmatrix} -10 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2.799 & 0.2007 \\ 0.2007 & 0.1338 \end{bmatrix} \left.\begin{array}{l} \Delta_1 = 2.799 > 0 \\ \Delta_2 = 0.334 > 0 \end{array}\right\} \text{def } +
$$

Donat que $\nabla_x^2 \mathcal{L}(x^*, \mu^*)$ és def + sobre $\mathbb{R}^2$, també ho serà sobre qualsevol subespai de $\mathbb{R}^2$. En particular, serà def + sobre el subespai tangent M. Així doncs, el punt x^* és un mínim local estricte de **(PNL)**.

4 Optimització amb constriccions lineals.

4.1 Mètode de Murtagh i Saunders

4.1.1 Procediment

Sigui el problema d'optimització no lineal amb constriccions lineals en forma estàndar **(NCL)** definit segons :

$$
(\mathbf{NCL}) \left\{ \begin{array}{rll} \min & f(x) & (4.1a) \\ \text{subj. to :} & Ax = b & (4.1b) \\ & 0 \le x \le u & (4.1c) \end{array} \right. \quad ; \quad \begin{array}{c} f\colon \mathbb{R}^n \rightarrow \mathbb{R}\,,\ f \in \mathcal{C}^2 \\ A \in \mathbb{R}^{m\times n}\,,\ \text{rang}(A) = m \\ x, u \in \mathbb{R}^n\,,\ b \in \mathbb{R}^m \end{array}
$$

on ja es té en compte la introducció de les folgues i escreixos necessaris per a obtenir un conjunt de m constriccions d'igualtat. S'indicarà la factibilitat d'un vector $x \in \mathbb{R}^n$ respecte de **(NCL)** amb $x \in \Omega_{NCL}$, sent Ω_{NCL} el conjunt de solucions factibles de **(NCL)** :

Definició 4.1 : $\Omega_{NCL} = \{x \in \mathbb{R}^n : Ax = b\,,\ 0 \le x \le u\}$

Donada una solució factible $x \in \Omega_{NCL}$, el vector $g \in \mathbb{R}^n$ denotarà el vector columna corresponent al gradient de la funció objectiu sobre x :

$$
g(x) = \nabla f(x)' \quad ; \quad g(x)_i = \frac{\partial f(x)}{\partial x_i} \quad , \quad i = 1, \ldots, n
$$

anàlogament, la matriu Hessiana de f s'indicarà :

$$
H(x) = \nabla^2 f(x) \quad ; \quad H(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad , \quad i, j = 1, \ldots, n
$$

L'algorisme de resolució de **(NCL)** presentat que, seguint la nomenclatura usada a [GILL81], anomenarem *Algorisme del Conjunt de Constriccions Actives*, va ser descrit inicialment per Murtagh i Saunders a [MUSA78]. Una descripció exhaustiva de l'algorisme es pot trobar descrita i demostrada amb detall a diversos llibres i articles de l'àrea (per exemple, [GILL81, LUEN84,MUSA78]).

4.1.2 Partició de les variables.

Donat $x \in \Omega_{NCL}$ es defineix la següent partició del conjunt de variables :

Definició 4.2 Variables bàsiques : conjunt de m variables de x amb valor $0 \leq x_i \leq u_i$ i associades a una submatriu de A no singular. Es defineix

$\mathcal{B}$: conjunt d'índexos de les variables bàsiques. Es dirà que x_i és variable bàsica si $i \in \mathcal{B}$. $|\mathcal{B}| = m$.

$x_{\mathcal{B}}$: vector de variables bàsiques ($x_{\mathcal{B}} \in \mathbb{R}^m$). Si la variable x_i és la p-èssima variable bàsica, aleshores : $\mathcal{B}_p = i$, $x_{\mathcal{B}p} = x_i$, $u_{\mathcal{B}p} = u_i$ i $g_{\mathcal{B}p} = g(x)_i$.

Definició 4.3 Variables superbàsiques : conjunt de variables de x tals que $x_i \notin \mathcal{B}$ i $0 < x_i < u_i$. Es defineix

$\mathcal{S}$: conjunt d'índexos de les variables superbàsiques. Es dirà que x_i és variable superbàsica si $i \in \mathcal{S}$. $|\mathcal{S}| = s$.

$x_{\mathcal{S}}$: vector de variables superbàsiques ($x_{\mathcal{S}} \in \mathbb{R}^s$). Si la variable x_i és la p-èssima variable superbàsica, aleshores : $\mathcal{S}_p = i$, $x_{\mathcal{S}p} = x_i$, $u_{\mathcal{S}p} = u_i$ i $g_{\mathcal{S}p} = g(x)_i$.

Definició 4.4 Variables no bàsiques : conjunt de variables de x tals que $x_i \notin \mathcal{B}$ i $x_i = 0$ o $x_i = u_i$. Es defineix :

$\mathcal{N}$: conjunt d'índexos de les variables no bàsiques. Es dirà que x_i és variable no bàsica si $i \in \mathcal{N}$. $|\mathcal{N}| = n - m - s$. Es distingirà entre $\mathcal{N}_u = \{i \in \mathcal{N} : x_i = u_i\}$ i $\mathcal{N}_0 = \{i \in \mathcal{N} : x_i = 0\}$

$x_{\mathcal{N}}$: vector de variables no bàsiques ($x_{\mathcal{N}} \in \mathbb{R}^{(n-m-s)}$). Si la variable x_i és la p-èssima variable no bàsica, aleshores $\mathcal{N}_p = i$, $x_{\mathcal{N}p} = x_i$ i $u_{\mathcal{N}p} = u_i$ i $g_{\mathcal{N}p} = g(x)_i$.

Definició 4.5 Solució degenerada : $x' = [x_{\mathcal{B}}' \mid x_{\mathcal{S}}' \mid x_{\mathcal{N}}']$ és una solució degenerada si $\exists i \in \mathcal{B}$ tq $x_i = u_i$ ó $x_i = 0$

Aquesta partició del conjunt de variables indueix una partició del conjunt de columnes de la matriu A en les submatrius *bàsica* B, sl superbàsica S i *no bàsica* N :

$$\begin{array}{cc} & \begin{array}{ccc} m & s & n-m-s \end{array} \\ A = & \left[\begin{array}{c|c|c} \quad B \quad & S & \qquad N \qquad \end{array}\right] \end{array} \tag{4.2}$$

sent B no singular, segons es desprén de la definició D4.2 . Es pot comprovar facilment que la partició (4.2) existeix sempre que A sigui de rang complet.

4.1.3 Matriu de constriccions actives i base Z del subespai nul.

Donat $x \in \Omega_{NCL}$, el conjunt total de constriccions actives (*"binding"*) vindrà determinat per les m constriccions lineals definides pel sistema $Ax = b$ més un nombre $n - m - s$, que depén del punt x, corresponent a les variables no bàsiques, que per definició es troben a una de les seves fites. La matriu de coeficients de les contriccions actives a x s'anomena *matriu de constriccions actives* :

Definició 4.6 Matriu de constriccions actives. : Sigui $x = \begin{bmatrix} x_{\mathcal{B}} \\ - \\ x_{\mathcal{S}} \\ - \\ x_{\mathcal{N}} \end{bmatrix} \in \Omega_{NCL}$ solució factible de **(NCL)**. Es defineix com la matriu de constriccions actives de **(NCL)** a x la matriu $n \times (n-s)$:

$$\begin{array}{cccc} & m \quad\; s & n-m-s & \\ \bar{M} = & \begin{array}{|c|c|c|} \hline B & S & N \\ \hline \multicolumn{2}{|c|}{\mathbf{0}} & \mathbf{I} \\ \hline \end{array} & & \begin{array}{c} m \\ (n-m-s) \end{array} \end{array} \tag{4.3}$$

La matriu $\bar{M}$ és de rang complet si A ho és. L'expressió de les constriccions actives a $x \in \Omega_{NCL}$ és :

$$\bar{M}x = \bar{b} \quad ; \quad \bar{b} = \begin{bmatrix} b \\ b_{\mathcal{N}} \end{bmatrix} \quad ; \quad b_{\mathcal{N}i} = \begin{cases} u_{\mathcal{N}i} & \text{si } \mathcal{N}_i \in \mathcal{N}_u \\ 0 & \text{si } \mathcal{N}_i \in \mathcal{N}_0 \end{cases} \tag{4.4}$$

on $b_{\mathcal{N}}$ és el *vector de fites actives.* El conjunt total de constriccions de **(NCL)**, expressat en funció de $\bar{M}$ és :

$$\bar{M}x = \bar{b} \tag{4.5a}$$

$$0 \leq x_i \leq u_i \quad \forall i \in \mathcal{B} \cup \mathcal{S} \tag{4.5b}$$

Donat que la matriu $\bar{M} \in \mathbb{R}^{(n-s)\times n}$ és de rang complet, les seves files generen un subespai de dimensió $n-s$ dit *espai de rang.* Interessa definir el seu complement ortogonal, dit *espai nul de $\bar{M}$* :

Definició 4.7 Espai nul de $\bar{M}$: $\mathrm{N}_{\bar{M}} = \{p \in \mathbb{R}^n \ : \bar{M}p = 0\}$

Donada una matriu $Z \in \mathbb{R}^{(n\times s)}$ les seves columnes formaràn una base de $\mathrm{N}_{\bar{M}}$ si i només si es satisfà :

$$i) \quad \mathrm{rang}(Z) = s \tag{4.6a}$$

$$ii) \quad \bar{M}Z = \mathbf{0} \tag{4.6b}$$

Hi ha diverses formes possibles de construïr la base Z. Una de les més emprades correspon a l'expressió :

$$Z = \begin{array}{r} \\ m\{ \\ s\{ \\ n-m-s\{ \end{array} \begin{array}{c} \overbrace{\phantom{-B^{-1}S}}^{s} \\ \begin{bmatrix} -B^{-1}S \\ \mathbf{I} \\ \mathbf{0} \end{bmatrix} \end{array} \tag{4.7}$$

la matriu definida a (4.7) és de rang complet i satisfà la condició (4.6b) :

$$\bar{M}Z = \begin{bmatrix} B & S & N \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} -B^{-1}S \\ \mathbf{I} \\ 0 \end{bmatrix} = \begin{bmatrix} -BB^{-1}S + S\mathbf{I} + N\mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} -S+S \\ \mathbf{0} \end{bmatrix} = [\mathbf{0}] \tag{4.8}$$

sent doncs, una base de $\mathrm{N}_{\bar{M}}$.

A partir del gradient $g(x)$, la Hessiana $H(x)$ i la base Z es poden calcular el gradient i la Hessiana *projectats*:

Definició 4.8 Gradient projectat : $g_z(x) = Z'g(x) = g_S(x) - S'(B^{-1})'g_B(x)$

Definició 4.9 Hessiana projectada : $H_z(x) = Z'H(x)Z$

$g_z(x)$ i $H_z(x)$ també són coneguts com el *gradient reduït* i *l'Hessià reduït*.

4.1.4 Condicions d'optimalitat de (NCL)

Donada una solució factible de **(NCL)** x, les condicions necessàries de primer ordre que s'han de satisfer a x per a poder ésser declarat minim local (*condicions de Karush-Kunh-Tucker (KKT)*) s'estableixen en el següent teorema :

Teorema 4.8 : Sigui $x \in \Omega_{NCL}$. Si x és mínim local de **(NCL)** aleshores existeixen $\pi^* \in \mathbb{R}^m$ i $\sigma_u^*, \sigma_0^* \in \mathbb{R}^n$, dits *Multiplicadors de Lagrange*, tals que :

$$i) \quad A'\pi^* + \sigma_0^* + \sigma_u^* = \nabla f(x) \tag{4.9a}$$

$$ii) \quad \sigma_{0i}^* x_i = 0 \ , \ i = 1, \dots, n \tag{4.9b}$$

$$\sigma_{ui}^*(u_i - x_i) = 0 \ , \ i = 1, \dots, n \tag{4.9c}$$

$$iii) \quad \sigma_{0i}^* \geq 0 \ , \ i = 1, \dots, n \tag{4.9d}$$

$$\sigma_{ui}^* \leq 0 \ , \ i = 1, \dots, n \tag{4.9e}$$

Les condicions (KKT) es poden redefinir tenint en compte la partició $x' = [\, x_B{}' \mid x_S{}' \mid x_N{}' \,]$ definida a D4.2 i D4.4 i el concepte de matriu de constriccions actives D4.6 :

Teorema 4.9 : Sigui $x' = [x_B' \mid x_S' \mid x_{\mathcal{N}}'] \in \Omega_{NCL}$. Sigui $\bar{M}$ la matriu de constriccions actives a x i $g' = [g_B' \mid g_S' \mid g_{\mathcal{N}}']$ el vector gradient de $f(x)$ a x. Si x és mínim local de (NCL) aleshores existeixen $\pi^* \in \mathbb{R}^m$ i $\sigma^*_{\mathcal{N}} \in \mathbb{R}^{(n-m-s)}$ tals que :

$$i)\quad \bar{M}' \begin{bmatrix} \pi^* \\ \sigma^*_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} B' & 0 \\ S' & 0 \\ N' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \pi^* \\ \sigma^*_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} g_B \\ - \\ g_S \\ - \\ g_{\mathcal{N}} \end{bmatrix} \tag{4.10a}$$

$$ii)\quad \sigma^*_{\mathcal{N}i} \geq 0\,,\ \forall \mathcal{N}_i \in \mathcal{N}_0 \tag{4.10b}$$

$$\sigma^*_{\mathcal{N}i} \leq 0\,,\ \forall \mathcal{N}_i \in \mathcal{N}_u \tag{4.10c}$$

Sobre un vector $x \in \Omega_{NCL}$ on el sistema (4.10a) sigui compatible s'obté, a partir de (4.10a)

$$B'\pi = g_B \quad ; \quad \pi' = g_B' B^{-1} \tag{4.11a}$$

$$N'\pi + \sigma_{\mathcal{N}} = g_{\mathcal{N}} \quad ; \quad \sigma_{\mathcal{N}}' = g_{\mathcal{N}}' - \pi' N \tag{4.11b}$$

$$S'\pi = g_S \quad ; \quad g_S - S'\pi = 0 \tag{4.11c}$$

Les equacions (4.11a) i (4.11b) poden ésser emprades per a calcular uns vectors π i $\sigma_{\mathcal{N}}$ a qualsevol solució $x \in \Omega_{NCL}$, peró només si el gradient projectat sobre x s'anula es satisfan les equacións (4.11c) . Com es veurà, els vectors π i $\sigma_{\mathcal{N}}$ són necessàris per a l'aplicació de l'algorisme.

4.1.5 Algorisme del Conjunt de Constriccions Actives.

Algorisme A4.1 : Algorisme del Conjunt de Constriccions Actives.

$\boxed{0}$ **Inicialització de l'algorisme :**
Donat $x^0 \in \Omega_{NCL}$ es defineixen :
$\mathcal{B}^0 \;;\; \mathcal{S}^0 \;;\; \mathcal{N}^0 \;;\; (\mathbf{SN}\bar{M})^0$
$Z^0 \;;\; g_z^0 := Z^{0\prime} g(x^0) \;;\; H_z^0 := Z^{0\prime} H(x^0) Z^0 \;;\; k := 0.$

$\boxed{1}$ **Comprovació de les condicions d'optimalitat de (SN$\bar{M}$) :**
Si $\|g_z^k\|_2 = 0 \Rightarrow x^k = x^{*\,k}_{\bar{M}}$: anada a $\boxed{2}$.
Si $\|g_z^k\|_2 \neq 0 \Rightarrow x^k \neq x^{*\,k}_{\bar{M}}$: anada a $\boxed{3}$.

$\boxed{2}$ $x^k = x^{*\,k}_{\bar{M}}$ **: comprovació de les condicions d'optimalitat de (NCL) :**
$\boxed{2.1}$ Càlcul $\sigma_{\mathcal{N}}^{k\prime} = g_{\mathcal{N}}^{k\prime} - \pi^{k\prime} N^k$.
$\boxed{2.2}$ Si $\exists\, q \in \mathcal{N}_0^k$ tq $\sigma_q^k < 0$ ó $\exists\, q \in \mathcal{N}_u^k$ tq $\sigma_q^k > 0$ llavors :
$\mathcal{B}^{k+1} := \mathcal{B}^k \;;\; \mathcal{S}^{k+1} := \mathcal{S}^k \cup \{q\} \;;\; \mathcal{N}^{k+1} := \mathcal{N}^k \setminus \{q\}$.
Actualització $(\mathbf{SN}\bar{M})^k$.
Anada a $\boxed{3}$.
Altrament, anar a $\boxed{4}$.

$\boxed{3}$ $x^k \neq x^{*\,k}_{\bar{M}}$ **: iteració a (SN$\bar{M}$)^k :**
$\boxed{3.1}$ **Càlcul d'una direcció factible de descens :**
Resolució de $H_z^k p_z^{\,k} = -g_z^k \;;\; p_S^k = p_z^{\,k} \;;\; p_B^k = -[B^{-1}]^k S^k p_S^k \;;\; p_{\mathcal{N}}^k = [0]$.
$\boxed{3.2}$ **Càlcul de la passa màxima :** $\bar{\alpha} = \min\{\bar{\alpha}_B, \bar{\alpha}_S\}$.

$\boxed{3.3}$ **Exploració lineal :** $\alpha^{*k} = \text{argmin}_{0<\alpha\leq\bar{\alpha}}\{f(x^k + \alpha p^k)\}$.
$\boxed{3.4}$ **Actualització del punt iterat :**
$x^{k+1} := x^k + \alpha^{*k} p^k$.
$g_z^{k+1} := Z^{k+1'} g(x^{k+1})$.
$H_z^{k+1} := Z^{k+1'} H(x^{k+1}) Z^{k+1}$.
$\boxed{3.5}$ **Pivotació** : si $\alpha^{*k} = \bar{\alpha}$ llavors:
Si $\alpha^{*k} = \bar{\alpha}_B$:
Seleccionar $l \in \mathcal{S}$ acceptable per a pivotar.
$\mathcal{B}^{k+1} := \mathcal{B}^k \setminus \{p\} \cup \{l\}$; $\mathcal{S}^{k+1} := \mathcal{S}^k \setminus \{l\}$; $\mathcal{N}^{k+1} := \mathcal{N}^k \cup \{p\}$.
Si $\alpha^{*k} = \bar{\alpha}_S$:
$\mathcal{B}^{k+1} := \mathcal{B}^k$; $\mathcal{S}^{k+1} := \mathcal{S}^k \setminus \{p\}$; $\mathcal{N}^{k+1} := \mathcal{N}^k \cup \{p\}$.
Actualització de $(\mathbf{SN}\bar{M})^k$: Z^{k+1}
$\boxed{3.6}$ $k := k + 1$. Anada a $\boxed{1}$.
$\boxed{4}$ **Acabament :** $x^k = x_{\bar{M}}^{*k}$ i satisfà (4.10b) -(4.10c) $\Rightarrow$ $\boxed{x^k \text{ òptim de } \mathbf{(NCL)}}$.

Es necessari fer alguns comentaris sobre l'ús de l'Hessià projectat H_z^k. El disseny de la versió especialitzada de l'algorisme que es presentarà als següents capítols considera que les segones derivades de la funció objectiu no estan disponibles. Això provoca que, en realitat, a les passes de l'algorisme A4.1 on s'indiqui H_z^k s'hagi de considerar que s'està fent referència a una certa aproximació de Hz^k. D'altra banda, el fet de no disposar de les segones derivades fa que les comprovacions d'optimalitat es limitin a assegurar les condicions necessàries de primer ordre si no es dispossa d'informació addicional sobre la convexitat de f.

4.1.6 Exercicis

Exercici 4.1. : pmusal/mètode de Murtagh-Saunders

Donada la minimització:

$$\min f(X) = \frac{x_1 {x_3}^2}{2} - x_1 x_5 + 3x_1 - 6x_2 - 4x_3 - 4x_4 x_6 + {x_6}^2 x_7 x_8$$

$$\text{subj.} \begin{bmatrix} 3 & 1 & 0 & -3 & 3 & 0 & 1 & 3 \\ -6 & 1 & 1 & 8 & 6 & 4 & 0 & -3 \\ 3 & -1 & 0 & -3 & 2 & -3 & 1 & 0 \end{bmatrix} X = \begin{bmatrix} 23 \\ 41 \\ 9 \end{bmatrix}$$

$$\underline{0} \leq X \leq \overline{X} \quad \text{amb} \quad \overline{X} = [5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5]'$$

determineu el gradient projectat pel procediment de Murtagh-Saunders a partir del punt factible $X_k = [2 \;\; 0 \;\; 1 \;\; 3 \;\; 5 \;\; 1 \;\; 5 \;\; 2]'$, reordenant, si cal, les variables i comproveu que la direcció de menys el gradient projectat a partir de X_k és una direcció factible de descens.

SOLUCIÓ

DETERMINACIÓ DE LES VARIABLES BÀSIQUES, SUPERBÀSIQUES I NO BÀSIQUES

Les variables bàsiques $\underline{0} \leq X_B \leq \overline{X}_B$, a ser possible estrictament entre fites, han de tenir imprescindiblement columnes d'A linealment independents. Les variables superbàsiques $\underline{0} < X_S < \overline{X}_S$ han d'estar estrictament entre fites. Les variables no bàsiques X_N han de tenir les seves components o a la seva fita superior: $x_{Nj} = \overline{x}_j$, $j \in \overline{I}$ o a la seva fita inferior: $x_{Nj} = 0$,

$j \in \underline{I}$, sent $\overline{I}$ i $\underline{I}$ els conjunt d'índexos de variables no bàsiques a fita superior i a fita inferior respectivament.

Per triar les variables bàsiques tenim que:

- x_2 està a la fita inferior i x_5 i x_7 estan a la superior (per tant, en principi, les excloem del conjunt de les bàsiques)
- x_1, x_3 i x_4 tenen les columnes linealment dependents (no poden ser simultàniament bàsiques)
- x_1, x_3 i x_6 poden ser les variables bàsiques X_B perquè tenen les columnes linealment independents:

$$B = \begin{bmatrix} 3 & 0 & 0 \\ -6 & 1 & 4 \\ 3 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 3 & & \\ -6 & 1 & \\ 3 & 0 & -3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 4 \\ & & 1 \end{bmatrix}$$

- donat que hem trobat variables bàsiques estrictament entre fites, totes les variables que estan a una fita seran no bàsiques: $X_N = \{x_2, x_5, x_7\}$, i la resta de variables (estrictament entre fites) seran les superbàsiques $X_S = \{x_4, x_8\}$. (Altres combinacions de bàsiques i superbàsiques són també possibles).

MATRIUS S, N I $\widehat{A}$ I CÀLCUL (NO NECESSARI) DE Z

S és doncs $S = \begin{bmatrix} -3 & 3 \\ 8 & -3 \\ -3 & 0 \end{bmatrix}$ i $N = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 6 & 0 \\ -1 & 2 & 1 \end{bmatrix}$ d'on la matriu de constriccions actives $\widehat{A}$ és:

$$\widehat{A} = \left[\begin{array}{c|c|c} B & S & N \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \tilde{\mathbf{1}} \end{array}\right] = \begin{bmatrix} 3 & 0 & 0 & -3 & 3 & 1 & 3 & 1 \\ -6 & 1 & 4 & 8 & -3 & 1 & 6 & 0 \\ 3 & 0 & -3 & -3 & 0 & -1 & 2 & 1 \\ & & & & & 1 & & \\ & & \mathbf{0} & & & & -1 & \\ & & & & & & & -1 \end{bmatrix}$$

La matriu Z, la qual **no és necessari** tenir emmagatzemada de forma explícita, fóra:

$$Z = \begin{bmatrix} -B^{-1}S \\ \mathbb{1} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -2 & 1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$ ja que $BV_1 = S_1$ dóna $V_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$ i $BV_2 = S_2$, $V_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

GRADIENT I GRADIENT PROJECTAT

$$\nabla f(X)' = \begin{bmatrix} \frac{x_3^2}{2} - x_5 + 3 \\ -6 \\ x_1 x_3 - 4 \\ -4x_6 \\ -x_1 \\ -4x_4 + 2x_6 x_7 x_8 \\ {x_6}^2 x_8 \\ {x_6}^2 x_7 \end{bmatrix} \quad \text{i } \nabla f(X_k)' = \begin{bmatrix} -\frac{3}{2} \\ -6 \\ -2 \\ -4 \\ -2 \\ 8 \\ 2 \\ 5 \end{bmatrix}$$

Tenint en compte la ordenació i distribució de variables bàsiques $X_B = \{x_1, x_3, x_6\}$ superbàsiques $X_S = \{x_4, x_8\}$ i no bàsiques $X_N = \{x_2, x_5, x_7\}$, tenim que:

$$G_B = \begin{bmatrix} -\frac{3}{2} \\ -2 \\ 8 \end{bmatrix} \quad G_S = \begin{bmatrix} -4 \\ 5 \end{bmatrix} \quad G_N = \begin{bmatrix} -6 \\ -2 \\ 2 \end{bmatrix} \text{ i que la solució de } B'\Pi = G_B \text{ és:}$$

$$\begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -6 & 3 \\ & 1 & 0 \\ & & -3 \end{bmatrix} \cdot \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ -2 \\ 8 \end{bmatrix} \longrightarrow \Pi = \begin{bmatrix} \frac{5}{6} \\ -2 \\ -\frac{16}{3} \end{bmatrix}$$

$$\text{i } Z'G = G_S - S'\Pi = \begin{bmatrix} -4 \\ 5 \end{bmatrix} - \begin{bmatrix} -3 & 8 & -3 \\ 3 & -3 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{5}{6} \\ -2 \\ -\frac{16}{3} \end{bmatrix} = \begin{bmatrix} -3/2 \\ -7/2 \end{bmatrix}$$

DIRECCIÓ DE MENYS EL GRADIENT PROJECTAT

Donat que $||Z'G||^2{=}14{,}5$ no està a prop de zero, haurem de fer una passa en la direcció $P_S{=}P_z{=}-Z'G$ (o en la que s'obtindria de resoldre $Z'H_k Z P_z{=}-Z'G$) i calcular $P_B{=}-B^{-1}SP_S$, és a dir, resoldre $BP_B = -SP_S$ i fer $P_N{=}\underline{0}$.

$$\text{Calculant } SP_s = \begin{bmatrix} -3 & 3 \\ 8 & -3 \\ -3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} 6 \\ 3/2 \\ -9/2 \end{bmatrix}$$

$$\text{i resolent } \begin{bmatrix} 3 & & \\ -6 & 1 & \\ 3 & 0 & -3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 4 \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} p_{B1} \\ p_{B2} \\ p_{B3} \end{bmatrix} = \begin{bmatrix} -6 \\ -3/2 \\ 9/2 \end{bmatrix} \text{ obtenim } P_B = \begin{bmatrix} -2 \\ 1/2 \\ -7/2 \end{bmatrix}$$

La derivada direccional $\nabla f(X_k)P{=}G'_B P_B{+}G'_S P_S{=}-29/2{<}0$ per tant queda comprovat que és una direcció de descens.

Exercici 4.2. : pmusa2/mètode de Murtagh-Saunders

En la minimització de:

$$\min f(X) = \frac{{x_1}^2 x_5}{10} + x_2 x_3 - 3x_4 {x_8}^2 + 2x_5 x_7 - 2x_6 + 3x_8$$

$$\text{subj. } \begin{bmatrix} 3 & 1 & 0 & -3 & 3 & 0 & 1 & 3 \\ -6 & 1 & 1 & 8 & 6 & 4 & 0 & -3 \\ 3 & -1 & 0 & -3 & 2 & -3 & 1 & 0 \end{bmatrix} X = \begin{bmatrix} 31 \\ 92/3 \\ 11 \end{bmatrix}$$

$$\underline{0} \le X \le \overline{X} \quad \text{amb} \quad \overline{X} = [5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5 \;\; 5]'$$

pel procediment de Murtagh-Saunders, s'ha trobat que en el punt X_0=[3 0 3 $\frac{1}{3}$ 5 4 5 1]' i partint de la subdivisió en variables bàsiques: $X_B=\{x_1,x_3,x_6\}$, superbàsiques: $X_S=\{x_4,x_8\}$ i no bàsiques: $X_N=\{x_2,x_5,x_7\}$, el gradient projectat val zero: $Z'\nabla f(X_0)'=\underline{0}$. Indiqueu si X_0 pot ser un òptim, i si no ho podés ser, quines variables haurien de canviar de valor i com calcularieu el seu canvi.

SOLUCIÓ

CÀLCUL DEL GRADIENT A X_0 I DELS MULTIPLICADORS Π

$$\nabla f(X)' = \begin{bmatrix} \frac{x_1 x_5}{5} \\ x_3 \\ x_2 \\ -3{x_8}^2 \\ \frac{{x_1}^2}{10}+2x_7 \\ -2 \\ 2x_5 \\ -6x_4x_8+3 \end{bmatrix} \quad \text{i} \quad G=\nabla f(X_0)' = \begin{bmatrix} 3 \\ 3 \\ 0 \\ -3 \\ \frac{109}{10} \\ -2 \\ 10 \\ 1 \end{bmatrix}$$

Tenint en compte l'ordenació i distribució de variables bàsiques $X_B = \{x_1,x_3,x_6\}$ superbàsiques $X_S=\{x_4,x_8\}$ i no bàsiques $X_N=\{x_2,x_5,x_7\}$, tenim que:

$$G_B = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} \quad G_S = \begin{bmatrix} -3 \\ 1 \end{bmatrix} \quad G_N = \begin{bmatrix} 3 \\ 109/10 \\ 10 \end{bmatrix}$$

i la matriu B de columnes bàsiques és: $B = \begin{bmatrix} 3 & 0 & 0 \\ -6 & 1 & 4 \\ 3 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 3 & & \\ -6 & 1 & \\ 3 & 0 & -3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 4 \\ & & 1 \end{bmatrix}$

La solució de $B'\Pi = G_B$ és:

$$\begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -6 & 3 \\ & 1 & 0 \\ & & -3 \end{bmatrix} \cdot \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} \quad \longrightarrow \quad \Pi = \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}$$

COMPROBACIÓ (NO NECESSÀRIA SEGONS L'ENUNCIAT) DE QUE $Z'\nabla f(X_0)'=\underline{0}$

La matriu Z és:

$$Z = \begin{bmatrix} -B^{-1}S \\ \mathbb{1} \\ \mathbf{0} \end{bmatrix} \quad \text{i} \quad Z'G = [-S'B^{-1'} \quad \mathbb{1} \quad \mathbf{0}] \cdot \begin{bmatrix} G_B \\ G_S \\ G_N \end{bmatrix} = G_S - SB^{-1'}G_B = G_S - S'\Pi$$

$$\text{on } S = \begin{bmatrix} -3 & 3 \\ 8 & -3 \\ -3 & 0 \end{bmatrix} \quad \text{així: } Z'G = \begin{bmatrix} -3 \\ 1 \end{bmatrix} - \begin{bmatrix} -3 & 8 & -3 \\ 3 & -3 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

TAXACIÓ DE LES VARIABLES NO BÀSIQUES

La matriu de constriccions actives $\widehat{A}$ és:

$$\widehat{A} = \left[\begin{array}{cc|c} B & S & N \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \tilde{\mathbf{1}} \end{array}\right] = \begin{bmatrix} 3 & 0 & 0 & -3 & 3 & 1 & 3 & 1 \\ -6 & 1 & 4 & 8 & -3 & 1 & 6 & 0 \\ 3 & 0 & -3 & -3 & 0 & -1 & 2 & 1 \\ & & & & & 1 & & \\ & & \mathbf{0} & & & & -1 & \\ & & & & & & & -1 \end{bmatrix}$$

L'equació de taxació (condició de 1$^{\text{er}}$ ordre) és: $\widehat{A}'\Lambda = G$ amb $\Lambda = \begin{bmatrix} \Pi \\ \Sigma \end{bmatrix}$ i cal que tots els components de Σ siguin positius perquè, sent $Z'G=\underline{0}$, siguem en un mínim.

$$\widehat{A}'\begin{bmatrix} \Pi \\ \Sigma \end{bmatrix} = \begin{bmatrix} G_B \\ G_S \\ G_N \end{bmatrix} \quad \longrightarrow \quad \begin{cases} B'\Pi = G_B \quad \Pi = \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} \quad \text{(ja trobat)} \\ S'\Pi = G_S \ \longrightarrow \ G_S - S'\Pi = 0 = Z'G \quad \text{(segons enunciat)} \\ N'\Pi + \tilde{\mathbf{1}}\Sigma = G_N \ \longrightarrow \ \Sigma = \tilde{\mathbf{1}}(G_N - N'\Pi) \end{cases}$$

Val la pena separar de la seva fita tota component x_{Nj} per a la qual $\sigma_j<0$

$$\sigma_2 = +1\left(3 - [1 \quad 1 \quad -1]\cdot\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}\right) = \frac{10}{3} > 0$$

$$\sigma_5 = -1\left(\frac{109}{10} - [3 \quad 6 \quad 2]\cdot\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}\right) = -\frac{257}{30} < 0 \quad (x_5\text{: candidata a entrar})$$

$$\sigma_7 = -1\left(10 - [1 \quad 0 \quad 1]\cdot\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix}\right) = -9 < 0 \quad (x_7\text{: candidata a entrar})$$

Donat que hi ha components de Σ negatius, **NO** som a l'òptim (malgrat que $Z'G=\underline{0}$)

ENTRADA D'UNA NO BÀSICA I DIRECCIÓ D'EXPLORACIÓ

De les dues candidates a entrar que tenim entrarem la de σ_j més negativa: x_7, la qual serà afegida, provisionalment, al conjunt de les superbàsiques: $\widetilde{X}_S=\{x_4, x_8, x_7\}$ i $\widetilde{X}_N=\{x_2, x_5\}$. Tenim així:

$$\widetilde{S} = \begin{bmatrix} -3 & 3 & 1 \\ 8 & -3 & 0 \\ -3 & 0 & 1 \end{bmatrix} \quad \widetilde{G}_S = \begin{bmatrix} -3 \\ 1 \\ 10 \end{bmatrix} \quad \widetilde{G}_N = \begin{bmatrix} 3 \\ \frac{109}{10} \end{bmatrix}$$ i el nou gradient projectat $\widetilde{Z}'G$ és:

$$\widetilde{Z}'G = \widetilde{G}_S - \widetilde{S}'\Pi = \begin{bmatrix} -3 \\ 1 \\ 10 \end{bmatrix} - \begin{bmatrix} -3 & 8 & -3 \\ 3 & -3 & 0 \\ 1 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 9 \end{bmatrix} = \begin{bmatrix} Z'G \\ \text{signe de}(\tilde{\mathbf{1}}_{x_7x_7})\sigma_7 \end{bmatrix}$$

Farem una passa en la direcció $P_S=P_z=-\widetilde{Z}'G$ i calcularem $P_B=-B^{-1}\widetilde{S}P_S$, és a dir, resoldrem $BP_B = -\widetilde{S}P_S$ i fem $P_N=\underline{0}$.

$$\text{Calculant } SP_S = \begin{bmatrix} -3 & 3 & 1 \\ 8 & -3 & 0 \\ -3 & 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 0 \\ 0 \\ -9 \end{bmatrix} = \begin{bmatrix} -9 \\ 0 \\ -9 \end{bmatrix}$$

$$\text{i resolent } \begin{bmatrix} 3 & & \\ -6 & 1 & \\ 3 & 0 & -3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 4 \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} p_{B1} \\ p_{B2} \\ p_{B3} \end{bmatrix} = \begin{bmatrix} 9 \\ 0 \\ 9 \end{bmatrix} \text{ obtenim } P_B = \begin{bmatrix} 3 \\ 18 \\ 0 \end{bmatrix}$$

Tenim doncs $P_B = \begin{bmatrix} 3 \\ 18 \\ 0 \end{bmatrix}$ i $P_S = \begin{bmatrix} 0 \\ 0 \\ -9 \end{bmatrix}$. Ara caldria determinar la passa màxima i fer exploració lineal per determinar el nou punt.

(Noti's que el component p_{S7}=-9<0 és negatiu, com correspon a la variació de x_7 que estava a la seva fita superior).

Exercici 4.3. : pmusa3/mètode de Murtagh-Saunders

Considereu el següent problema d'optimització no lineal:

$$(\mathbf{P}) \begin{cases} \min\limits_{x \in \Re^7} & f(x) \\ \text{subj. a :} & Ax = b \\ & [0] \le x \le \bar{x} \end{cases}$$

amb:

$$f(x) = 3x_1^2 + x_2^2 - 4x_3^2 - 2x_4^2 + x_5^2 - 2x_6^2 + 3x_7^2 + x_1x_2 + x_3x_4 - x_3x_7 + 4x_5 - 2x_6$$

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 2 & -3 \\ -1 & 0 & 4 & -2 & 2 & 4 & -4 \\ 0 & -3 & 0 & 1 & 3 & -3 & -2 \end{bmatrix} \quad ; \quad \bar{x}' = [3 \quad 2 \quad 3 \quad 2 \quad 3 \quad 3 \quad 2]$$

Durant la resolució d'aquest problema amb l'algorisme de Murtagh-Saunders s'ha obtingut l'iterat $x^{k'} = [2 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2 \quad 0]$ amb la partició de variables següent: $x_{\mathcal{B}}^{k'} = [x_1 \quad x_2 \quad x_5]$, $x_{\mathcal{S}}^{k'} = [x_3 \quad x_6]$, $x_{\mathcal{N}}^{k\,'} = [x_4 \quad x_7]$.

4.3.1.– Realitzeu una nova iteració de l'algorisme Murtagh-Saunders a partir de x^k.

4.3.2.– És x^k un mínim local de problema? En cas que no ho sigui, podrieu indicar algún procediment que permetés trobar una direcció de descens i continuar iterant?

Exercici 4.4. : pmusa3/mètode de Murtagh-Saunders

Considereu el següent problema d'optimització no lineal:

$$(\mathbf{P}) \begin{cases} \min\limits_{x \in \Re^7} & f(x) \\ \text{subj. a :} & Ax = b \\ & [0] \le x \le \bar{x} \end{cases}$$

amb:

$$f(x) = 3x_1^2 + x_2^2 - 4x_3^2 - 2x_4^2 + x_5^2 - 2x_6^2 + 3x_7^2 + x_1x_2 + x_3x_4 - x_3x_7 + 4x_5 - 2x_6$$

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 2 & -3 \\ -1 & 0 & 4 & -2 & 2 & 4 & -4 \\ 0 & -3 & 0 & 1 & 3 & -3 & -2 \end{bmatrix} \quad ; \quad \bar{x}' = [3 \quad 2 \quad 3 \quad 2 \quad 3 \quad 3 \quad 2]$$

Durant la resolució d'aquest problema amb l'algorisme de Murtagh-Saunders s'ha obtingut l'iterat $x^{k'} = [2 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2 \quad 0]$ amb la partició de variables següent: $x_{\mathcal{B}}^{k'} = [x_1 \quad x_2 \quad x_5]$, $x_{\mathcal{S}}^{k'} = [x_3 \quad x_6]$, $x_{\mathcal{N}}^{k'} = [x_4 \quad x_7]$.

4.1.– Realitzeu una nova iteració de l'algorisme Murtagh-Saunders a partir de x^k.

4.2.– És x^k un minim local de problema? En cas que no ho sigui, podrieu indicar algún procediment que permetés trobar una direcció de descens i continuar iterant?

Exercici 4.5. : pmusa4/mètode de Murtagh-Saunders

Considereu el següent problema d'optimització no lineal:

$$
(\mathbf{P}) \left\{ \begin{array}{ll} \min\limits_{x \in \Re^5} & f(x) = \dfrac{x_1 x_2}{x_3 + 1} - x_4 x_5 \\ \text{subj. a :} & \begin{bmatrix} 2 & 4 & 4 & 0 & 3 \\ 1 & 9 & 2 & 5 & 0 \end{bmatrix} x = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\ & [0]' \leq x \leq \bar{x} = [4 \quad 4 \quad 3 \quad 4 \quad 4] \end{array} \right.
$$

Aplicant la fase I del símplex s'ha obtingut la solució inicial factible $x^0 = [1 \quad 4 \quad 3 \quad 0 \quad 2]'$. Efectueu la primera iteració de l'algorisme de Murtagh i Saunders aplicat sobre $(\mathbf{P})$ a partir de x^0. En concret:

4.5.1.– Determineu els conjunts $\mathcal{B}^0$, $\mathcal{S}^0$ i $\mathcal{N}^0$.

4.5.2.– Comproveu la condició d'aturada de l'algorisme sobre x^0, realitzant, si cal, les actualitzacions de $\mathcal{B}^0$, $\mathcal{S}^0$ i $\mathcal{N}^0$.

4.5.3.– Calculeu la direcció de descens.

4.5.4.– Efectueu l'actualització de les variables prenent com a passa òptima α^* la passa màxima permessa $\bar{\alpha}$, indicant com s'actualitzarien les llistes $\mathcal{B}^0$, $\mathcal{S}^0$ i $\mathcal{N}^0$.

Exercici 4.6. : pmusa5/mètode de Murtagh-Saunders

Considereu el següent problema d'optimització no lineal amb constriccions lineals:

$$
(\mathbf{NCL}) \left\{ \begin{array}{ll} \min & f(x) \\ \text{subj. to :} & Ax = b \\ & 0 \leq x \leq u \end{array} \right. ; \left. \begin{array}{l} f\colon \mathbb{R}^n \longrightarrow \mathbb{R}\ ,\ f \in \mathcal{C}^2 \\ A \in \mathbb{R}^{m \times n}\ ,\ \text{rang}(A) = m \\ x, u \in \mathbb{R}^n\ ,\ b \in \mathbb{R}^m \end{array} \right\}
$$

En la resolució d'aquest problema mitjançant l'algorisme de Murtagh i Saunders s'ha arribat a una solució factible X^k amb gradient reduït nul ($g_z^k = 0$). Un cop avaluats els multiplicadors de Lagrange $\sigma_{\mathcal{N}}$ s'observa que la i-èssima variable no bàsica $x_{\mathcal{N}i} = 0$ té multiplicador de Lagrange associat $\sigma_{\mathcal{N}i} < 0$. Sigui p la direcció definida per la relació $\bar{A}^k p = e_{m+i}$, amb $\bar{A}^k$ matriu de constriccions actives a X^k i $e_{m+i} \in \mathbb{R}^{n-s}$ vector unitari amb $+1$ a la posició $m + i$. Demostreu que la direcció p és factible i de descens.

Exercici 4.7. : pmusa6/mètode de Murtagh-Saunders

Considereu el següent problema d'optimització no lineal:

$$
(\mathrm{P}) \quad \begin{cases} \min\limits_{x \in \Re^4} & f(x) = -x_1^2 + x_3 x_2^2 - x_1 x_4 \\ \text{subj. a :} & \begin{bmatrix} 2 & 2 & 1 & 4 \\ 3 & 0 & 0 & -1 \end{bmatrix} x = \begin{bmatrix} 10.75 \\ 1.25 \end{bmatrix} \\ & [0]' \le x \le \bar{x} = [2 \quad 2 \quad 2 \quad 2] \end{cases}
$$

4.7.1.– Apliqueu dues iteracions de l'algorisme de Murtagh i Saunders a partir del punt $X_k = [1, 0, 1.75, 1.75]'$ amb la següent partició de variables: $X_B = \{x_1, x_3\}$, $X_S = \{x_4\}$ i $X_N = \{x_2\}$. Preneu com a longitud de pas òptima la longitud de pas màxima permessa.

4.7.2.– Estudieu el càracter de mínim local de **(P)** del punt obtingut a l'apartat anterior.

5 Optimització amb constriccions no lineals.

5.1 Mètodes de Lagrangians projectats

5.1.1 Procediment

5.1.1.1 Lagrangians Projectats.

Consideri's el problema estàndar d'optimització No lineal amb Constriccions No lineals **(NCN)**:

$$
(\mathbf{NCN})\left\{\begin{array}{rcl} \min & f(x) & (5.1a) \\ \text{subj. a}: & g(x)=0 & (5.1b) \\ & 0\le x\le u & (5.1c) \end{array}\right. \quad ; \quad \begin{array}{l} f\colon \mathbb{R}^n \to \mathbb{R}\,,\ f\in C^2 \\ g\colon \mathbb{R}^n \to \mathbb{R}^m\,,\ g\in C^2 \\ x,u\in \mathbb{R}^n\,,\ b\in \mathbb{R}^m \end{array}
$$

on les possibles constriccions de desigualtat de la formulació original s'han transformat a la forma estàndar amb la inclusió de les folgues pertinents. Acceptarem que totes les funcions que apareixen a la formulació de **(NCN)** són dues vegades diferenciables amb derivades continues, i que les hessianes $\nabla^2 f(x)$ i $\nabla^2 g_i(x)$, $i=1,\ldots,m$ són afitades. Considerarem també que existeix un punt x^* que satisfà les condicions de primer i segon ordre de mínim local de **(NCN)**, amb multiplicadors de Lagrange λ^* i σ^* associats, respectivament a les constriccions $g(x)=0$ i a les fites actives.

Les tècniques del tipus Lagrangià Projectat permeten resoldre problemes tipus **(NCN)** mitjançant la resolució d'una successió de subproblemas $(\mathbf{SP}^k)$ associats a la linealització $g(x)\approx g^k(x)=g(x^k)+J^k(x-x^k)$ de les restriccions no lineales (5.1b) sobre els iterats x^k. Aquests subproblemas es formulen com:

$$
(\mathbf{SP}^k)\left\{\begin{array}{rcl} & \min \Phi^k(x;\cdot) & (5.2a) \\ \text{subj. to}: & J^k x = -g(x^k)+J^k x^k & (5.2c) \\ & 0\le x\le u & (5.2e) \end{array}\right.
$$

on J^k és el Jacobià de $g(x)$ a x^k. Noti's que $(\mathbf{SP}^k)$ és un probla lineals que pot ser resolt aplicant l'algorisme A4.1. El detall de l'expresió de la funció objetiu $\Phi_k(x;\cdot)$, que està sempre relacionada d'alguna forma amb la funció Lagrangiana $\mathcal{L}(x,\lambda)=f(x)-\lambda' g(x)$, varia segons els autors. L'expressió usada al paquet Minos consisteix en el *Lagrangià Augmentat Modificat:*

$$
\Phi^k(x;x^k,\lambda^k,\rho)=f(x)-\lambda^{k\prime}(g(x)-g^k(x))+\frac{1}{2}\rho\|g(x)-g^k(x)\|_2^2 \tag{5.3}
$$

aquesta aproximació es basa en el treball de Robinson [ROB72] on es demostra pel cas $\rho = 0$ que, si x^k i λ^k són, respectivament, la solució i els multiplicadors de Lagrange de les constriccions lineals del problema $(\mathbf{SP}^{k-1})$, la seqüència $\{(x^k, \lambda^k)\}$ convergirà, sota certes circumstàncies, a una solució $\{(x^*, \lambda^*)\}$ del problema **(NCN)**, amb ordre de convergència quadràtic. Tanmateix, la convergència només està assegurada si el punt inicial (x^0, λ^0) està prou a prop de (x^*, λ^*).

L'algorisme implementat a Minos es basa en una seqüència de *iteracions principals* (*"major iterations"*), cadascuna d'elles consistent el la resolució d'un subproblema $(\mathbf{SP}^k)$ amb Φ^k donada per (5.3) . Cada iteració del procés de resolució de $(\mathbf{SP}^k)$ (iteracions de l'algorisme A4.1) es denomina *iteracions menors* (*"minor iteration"*). El terme de penalització ρ es selecciona de forma que la Hessiana de $\Phi^k(x; x^k, \lambda^k, \rho)$ sigui definida positiva sobre un cert subespai. A més, evita discrepàncies excessives entre $g(x)$ i $g^k(x)$, fent que l'òptim de cada subproblema $(\mathbf{SP}^k)$ romangui a prop de x^k si la corbatura de les constriccions és important. La selecció del valor de ρ es realitza heurísticament, incrementant-se si es detecta pèrdua de convergència, i intentant anular-lo a prop de x^* per tal d'aprofitar la convergència quadràtica del mètode de Robinson.

5.1.1.2 Criteri de convergència.

Es pot demostrar que si x^k satisfà les condicions de mínim de $(\mathbf{SP}^k)$ també satisfà les del problema original **(NCN)**, tret, possiblement de la condició de factibilitat primal $g(x^k) = 0$. Si aquesta darrera condició també es verifica, llavors x^k serà el mínim buscat. Segons això, es considera que el punt x^k és la solució del problema **(NCN)** si:

1.- x^k satisfà les constriccions no lineals $g(x) = 0$ dins d'una certa tolerància ϵ_r, és a dir, si $\epsilon_2 < \epsilon_r$.

2.- x^k satisfà les condicions de KKT de primer ordre del problema $(\mathbf{SP}^k)$.

5.1.1.3 Problemes $(\mathbf{SP}^k)$ infactibles

Si el punt x^k usat en la definició de $(\mathbf{SP}^k)$ és factible per **(NCN)** llavors $(\mathbf{SP}^k)$ serà factible. En general, però, x^k violarà en cert grau les constriccions no lineals de **(NCN)**, i això farà que $(\mathbf{SP}^k)$ pugui ser infactible. En particular, si $g(x^k) \neq 0$ llavors x^k serà infactible per $(\mathbf{SP}^k)$. Tanmateix, x^k és l'òptim de $(\mathbf{SP}^{k-1})$ i, per tant, factible $(\mathbf{SP}^{k-1})$. Tenint en compte aquest fet, es pot comprovar facilment que x^k satisfarà la següent pertorbació de les constriccions linealitzades (5.2c) :

$$J^k x = -g(x^k) + J^k x^k + \gamma q \tag{5.4}$$

amb $\gamma = 1$ i $q = g(x^k) - g^{k-1}(x^k)$. El procediment seguit per tractar amb problemes $(\mathbf{SP}^k)$ infactibles consisteix en plantejar inicialment el problema $(\mathbf{SP}^k)$ amb les constriccions pertorbades (5.4) amb $\gamma = 0$. Si es detecta que $(\mathbf{SP}^k)$ és infactible, s'afegeix $\frac{1}{2}q$ al terme independent de (5.4) i es continua amb el procés d'optimització. Si la situació d'infactibilitat es repeteix, s'afegeixen successivament els valors $\frac{1}{4}q$, $\frac{1}{8}q$, etc. Aquest procediment simula la seqüència $\gamma = \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots$ que tendeix a 1. Si aquest procediment no proporciona un subproblema $(\mathbf{SP}^k)$ factible després de 10 modificacions, o no és aplicable (per exemple, quan $k = 0$ o el problema previ era infactible), es realitza una nova linealització sobre el punt de l'última iteració menor realitzada. Altrament, es considera que el problema original **(NCN)** era infactible.

5.1.1.4 Algorisme del Lagrangià Projectat.

El procediment de resolució de **(NCN)** implementat a Minos, i que recull el desenvolupament de les seccions anteriors és:

Algorisme A5.1 : Algorisme del Lagrangià Projectat.

$\boxed{0}$ **Inicialitzacions:**

$\boxed{0.1}$ Seleccionar $(x^0, \lambda^0).k = 0$
$\boxed{0.2}$ Inicialització de toleràncies:
Tolerància de constriccions: $\epsilon_r > 0$
Radi de convergència $\epsilon_c > 0$
Tolerància de multiplicadors: $\epsilon_{ml} > 0$
Paràmetre de penalització $\rho \geq 0$
Factor de canvi de ρ: $\beta_\rho > 1$
Nombre màxim d'iteracions principals: n_P

$\boxed{1}$ **Resolució de $(\mathbf{SP}^k)$:**
$\boxed{1.3}$ Resolució del problema $(\mathbf{SP}^k)$ definit a partir de (x^k, λ^k) i ρ: obtenció de (x^{k+1}, λ^{k+1}).
$\boxed{1.4}$ Càlcul del canvi relatiu de multiplicadors de Lagrange ϵ_1 i de la violació de les constriccions ϵ_2:
$\epsilon_1 := \|\lambda^k - \lambda^{k+1}\|_2/(1 + \|\lambda^{k+1}\|_2)$
$\epsilon_2 := \|g(x^{k+1})\|_2/(1 + \|x^{k+1}\|_2)$

$\boxed{2}$ **Test de convergència:**
Si $\epsilon_2 \leq \epsilon_r$ i x^{k+1} satisfà les condicions de KKT de $(\mathbf{SP}^{k+1})$ llavors
$(x^*, \lambda^*) \equiv (x^{k+1}, \lambda^{k+1})$. **STOP**
Altrament Si $k = n_P$ llavors
STOP : nombre màxim d'iteracions excedit.
FiSi

$\boxed{3}$ **Actualització de ρ:**
Si $\epsilon_1 \leq \epsilon_c$ i $\epsilon_2 \leq \epsilon_c$ llavors
Anular el paràmetre de penalització:$\rho := 0$
Altrament Si $\epsilon_1 > \epsilon_{ml}$ llavors
Incrementar ρ : $\rho := \beta_\rho \rho$
FiSi

$\boxed{4}$ $k := k + 1$, Anada a $\boxed{1}$

Els valors adoptats per Minos per defecte d'algunes de les toleràncies són: $\epsilon_r = 10^{-6}$, $\epsilon_c = 10^{-2}$, $\rho = 100/m$.

5.1.2 Exercicis

Exercici 5.1. : plagp1/lagrangians projectats

Donada la minimització:

$$\min f(X) = \frac{x_1 x_3{}^2}{2} - x_1 x_5 + 3x_1 - 6x_2 - 4x_3 - 4x_4$$
$$\text{subj. } \widehat{C}(X) = \begin{bmatrix} \frac{x_1{}^2}{2} + x_2 - x_3 + 2x_4 x_5 - x_5 + 2 \\ \frac{3x_1{}^2}{2} + x_2 x_3 - 2x_2 + \frac{3x_4{}^2}{2} + x_5 - \frac{7}{2} \end{bmatrix} = \underline{0}$$
$$\underline{0} \leq X \leq \overline{X} \quad \text{amb} \quad \overline{X} = [4 \quad 4 \quad 4 \quad 4 \quad 4]'$$

efectueu un pas complet pel procediment de la formulació tipus "Minos" dels lagrangians projectats a partir del punt factible $X_0 = [1 \quad 0 \quad 3 \quad 1 \quad \frac{1}{2}]'$, reordenant, si cal, les variables i comproveu que la direcció de menys el gradient projectat a partir de X_0 és una direcció factible de descens.

(La funció objectiu $\Phi_{cl}(X)$, subjecta a constriccions lineals i fites, considerada per "Minos" és del tipus: $\Phi_{cl}(X) = f(X) - \Lambda_k' \left(\widehat{C}(X) - \widehat{C}_l(X)\right) + \frac{\rho}{2}\left(\widehat{C}(X) - \widehat{C}_l(X)\right)' \left(\widehat{C}(X) - \widehat{C}_l(X)\right)$, sent $\widehat{C}_l(X) = \widehat{C}(X_0) + \nabla\widehat{C}(X_0)(X - X_0)$ l'aproximació de primer ordre de $\widehat{C}(X)$ al voltant de X_0. Λ_k és una aproximació dels multiplicadors de Lagrange de les constriccions no lineals d'igualtat i ρ és un factor de penalització que suposarem fixat inicialment a ρ=10).

SOLUCIÓ

JACOBIÀ A X_0

$$\nabla\widehat{C}(X) = \begin{bmatrix} x_1 & 1 & -1 & 2x_5 & 2x_4 - 1 \\ 3x_1 & x_3 - 2 & x_2 & 3x_4 & 1 \end{bmatrix}, \quad \nabla\widehat{C}(X_0) = A_0 = \begin{bmatrix} 1 & 1 & -1 & 1 & 1 \\ 3 & 1 & 0 & 3 & 1 \end{bmatrix}$$

DETERMINACIÓ DE LES VARIABLES BÀSIQUES, SUPERBÀSIQUES I NO BÀSIQUES

Les variables bàsiques $\underline{0} \le X_B \le \overline{X}_B$, a ser possible estrictament entre fites, han de tenir imprescindiblement columnes d'A linealment independents. Les variables superbàsiques $\underline{0} < X_S < \overline{X}_S$ han d'estar estrictament entre fites. Les variables no bàsiques X_N han de tenir les seves components o a la seva fita superior: $x_{Nj} = \overline{x}_j$, $j \in \overline{I}$ o a la seva fita inferior: $x_{Nj} = 0$, $j \in \underline{I}$, sent $\overline{I}$ i $\underline{I}$ els conjunt d'índexos de variables no bàsiques a fita superior i a fita inferior respectivament.

Per triar les variables bàsiques tenim que:

- x_2 està a la fita inferior (per tant, en principi, la excloem del conjunt de les bàsiques)
- x_1, i x_3 poden ser les variables bàsiques X_B perquè tenen les columnes linealment independents:
$$B = \begin{bmatrix} 1 & -1 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 1 & \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ & 3 \end{bmatrix}$$
- donat que hem trobat variables bàsiques estrictament entre fites, la variable x_2 , la qual està a una fita serà no bàsica: $X_N = \{x_2\}$, i la resta de variables (estrictament entre fites) seran les superbàsiques $X_S = \{x_4, x_5\}$. (Altres combinacions de bàsiques i superbàsiques són també possibles).

MATRIUS S, N I $\widehat{A}$

S és doncs $S = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}$ i $N = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ d'on la matriu de constriccions actives $\widehat{A}$ és:

$$\widehat{A} = \left[\begin{array}{cc|c} B & S & N \\ \hline \multicolumn{2}{c|}{\mathbf{0}} & \tilde{I} \end{array}\right] = \begin{bmatrix} 1 & -1 & 1 & 1 & 1 \\ 3 & 0 & 3 & 1 & 1 \\ & & \mathbf{0} & & 1 \end{bmatrix}$$

FUNCIÓ OBJECTIU $\Phi_{cl}(X)$ PER A CONSTRICCIONS LINEALITZADES

El subproblema a resoldre és:

$$\begin{aligned} \min \quad & \Phi_{cl}(X) \\ \text{subj. a :} \quad & \nabla\widehat{C}(X_0)X + \left(\widehat{C}(X_0) - \nabla\widehat{C}(X_0)X_0\right) = 0 \\ & \underline{0} \le X \le \overline{X} \end{aligned}$$

amb:

$$\Phi_{cl}(X) = f(X) - \Lambda'_k\left(\widehat{C}(X) - \widehat{C}_l(X)\right) + \frac{\rho}{2}\left(\widehat{C}(X) - \widehat{C}_l(X)\right)'\left(\widehat{C}(X) - \widehat{C}_l(X)\right)$$

Cal doncs primer estimar Λ_0. Podriem fer-ho fent els mínims quadrats $\begin{bmatrix} B'_0 \\ S'_0 \end{bmatrix}\Lambda_0 = \begin{bmatrix} G_{B0} \\ G_{S0} \end{bmatrix}$, però també podem obtenir una primera aproximació resolent $B'\Lambda_0 = [\nabla f(X_0)']_B = G_{B0}$. Determinem primer $\nabla f(X_0)'$ i G_{B0} (sabent que $X_B = \{x_1, x_3\}$).

$$\nabla f(X)' = \begin{bmatrix} \frac{x_3^2}{2} - x_5 + 3 \\ -6 \\ x_1x_3 - 4 \\ -4 \\ -x_1 \end{bmatrix},\ \nabla f(X_0)' = \begin{bmatrix} 7 \\ -6 \\ -1 \\ -4 \\ -1 \end{bmatrix} \text{ i } G_{B0} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \\ -1 & 3 \end{bmatrix}\cdot\begin{bmatrix} 1 & 3 \\ & 1 \end{bmatrix}\cdot\begin{bmatrix} \lambda_{01} \\ \lambda_{02} \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \end{bmatrix} \quad\longrightarrow\quad \Lambda_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

L'expressió de la linealització de $\widehat{C}(X)$ sobre X_0 és :

$$\widehat{C}_l(X) = \begin{bmatrix} \hat{c}_{l_1}(X) \\ \hat{c}_{l_2}(X) \end{bmatrix} = \widehat{C}(X_0) + \nabla\widehat{C}(X_0)(X - X_0) = \widehat{C}(X_0) + \nabla\widehat{C}(X_0)X - \nabla\widehat{C}(X_0)X_0 =$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 & -1 & 1 & 1 \\ 3 & 1 & 0 & 3 & 1 \end{bmatrix}X + \begin{bmatrix} 1 & 1 & -1 & 1 & 1 \\ 3 & 1 & 0 & 3 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 3 \\ 1 \\ 1/2 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \begin{bmatrix} \hat{c}_{l_1}(X) \\ \hat{c}_{l_2}(X) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 - x_3 + x_4 + x_5 + 1/2 \\ 3x_1 + x_2 + 3x_4 + x_5 - 6,5 \end{bmatrix}$$

Substituint-hi els valors de Λ_0 i ρ, i anomenant $\hat{c}_1(X) = \frac{x_1^2}{2} + x_2 - x_3 + 2x_4x_5 - x_5 + 2$ i $\hat{c}_2(X) = \frac{3x_1^2}{2} + x_2x_3 - 2x_2 + \frac{3x_4^2}{2} + x_5 - \frac{7}{2}$, La funció a minimitzar és doncs:

$$\Phi_{cl}(X) = \frac{x_1x_3^2}{2} - x_1x_5 + 3x_1 - 6x_2 - 4x_3 - 4x_4 - (\hat{c}_1(X) - \hat{c}_{l_1}(X)) - 2(\hat{c}_2(X) - \hat{c}_{l_2}(X)) +$$
$$+ \frac{\rho}{2}\left[(\hat{c}_1(X) - \hat{c}_{l_1}(X))^2 + (\hat{c}_2(X) - \hat{c}_{l_2}(X))^2\right]$$

GRADIENT DE LA FUNCIÓ OBJECTIU $\Phi_{cl}(X)$

El gradient de la funció $\Phi_{cl}(X)$ ve donat per l'expressió:

$$\nabla\Phi_{cl}(X) = \nabla f(X) - \Lambda'_0\left(\nabla\widehat{C}(X) - \nabla\widehat{C}(X_0)\right) + \rho\left(\widehat{C}(X) - \widehat{C}_l(X)\right)'\left(\nabla\widehat{C}(X) - \nabla\widehat{C}(X_0)\right)$$

que, sobre l'iterat actual X_0 coincideix amb el gradient de la funció objectiu:

$$\nabla\Phi_{cl}(X_0) = \nabla f(X_0) = [7 \quad -6 \quad -1 \quad -4 \quad -1]$$

Aquesta coinicidència es produirà sempre a la primera iteració després de linealitzar les constriccions (inici d'una "Major Iteration"). Tenint en compte l'ordenació i distribució de variables bàsiques $X_B=\{x_1, x_3\}$, superbàsiques $X_S=\{x_4, x_5\}$ i no bàsiques $X_N=\{x_2\}$, tenim: $G_{Bcl} = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$ $G_{Scl} = \begin{bmatrix} -4 \\ -1 \end{bmatrix}$ $G_{Ncl} = [-6]$ i que la solució de $B'\Pi = G_{Bcl}$ és $\Pi=\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, que coincideix amb Λ_0. El gradient reduït és doncs: $G_z = Z'G = G_S - S'\Pi = \begin{bmatrix} -11 \\ -4 \end{bmatrix}$

DIRECCIÓ DE MENYS EL GRADIENT PROJECTAT

Donat que $||Z'G||^2=137$ no està a prop de zero, haurem de fer una passa en la direcció $P_S=P_z=-Z'G$ (o en la que s'obtindria de resoldre $Z'H_kZP_z=-Z'G$) i calcular $P_B=-B^{-1}SP_S$, és a dir, resoldre $BP_B = -SP_S$ i fer $P_N=\underline{0}$.

Calculant $SP_s = \begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 11 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 37 \end{bmatrix}$

i resolent $\begin{bmatrix} 1 & \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ & 3 \end{bmatrix} \cdot \begin{bmatrix} p_{B1} \\ p_{B2} \end{bmatrix} = \begin{bmatrix} -15 \\ -37 \end{bmatrix}$ obtenim $P_B = \begin{bmatrix} -37/3 \\ 8/3 \end{bmatrix}$

La derivada direccional $\nabla\Phi_{cl}(X_0)P=G'_{Bcl}P_B+G'_{Scl}P_S=-137<0$ per tant queda comprovat que és una direcció de descens.

Exercici 5.2. : plagp2/Lagrangians Projectats

Considereu el següent problema d'optimització no lineal:

$$(\mathbf{P}) \quad \begin{cases} \min\limits_{x\in\Re^4} & f(x) = \dfrac{e^{x_1x_2}}{1+x_3+x_4} \\ \text{subj. a :} & x_1^2 + 3x_1x_2 - 7 = 0 \\ & -3x_2 + x_3 + x_4 + 4 = 0 \\ & [0]' \le x \le u = [2 \quad 2 \quad 2 \quad 2] \end{cases}$$

aplicant l'algorisme del Lagrangià Projectat a aquest problema s'ha arribat a la iteració k-éssima amb el parell (x^k, λ^k) de valor $x^k = [1 \quad 2 \quad 1 \quad 1]$, $\lambda^k = [1/2 \quad -1]$ amb un factor de penalització $\rho = 0.01$.

5.2.1.– Formuleu el subproblema linealitzat $(\mathbf{SP}^k)$ associat a x^k, sense desenvolupar amb detall la funció objectiu $\Phi^k(x; x^k, \lambda^k, \rho)$.

5.2.2.– Efectueu una iteració de l'algorisme del Lagrangià Projectat a partir de (x^k, λ^k) obtenint el nou parell (x^{k+1}, λ^{k+1}) a partir de la solució de $(\mathbf{SP}^k)$.

5.2.3.– Comproveu si el punt obtingut a l'apartat anterior satisfà les condicions suficients de segon ordre de mínim local de $(\mathbf{P})$.

Exercici 5.3. : plagp3/Lagrangians Projectats

Considereu el següent problema d'optimització no lineal:

$$(\mathbf{P}) \quad \begin{cases} \min\limits_{x\in\Re^4} & f(x) = (x_1 - x_4)^2 - x_2x_3x_4 \\ \text{subj. a :} & g_1(x) = x_1^2 - x_2x_3 = 0 \\ & g_2(x) = x_4^2 - x_1x_2 = 0 \\ & [0] \le x \le u = [2 \quad 2 \quad 2 \quad 2]' \end{cases}$$

aplicant l'algorisme del Lagrangià Projectat segons la formulació tipus Minos a aquest problema s'ha arribat a la iteració k-éssima amb el parell (x^k, λ^k) de valor $x^k = [1 \quad 1 \quad 1 \quad 1]'$, $\lambda^k = [1/3 \quad 2/3]'$ amb un factor de penalització $\rho = 2$.

5.3.1.– Formuleu el subproblema linealitzat $(\mathbf{SP}^k)$ associat a (x^k, λ^k), indicant l'expressió de la funció objectiu $\Phi^k(x; x^k, \lambda^k, \rho)$.

5.3.2.– En el procés de resolució del subproblema linealitzat $(\mathbf{SP}^k)$ s'ha arribat al punt $x_c = [1, 2, 0, 3/2]'$. Calculeu la direcció de moviment que l'algorisme prendria a partir del punt x_c. Recordeu que $\nabla\Phi^k(x; x^k, \lambda^k, \rho) = \nabla f(x) - \lambda^{k'}\left(\nabla g(x) - \nabla g(x^k)\right) + \rho\left(g(x) - g^k(x)\right)'\left(\nabla g(x) - \nabla g(x^k)\right)$, on $g^k(x)$ és la linealització de les constriccions $g(x)$ sobre x^k.

Bibliografia.

[BERT95] Bertsekas, D. *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts. 1995.

[DESCH83] Dennis Jr., J.E. and R.B. Schnabel *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.*Prentice-Hall, Inc., Englewood Cliffs, NJ 07632, USA. 1983.

[GILL74] Gill, P.E. & W. Murray. *Newton–type mrthods for unconstrained and linearly constrained optimization. Mathematical Programming*, vol. 28, pp. 311-350. 1974

[GILL81] Gill, P.E., W. Murray, and Wright, M.H., *Practical Optimization*, Academic Press, London, UK, 1981.

[LUEN84] Luenberger, D.G. *Linear and Nonlinear Programming.* Addison-Wesley Publ. Co, Reading, MA., USA. 1984.

[MUSA78] Murtagh, B.A., and Saunders, M.A., "Large-scale linearly constrained optimization", *Mathematical Programming*, 14 (1978) 41–72.

[MUSA82] Murtagh, B.A., and Saunders, M.A., "A projected lagrangian algorithm and its implementaton for sparse nonlinear constraints", *Mathematical Programming Study*, 16 (1982) 84–117.

[ROB72] Robinson, S. M., "A quadratically convergent algorithm for general nonlinear programming problems", *Mathematical Programming*, 7 (1972) 145–156.

EST APRIL