Curs d'Investigació Operativa

Aureli Alabert

Departament de Matemàtiques
Universitat Autònoma de Barcelona

This document was originally conceived for internal use in the courses “Operations Research I” and “Optimization” for the Statistics degree in the Universitat Autònoma de Barcelona.

Released to public domain by the author, February 2004.

Índex

0. Introducció

Què vol dir "Investigació Operativa"? Una definició temptativa podria ser la següent:

0.0.1 Definició temptativa

La Investigació Operativa és un conjunt de *tècniques variades* que pretenen facilitar la *presa de decisions.* □

Evidentment, és una definició molt vaga. Cap al final del curs podrem tenir una idea més específica del que vol dir, un cop vistos molts exemples concrets. De moment, ens conformem amb precisar una mica les expressions subratllades:

- "Presa de decisions": Ens indica que s'analitzaran situacions en què és possible optar per diferents alternatives, però es busca la millor segons un criteri prefixat.
- "Tècniques variades": Hi ha molts tipus de problemes de "presa de decisions", i això dóna lloc a tècniques molt diverses. Enumerar-les o classificar-les és llarg i difícil.

L'esquema general de resolució d'un problema real mitjançant models (un model és una representació simplificada d'un problema real) és el següent:

$$\begin{tikzcd}
\boxed{\text{Problema real}} \arrow[d, "\text{1. Construcció del model}"] \\
\text{Model} \arrow[d, "\text{2. Aplicació de tècniques adequades al model}"] \\
\text{Resolució del model} \arrow[d, "\text{3. Interpretació de la solució del model}"] \\
\text{Solució del problema real}
\end{tikzcd}$$

El pas difícil (en general, i en la Investigació Operativa en particular) és el 1. Un dels objectius principals del curs és aprendre a fer aquest pas, que és essencialment creatiu, en contraposició amb el 2, que és més rutinari.

Comencem amb un exemple senzill de problema de investigació operativa.

0.0.2 Exemple

Una empresa fabrica dos tipus de llumins de fusta: Llumins llargs i llumins curts. Els llumins es venen embalats en caixes de cartró. Per cada caixa de llumins llargs, l'empresa fa un benefici de 30 euros, mentre que per cada caixa de llumins curts, el benefici és de 20 euros.

Es disposa d'una sola màquina, que pot produir tant llumins llargs com llumins curts. La màquina pot produir en un any un màxim de 9 000 caixes de llumins, entre llargs i curts.

La matèria primera que es necessita per la fabricació és fusta i caixes de cartró. Per cada caixa de llumins llargs són necessaris 3 metres cúbics de fusta. Per cada caixa de llumins curts, 1 metre cúbic de fusta. Pel proper any, l'empresa podrà disposar d'un total de 18 000 metres cúbics de fusta. Quant a les caixes de cartró, es disposarà de 7 000 caixes per a llumins llargs, i 6 000 caixes per a llumins curts.

Se suposa que es podrà vendre la totalitat de la producció.

Quina és l'estratègia òptima de producció? És a dir: Quantes caixes de cada tipus de llumí cal produir per tal que el benefici de l'empresa sigui màxim? □

La construcció d'un model per aquest problema seria com segueix: Hi ha diversos elements a identificar, que anomenarem *variables*, *objectiu* i *restriccions*.

- **Variables:**

  x_1: Quantitat de caixes de llumins llargs produïdes (en milers).
  x_2: Quantitat de caixes de llumins curts produïdes (en milers).

- **Objectiu:**
  Fer màxim el benefici $z = 3x_1 + 2x_2$ desenes de milers de euros.

- **Restriccions:**

  1) Capacitat de la màquina:

  $$x_1 + x_2 \leqslant 9 \text{ milers de caixes.}$$

  2) Disponibilitat de fusta:

  $$3x_1 + x_2 \leqslant 18 \text{ milers de m}^3 \text{ de fusta.}$$

  3) Disponibilitat de caixes de cartró:

  $$x_1 \leqslant 7 \text{ milers de caixes.}$$
  $$x_2 \leqslant 6 \text{ milers de caixes.}$$

  4) Positivitat:

  $$x_1 \geqslant 0\ .$$
  $$x_2 \geqslant 0\ .$$

En resum, el model resultant és el problema matemàtic que s'expressa com:

$$\begin{array}{l} \max\ z = 3x_1 + 2x_2 \\ \text{subj. a:} \\ \left.\begin{array}{r} x_1 + x_2 \leqslant 9 \\ 3x_1 + x_2 \leqslant 18 \\ x_1 \qquad \leqslant 7 \\ x_2 \leqslant 6 \\ x_1 \geqslant 0, \quad x_2 \geqslant 0 \end{array}\right\} \end{array}$$

La resolució del model es pot fer gràficament, segons es mostra a les figures següents. Les restriccions de positivitat determinen que la solució ha d'estar en el primer quadrant (Figura 1); les altres restriccions delimiten semiplans (regions a l'esquerra de les rectes (1) a (4)), i s'obté la "regió factible" de la Figura 2; finalment, dibuixem unes quantes "corbes de nivell de z", o sigui el feix de rectes paral·leles que resulten de donar diferents valors a "la variable z", i observem que augmentant z la recta es mou en la direcció de les fletxes (Figura 3). Això vol dir que el "valor òptim" (el màxim de z dins la regió factible) s'assoleix en la intersecció de les rectes (1) i (2); per tant, en el punt $x_1 = 4.5$, $x_2 = 4.5$.

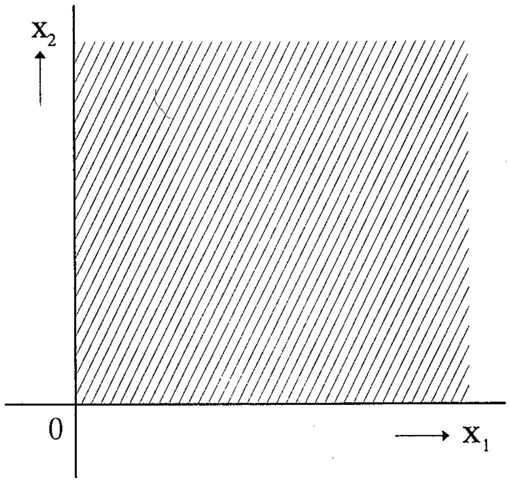


Figura 1

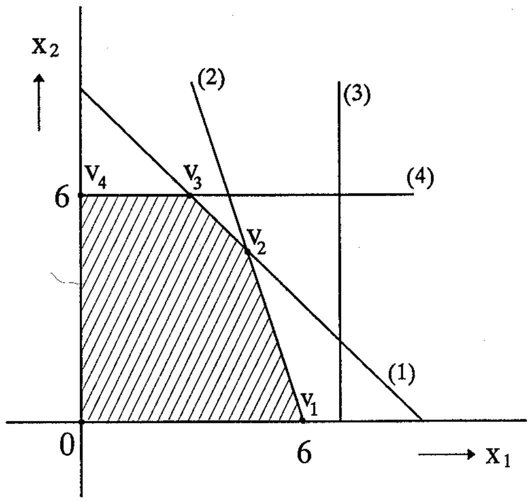


Figura 2

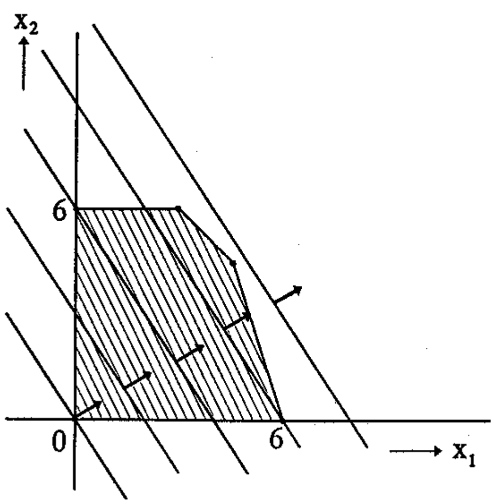


Figura 3

Observeu que canviant les unitats en una restricció, res no canvia. I si canviem les unitats en la "funció objectiu" z, el valor òptim canvia en consonància, però el punt on s'assoleix aquest valor òptim és el mateix.

Aquest problema és fàcil i es pot resoldre gràficament. Però un es pot preguntar:

- I si la funció objectiu no fos lineal? (I per tant les corbes de nivell de z no fossin rectes).
- I si les restriccions no fossin lineals?
- I si hi haguessin més variables?

1. Programació lineal

1.1 Exemples

1.1.1 Exemple

Uns alts forns tenen l'encàrrec de fer 1000 Kg d'un aliatge de ferro contenint almenys un 0.45% de manganès i entre un 3.25% i un 5.50% de silici.

Es disposa de 3 tipus de ferro en brut A, B, C, en quantitats essencialment il·limitades, amb les propietats següents:

	A	B	C
Silici	4%	1%	0.6%
Manganès	0.45%	0.5%	0.4%

A més, el procés de producció és tal que es pot afegir directament manganès pur a la barreja.

Els costos són: 0.26 euros/Kg de A, 0.30 euros/Kg de B, 0.20 euros/Kg de C, i 0.08 euros/gram de manganès.

Com cal barrejar els quatre productes per tal que el cost sigui mínim? □

Anem a construir un model per aquest problema:

- **Variables:**
  x_1: Kg de ferro A usats.
  x_2: Kg de ferro B usats.
  x_3: Kg de ferro C usats.
  x_4: Grams de manganès usats.
- **Objectiu:**
  Fer mínim el cost total, que és la suma de

  0.26 euros per Kg de A usat,
  0.30 euros per Kg de B usat,
  0.20 euros per Kg de C usat,
  0.08 euros per gram de manganès usat,

  o sigui
  $$z = 0.26x_1 + 0.30x_2 + 0.20x_3 + 0.08x_4 \ .$$
- **Restriccions:**

1) Produir 1000Kg:
$$x_1 + x_2 + x_3 + 0.001x_4 = 1000$$
(estem suposant aquí que la massa del producte final és igual a la massa total de matèria primera emprada).
2) El % mínim de Mn és 0.45%; per tant, la quantitat mínima de Mn en 1000Kg és 4.5Kg=4500g. Obtenim
$$4.5x_1 + 5x_2 + 4x_3 + x_4 \geqslant 4500 \text{ grams} .$$
Els coeficients de x_1, x_2 i x_3 venen de la quantitat en grams de manganès en cadascun dels ferros en brut.
3) El % mínim de Si és 3.25%; per tant, la quantitat mínima de Si en 1000Kg és 32.5Kg. Obtenim
$$0.04x_1 + 0.01x_2 + 0.006x_3 \geqslant 32.5 \text{ Kg} .$$
4) El % màxim de Si és 5.50%; per tant, la quantitat màxima de Si en 1000Kg és 55Kg. Obtenim
$$0.04x_1 + 0.01x_2 + 0.006x_3 \leqslant 55 \text{ Kg} .$$
5) Positivitat:
$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0,\ x_4 \geqslant 0.$$

El model matemàtic resultant és doncs:

$$\min\ z = 26x_1 + 30x_2 + 20x_3 + 8x_4$$
$$\text{subj. a:} \quad \left.\begin{array}{rcl} x_1 + x_2 + x_3 + 0.001x_4 &=& 1000 \\ 4.5x_1 + 5x_2 + 4x_3 + x_4 &\geqslant& 4500 \\ 0.04x_1 + 0.01x_2 + 0.006x_3 &\geqslant& 32.5 \\ 0.04x_1 + 0.01x_2 + 0.006x_3 &\leqslant& 55 \\ x_1, x_2, x_3, x_4 \geqslant 0 && \end{array}\right\}$$

1.1.2 Exemple

L'encarregat de cartera d'un banc disposa de 10 milions d'euros per invertir en diversos tipus de bons emesos per diferents administracions.

Concretament, els bons a l'abast junt amb llurs dades són els de la taula següent:

Nom	Emissor	Qualificació	Anys venciment	Rendiment
A	Municipi	2	9	4.3%
B	Autonomia	2	15	2.7%
C	Estat	1	4	2.5%
D	Estat	1	3	2.2%
E	Municipi	5	2	4.5%

El banc posa les limitacions següents a les accions de l'encarregat:

1. Els bons estatals i autonòmics han de totalitzar almenys 4 milions d'euros.
2. La qualitat mitjana de la cartera no pot excedir 1.4 en l'escala de qualificacions del banc (nombres més baixos indiquen més qualitat).
3. Els anys de venciment mitjà de la cartera no ha de superar 5 anys.

Com cal fer les inversions per fer màxim el rendiment? □

Model per aquest problema:

- **Variables:**

  A: Milions d'euros invertits en el valor A.
  B: Milions d'euros invertits en el valor B.
  C: Milions d'euros invertits en el valor C.
  D: Milions d'euros invertits en el valor D.
  E: Milions d'euros invertits en el valor E.

- **Objectiu:**
  Fer màxim
  $$z = 0.043A + 0.027B + 0.025C + 0.022D + 0.045E \ .$$

- **Restriccions:**

  1) Diners disponibles:
  $$A + B + C + D + E \leqslant 10 \ .$$

  2) Limitació 1 del banc:
  $$B + C + D \geqslant 4 \ .$$

  3) Limitació 2 del banc:
  $$\frac{2A + 2B + C + D + 5E}{A + B + C + D + E} \leqslant 1.4$$
  o el que és el mateix
  $$0.6A + 0.6B - 0.4C - 0.4D + 3.6E \leqslant 0 \ .$$

  4) Limitació 3 del banc:
  $$\frac{9A + 15B + 4C + 3D + 2E}{A + B + C + D + E} \leqslant 5$$
  o el que és el mateix
  $$4A + 10B - C - 2D - 3E \leqslant 0 \ .$$

  4) Positivitat:
  $$A \geqslant 0,\ B \geqslant 0,\ C \geqslant 0,\ D \geqslant 0,\ E \geqslant 0.$$

Hem obtingut el problema matemàtic següent:

$$\max\ z = 0.043A + 0.027B + 0.025C + 0.022D + 0.045E$$

subj. a:

$$\left.\begin{array}{rrrrrcr} A+ & B+ & C+ & D+ & E & \leqslant & 10 \\ & B+ & C+ & D+ & & \geqslant & 4 \\ 0.6A+ & 0.6B- & 0.4C- & 0.4D+ & 3.6E & \leqslant & 0 \\ 4A+ & 10B- & C- & 2D- & 3E & \leqslant & 0 \\ & & \multicolumn{3}{r}{A,B,C,D,E} & \geqslant & 0 \end{array}\right\}$$

1.2 Definicions

1.2.1 Definició

Un problema de *Programació Matemàtica* (PM) és el que es pot enunciar com:

$$
\begin{aligned}
&\min \ f(x) \\
&\text{subj. a:} \\
&\left.\begin{aligned} h_i(x) = 0, &\quad i = 1, 2, \ldots, p \\ g_i(x) \leqslant 0, &\quad i = p+1, \ldots, m \end{aligned}\right\}
\end{aligned}
$$

on

$$
\begin{aligned}
&x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ , \\
&f \colon \mathbb{R}^n \to \mathbb{R} \\
&h_i \colon \mathbb{R}^n \to \mathbb{R} \ , \quad i = 1, \ldots, p \ , \\
&g_i \colon \mathbb{R}^n \to \mathbb{R} \ , \quad i = p+1, \ldots, m \ ,
\end{aligned}
$$

essent m, n, p nombres naturals (inclòs el 0).

$x_1, \ldots, x_n$ són les *variables*, f s'anomena *funció objectiu* i les igualtats i desigualtats són les *restriccions*. $\square$

1.2.2 Definició

Donat un problema de PM, s'anomena *punt factible* (o *solució factible*, o *programa*) a tot $x \in \mathbb{R}^n$ que compleixi totes les restriccions. El conjunt de punts factibles és la *regió factible* del problema.

Si denotem per S la regió factible, un punt $x^* \in S$ és un *punt optimal* (o *solució optimal*, o *programa optimal*) si

$$f(x^*) = \min \ \{f(x) : \ x \in S\} \ .$$

El nombre $f(x^*)$ és el *valor òptim* de la funció objectiu. $\square$

1.2.3 Observació

En lloc de 'min' (minimitzar, trobar el mínim) pot ser 'max' (maximitzar, trobar el màxim), i també podrien haver restriccions del tipus $g_i(x) \geqslant 0$; però:

- $\max\{f(x) : \ x \in S\} = -\min \ \{-f(x) : \ x \in S\}$ i $\max f$ i $\min \ -f$ s'assoleixen en els mateixos punts.
- $g_i(x) \geqslant 0$ equival a $-g_i(x) \leqslant 0$.

Per tant, sempre podem escriure-ho exactament com a la definició. $\square$

Els problemes dels Exemples 1.1.1 i 1.1.2 els hem modelats com a problemes de PM. Però de fet són del tipus particular que veurem a la Definició ■.

1.2.4 Definició

Un problema de *Programació Lineal* (PL) és un problema de PM en què les funcions f, h_i, g_i són totes afins. És a dir, és el que es pot expressar com:

$$
\begin{aligned}
&\min \ z = \sum_{j=1}^{n} c_j x_j + c_0 \\
&\text{subj. a:} \\
&\left.\begin{aligned} \sum_{j=1}^{n} a_{ij} x_j = d_i, &\quad i = 1, \ldots, p \\ \sum_{j=1}^{n} a_{ij} x_j \leqslant d_i, &\quad i = p+1, \ldots, m \end{aligned}\right\}
\end{aligned}
$$

on les constants c_j, a_{ij}, d_i s'anomenen coeficients. $\square$

1.2.5 Definició

Un problema de PL està en *forma estàndard* si s'expressa com:

$$\begin{aligned} &\min\ z = \sum_{j=1}^{n} c_j x_j + c_0 \\ &\text{subj. a:} \\ &\left.\begin{aligned} \sum_{j=1}^{n} a_{ij} x_j = d_i, &\quad i = 1, \dots, m \\ x_j \geqslant 0, &\quad j = 1, \dots, n \end{aligned}\right\} , \end{aligned}$$

on la matriu

$$A := \begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{mn} & \dots & \dots & a_{mm} \end{pmatrix}$$

té rang màxim m (amb $m < n$), i el vector $d = (d_1, \dots, d_m)$ té totes les components positives.

És a dir: Les úniques desigualtats són les restriccions de positivitat de les variables; no hi ha equacions redundants; hi ha menys equacions que incògnites. $\square$

1.2.6 Notació

Si posem, a més, $x := (x_1, \dots, x_n)$, $c := (c_1, \dots, c_n)$, podem expressar el problema abreujadament

$$\begin{aligned} &\min\ z(x) = c \cdot x + c_0 \\ &\text{subj. a:} \\ &\left.\begin{aligned} Ax = d \\ x \geqslant 0 \end{aligned}\right\} \quad \square \end{aligned}$$

Tot problema de PL es pot "passar" a forma estàndard. El procediment és el següent:

1. **Eliminació de variables lliures**
   x_j és una *variable lliure* si la restricció "$x_j \geqslant 0$" no hi és. Per desfer-se d'una variable lliure hi ha dues possibilitats:

   1- Si x_j apareix en alguna restricció d'igualtat, l'aïllem i substituïm a totes les altres restriccions i a la funció objectiu. Un cop resolt el problema en forma estàndard, recuperem el valor de x_j mitjançant la restricció que hem usat per eliminar-la.

   2- En tot cas, sempre es pot posar $x_j = x_j^1 - x_j^2$ a tot arreu on aparegui x_j, i afegir les restriccions $x_j^1 \geqslant 0$ i $x_j^2 \geqslant 0$. (Tot nombre real es pot posar com a diferència de nombres positius.) Després, un cop resolt el problema en forma estàndard, obtenim x_j restant $x_j^1 - x_j^2$.

2. **Conversió de les desigualtats en igualtats**
   Afegim una nova variable a cada restricció del tipus

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant d_i\ , \quad i = p+1, \dots, m\ ,$$

obtenint

$$\sum_{j=1}^{n} a_{p+1,j} x_j + x_{n+1} = d_{p+1}$$
$$\sum_{j=1}^{n} a_{p+2,j} x_j + x_{n+2} = d_{p+2}$$
$$\vdots$$
$$\sum_{j=1}^{n} a_{m,j} x_j + x_{n+m-p} = d_m$$

i afegim també $x_{n+1} \geqslant 0, \dots, x_{n+m-p} \geqslant 0$.
Les noves variables $x_{n+1}, \dots, x_{n+m-p}$ s'anomenen *variables de folga* o *variables de separació*, i tenen una interpretació fàcil: Donat un punt factible del problema original, les variables de folga mesuren fins a quin punt una restricció es compleix folgadament o no en aquest punt. Un cop obtinguda la solució optimal del problema en forma estàndard, només cal oblidar les variables de folga i el que ens queda és una solució optimal del problema original.

3. **Positivització del vector de termes independents**
   Només cal multiplicar per -1 cada equació que ho necessiti. Òbviament, res no canvia.
4. **Eliminació d'equacions redundants**
   Simplement se suprimeixen. Res no canvia.

Les definicions següents es refereixen només a problemes de PL en forma estàndard.

1.2.7 Definicions

Sigui

$$\min \ z = c \cdot x + c_0$$
$$\text{subj. a:} \quad \left.\begin{aligned} Ax &= d \\ x &\geqslant 0 \end{aligned}\right\} ,$$

amb $x \in \mathbb{R}^n$ i A una matriu $m \times n$, un problema de PL en forma estàndard. S'anomena *base* a tot conjunt de m columnes de A que constitueixin un menor de rang m. En altres paraules, una base és una submatriu B de A, $m \times m$, invertible.

Sabem que existeix sempre una base per la hipòtesi rang $A = m$. Reordenant si cal les columnes de A (és a dir, reordenant les variables del problema), el sistema d'equacions $Ax = d$ es pot expressar com

$$\big(B \,\big|\, N \big) x = d \ ,$$

on B és $m \times m$ (la base) i N és $m \times (n - m)$ (la resta de les columnes de A).

S'anomenen *variables bàsiques* les associades a les columnes de B, i *variables no-bàsiques* a les demés.

Si denotem per x_B el vector de variables bàsiques i per x_N el de variables no-bàsiques, podem escriure el sistema així:

$$B x_B + N x_N = d \ ,$$

o encara, multiplicant per B^{-1},

$$I x_B + B^{-1} N x_N = B^{-1} d \ ,$$

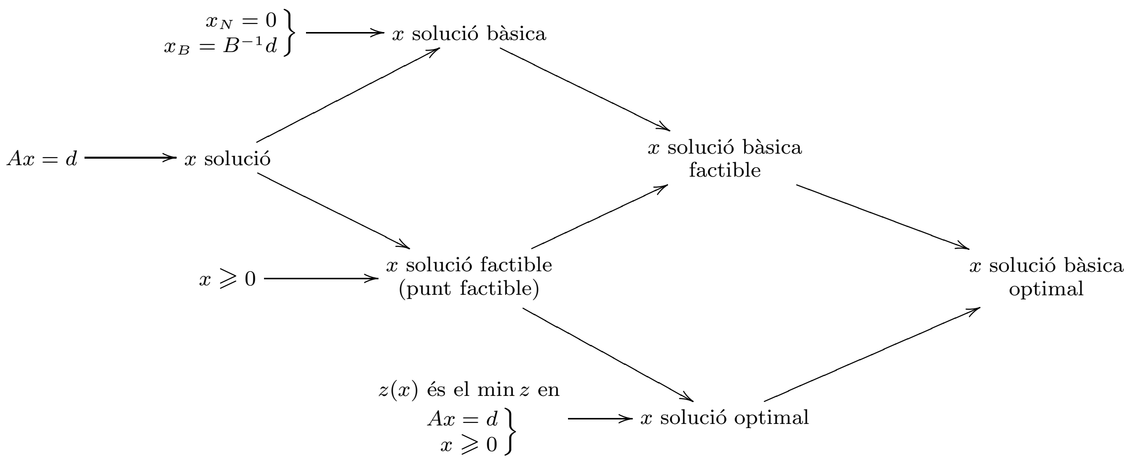
on I representa la matriu identitat $m \times m$.

S'anomena *solució bàsica* associada a B a la que resulta de donar valor zero a totes les variables no-bàsiques. És a dir, és el punt

$$\left.\begin{aligned} x_B &= B^{-1} d \\ x_N &= 0 \end{aligned}\right\} .$$

La paraula 'solució' vol dir aquí simplement 'solució del sistema $Ax = d$'. Pot ser un punt que no compleixi '$x \geqslant 0$'. Si compleix a més les restriccions de positivitat, aleshores és una *solució bàsica factible*. $\square$

El diagrama següent pot ajudar a entendre totes aquestes definicions.



1.3 Mètode del Símplex. Introducció

Es tracta d'estudiar un mètode sistemàtic per resoldre qualsevol problema de PL en forma estàndard (i per tant, segons hem vist, per resoldre qualsevol problema de PL, en la forma que sigui, una vegada convertit a forma estàndard).

1.3.1 Exemple 1

Considerem el problema de PL en forma estàndard següent:

$$
\begin{aligned}
&\min \;\; z = 0x_1 + 0x_2 + 3x_3 + x_4 + 20 \\
&\text{subj. a:} \\
&\left.\begin{aligned}
x_1 \quad - 3x_3 + 3x_4 &= 6 \\
x_2 - 8x_3 + 4x_4 &= 4 \\
x_j \geqslant 0,\ j = 1,2,3,4 &
\end{aligned}\right\}
\end{aligned}
$$

Apart d'estar en forma estàndard, té dues característiques especials:

(C1) Hi ha una submatriu identitat dins A (primera i segona columnes).

(C2) Les variables d'aquesta submatriu tenen coeficient 0 en la funció objectiu.

El sistema $Ax = d$ es pot expressar

$$
\begin{pmatrix} 1 & 0 & -3 & 3 \\ 0 & 1 & -8 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}
$$

o també

$$
\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -3 & 3 \\ -8 & 4 \end{pmatrix} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}
$$

Les variables x_1, x_2 són les variables bàsiques respecte la matriu $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

La solució bàsica associada és

$$
x_1 = 6,\ x_2 = 4,\ x_3 = 0,\ x_4 = 0,
$$

amb valor de la funció objectiu

$$z = 20 \ .$$

En la situació 'forma estàndard + (C1)+(C2)', la solució bàsica és automàticament factible, ja que $d \geqslant 0$ i estem fent $x_1 = d_1$, $x_2 = d_2$; per tant es compleixen les restriccions de positivitat.

Ara ens preguntem: És optimal aquesta solució? O bé hi ha altres punts factibles amb $z < 20$?

Observem que:

1) En aquesta solució, $x_3 = 0$ i $x_4 = 0$. En qualsevol altre solució factible, $x_3 \geqslant 0$ i $x_4 \geqslant 0$, per les restriccions de positivitat.
2) Els coeficients de x_3 i x_4 en z són positius.
3) Els valors de x_1 i x_2 no afecten z.

D'aquestes observacions deduïm fàcilment que si canviem x_3 i/o x_4, llavors z augmentarà. Conclusió: La solució $x_1 = 6$, $x_2 = 4$, $x_3 = 0$, $x_4 = 0$ és optimal i el problema s'ha acabat. □

De l'Exemple 1.3.1 es desprèn el següent criteri d'optimalitat.

1.3.2 Criteri d'optimalitat

Si en un problema de PL en forma stàndard complint les condicions (C1) i (C2) tota variable no-bàsica té un coeficient positiu a la funció objectiu, aleshores la solució bàsica associada és optimal. □

1.3.3 Exemple 2

$$\begin{array}{l} \min \ z = 0x_1 + 0x_2 - 3x_3 + x_4 + 20 \\ \text{subj. a:} \\ \left.\begin{array}{r} x_1 \quad - 3x_3 + 3x_4 = 6 \\ x_2 - 8x_3 + 4x_4 = 4 \\ x_j \geqslant 0,\ j = 1,2,3,4 \end{array}\right\} \end{array}$$

Partim de la solució bàsica $x_1 = 6$, $x_2 = 4$, $x_3 = 0$, $x_4 = 0$. Seguint els raonaments d'abans, ara trobem que augmentant la variable x_3 la funció objectiu disminueix. Per tant, sembla que la solució actual no és optimal. Arem a millorar-la:

Posem que mantenim $x_4 = 0$ i augmentem x_3 fins a un valor t. Com canvia la solució?

$$\left.\begin{array}{l} x_1 = 6 + 3t \\ x_2 = 4 + 8t \\ x_3 = t > 0 \\ x_4 = 0 \end{array}\right\}$$

Hem obtingut una nova solució factible, i la funció objectiu ha millorat:

$$z = -3t + 20 < 20 \ .$$

Com més gran sigui t, més millora la funció objectiu. Podem augmentar t indefinidament, sense violar les restriccions '$x_j \geqslant 0$'? És obvi que sí. Conclusió: La funció objectiu no està acotada inferiorment sobre la regió factible. □

1.3.4 Definició

En la situació anterior, diem que el problema és *no acotat*. □

Hem obtingut el següent criteri de no-acotació:

1.3.5 Criteri de no-acotació

Si en un problema de PL en forma estàndard complint (C1) i (C2) hi ha una variable no-bàsica amb coeficient estrictament negatiu a la funció objectiu i coeficients negatius a totes les igualtats, aleshores el problema és no-acotat. □

1.3.6 Exemple 3

$$\begin{array}{l} \min\ z = 0x_1 + 0x_2 + 3x_3 - x_4 + 20 \\ \text{subj. a:} \\ \left.\begin{array}{r} x_1 \quad - 3x_3 + 3x_4 = 6 \\ x_2 - 8x_3 + 4x_4 = 4 \\ x_j \geqslant 0,\ j = 1,2,3,4 \end{array}\right\} \end{array}$$

Intentem reproduir els raonaments de l'Exemple 1.3.3. Partim de la solució bàsica

$$x_1 = 6,\ x_2 = 4,\ x_3 = 0,\ x_4 = 0\ .$$

Ara convé augmentar x_4, perquè té signe '$-$' a la funció objectiu. Igual que abans, mantenim $x_3 = 0$ i posem $x_4 = t > 0$. El nou punt és:

$$\left.\begin{array}{l} x_1 = 6 - 3t \\ x_2 = 4 - 4t \\ x_3 = 0 \\ x_4 = t > 0 \end{array}\right\}$$

Obtenim la millora $z = -t + 20 < 20$.

Podem augmentar t indefinidament, com abans, sense violar les restriccions de positivitat? Ara és clar que no, si mirem les expressions per x_1 i x_2. Fins a quin valor de t podem arribar?

$$x_1 \geqslant 0 \Leftrightarrow 6 - 3t \geqslant 0 \Leftrightarrow t \leqslant 2\ ,$$
$$x_2 \geqslant 0 \Leftrightarrow 4 - 4t \geqslant 0 \Leftrightarrow t \leqslant 1\ .$$

Per tant, com a màxim podem augmentar t fins a 1. La nova solució factible serà:

$$x_1 = 3,\ x_2 = 0,\ x_3 = 0,\ x_4 = 1.$$

El nou valor de la funció objectiu és $z = 19$.

Ara cal tornar a investigar aquesta solució, a veure si es pot millorar o és optimal.

Per això, necessitem posar-nos altre cop en la situació inicial, és a dir, que la solució que tenim ara sigui la solució bàsica d'un problema de PL en forma stàndard complint (C1) i (C2). Vegem com:

Si observem la solució que tenim veiem que les variables no-bàsiques haurien de ser x_2 i x_3, doncs són les que estan a zero, i per tant les bàsiques han de ser x_1 i x_4. Manipulem el sistema per aconseguir-ho:

$$\left.\begin{array}{r} x_1 \quad - 3x_3 + 3x_4 = 6 \\ x_2 - 8x_3 + 4x_4 = 4 \end{array}\right\}$$

$$\Big\downarrow \text{Coef. 1 per a } x_4 \text{ a la segona equació}$$

$$\left.\begin{array}{r} x_1 \quad - 3x_3 + 3x_4 = 6 \\ \frac{1}{4}x_2 - 2x_3 + \ \ x_4 = 1 \end{array}\right\}$$

$$\Big\downarrow \begin{array}{l}\text{Eliminar } x_4 \text{ de la primera equació} \\ \text{(restant la segona multiplicada per 3)}\end{array}$$

$$\left.\begin{array}{r} x_1 - \frac{3}{4}x_2 + 3x_3 \qquad = 3 \\ \frac{1}{4}x_2 - 2x_3 + x_4 = 1 \end{array}\right\}$$

Ja estem altre cop en la situació (C1) per a les variables x_1 i x_4.

Manipulem ara la funció objectiu:

$$z = 0x_1 + 0x_2 + 3x_3 - x_4 + 20$$

$$\Big\downarrow$$

$$z - 3x_3 + x_4 = 20$$

$$\Big\downarrow \begin{array}{l}\text{Posar } x_4 \text{ a coeficient 0} \\ \text{(restant la segona equació)}\end{array}$$

$$z - \tfrac{1}{4}x_2 - x_3 = 19$$

$$\Big\downarrow$$

$$z = 0x_1 + \tfrac{1}{4}x_2 + x_3 + 0x_4 + 19$$

Hem arribat a:

$$\left.\begin{array}{l} \min\ z = 0x_1 + \dfrac{1}{4}x_2 + x_3 + 0x_4 + 19 \\ x_1 - \frac{3}{4}x_2 + 3x_3 \qquad\ = 3 \\ \qquad \frac{1}{4}x_2 - 2x_3 + x_4 = 1 \\ x_j \geqslant 0,\ j = 1,2,3,4 \end{array}\right\},$$

problema equivalent a l'original, complint (C1) i (C2) amb variables diferents a les d'abans. Tenim ara la solució bàsica factible

$$x_1 = 3,\ x_2 = 0,\ x_3 = 0,\ x_4 = 1,$$

amb valor de la funció objectiu $z = 19$.

Ara tornem al principi. En aquest cas, tots els coeficients de variables no-bàsiques en z són positius i el criteri d'optimalitat s'aplica. Aquesta és la solució optimal. $\square$

1.3.7 Definició

El procediment per reescriure el problema de PL de forma que es compleixin (C1) i (C2) en una altra base s'anomena *pivotació*. A l'Exemple 3 anterior, diríem que "hem pivotat sobre x_4 a la segona equació". $\square$

Hem obtingut el tercer i últim criteri del mètode del símplex:

1.3.8 Criteri de millora de la solució

Suposem que en un problema PL en forma estàndard complint (C1) i (C2) hi ha alguna variable no-bàsica amb coeficient estrictament negatiu a la funció objectiu i coeficient estrictament positiu en alguna equació.

Llavors, s'obté una nova solució bàsica factible que millora (millor dit, que no empitjora) la funció objectiu, fent pivotació sobre aquesta variable.

A quina equació hem de pivotar? Per a cada equació on aparegui la variable amb coeficient estrictament positiu, es calcula el quocient entre el terme independent i el coeficient. El mínim d'aquests quocients dóna l'equació en la qual cal pivotar. $\square$

1.3.9 Observació

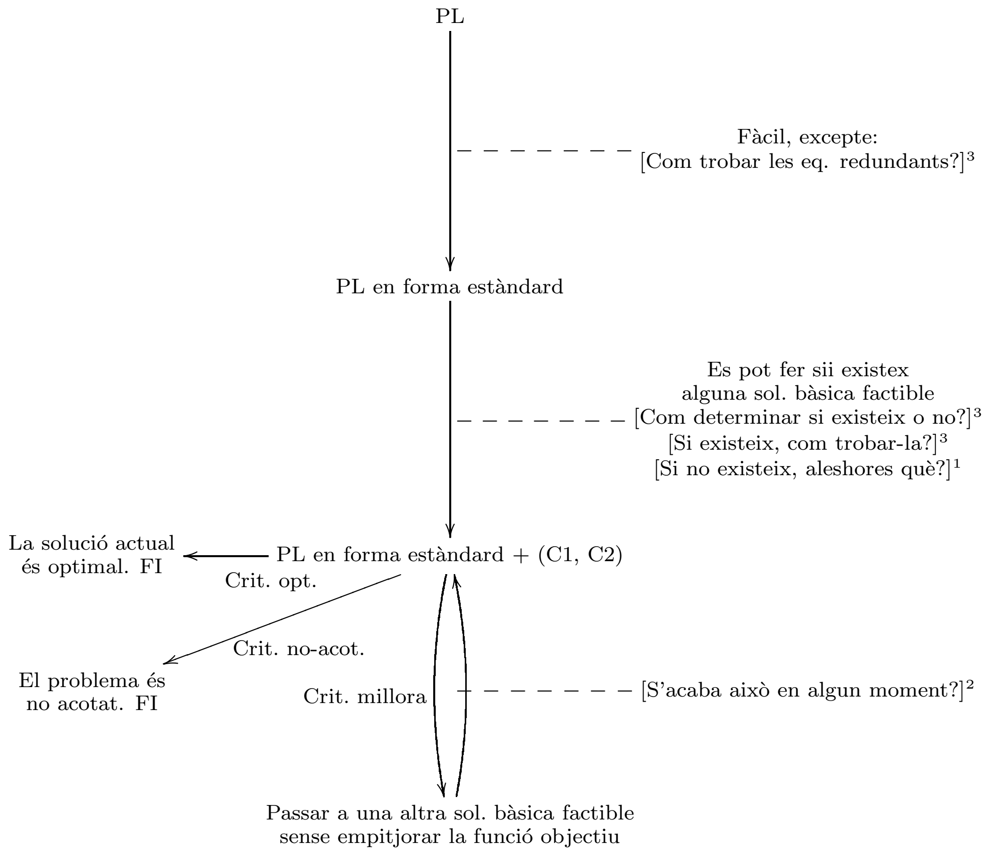
Si el quocient mínim anterior és zero, podem pivotar però la funció objectiu mantindrà el mateix valor. En principi, no ens hem de preocupar si es produeix aquesta circumstància. Tornarem a parlar-ne més endavant. $\square$

1.3.10 Observació

Els criteris d'optimalitat, no-acotació i millora de la solució es poden aplicar si es compleixen les condicions (C1) i (C2). Quan es pot passar de 'forma estàndard' a 'forma estàndard + (C1),(C2)'?

La resposta és fàcil: Es pot fer si i només si existeix alguna solució bàsica factible, perquè, si B és la base associada, podrem escriure el sistema $Ax = d$ com $x_B + B^{-1}Nx_N = B^{-1}d$ (es compleix (C1)), i aïllar en aquesta expressió les variables bàsiques x_B per substituir-les a la funció objectiu, quedant aquesta en funció només de x_N (condició (C2)). □

El gràfic següent resumeix el que hem fet fins aquí i planteja les preguntes a les que ens haurem d'adreçar ara:



La resposta a la pregunta marcada amb [1] la respondrà el *Teorema fonamental de la programació lineal (1a part)*; la resposta a la [2] sortirà del *Teorema fonamental de la programació lineal (2a part)*, el *Teorema de convergència per problemes no degenerats* i la *Regla de Bland*; finalment, les preguntes [3] quedaran contestades totes juntes amb el *Mètode de les dues fases.*

1.3.11 Teorema fonamental de la programació lineal (1a part)

Sigui

$$
\begin{aligned}
&\min\ \ z = c \cdot x + c_0 \\
&\textit{subj. a:} \\
&\left.\begin{aligned} Ax &= d \\ x &\geqslant 0 \end{aligned}\right\}
\end{aligned}
$$

un problema de programació lineal en forma estàndard, on A és una matriu $m \times n$.

Aleshores: Si existeix alguna solució factible, existeix una solució bàsica factible. □

Conseqüència: Si no existeix cap solució bàsica factible, la regió factible és buida. Diem que el problema és *impossible.* □

Demostració: Sigui $x = (x_1, \ldots, x_n)$ una solució factible. En particular, x satisfà el sistema $Ax = d$, que també es pot expressar

$$x_1 A^1 + \cdots + x_n A^n = d \ ,$$

on A^j denota la columna j de la matriu A. Suposem que hi ha exactament p dels valors $x_1, \ldots, x_n$ que són diferents de zero. Suposem, sense pèrdua de generalitat, que són els p primers. Llavors

$$x_1 A^1 + \cdots + x_p A^p = d \ . \tag{1.3.1}$$

Distingim dos casos:

- Cas 1: $A^1, \ldots, A^p$ són linealment independents.
  Això implica en particular que $p \leqslant m$. Si $p = m$, aleshores $A^1, \ldots, A^m$ formen base de $\mathbb{R}^m$; si $p < m$, com que $\operatorname{rang} A = m$, podem trobar entre les columnes $A^{p+1}, \ldots, A^n$ els vectors necessaris per completar una base. Suposem, sense pèrdua de generalitat, que són $A^{p+1}, \ldots, A^m$. En totes dues situacions, la solució de què haviem partit era ja solució bàsica, associada a la base $B = (A^1, \ldots, A^m)$.
- Cas 2: $A^1, \ldots, A^p$ són linealment dependents.
  Existiran aleshores escalars $\alpha_1, \ldots, \alpha_p$, amb almenys algun j tal que $\alpha_j > 0$, tals que

  $$\alpha_1 A^1 + \cdots + \alpha_p A^p = 0. \tag{1.3.2}$$

  Sigui $t \in \mathbb{R}$. Multiplicant (1.3.2) per t i restant de (1.3.1), obtenim

  $$(x_1 - t\alpha_1) A^1 + \cdots + (x_p - t\alpha_p) A^p = d.$$

  Això vol dir que, per a qualsevol $t \in \mathbb{R}$, el punt

  $$y = (x_1 - t\alpha_1, x_2 - t\alpha_2, \ldots, x_p - t\alpha_p, 0, \ldots, 0)$$

  satisfà el sistema $Ax = d$.
  Per $t = 0$ tenim la solució de què havíem partit. Si anem augmentant el valor de t des de zero, hi haurà almenys una component del vector y que decreixerà, perquè existeix $\alpha_j > 0$. Augmentem t fins que alguna component es fa zero. O sigui, prenem

  $$t = \min \left\{ \frac{x_j}{\alpha_j} : \ \alpha_j > 0 \right\}.$$

  Per aquest t, tenim una solució factible y amb $p - 1$ variables diferents de zero com a molt. Anem repetint aquest procediment fins que puguem aplicar el cas 1. □

1.3.12 Teorema fonamental de la programació lineal (2a part)

En la mateixa situació de la 1a part, si existeix alguna solució optimal, llavors existeix una solució bàsica optimal. □

Conseqüència: Per trobar una solució optimal, és suficient buscar entre les solucions bàsiques. Donat que el mètode del símplex va passant de solució bàsica a solució bàsica, això demostra que no estem perdent res amb aquest mètode. □

Demostració: Sigui $x = (x_1, \ldots, x_n)$ una solució factible optimal. Es procedeix igual que abans, distingint els mateixos dos casos. El cas 1 és idèntic i s'arriba a la conclusió que la solució de la qual hem partit era ja bàsica. El cas 2 també es fa igual, llevat que s'ha de demostrar que la solució amb com a molt $p - 1$ variables diferents de zero és també optimal. Vegem-ho:

$$z(y) = c \cdot y + c_0 = c \cdot x - t \cdot c \cdot (\alpha_1, \ldots, \alpha_p, 0, \ldots, 0)^{\perp} + c_0 \ .$$

Si veiem que $K := c \cdot (\alpha_1, \dots, \alpha_p, 0, \dots, 0) = 0$ ja haurem acabat. Suposem que $K \neq 0$. Aleshores, prenent t prou proper a zero i amb el mateix signe que K, el punt y seguirà essent factible i $z(y) < z(x)$, contradient que x és optimal. □

1.3.13 Observació

Acabem de deduir que per trobar un punt optimal és suficient buscar només entre les solucions bàsiques. Quantes solucions bàsiques hi ha? N'hi ha tantes com maneres hi ha d'escollir una base B dins la matriu A. Com que escollir B és escollir m columnes d'una matriu de n columnes, hi ha com a molt $\binom{n}{m}$ bases possibles. Per tant,

$$\text{nombre de solucions bàsiques factibles} \leqslant \binom{n}{m} < \infty.$$

Això suggereix que un possible mètode de resolució seria trobar totes les bases (això no és fàcil), seleccionar les factibles, i avaluar en elles la funció objectiu, a veure quina dóna el valor mínim. Aquest mètode és altament ineficient.

1.3.14 Definicions

Una solució bàsica factible es diu que és *degenerada* si alguna variable bàsica val zero. En cas contrari és *no-degenerada*.

Si tota solució bàsica factible és no-degenerada, diem que el problema és *no-degenerat*. □

1.3.15 Teorema de convergència del mètode del símplex

Si un problema de PL és no-degenerat, llavors el mètode del símplex acaba. És a dir, partint d'una solució bàsica factible i aplicant reiteradament el criteri de millora de la solució, s'arriba a una base en la qual és aplicable el criteri d'optimalitat o bé és aplicable el criteri de no-acotació.

Demostració: Recordeu que, en el criteri de millora de la solució, escollíem

$$t = \min_i \left\{ \frac{d_i}{A_i^j} : \ A_i^j > 0 \right\} ,$$

on A_i^j és l'element de la columna j i la fila i de la matriu A, i el nou punt y complia

$$z(y) = z(x) + t \cdot c_j \leqslant z(x) \ ,$$

perquè $t \geqslant 0$ i $c_j < 0$.

Sota la hipòtesi de no-degeneració, $d_i > 0$, $\forall i$, perquè els d_i són els valors de les variables bàsiques, que no són zero. Per tant, $t > 0$, d'on obtenim que

$$z(y) < z(x) \ ,$$

i la funció objectiu millora efectivament a cada aplicació del criteri.

Com que hi ha un nombre finit de bases factibles, i no podrem repetir-ne, arribarà un moment en què el criteri de millora no serà aplicable, i per tant ho serà algun dels altres dos. □

A la pràctica, els problemes solen ser degenerats, i a més és difícil saber a priori si un problema és degenerat o no. Què passa aleshores amb els problemes degenerats? Pot passar que apliquem diverses vegades el criteri de millora de la solució sense que el valor de z millori de veritat, i, eventualment, tornem a una base per la qual ja hem passat. Llavors es diu que el mètode del símplex entra en un *cicle*.

Això és rar, però per estar segurs que no passarà, podem aplicar la següent regla d'elecció dels pivots, que es coneix com a *regla de Bland.*

1.3.16 Regla de Bland

Observem que en el criteri de millora de la solució hi ha llibertat per escollir la variable que entra a la base entre totes aquelles que tenen coeficient estrictament negatiu a la funció objectiu. I que després, si hi ha empat entre els quocients que s'han de calcular, es pot escollir la fila que es vulgui entre les que estan empatades.

La Regla de Bland dóna una manera sistemàtica d'escollir la columna i la fila de pivotació de manera que es pot demostrar que l'algorisme del símplex mai entra en un cicle. Consisteix en:

- Entre totes les variables candidates a entrar a la base, agafar sempre la d'índex més petit.
- Entre totes les files que empatin, agafar la de més amunt.

De tota manera, al resoldre un problema petit a mà, no ens preocupem mai d'aplicar la regla de Bland, sinó que escollim el pivot segons ens sembla que les operacions següents seran més fàcils. □

1.4 Mètode del Símplex. Algorisme i taula

Anem a escriure de manera sistemàtica l'algorisme del mètode del símplex que hem introduït a l'Apartat 1.3.

Partim del problema de PL en forma estàndard

$$\begin{aligned} &\min\ z = c\cdot x + c_0 \\ &\text{subj. a:} \\ &\left.\begin{aligned} Ax &= d \\ x &\geqslant 0 \end{aligned}\right\} \end{aligned}$$

del qual *suposem* que existeix una base B amb solució bàsica associada factible, i que la tenim trobada.

1.4.1 Notació

Si abreviem

$$R := B^{-1}N\ , \quad r := B^{-1}d\ ,$$

aleshores

$$Ax = d \Leftrightarrow Bx_B + Nx_N = d \Leftrightarrow Ix_B + Rx_N = r\ .$$

Si posem, a més,

$$u_N := (c_N - c_B R)\ , \quad u_0 := c_0 + c_B r\ ,$$

aleshores

$$z = cx + c_0 = c_B x_B + c_N x_N + c_0 = c_B(r - Rx_N) + c_N x_N + c_0 = u_N x_N + u_0\ .$$

Això que acabem de fer no és més que passar de forma estàndard a un problema que segueix estant en forma estàndard i a més compleix (C1)+(C2).

Denotarem també:

$$\begin{aligned} R^j &= \text{ columna } j \text{ de la matriu } R. \\ R_i^j &= \text{ element de la fila } i \text{ i columna } j \text{ de } R. \quad \square \end{aligned}$$

1.4.2 Algorisme

0. Determinar una base B factible.
Escriure

$$\begin{array}{l} \min\ z = u_N x_N + u_0 \\ \text{subj. a:} \\ \left.\begin{array}{r} x_B + R x_N = r \\ x \geqslant 0 \end{array}\right\} \end{array}$$

La solució bàsica associada a la base B és $x_B = r$, $x_N = 0$.

1. Si $u_N \geqslant 0$, FI. La solució actual és optimal.

2. Seleccionar $j \in N$ tal que $u_j < 0$.
(x_j serà la variable que entrarà a la base; pivotarem sobre la columna j).

3. Si $R^j \leqslant 0$, FI. El problema és no-acotat.

4. Calcular $\dfrac{r_i}{R_i^j}$ per a tots els $R_i^j > 0$.
Seleccionar i tal que $\dfrac{r_i}{R_i^j}$ és el mínim d'aquests quocients. En cas d'empat, s'agafa qualsevol.
(Pivotarem sobre la fila i.)

5. Pivotar sobre R_i^j.
Tornar al pas 1. □

La *taula del símplex* és una manera convenient de representar el problema per executar l'algorisme del símplex. Ho veurem sobre un exemple:

1.4.3 Exemple

Considereu el problema

$$\begin{array}{l} \max\ z = 3x_1 + x_2 + 3x_3 \\ \text{subj. a:} \\ \left.\begin{array}{r} 2x_1 + x_2 + x_3 \leqslant 2 \\ x_1 + 2x_2 + 3x_3 \leqslant 5 \\ 2x_1 + 2x_2 + x_3 \leqslant 6 \\ x_j \geqslant 0, \quad j = 1,2,3 \end{array}\right\} \end{array}$$

En primer lloc el convertim en un problema de minimitzar. Després el passem a forma estàndard afegint-hi variables de folga. Ens queda:

$$\begin{array}{l} \min\ z = -3x_1 - x_2 - 3x_3 \\ \text{subj. a:} \\ \left.\begin{array}{r} 2x_1 + x_2 + x_3 + x_4 \qquad\qquad = 2 \\ x_1 + 2x_2 + 3x_3 \qquad + x_5 \qquad = 5 \\ 2x_1 + 2x_2 + x_3 \qquad\qquad + x_6 = 6 \\ x_j \geqslant 0, \quad j = 1, \ldots, 6 \end{array}\right\} \end{array}$$

Veiem que, casualment, hi ha una base factible que es veu a ull, amb variables bàsiques x_4, x_5, x_6, $B = I$, i els coeficients corresponents a la funció objectiu són zero. Podem escriure la primera taula del símplex:

x_1	x_2	x_3	x_4	x_5	x_6	1	
2	1	1	1	0	0	-2	$= 0$
1	2	3	0	1	0	-5	$= 0$
2	2	1	0	0	1	-6	$= 0$
-3	-1	-3	0	0	0	0	$= z$

1. $u_N \geqslant 0$? No.

2. Escollim, per exemple, $j = 2$.

3. $R^2 \leqslant 0$? No.

4. $\dfrac{r_1}{R_1^2} = 2$, $\dfrac{r_2}{R_2^2} = 5/2$, $\dfrac{r_3}{R_3^2} = 3$.
Pivotarem sobre R_1^2. La variable x_2 entra a la base. Surt la variable x_4.

5. Pivotem:

x_1	x_2	x_3	x_4	x_5	x_6	1	
2	1	1	1	0	0	-2	$= 0$
-3	0	1	-2	1	0	-1	$= 0$
-2	0	-1	-2	0	1	-2	$= 0$
-1	0	-2	1	0	0	-2	$= z$

1. $u_N \geqslant 0$? No.

2. Escollim, per exemple, $j = 3$.

3. $R^3 \leqslant 0$? No.

4. $\dfrac{r_1}{R_1^3} = 2$, $\dfrac{r_2}{R_2^3} = 1$.
Pivotarem sobre R_2^3. Entra x_3, surt x_5.

5. Pivotem:

x_1	x_2	x_3	x_4	x_5	x_6	1	
5	1	0	3	-1	0	-1	$= 0$
-3	0	1	-2	1	0	-1	$= 0$
-5	0	0	-4	1	1	-3	$= 0$
-7	0	0	-3	2	0	-4	$= z$

1. $u_N \geqslant 0$? No.

2. Escollim, per exemple, $j = 1$.

3. $R^1 \leqslant 0$? No.

4. Forçosament hem de pivotar sobre R_1^1.

5. Pivotem:

x_1	x_2	x_3	x_4	x_5	x_6	1	
1	1/5	0	3/5	-1/5	0	-1/5	$= 0$
0	3/5	1	-1/5	2/5	0	-8/5	$= 0$
0	1	0	-1	0	1	-4	$= 0$
0	7/5	0	6/5	3/5	0	-27/5	$= z$

1. $u_N \geqslant 0$? Sí. Hem acabat. L'actual solució bàsica és optimal:

$$x_1 = 1/5, \quad x_2 = 0, \quad x_3 = 8/5, \quad x_4 = 0, \quad x_5 = 0, \quad x_6 = 4.$$

El valor òptim de la funció objectiu és $z = -27/5$.

Ara retornem al problema original, oblidant les folgues i canviant de signe z:

$$x_1 = 1/5, \quad x_2 = 0, \quad x_3 = 8/5, \quad z = 27/5.$$

1.5 Mètode del Símplex. Determinació d'una base inicial

Tenim pendents les preguntes següents:

- Com determinar si existeix o no una base factible?
- Si existeix, com trobar-la?
- Com trobar equacions redundants en el sistema $Ax = d$?

Les respondrem totes a la vegada, amb el *mètode de les dues fases.*

Suposem que tenim un problema de PL en forma estàndard (llevat que no sabem si $\operatorname{rang} A = m$)

$$P_z : \left[\begin{array}{l} \quad\min\ z = cx + c_0 \\ \text{subj. a:} \\ \left.\begin{array}{r} Ax = d \\ x \geqslant 0 \end{array}\right\} \end{array}\right.$$

i suposem que no està a la vista una base inicial amb solució bàsica factible.

Considerem apart el problema

$$P_w : \left[\begin{array}{l} \quad\min\ w = y_1 + \cdots + y_m \\ \text{subj. a:} \\ \left.\begin{array}{r} Ax + Iy = d \\ x \geqslant 0,\ y \geqslant 0 \end{array}\right\} \quad \text{on } y = (y_1, \ldots, y_m). \end{array}\right.$$

1.5.1 Observacions

1) P_w sempre està acotat i té solucions factibles. En efecte, w no pot ser mai més petita que zero, i per altre banda si posem totes les variables x a zero s'obté una solució factible.
2) P_w està gairebé preparat per començar el símplex (està en forma estàndard + (C1)). Només falta arreglar la funció objectiu.

1.5.2 Definició

Les variables y del problema P_w s'anomenen *variables artificials* pel problema P_z. □

Els dos problemes P_z i P_w estan relacionats de la manera següent:

1.5.3 Proposició

P_z té algun punt factible $\Leftrightarrow$ El valor òptim de P_w és $w = 0$.

Demostració:

$\Rightarrow$) Si $x = x^0$ és factible de P_z, llavors $\left.\begin{array}{c} x = x^0 \\ y = 0 \end{array}\right\}$ és factible de P_w. El valor de w en aquest punt és $w = 0$, que no es pot millorar sense violar la positivitat d'alguna de les y.

$\Leftarrow$) Si l'òptim de P_w és $w = 0$, el punt on s'assoleix és necessàriament de la forma $\left.\begin{array}{c} x = x^0 \\ y = 0 \end{array}\right\}$ per algun $x^0 \in \mathbb{R}^n$. Per tant, $x = x^0$ és un punt factible de P_z. □

El mètode de les dues fases consisteix en el següent:

Fase I: Resoldre P_w. Possibilitats:

1) min $w \neq 0$. Llavors P_z és impossible.
2) min $w = 0$ i no hi ha cap variable artificial a la base. Passar a Fase II.
3) min $w = 0$ i hi ha alguna variable artificial a la base. Per a cada una d'elles, intentar treure-la de la base pivotant. Si no es pot perquè no podem pivotar enlloc, suprimir les equacions que la continguin (són redundants). Passar a Fase II.

Fase II: A partir de la base trobada a la Fase I (on no hi haurà variables artificials), resoldre P_z.

1.5.4 Exemple

$$\min\ z = 3x_1 - 3x_2 - 2x_3 + 2x_4 + x_5 - 4x_6$$

subj. a:

$$\left.\begin{array}{rl} x_1 - \ x_2 + x_3 - x_4 - 4x_5 + 2x_6 \geqslant & 4 \\ -3x_1 + 3x_2 + x_3 - x_4 - 2x_5 \qquad \leqslant & 6 \\ x_3 - x_4 \qquad - \ x_6 = & -1 \\ x_1 - \ x_2 + x_3 - x_4 - \ x_5 \qquad = & 0 \\ x_j \geqslant 0, \quad j = 1, \ldots, 6 & \end{array}\right\}$$

Operacions prèvies:

1) Convertir el problema a forma estàndard (llevat que no sabem si $\operatorname{rang} A = m$).

$$\min\ z = 3x_1 - 3x_2 - 2x_3 + 2x_4 + x_5 - 4x_6$$

subj. a:

$$\left.\begin{array}{rl} x_1 - \ x_2 + x_3 - x_4 - 4x_5 + 2x_6 - x_7 & = 4 \\ -3x_1 + 3x_2 + x_3 - x_4 - 2x_5 \qquad + x_8 & = 6 \\ - x_3 + x_4 \qquad + \ x_6 & = 1 \\ x_1 - \ x_2 + x_3 - x_4 - \ x_5 & = 0 \\ x_j \geqslant 0, \quad j = 1, \ldots, 8 & \end{array}\right\}$$

2) Afegir variables artificials. (Observació: No cal una per cada equació. És suficient posar només les que calguin per tal que es "vegi" una base.)

$$\min\ z = 3x_1 - 3x_2 - 2x_3 + 2x_4 + x_5 - 4x_6$$

subj. a:

$$\left.\begin{array}{rl} x_1 - \ x_2 + x_3 - x_4 - 4x_5 + 2x_6 - x_7 \qquad + x_9 & = 4 \\ -3x_1 + 3x_2 + x_3 - x_4 - 2x_5 \qquad + x_8 & = 6 \\ - x_3 + x_4 \qquad + \ x_6 \qquad + x_{10} & = 1 \\ x_1 - \ x_2 + x_3 - x_4 - \ x_5 \qquad + x_{11} & = 0 \\ x_j \geqslant 0, \quad j = 1, \ldots, 11 & \end{array}\right\}$$

Fase I:

Hem de Resoldre P_w amb $w = x_9 + x_{10} + x_{11}$.

1) Eliminar de la funció objectiu les variables bàsiques:

$$w = -2x_1 + 2x_2 - x_3 + x_4 + 5x_5 - 3x_6 + x_7 + 5.$$

2) Resoldre P_w:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	1	
1	-1	1	-1	-4	2	-1	0	1	0	0	-4	$= 0$
-3	3	1	-1	-2	0	0	1	0	0	0	-6	$= 0$
0	0	-1	1	0	$\boxed{1}$	0	0	0	1	0	-1	$= 0$
1	-1	1	-1	-1	0	0	0	0	0	1	0	$= 0$
3	-3	-2	2	1	-4	0	0	0	0	0	0	$= z$
-2	2	-1	1	5	-3	1	0	0	0	0	5	$= w$
					$\uparrow$							

Observació: És opcional afegir la fila de la z a la taula. Si ho fem, al acabar la fase I ja tindrem la taula preparada per començar la fase II. Si no, abans de començar la fase II caldrà eliminar de z les variables que en aquell moment siguin bàsiques.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	1	
1	-1	3	-3	-4	0	-1	0	1	0	-2	$=0$
-3	3	1	-1	-2	0	0	1	0	0	-6	$=0$
0	0	-1	1	0	1	0	0	0	0	-1	$=0$
1	-1	$\boxed{1}$	-1	-1	0	0	0	0	1	0	$=0$
3	-3	-6	6	1	0	0	0	0	0	-4	$=z$
-2	2	-4	4	5	0	1	0	0	0	2	$=w$

Observació: Les variables artificials que surten de la base les esborrem de la taula, perquè no tornin a entrar.
Observació: Aquesta iteració serà "degenerada", perquè el mínim dels quocients ha estat 0 (hi havia una variable bàsica a zero).

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	1	
-2	$\boxed{2}$	0	0	-1	0	-1	0	1	-2	$=0$
-4	4	0	0	-1	0	0	1	0	-6	$=0$
1	-1	0	0	-1	1	0	0	0	-1	$=0$
1	-1	1	-1	-1	0	0	0	0	0	$=0$
9	-9	0	0	-5	0	0	0	0	-4	$=z$
2	-2	0	0	1	0	1	0	0	2	$=w$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	1	
-1	1	0	0	-1/2	0	-1/2	0	-1	$=0$
0	0	0	0	1	0	2	1	-2	$=0$
0	0	0	0	-3/2	1	-1/2	0	-2	$=0$
0	0	1	-1	-3/2	0	-1/2	0	-1	$=0$
0	0	0	0	-19/2	0	-9/2	0	-13	$=z$
0	0	0	0	0	0	0	0	0	$=w$

Òptim trobat (tots els coeficients de w són $\geqslant 0$). min $w = 0$. No hi ha cap variable artificial a la base. Fi de la fase I.

Fase II:

Resoldre P_z a partir de l'última taula:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	1	
-1	1	0	0	-1/2	0	-1/2	0	-1	$=0$
0	0	0	0	$\boxed{1}$	0	2	1	-2	$=0$
0	0	0	0	-3/2	1	-1/2	0	-2	$=0$
0	0	1	-1	-3/2	0	-1/2	0	-1	$=0$
0	0	0	0	-19/2	0	-9/2	0	-13	$=z$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	1	
-1	1	0	0	0	0	1/2	1/2	-2	$=0$
0	0	0	0	1	0	2	1	-2	$=0$
0	0	0	0	0	1	5/2	3/2	-5	$=0$
0	0	1	-1	0	0	5/2	3/2	-4	$=0$
0	0	0	0	0	0	29/2	19/2	-32	$=z$

Tots els coeficients de z són $\geqslant 0$. Solució optimal trobada:

$$x_1 = 0,\ x_2 = 2,\ x_3 = 4,\ x_4 = 0,\ x_5 = 2,\ x_6 = 5.$$

1.5.5 Exemple

Suposem que el problema P_w de la Fase I té la forma següent (x_4, x_5, x_6 són les variables artificials):

$$\begin{aligned}
&\min\ w = x_4 + x_5 + x_6\\
&\text{subj. a:}\\
&\left.\begin{array}{l}
x_1 - 2x_2 \qquad + x_4 \qquad\qquad = 2\\
x_1 - 3x_2 - x_3 \qquad + x_5 \qquad = 1\\
x_1 - \ \ x_2 \qquad\qquad\qquad + x_6 = 3\\
\qquad x_j \geqslant 0, \quad j = 1, \dots, 6
\end{array}\right\}
\end{aligned}$$

Pivotant primer sobre x_1 a la segona fila, i després sobre x_2 a la primera fila, s'arriba a:

x_1	x_2	x_3	x_4	x_5	x_6	1	
0	1	1	1	-1	0	-1	$=0$
1	0	2	3	-2	0	-4	$=0$
0	0	-1	-2	1	1	0	$=0$
0	0	1	3	0	0	0	$=w$

Hem acabat, però la variable artificial x_6 encara és a la base. La fem sortir pivotant sobre x_3 a la tercera fila. Aquest és l'únic cas en què es permet pivotar sobre un nombre negatiu; no hi ha perill de perdre la factibilitat, perquè en aquesta situació el terme independent sempre és zero.

x_1	x_2	x_3	x_4	x_5	x_6	1	
0	1	0	-1	0	1	-1	$=0$
1	0	0	-1	0	2	-4	$=0$
0	0	1	2	-1	-1	0	$=0$
0	0	0	1	1	1	0	$=w$

Ara ja no hi ha variables artificials a la base, i podem procedir a la Fase II.

1.5.6 Exemple Modifiquem lleugerament l'Exemple 1.5.5:

$$\begin{aligned}
&\min\ w = x_4 + x_5 + x_6\\
&\text{subj. a:}\\
&\left.\begin{array}{l}
x_1 - 2x_2 \qquad + x_4 \qquad\qquad = 2\\
x_1 - 3x_2 - x_3 \qquad + x_5 \qquad = 1\\
x_1 - \ \ x_2 + x_3 \qquad\qquad + x_6 = 3\\
\qquad x_j \geqslant 0, \quad j = 1, \dots, 6
\end{array}\right\}
\end{aligned}$$

Procedint igual que abans, s'arriba a:

x_1	x_2	x_3	x_4	x_5	x_6	1	
0	1	1	1	-1	0	-1	$=0$
1	0	2	3	-2	0	-4	$=0$
0	0	0	-2	1	1	0	$=0$
0	0	0	3	0	0	0	$=w$

Ara no podem pivotar sobre x_3 a la tercera fila, perquè no es pot pivotar sobre un zero. Però s'observa que traient x_4 i x_5 de la taula (són variables no-bàsiques i per tant estan a zero), l'última equació és $x_6 = 0$. Podem eliminar-la, juntament amb la variable x_6. Aquesta situació es dóna quan el sistema inicial era redundant.

Començarem la fase II amb:

x_1	x_2	x_3	1	
0	1	1	-1	$= 0$
1	0	2	-4	$= 0$
·	·	·	·	$= z$

2. Programació no lineal

Al capítol anterior vam definir *problema de programació matemàtica* (PM) com $\min_{x \in S} f(x)$, on el conjunt factible $S \subset \mathbb{R}^n$ és una regió determinada per igualtats i desigualtats

$$S = \begin{cases} h_i(x) = 0, & i = 1, \dots, p \\ g_i(x) \leqslant 0, & i = p+1, \dots, m \ . \end{cases}$$

Si totes les funcions involucrades són afins, es tracta d'un *problema de programació lineal* (PL), per al qual tenim una bona teoria, basada en àlgebra lineal, i un bon mètode de resolució (símplex). Si no, diem que tenim un problema de *Programació No Lineal* (PNL), que és més complicat. Podem dir que els problemes de programació no lineal són els problemes de programació matemàtica que no són lineals:

$$\boxed{\text{PNL} = \text{PM} - \text{PL}}$$

2.1 Teoria d'extrems en $\mathbb{R}^n$

En aquest primer apartat veurem unes quantes qüestions teòriques que són d'interès en PNL. La resta d'apartats del capítol tractaran alguns mètodes efectius de resolució.

2.1.1 El problema general d'optimització

Veurem primer què es pot dir del *problema general d'optimització en* $\mathbb{R}^n$, $\min_{x \in S} f(x)$, on no se suposa cap forma particular del conjunt S.

2.1.1 Definició

Un punt $x^* \in S$ és un *mínim global* (o *absolut*) de f en S si

$$\forall x \in S,\ f(x) \geqslant f(x^*) \ .$$

El mínim global s'anomena *estricte* si

$$\forall x \in S,\ x \neq x^* \Rightarrow f(x) > f(x^*) \ .$$

(Està clar que no pot haver-hi més d'un mínim global estricte.)

2.1.2 Definició

Un punt $x^* \in S$ és un *mínim local* (o *relatiu*) de f en S si

$$\exists \varepsilon > 0\colon\ \forall x \in S,\ \|x - x^*\| < \varepsilon \Rightarrow f(x) \geqslant f(x^*) \ .$$

Un mínim local és *estricte* si

$$\exists \varepsilon > 0\colon\ \forall x \in S,\ \|x - x^*\| < \varepsilon,\ x \neq x^* \Rightarrow f(x) > f(x^*)\ . \quad \square$$

La definició de mínim global coincideix amb la de "punt optimal" d'un problema de programació matemàtica, vista al capítol anterior. En principi, doncs, quan escrivim

$$\min_{x \in S} f(x)$$

el que voldríem és trobar un mínim global de la funció f sobre el conjunt S. Malauradament aquest és un problema molt difícil. Tant la teoria com els mètodes que veurem estan enfocats cap a la cerca d'un mínim local, que és un objectiu més modest.

D'entrada, a més, un es pot demanar si existeix o no el mínim global de f en S. La resposta més senzilla ve donada pel Teorema de Weierstrass:

2.1.3 Teorema de Weierstrass

Si f és contínua i S és compacte, llavors existeix mínim global de f en S. $\square$

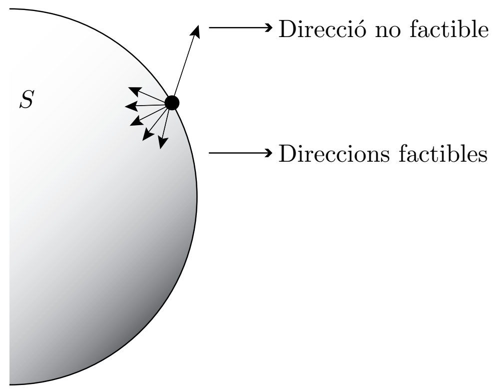
Veurem que, de manera similar al mètode del símplex, els mètodes de resolució de problemes de PNL consisteixen en anar-nos movent d'un punt factible a un altre punt factible que millori la funció objectiu, tantes vegades com calgui. Això motiva la definició següent:

2.1.4 Definició

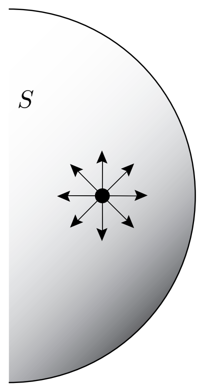
Donat $x \in S$, un vector $v \in \mathbb{R}^n$ és una *direcció factible* en x si

$$\exists \bar{\alpha} > 0 :\ \forall \alpha,\ 0 \leqslant \alpha \leqslant \bar{\alpha}\ \Rightarrow\ x + \alpha v \in S\ .$$

(En paraules: "Per $\alpha > 0$ prou petit, $x + \alpha v$ encara pertany a S".) Per exemple:



Observeu que si $x \in \overset{\circ}{S}$ (l'*interior* de S), tot vector $v \in \mathbb{R}^n$ és una direcció factible:



$\square$

2.1.5 Notacions

Denotarem per $\nabla f(x)$ el *vector gradient* de f en el punt x; és a dir:

$$\nabla f(x) := \Big(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x)\Big) \ .$$

Recordem que, si v és un vector de $\mathbb{R}^n$, el producte escalar $\nabla f(x) \cdot v$ coincideix amb la derivada direccional de f en x en la direcció v:

$$\nabla f(x) \cdot v = \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h} \ .$$

Denotarem per $\nabla^2 f(x)$ la *matriu hessiana* de f en el punt x; és a dir:

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix} \ . \ \square$$

2.1.6 Proposició (Condicions necessàries de mínim local)

Suposem $f \in C^2$.

Si x^* és mínim local de f en S, llavors per a tota direcció factible v en x^* es té:

i) $\nabla f(x^*) \cdot v \geqslant 0$.

ii) Si $\nabla f(x^*) \cdot v = 0$, llavors $v \cdot \nabla^2 f(x^*) \cdot v \geqslant 0$.

2.1.7 Corol·lari (Condicions necessàries de mínim local, cas interior)

En la situació anterior, si $x^* \in \overset{\circ}{S}$, llavors:

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

ii) $\nabla^2 f(x^*)$ *és semidefinida positiva.*

Demostració:

i) $\nabla f(x^*) \cdot v \geqslant 0$, $\forall v \in \mathbb{R}^n \Rightarrow \nabla f(x^*) = 0$. En efecte, només cal agafar com a direcció v els elements de la base canònica i després ells mateixos canviats de signe.

ii) $v \cdot \nabla^2 f(x^*) \cdot v \geqslant 0$, $\forall v \in \mathbb{R}^n$ és precisament la definició de *semidefinida positiva*.

2.1.8 Observació

Els apartats i) de la Proposició i el Corol·lari anteriors són certs suposant només $f \in C^1$.

2.1.9 Proposició (Condicions suficients de mínim local, cas interior)

Suposem $f \in C^2$.

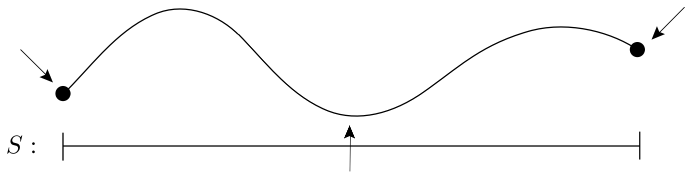
Sigui $x^* \in \overset{\circ}{S}$. Si

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

b) $\nabla^2 f(x^*)$ *és definida positiva (és a dir, $v \cdot \nabla^2 f(x^*) \cdot v > 0$, $\forall v \in \mathbb{R}^n$, $v \neq 0$),* *aleshores x^* és mínim local de f en S (estricte, a més).* □

Aquesta última proposició no la formulem en general perquè per x^* a la frontera de S és difícil d'escriure i no té prou interès l'esforç. Més endavant ho farem per al cas en què S ve donat per restriccions funcionals.

És interessant interpretar i visualitzar les proposicions anteriors per a funcions d'una variable sobre un interval. A la figura següent, els punts assenyalats són mínims locals de la funció:



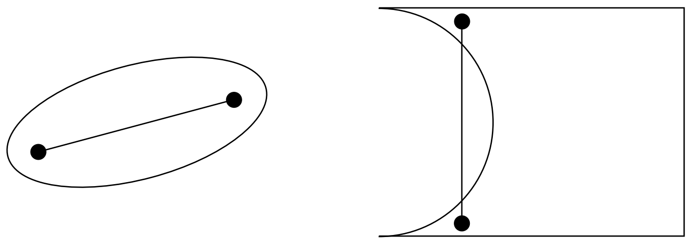
□

2.1.2 Convexitat

2.1.10 Definicions

Un conjunt $S \subset \mathbb{R}^n$ és *convex* si

$$\forall x_1, x_2 \in S,\ \forall \lambda \in]0,1[,\ \lambda x_1 + (1-\lambda)x_2 \in S\ .$$



convex no-convex

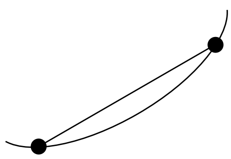
Sigui $S \subset \mathbb{R}^n$ convex. Aleshores:

a) $f \colon S \to \mathbb{R}$ és una *funció convexa* si

$$\forall x_1, x_2 \in S,\ \forall \lambda \in]0,1[,\ f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2)\ .$$

b) $f \colon S \to \mathbb{R}$ és *estrictament convexa* si

$$\forall x_1, x_2 \in S,\ x_1 \neq x_2,\ \forall \lambda \in]0,1[,\ f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2)\ .$$



c) $f \colon S \to \mathbb{R}$ és *còncava* o *estrictament còncava* si $-f$ és convexa o estrictament convexa, respectivament. (Equivalentment, les mateixes definicions de convexa i estrictament convexa invertint les desigualtats.) □

Per a funcions de classe C^2, la convexitat té una caracterització útil en termes de les derivades segones: La matriu hessiana ha de ser semidefinda positiva, almenys en les direccions factibles. Per demostrar aquesta caracterització, necessitem primer establir-ne una altra, que val per a funcions de classe C^1, i té una interpretació gràfica fàcil.

2.1.11 Lema

Siguin $S \subset \mathbb{R}^n$ convex i $f\colon S \to \mathbb{R}$ de classe C^1. Aleshores:

f és convexa en $S \Leftrightarrow \forall x, y \in S,\ f(y) \geqslant f(x) + \nabla f(x) \cdot (y - x)$.

Demostració:

$\Rightarrow$) Per $\lambda \in]0,1[$,

$$\begin{aligned} &f(\lambda y + (1-\lambda)x) \leqslant \lambda f(y) + (1-\lambda) f(x)\ , \Rightarrow \\ &f(x + \lambda(y-x)) \leqslant f(x) + \lambda(f(y) - f(x))\ , \Rightarrow \\ &\frac{f(x+\lambda(y-x)) - f(x)}{\lambda} \leqslant f(y) - f(x)\ . \end{aligned}$$

Fent $\lambda \to 0$, queda

$$\nabla f(x) \cdot (y-x) \leqslant f(y) - f(x)\ .$$

$\Leftarrow$) Fixem $x_1, x_2 \in S$ i $\lambda \in]0,1[$.
Fem $x := \lambda x_1 + (1-\lambda)x_2$.
Prenent $y = x_1$, obtenim $f(x_1) \geqslant f(x) + \nabla f(x) \cdot (x_1 - x)$.
Prenent $y = x_2$, obtenim $f(x_2) \geqslant f(x) + \nabla f(x) \cdot (x_2 - x)$.
Multiplicant per λ la primera expressió, per $(1-\lambda)$ la segona, i sumant,

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geqslant f(\lambda x_1 + (1-\lambda)x_2) + 0\ . \quad \square$$

2.1.12 Proposició

Siguin $S \subset \mathbb{R}^n$ convex, i $f\colon S \to \mathbb{R}$ de classe C^2. Aleshores:

f és convexa en $S \Leftrightarrow \forall x \in S$, $\forall v$ direcció factible en x, $v \cdot \nabla^2 f(x) \cdot v \geqslant 0$.

Demostració:

$\Leftarrow$) Donats $x, y \in S$, existeix $\lambda \in [0,1]$ tal que (Teorema de Taylor)

$$f(y) = f(x) + \nabla f(x) \cdot (y-x) + \frac{1}{2}(y-x) \cdot \nabla^2 f(x + \lambda(y-x)) \cdot (y-x)\ .$$

L'últim sumand és positiu, per la hipòtesi. Per tant,

$$f(y) \geqslant f(x) + \nabla f(x) \cdot (y-x)\ ,$$

que és equivalent a la convexitat, pel Lema 2.1.11.

$\Rightarrow$) Suposem que per certs punts $x, y \in S$ es té

$$(y-x) \cdot \nabla^2 f(x) \cdot (y-x) < 0\ .$$

Per continuïtat de les derivades segones, podem seleccionar y prou proper a x de manera que

$$(y-x) \cdot \nabla f(x + \lambda(y-x)) \cdot (y-x) < 0,\ \forall \lambda \in [0,1]\ .$$

Això vol dir, usant la fórmula de Taylor d'abans, que

$$f(y) < f(x) + \nabla f(x) \cdot (y-x)\ ,$$

i, pel Lema 2.1.11, la funció no serà convexa. $\square$

2.1.13 Teorema (Local-Global per a funcions convexes)

Sigui $S \subset \mathbb{R}^n$ convex, i $f\colon S \to \mathbb{R}$ convexa.

Si x és mínim local de f en S, llavors x és mínim global de f en S.

Demostració: Si x no és mínim global de f en S, existirà $y \in S$ tal que $f(y) < f(x)$. Aleshores, per $\forall\lambda \in]0,1[$,

$$f(\lambda y + (1-\lambda)x) \leqslant \lambda f(y) + (1-\lambda)f(x) < f(x),$$

i això contradiu que x sigui mínim local, ja que λ pot ser arbitràriament proper a zero. □

2.1.14 Teorema ("les condicions necessàries són suficients")

Siguin $S \subset \mathbb{R}^n$ convex, i $f\colon S \to \mathbb{R}$ convexa i de classe C^1.

Si $x \in S$ compleix $\nabla f(x) \cdot v \geqslant 0$, $\forall v$ direcció factible en x, llavors x és mínim (global!) de f en S.

En particular,

$$\nabla f(x) = 0 \Rightarrow x \text{ és mínim global.}$$

Demostració: Per qualsevol $y \in S$, $y - x$ és direcció factible gràcies a la convexitat. Per tant, $f(y) \geqslant f(x) + \nabla f(x) \cdot (y-x) \geqslant f(x)$. □

2.1.3 Cas de restriccions funcionals

Considerarem ara el cas en què la regió factible S ve donada per igualtats i desigualtats funcionals (el que pròpiament anomenem "problema de PNL"):

$$S : \left\{ \begin{array}{cc} h_1(x) = 0 & g_{p+1}(x) \leqslant 0 \\ \vdots & \vdots \\ h_p(x) = 0 & g_m(x) \leqslant 0 \end{array} \right\}$$

Com que el problema té més estructura, es d'esperar que puguem afinar la teoria del Subapartat 2.1.1.

2.1.15 Definicions

Donat un punt factible x^*, una restricció $g_i(x) \leqslant 0$ es diu que és *activa* en x^* si $g_i(x^*) = 0$. És *inactiva* si $g_i(x^*) < 0$.

Les restriccions $h_i(x) = 0$ són sempre actives en tot punt (per definició!).

2.1.16 Definicions

Si tenim una funció $h\colon \mathbb{R}^n \to \mathbb{R}$, l'equació $h(x) = 0$ determina una *superfície* de *dimensió* $n-1$.

Una funció $h\colon \mathbb{R}^n \to \mathbb{R}^p$ es pot pensar com un vector de funcions reals $h(x) := \big(h_1(x), \dots, h_p(x)\big)$, i l'equació vectorial $h(x) = 0$ no és més que el sistema de p equacions escalars

$$\left. \begin{array}{c} h_1(x) = 0 \\ \vdots \\ h_p(x) = 0 \end{array} \right\} .$$

Aquest sistema determina una *superfície* de *dimensió* $n-p$.

Les superfícies de dimensió 1 s'anomenen també *corbes*, i les de dimensió $n-1$, *hipersuperfícies*. Si h és una funció afí, es parla de *rectes* i d'*hiperplans*, respectivament.

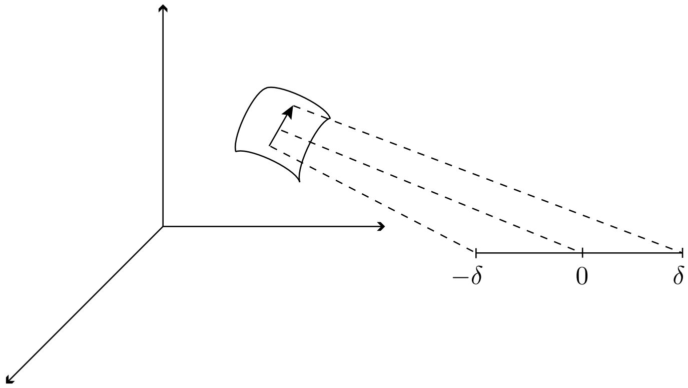
Una corba també es pot donar com una aplicació d'un interval $I \subset \mathbb{R}$ en $\mathbb{R}^n$:

$$\begin{aligned} \alpha\colon I &\longrightarrow \mathbb{R}^n \\ t &\longmapsto (\alpha_1(t), \dots, \alpha_n(t)) \end{aligned} \qquad \square$$

Suposarem a partir d'ara que les funcions h que defineixen les superfícies, i les funcions α que defineixen corbes, tenen totes les derivades que necessitem.

2.1.17 Definició

L'*espai tangent* a una superfície S en un punt $x \in S$ és l'espai vectorial format pels vectors de la forma $(\alpha_1'(0), \dots, \alpha_n'(0))$, per a totes les corbes $\alpha\colon \mathbb{R} \to S$ infinitament derivables tals que $\alpha(0) = x$.



$\square$

2.1.18 Definició

Sigui $h\colon \mathbb{R}^n \to \mathbb{R}^p$.

Un punt x^*, solució de $h(x) = 0$, és un punt *regular* de l'equació si $\nabla h_1(x^*), \dots, \nabla h_p(x^*)$ són vectors linealment independents. $\square$

El concepte de punt regular no és intrínsec de la superfície determinada per $h = 0$, sinó de la seva representació concreta en termes de la funció h. Per exemple, en $\mathbb{R}^2$, les funcions $h(x_1, x_2) = x_1$ i $h(x_1, x_2) = x_1^2$ determinen la mateixa recta $x_1 \equiv 0$, al posar $h = 0$; tots els punts de la recta són regulars per la primera representació, però el $(0, 0)$ no ho és per la segona.

2.1.19 Notació

Si $h\colon \mathbb{R}^n \to \mathbb{R}^p$, denotarem per $\nabla h(x)$ la *matriu jacobiana* de h en el punt x, és a dir,

$$\nabla h(x) := \begin{pmatrix} \dfrac{\partial h_1}{\partial x_1}(x) & \dots & \dfrac{\partial h_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \dfrac{\partial h_p}{\partial x_1}(x) & \dots & \dfrac{\partial h_p}{\partial x_n}(x) \end{pmatrix}$$

Observeu que no hi ha conflicte amb la notació per al gradient que hem introduït a 2.1.5. El gradient no és més que la matriu jacobiana d'una funció $h\colon \mathbb{R}^n \to \mathbb{R}$. $\square$

La següent proposició dóna una caracterització útil de l'espai tangent en un punt regular.

2.1.20 Proposició

Si x^* és un punt regular de l'equació $h(x) = 0$, l'espai tangent a la superfície associada en x^* està format pels vectors ortogonals al gradient de h en x^*; o sigui, és

$$\{y \in \mathbb{R}^n : \ \nabla h(x^*) \cdot y = 0\} \, . \qquad \square$$

2.1.21 Teorema (condicions necessàries de mínim local)

Siguin

$$S : \left\{ \begin{array}{cc} h_1(x) = 0 & g_{p+1}(x) \leqslant 0 \\ \vdots & \vdots \\ h_p(x) = 0 & g_m(x) \leqslant 0 \end{array} \right\}$$

i $f\colon S \to \mathbb{R}$, i suposem que totes les funcions són de classe C^2.

Si x és un mínim local de f en S i és un punt regular de les restriccions actives en x, llavors existeixen $\lambda = (\lambda_1, \dots, \lambda_p)$ i $\mu = (\mu_{p+1}, \dots, \mu_m) \geqslant 0$ tals que

1)

$$\left. \begin{array}{r} \nabla f(x) + \lambda \cdot \nabla h(x) + \mu \cdot \nabla g(x) = 0 \\ \text{Per } p+1 \leqslant i \leqslant m, \quad \mu_i g_i(x) = 0 \end{array} \right\} .$$

2) La matriu

$$M(x) := \nabla^2 f(x) + \sum_{i=1}^{p} \lambda_i \cdot \nabla^2 h_i(x) + \sum_{i=p+1}^{m} \mu_i \cdot \nabla^2 g_i(x)$$

compleix $v \cdot M(x) \cdot v \geqslant 0$, $\forall v$ de l'espai tangent a les restriccions actives en x. □

Les constants $\lambda_1, \dots, \lambda_p$ s'anomenen *multiplicadors de Lagrange*, i les constants $\mu_{p+1}, \dots, \mu_m$ s'anomenen *multiplicadors de Kuhn-Tucker*. La condició 1) és una *condició de 1r ordre*, perquè només involucra derivades primeres; la 2) és de *2n ordre*, perquè hi intervenen derivades segones.

Per a la demostració, distingirem dos casos:

1r cas: No hi ha restriccions de tipus $\leqslant$. Llavors:

1) La condició de primer ordre queda $\nabla f(x) + \lambda \nabla h(x) = 0$.

2) La matriu $M(x)$ és $\nabla^2 f(x) + \sum_{i=1}^{p} \lambda_i \nabla^2 h_i(x)$.

2n cas: Cas general.

Per al primer cas, necessitem el següent lema previ:

2.1.22 Lema

Si x és un mínim local de f sobre

$$S : \left\{ \begin{array}{c} h_1(x) = 0 \\ \vdots \\ h_p(x) = 0 \end{array} \right.$$

aleshores $\nabla f(x)$ és ortogonal a l'espai tangent $T(x)$ a la superfície S en x.

Demostració del Lema:

Sigui $v \in T(x)$.

Sigui $\alpha\colon \mathbb{R} \to S$ una corba tal que $\alpha(0) = x$, i $\alpha'(0) = v$.

Si f té un mínim local en x, per força $f \circ \alpha\colon \mathbb{R} \to \mathbb{R}$ ha de tenir un mínim local en zero. Això implica

$$0 = (f \circ \alpha)'(0) = \nabla f(\alpha(0))\alpha'(0) = \nabla f(x)v \ .$$

Com que això val per a tot $v \in T(x)$, tenim l'ortogonalitat. □

Demostració del 1r cas; condició de primer ordre: De la caracterització d'espai tangent en termes de $\nabla h(x)$ (Proposició 2.1.20), tenim que una base del seu ortogonal és $\nabla h_1(x), \dots, \nabla h_p(x)$ (recordem que aquests vectors són linealment independents).

Pel Lema 2.1.22, $\nabla f(x)$ pertany a aquest espai. Per tant, existiran constants $\lambda_1, \dots, \lambda_p$ tals que

$$\nabla f(x) + \lambda_1 \nabla h_1(x) + \lambda_2 \nabla h_2(x) + \dots + \lambda_p \nabla h_p(x) = 0 \ . \quad \square$$

Demostració del 1r cas; condició de segon ordre:

Sigui $v \in T(x)$.

Sigui $\alpha \colon \mathbb{R} \to S$ una corba tal que $\alpha(0) = x$, i $\alpha'(0) = v$.

Igual que en el Lema 2.1.22, deduïm que $(f \circ \alpha)'(0) = 0$ i $(f \circ \alpha)''(0) \geqslant 0$.

Calculem

$$\begin{aligned}
(f \circ \alpha)'(t) &= \nabla f(\alpha(t)) \cdot \alpha'(t) \\
(f \circ \alpha)''(t) &= \alpha'(t) \cdot \nabla^2 f(\alpha(t)) \cdot \alpha'(t) + \nabla f(\alpha(t)) \cdot \alpha''(t) \\
(f \circ \alpha)''(0) &= \alpha'(0) \cdot \nabla^2 f(x) \cdot \alpha'(0) + \nabla f(x) \cdot \alpha''(0) \geqslant 0 \ .
\end{aligned}$$

Per altra banda, derivem dues vegades $\lambda \cdot h(\alpha(t))$, que és la funció idènticament nul·la perquè h és zero sobre la corba α:

$$\begin{aligned}
\Big(\sum_{i=1}^p \lambda_i \cdot (h_i \circ \alpha)\Big)'(t) &= \sum_{i=1}^p \lambda_i \cdot (h_i \circ \alpha)'(t) = \sum_{i=1}^p \lambda_i \cdot \nabla h_i(\alpha(t)) \cdot \alpha'(t) \\
\Big(\sum_{i=1}^p \lambda_i \cdot (h_i \circ \alpha)\Big)''(t) &= \sum_{i=1}^p \lambda_i \cdot \Big(\alpha'(t) \cdot \nabla^2 h(\alpha(t)) \cdot \alpha'(t) + \nabla h_i(\alpha(t)) \cdot \alpha''(t)\Big) \\
\Big(\sum_{i=1}^p \lambda_i \cdot (h_i \circ \alpha)\Big)''(0) &= \alpha'(0) \cdot \Big(\sum_{i=1}^p \lambda_i \nabla^2 h_i(x)\Big) \cdot \alpha'(0) + \sum_{i=1}^p \lambda_i \cdot \nabla h_i(x) \cdot \alpha''(0) = 0 \ .
\end{aligned}$$

Sumant,

$$\begin{aligned}
0 \leqslant \alpha'(0) \cdot &\Big(\nabla^2 f(x) + \sum_{i=1}^p \lambda_i \cdot \nabla^2 h_i(x)\Big) \cdot \alpha'(0) \\
&+ \underbrace{\Big(\nabla f(x) + \sum_{i=1}^p \lambda_i \cdot \nabla h_i(x)\Big)}_{=\vec{0}} \cdot \alpha''(0) \ ,
\end{aligned}$$

és a dir

$$v \cdot \Big(\nabla^2 f(x) + \sum_{i=1}^p \lambda_i \cdot \nabla^2 h_i(x)\Big) \cdot v \geqslant 0 \ . \quad \square$$

Ara tractarem el cas general, en què S ve determinada per igualtats i desigualtats.

Demostració del 2n cas; condició de primer ordre:

Suposem que $h_1 = 0, \dots, h_p = 0$, $g_{p+1} \leqslant 0, \dots, g_r \leqslant 0$ són les restriccions actives en el mínim local x.

x serà també mínim local de f en la regió

$$\left.\begin{array}{cc}
h_1(x) = 0 & g_{p+1}(x) = 0 \\
\vdots & \vdots \\
h_p(x) = 0 & g_r(x) = 0
\end{array}\right\} .$$

Per tant, el resultat del 1r cas s'aplica a aquest problema que només té restriccions d'igualtat, i obtenim l'existència de les constants $\lambda_1, \dots, \lambda_p, \mu_{p+1}, \dots, \mu_r$ complint la condició necessària de primer ordre. Posem a més $\mu_{r+1} = \cdots = \mu_m = 0$, i ja tenim la primera de les igualtats que hem de comprovar.

La segona també se satisfà, doncs $g_i(x) < 0 \Rightarrow \mu_i = 0$.

Només falta provar que $\mu \geqslant 0$.

Suposem que $\mu_i < 0$ per algun i, $p+1 \leqslant i \leqslant r$.

Traiem la restricció $g_i(x) = 0$ i considerem la superfície $\hat{S}$ formada per les restants restriccions actives. Sigui $\hat{T}(x)$ l'espai tangent a $\hat{S}$ en x.

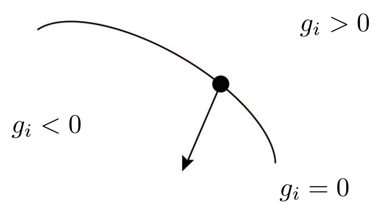
$\nabla g_i(x)$ no és ortogonal a $\hat{T}(x)$: Si ho fos, seria linealment dependent del conjunt $\nabla h_1(x), \dots, \nabla h_p(x)$, $\nabla g_1(x), \dots, \nabla g_r(x)$ (sense el $\nabla g_i(x)$) i això no pot ser per la hipòtesi de punt regular de x. Per tant $\exists\, y \in \hat{T}(x)$ tal que $\nabla g_i(x) \cdot y < 0$.

Llavors

$$\nabla f(x) \cdot y + \lambda \underbrace{\nabla h(x) \cdot y}_{=0} + \underbrace{\mu \nabla g(x) \cdot y}_{>0} = 0, \ \Rightarrow \nabla f(x) \cdot y < 0 \ .$$

(El producte $\nabla h(x) \cdot y$ és zero perquè $y \in \hat{T}(x)$; per altra banda, $\nabla g_k(x) \cdot y = 0$ per tota g_k, excepte per g_i, que dóna $\nabla g_i(x) \cdot y < 0$.)

En definitiva: En la direcció y la funció f disminueix, perquè la derivada direccional de f en la direcció y és negativa; a més aquesta direcció és factible, perquè les derivades direccionals de g en la direcció y són negatives, i per tant ens mantenim dins de les restriccions originals.



□

Demostració del 2n cas; condició de segon ordre:

Si x és un mínim local del problema, també ho és del problema agafant només les restriccions actives. Apliquem doncs la condició anàloga per a restriccions d'igualtat. □

2.1.23 Teorema (condicions suficients de mínim local)

Siguin S i f com en el Teorema 2.1.21.

Donat un punt $x \in S$, si existeixen $\lambda = (\lambda_1, \dots, \lambda_p)$ i $\mu = (\mu_{p+1}, \dots, \mu_m) \geqslant 0$ tals que

a)

$$\left.\begin{aligned} \nabla f(x) + \lambda \cdot \nabla h(x) + \mu \cdot \nabla g(x) = 0 \\ \textit{Per } p+1 \leqslant i \leqslant m, \quad \mu_i g_i(x) = 0 \end{aligned}\right\} ,$$

b) La matriu $M(x)$ del Teorema 2.1.21 compleix $v \cdot M(x) \cdot v > 0$, $\forall v \neq 0$ del subespai

$$\{v : \ \nabla h(x) \cdot v = 0; \ \nabla g_i(x) \cdot v = 0, \ \forall i \textit{ tal que } g_i(x) = 0 \textit{ i } \mu_i > 0\} ,$$

aleshores x és mínim local (estricte) de f en S. □

Observeu que la condició *b)* diu essencialment que $M(x)$ ha de ser definida positiva sobre l'ortogonal de les restriccions actives en x.

2.1.24 Exemple

Vegem un exemple senzill en què la teoria que hem vist és aplicable de manera directa:

$$\begin{array}{l} \min\ f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subj. a:} \\ \left.\begin{array}{r} x_1^2 + x_2^2 \leqslant 5 \\ 3x_1 + x_2 \leqslant 6 \end{array}\right\} . \end{array}$$

Imposem les condicions necessàries de 1r ordre, per trobar candidats a mínim local.

$$\nabla f(x) = (4x_1 + 2x_2 - 10, 2x_1 + 2x_2 - 10)$$
$$\nabla g(x) = \begin{pmatrix} 2x_1 & 2x_2 \\ 3 & 1 \end{pmatrix} .$$

$$\left.\begin{array}{r} 4x_1 + 2x_2 - 10 + \mu_1 \cdot 2x_1 + \mu_2 \cdot 3 = 0 \\ 2x_1 + 2x_2 - 10 + \mu_1 \cdot 2x_2 + \mu_2 \quad\ = 0 \\ \mu_1 \cdot (\ x_1^2 + x_2^2 - 5) = 0 \\ \mu_2 \cdot (3x_1 + x_2 - 6) = 0 \\ \mu_1 \geqslant 0,\ \mu_2 \geqslant 0 \end{array}\right\}$$

- Si $g_1(x) \leqslant 0$ i $g_2(x) \leqslant 0$ són inactives, llavors $\mu_1 = 0$ i $\mu_2 = 0$ i queda

  $$\left.\begin{array}{r} 4x_1 + 2x_2 - 10 = 0 \\ 2x_1 + 2x_2 - 10 = 0 \end{array}\right\} \longrightarrow x_1 = 0,\ x_2 = 5$$

  Satisfà $(0, 5)$ les restriccions? No. Per tant, aquest punt no val.
  (Observeu que aquest cas correspon a resoldre el problema inicial sense restriccions.)
- Si $g_1(x) \leqslant 0$ és activa i $g_2(x) \leqslant 0$ és inactiva, llavors $\mu_2 = 0$, i queda

  $$\left.\begin{array}{r} 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 = 0 \\ 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 = 0 \\ x_1^2 + x_2^2 - 5 = 0 \end{array}\right\} .$$

La resolució d'aquest sistema no és fàcil, però encara es pot fer raonablement. Aïllem μ_1 de la primera equació:

$$\mu_1 = \frac{-4x_1 - 2x_2 + 10}{2x_1} .$$

(Podem suposar que $x_1 \neq 0$, perquè si no es veu ràpidament que queda un sistema incompatible.) Substituint a la segona equació,

$$\begin{array}{l} 2x_1 + 2x_2 - 10 + \dfrac{-4x_1 - 2x_2 + 10}{x_1} x_2 = 0,\ \Rightarrow \\ 2x_1^2 - 2x_1x_2 - 2x_2^2 - 10x_1 + 10x_2 = 0\ . \end{array}$$

De

$$x_1^2 + x_2^2 - 5 = 0$$

obtenim

$$x_2 = \pm\sqrt{5 - x_1^2}\ ,$$

que substituït a l'altre equació i simplificant (això involucra elevar al quadrat l'equació) dóna

$$5x_1^4 - 30x_1^3 + 25x_1^2 + 100x_1 - 100 = 0.$$

Aquest polinomi resulta tenir dues arrels reals:

$$x_1 = \begin{cases} 1 \\ \frac{1}{3}\left[5 - (145 + 30\sqrt{6})^{1/3} - (145 - 30\sqrt{6})^{1/3}\right] \end{cases} .$$

Per la primera,

$$x_2 = \begin{cases} 2 \\ -2 \end{cases} .$$

Per la segona,

$$x_2 = \begin{cases} \sqrt{5 - x_1^2} \\ -\sqrt{5 - x_1^2} \end{cases} .$$

Per al punt $(1, 2)$, s'obté $\mu_1 = \dfrac{-4 - 4 + 10}{2} = 1 > 0$.

Per al punt $(1, -2)$ s'obté $\mu_1 = \dfrac{-4 + 4 + 10}{2} = 5 > 0$.

Per a les altres dues solucions, $\mu_1 \approx -4.07395$, $\mu_1 \approx -8.62383$, que no són admissibles.

Però $x_1 = 1$, $x_2 = -2$, $\mu_1 = 5$ no és solució del sistema. (És una solució espúria afegida al elevar al quadrat per obtenir l'equació de 4rt grau.) Només ens queda un únic candidat: $x_1 = 1$, $x_2 = 2$, $\mu_1 = 1$, que sí és solució del sistema.

Investiguem les condicions suficients:

$$\nabla^2 f(x) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}, \quad \nabla^2 g_1(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \nabla^2 g_2(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \forall x .$$

$$M(x) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}, \quad \forall x .$$

M és definida positiva en tot l'espai. Per tant, $(1, 2)$ és mínim local estricte del problema. De fet, per arribar a aquesta conclusió seria suficient que M fos definida positiva sobre

$$\{v : \ \nabla g_1(1, 2) \cdot v = 0\} = \{v : \ (2, 4) \cdot v = 0\} = \{v : \ 2v_1 + 4v_2 = 0\},$$

subespai vectorial que té per base el vector $(2, -1)$, per exemple. Per tant, seria suficient comprovar que

$$(2, -1) \cdot \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} = (10, 0) \cdot \begin{pmatrix} 2 \\ -1 \end{pmatrix} = 20 > 0 .$$

Observeu, a més, que:

- f és convexa (estrictament), perquè $\nabla^2 f(x) \equiv \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$ és definida positiva.
- El conjunt factible és convex:

$$\left.\begin{array}{rcl} x_1^2 + x_2^2 \leqslant 5 & \leftarrow & \text{cercle (convex)} \\ 3x_1 + x^2 \leqslant 6 & \leftarrow & \text{semiespai (convex)} \end{array}\right\} \text{Intersecció de convexos és convex.}$$

Pel Teorema local-global 2.1.13, $(1, 2)$ és mínim global estricte del problema. Ja no cal continuar. Si haguéssim de continuar, faríem ara els casos:

- Si $g_1(x) \leqslant 0$ és inactiva i $g_2(x) \leqslant 0$ és activa, llavors $\mu_1 = 0$ i queda

$$\left.\begin{array}{r} 4x_1 + 2x_2 - 10 + 3\mu_2 = 0 \\ 2x_1 + 2x_2 - 10 + \ \ \mu_2 = 0 \\ 3x_1 + x_2 - 6 = 0 \end{array}\right\} .$$

  Dóna una solució no admissible: $x_1 = \frac{6}{15}$, $x_2 = \frac{72}{15}$, $\mu_2 = \frac{-6}{15}$.

- Si totes dues són actives, hauríem de resoldre

$$\left.\begin{array}{r} 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0 \\ 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \ \ \mu_2 = 0 \\ x_1^2 + x_2^2 - 5 = 0 \\ 3x_1 + x_2 - 6 = 0 \end{array}\right\} .$$

2.2 Optimització en $\mathbb{R}$

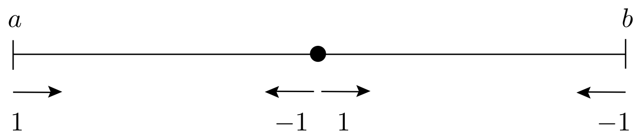
Estudiarem en aquest apartat el problema

$$\min_{x \in S} f(x)$$

en el cas unidimensional ($x \in \mathbb{R}$), essent S un interval $[a,b] \subset \mathbb{R}$. Repassarem primer el que ens diu la teoria de l'apartat anterior en aquest cas.

2.2.1 Aplicació de la teoria d'extrems

- Direccions factibles:



  En qualsevol punt de l'interval obert $]a,b[$, hi ha dues direccions factibles, que corresponen als vectors unidimensionals $(+1)$ (anar cap a la dreta) i (-1) (anar cap a l'esquerra). En el punt a només és factible la direcció $(+1)$ (cap a la dreta), i en el b la direcció (-1) (cap a l'esquerra).

- Condicions necessàries:
  Es dedueixen de la teoria general del Subapartat 2.1.1 (concretament, dels enunciats 2.1.6, 2.1.7) els fets següents:

  $x \in]a,b[$ mínim local $\Rightarrow f'(x) = 0$ i $f''(x) \geqslant 0$.
  a mínim local $\Rightarrow f'(a) \geqslant 0$ i $[f'(a) = 0 \Rightarrow f''(a) \geqslant 0]$.
  b mínim local $\Rightarrow f'(b) \leqslant 0$ i $[f'(b) = 0 \Rightarrow f''(b) \geqslant 0]$.

- Condicions suficients:
  Es dedueix de la teoria general del Subapartat 2.1.1 (concretament, de l'enunciat 2.1.9) que:

  $x \in]a,b[$, $f'(x) = 0$ i $f''(x) > 0 \Rightarrow x$ mínim local.

  Es dedueixen de les condicions suficients del Subapartat 2.1.3 (concretament, de l'enunciat 2.1.23):

  $f'(a) > 0$ o $[f'(a) = 0$ i $f''(a) > 0] \Rightarrow a$ mínim local.
  $f'(b) < 0$ o $[f'(b) = 0$ i $f''(b) > 0] \Rightarrow b$ mínim local.

Anem a detallar la deducció d'aquests dos últims fets. Ho farem pel primer; el segon és anàleg.

2.2.1 Proposició

Sigui $f\colon [a,b] \to \mathbb{R}$ de classe C^2.

Si $f'(a) > 0$ o bé $[f'(a) = 0$ i $f''(a) > 0]$, aleshores a és mínim local.

Demostració: El problema

$$\min_{x \in [a,b]} f(x)$$

es pot formular com

$$\begin{aligned} &\min \ f(x) \\ &\text{subj. a:} \\ &\left.\begin{aligned} g_1(x) &:= x - b \leqslant 0 \\ g_2(x) &:= -x + a \leqslant 0 \end{aligned}\right\} \end{aligned}$$

Segons el Teorema de condicions suficients de mínim local (Teorema 2.1.23), si existeixen $\mu_1 \geqslant 0$, $\mu_2 \geqslant 0$ tals que

a)

$$\left.\begin{aligned} f'(a) + \mu_1 g_1'(a) + \mu_2 g_2'(a) = 0 \\ \mu_1 g_1(a) = 0 \\ \mu_2 g_2(a) = 0 \end{aligned}\right\}$$

b) $M(a) := f''(a) + \mu_1 g_1''(a) + \mu_2 g_2''(a)$ compleix $v \cdot M(a) \cdot v > 0$, $\forall v \neq 0$ del subespai

$$\{v : \ g_i'(a) \cdot v = 0, \ \forall i \text{ tal que } g_i(a) = 0 \text{ i } \mu_i > 0\},$$

aleshores a és mínim local.

De la segona equació de *a)*, es dedueix $\mu_1 = 0$, i la primera queda $f'(a) + \mu_2 \cdot (-1) = 0$, d'on obtenim $\mu_2 = f'(a)$.

En *b)*, tenim que $M(a) = f''(a)$. Hi ha dos casos:

1. Si $\mu_2 > 0$, queda la condició

$$v \cdot f''(a) \cdot v > 0, \ \forall v \in \{v \neq 0 : \ (-1) \cdot v = 0\} = \emptyset \ ,$$

   és a dir, cap condició. És el cas corresponent a la possibilitat $f'(a) > 0$.

2. Si $\mu_2 = 0$, queda

$$v \cdot f''(a) \cdot v > 0, \ \forall v \in \{v \neq 0 : \ —\} = \mathbb{R} - \{0\} \ ,$$

   és a dir, $v \cdot f''(a) \cdot v > 0$, $\forall v \neq 0$, és a dir, $f''(a) > 0$. És el cas corresponent a $f'(a) = 0$ i $f''(a) > 0$. □

2.2.2 Mètodes iteratius

La teoria de condicions necessàries i suficients ens diu doncs que, per trobar tots els mínims locals d'una funció en un interval $[a, b]$, el que cal fer és:

1. Calcular f'.
2. Trobar les solucions de $f'(x) = 0$ en $]a, b[$.
3. Quedar-nos amb aquelles solucions x^* tals que $f''(x^*) > 0$.
4. Estudiar apart els punts frontera a i b.

Normalment això no funciona, perquè no es poden trobar exactament les solucions de l'equació $f'(x) = 0$. També pot passar que $f'(x^*) = 0$ i $f''(x^*) = 0$, i la teoria no decideix si x^* és un mínim local o no.

El que es fa és utilitzar procediments iteratius que van obtenint, a partir d'un punt factible, un altre punt millor, a partir d'aquest un altre millor, i així successivament.

El mètode del símplex és un procediment iteratiu que vam demostrar que conduïa a la solució del problema després d'una quantitat finita d'iteracions. Els mètodes iteratius que s'apliquen en PNL no donen, generalment, la solució en un nombre finit d'iteracions, sinó que produeixen una successió de solucions aproximades que convergeix a la solució exacta en el límit.

Això fa que a la pràctica hàgim de conformar-nos sempre amb solucions aproximades. El càlcul s'acaba quan veiem que la nostra solució aproximada ja no millora gaire entre una iteració i la següent, i aleshores decidim parar.

Veurem dos mètodes iteratius d'optimització unidimensional: Un d'ells (el mètode de Newton) requereix que la funció es pugui derivar dues vegades; l'altre (l'ajust quadràtic) es pot aplicar a qualsevol funció contínua.

2.2.2 Newton

Considerem el problema $\min_{x \in \mathbb{R}} f(x)$, amb $f \in C^2$.

Partim d'un punt inicial x_1.

Aproximem f al voltant del punt x_1, per un polinomi de segon grau, usant els valors $f(x_1)$, $f'(x_1)$, $f''(x_1)$ (polinomi de Taylor):

$$P(x) = f(x_1) + f'(x_1) \cdot (x - x_1) + \frac{1}{2} f''(x_1) \cdot (x_1 - x)^2.$$

Calculem el mínim de $P(x)$:

$$x_2 = x_1 - \frac{f'(x_1)}{f''(x_1)} \ .$$

A partir d'aquest,

$$x_3 = x_2 - \frac{f'(x_2)}{f''(x_2)} \ .$$

En general, obtenim la fórmula iterativa

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \ . \qquad \square$$

Si el problema és $\min\limits_{x \in S} f(x)$, on S és un interval de $\mathbb{R}$, aleshores pot passar que al fer una iteració sortim de S. Aleshores hem de considerar investigar si l'extrem de l'interval és un mínim relatiu, o tornar a començar el mètode de Newton o partir d'un punt prop de l'extrem.

El mètode de Newton convergeix molt ràpid si partim d'un punt que està prop del mínim.

2.2.3 Ajust quadràtic de tres punts

Considerem ara el problema $\min\limits_{x \in S} f(x)$, amb f contínua, i S un interval (potser il·limitat).

Com en el mètode de Newton, es tracta d'aproximar la funció per una paràbola, però en lloc d'usar un sol punt i el valor de f, f', f'' en aquest punt, s'utilitzen tres punts, i només el valor de f en ells.

Procediment:

0) Trobar tres punts per començar $x_1 < x_2 < x_3$ tals que
$$f(x_1) \geqslant f(x_2) \leqslant f(x_3) \ .$$
(Això implica l'existència d'un mínim local entre x_1 i x_3.)

1) Trobar el polinomi $P(x) = ax^2 + bx + c$ tal que $P(x_1) = f(x_1)$, $P(x_2) = f(x_2)$, $P(x_3) = f(x_3)$. Per trobar-lo cal resoldre el sistema en (a, b, c)
$$\left.\begin{aligned} ax_1^2 + bx_1 + c &= f(x_1) \\ ax_2^2 + bx_2 + c &= f(x_2) \\ ax_3^2 + bx_3 + c &= f(x_3) \end{aligned}\right\} \ ,$$
o bé usar la fórmula directa
$$P(x) = f(x_1) \cdot \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + f(x_2) \cdot \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_1)} + f(x_3) \cdot \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} \ .$$

2) Trobar el mínim de P:
$$x = \frac{-b}{2a} \ .$$

3) Amb el x trobat i dos dels punts x_1, x_2, x_3, construir una nova terna de punts x_1, x_2, x_3 amb $f(x_1) \geqslant f(x_2) \leqslant f(x_3)$. Tornar al pas 1.

Detallem l'últim pas: Suposem que $x_1 \leqslant x \leqslant x_2 < x_3$; aleshores

a) Si $f(x) \leqslant f(x_2)$, agafem com a nova terna x_1, x, x_2.

b) Si $f(x) \geqslant f(x_2)$, agafem com a nova terna x, x_2, x_3.

Si $x_1 < x_2 \leqslant x \leqslant x_3$, es fa anàlogament. $\square$

Aquest mètode és més lent que el mètode de Newton, però té l'avantatge que no cal calcular derivades. Pot ser una bona idea combinar ambdós mètodes: Començar fent ajust quadràtic i, quan sospitem que estem prop del mínim, usar el mètode de Newton.

2.3 Optimització sense restriccions a $\mathbb{R}^n$

Considerem ara el problema

$$\min \ f(x) \ ,$$

amb $f\colon \mathbb{R}^n \to \mathbb{R}$, sense restriccions, és a dir, amb $S = \mathbb{R}^n$ com a conjunt factible.

Anàlogament al cas unidimensional de l'Apartat 2.2, la teoria de condicions necessàries i suficients ens diu que, en aquest cas, per tal de trobar els mínims locals cal, en principi:

1. Calcular ∇f.
2. Trobar les solucions del sistema $\nabla f(x) = 0$.
3. Quedar-nos amb aquelles solucions x tals que $\nabla^2 f(x)$ és una matriu definida positiva.

Els problemes per aplicar aquest procediment són també anàlegs al cas d'una variable. Essencialment, el que passa és que el sistema $\nabla f(x) = 0$, o sigui,

$$\begin{cases} \dfrac{\partial f}{\partial x_1}(x_1, \dots, x_n) = 0 \\ \qquad\vdots \\ \dfrac{\partial f}{\partial x_n}(x_1, \dots, x_n) = 0 \ , \end{cases}$$

no es pot resoldre de manera exacta, llevat de casos molt senzills. Per tant, gairebé sempre cal utilitzar un mètode iteratiu per aproximar el punt optimal.

Hi ha una classe de mètodes per a l'optimització sense restriccions a $\mathbb{R}^n$ que responen a l'esquema general següent:

0. Partir d'un punt inicial x.
1. Escollir una direcció de moviment.
2. Trobar el mínim de la funció objectiu restringida a la recta determinada pel punt x i la direcció de moviment.
3. Prendre el punt on s'assoleix el mínim com a nou punt, i tornar a 1.

Tots els mètodes que responen a aquest esquema es distingeixen entre si en l'elecció de la direcció de moviment. El pas 2 és una part comuna a tots ells. Aquest pas és una optimització a $\mathbb{R}$ que s'anomena *exploració lineal*.

Més concretament: Suposem que tenim un punt factible x i hem escollit la direcció de moviment v. La funció f restringida a la recta que passa per x i té la direcció v es pot escriure com una funció d'una variable α:

$$g(\alpha) = f(x + \alpha \cdot v) \ .$$

Si donem α^* com a mínim de g, el nou punt serà $x + \alpha^* v$.

Veurem dos mètodes molt senzills sota aquest esquema, el de descens de més gran pendent i el de descens per coordenades. Després veurem el mètode de Newton, que es d'estil diferent.

2.3.1 Mètode del descens de més gran pendent

Sigui $f \in C^1$. El mètode del descens de més gran pendent es basa en el fet que el gradient de f en un punt x sempre assenyala la direcció d'ascens de més gran pendent. Això es pot comprovar resolent, mitjançant la teoria de condicions necessàries i suficients, el problema

$$\max_{\|v\|=1} \nabla f(x) \cdot v \ .$$

El màxim s'assoleix en $v = \frac{\nabla f(x)}{\|\nabla f(x)\|}$.

Per tant, $-\nabla f(x)$ assenyala la direcció de descens de més gran pendent. Aquesta és la direcció de moviment que escollim.

En definitiva, el mètode consisteix en generar una successió de solucions aproximades amb la fórmula iterativa

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

on α_k és la solució del problema

$$\min_{\alpha \geqslant 0} g(\alpha) := f(x_k - \alpha \nabla f(x_k)) \ . \qquad \square$$

2.3.2 El descens de més gran pendent en el cas quadràtic

Una *funció quadràtica* en $\mathbb{R}^n$ (polinomi de segon grau en n variables), es pot escriure sempre així:

$$f(x) = \frac{1}{2} x \cdot Q \cdot x + b \cdot x + c \ ,$$

on Q és una matriu simètrica.

Es té que

$$\nabla f(x) = Q \cdot x + b \ , \quad \nabla^2 f(x) = Q \ .$$

Si Q és semidefinida positiva aleshores la funció és convexa i té mínim global. Si Q és definida positiva, aleshores la funció és estrictament convexa i hi ha un únic mínim global.

En aquest cas quadràtic l'exploració lineal es pot fer exacta: En efecte, si posem per abreviar $h_k := \nabla f(x_k) = Qx_k + b$, aleshores

$$\begin{aligned}
g(\alpha) &= f(x_k - \alpha h_k) = \frac{1}{2}(x_k - \alpha h_k) Q (x_k - \alpha h_k) + b(x_k - \alpha h_k) + c \ . \\
g'(\alpha) &= \alpha h_k Q h_k - \frac{1}{2} h_k Q x_k - \frac{1}{2} x_k Q h_k - b h_k \\
&= \alpha h_k Q h_k - h_k Q x_k - h_k b \\
&= \alpha h_k Q h_k - h_k h_k \ . \\
g'(\alpha) &= 0 \Leftrightarrow \alpha = \frac{h_k h_k}{h_k Q h_k} \ .
\end{aligned}$$

Ens queda la fórmula iterativa

$$x_{k+1} = x_k - \frac{h_k h_k}{h_k Q h_k} \cdot h_k \ . \qquad \square$$

2.3.3 Mètode del descens per coordenades

En el mètode de descens per coordenades les direccions de moviment són les dels eixos de coordenades: Si estem en un punt $x^* = (x_1^*, \dots, x_n^*)$, fer un descens en la coordenada i-èsima és trobar

$$\min_{x_i} \ f(x_1^*, \dots, x_{i-1}^*, x_i, x_{i+1}^*, \dots, x_n^*)$$

(fixem totes les coordenades excepte la i-èsima i calculem el mínim unidimensional respecte aquesta).

El mètode del descens per coordenades consisteix en aplicar el descens respecte la coordenada i-èsima per $i = 1$, $i = 2$, $i = 3, \dots$, $i = n$; $i = 1$, $i = 2$, $i = 3, \dots$, $i = n$; $i = 1, \dots$ cíclicament.

L'avantatge respecte el descens de més gran pendent és que no cal calcular el gradient. Però convergeix més lentament. $\square$

2.3.4 Mètode de Newton

El mètode de Newton no es basa en l'elecció d'una direcció de moviment i posterior exploració lineal, sinó que produeix directament un punt a partir de l'anterior. És la generalització directa del mètode de Newton que hem vist per al cas unidimensional. Es necessita $f \in C^2$.

Considerem el polinomi de Taylor de grau 2 de f en el punt x_k:

$$P(x) = f(x_k) + \nabla f(x_k)(x - x_k) + \frac{1}{2}(x - x_k)\nabla^2 f(x)(x - x_k) \ .$$

El nou punt x_{k+1} serà el mínim de P: Derivem P,

$$\nabla P(x) = \nabla f(x_k) + \nabla^2 f(x)(x - x_k) \ ,$$

igualem a zero i aïllem x:

$$x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k) \ .$$

Com en el cas de dimensió 1, l'algorisme és molt ràpid quan estem prop de l'òptim. L'inconvenient és haver d'avaluar derivades segones i invertir una matriu a cada iteració.

2.4 Optimització amb restriccions a $\mathbb{R}^n$

Aquest tema és molt ampli, però ens restringirem a estudiar *un* mètode, enfocat al problema particular

$$\begin{aligned} \min \ f(x)& \\ \left.\begin{aligned} Ax = d \\ 0 \leqslant x \leqslant u \end{aligned}\right\}& \end{aligned}$$

on $x \in \mathbb{R}^n$, f és no lineal, les restriccions són lineals, hi ha condicions de no negativitat i cotes superiors per a totes les variables, el vector de termes independents d és positiu, i la matriu A és $m \times n$ amb rang m. Eventualment, totes o algunes de les cotes u_j poden ser $+\infty$, el que equival a dir que no hi ha cotes superiors.

En programació lineal, el mètode del símplex es pot adaptar al problema amb restriccions $0 \leqslant x \leqslant u$, de manera que les cotes superiors es tracten de manera implícita, com les de positivitat. El que farem aquí il·lustra quines són les modificacions que caldria fer en el símplex.

En cas d'haver-hi variables lliures i restriccions lineals de desigualtat, podem convertir el problema a un amb la forma anterior, exactament igual que en el cas de PL.

El mètode que descriurem s'anomena *mètode del gradient reduït.*

2.4.1 Notació

Sigui B una base formada per columnes de A. Qualsevol punt factible x es pot expressar com

$$x = (x_B, x_S, x_N)$$

amb

$$\begin{aligned} x_j = 0 \text{ o } u_j,\ \forall j \in \mathrm{N} \ , \\ 0 < x_j < u_j,\ \forall j \in \mathrm{S} \ , \end{aligned}$$

i x_B determinades a partir de x_S i x_N:

$$Bx_B + Sx_S + Nx_N = d \implies x_B = B^{-1}d - B^{-1}Sx_S - B^{-1}Nx_N \ .$$

Les variables de B s'anomenen *bàsiques.*

Les variables de N s'anomenen *no-bàsiques.*

Les variables de S s'anomenen *superbàsiques.* □

La idea del mètode és la següent:

Amb les variables no-bàsiques fixades a una de les seves cotes, fer variar les superbàsiques fins a tenir un òptim de f sota aquestes restriccions addicionals (usant la mateixa idea del descens de més gran pendent).

Un cop trobat aquest òptim, passem una variable no-bàsica a superbàsica, relaxant així les restriccions addicionals.

També, sota certes circumstàncies, caldrà passar una variable superbàsica a no-bàsica, o intercanviar una bàsica i una no-bàsica.

2.4.2 Càlculs previs

Com que x_B queden determinades per x_S i x_N, podem definir una funció g de la manera següent:

$$f(x_B, x_S, x_N) = f(x_B(x_S, x_N), x_S, x_N) =: g(x_S, x_N) \ ,$$

i calcular el seu gradient en termes del de f:

$$\begin{aligned}\nabla g = (\nabla_S g, \nabla_N g) &= \big(\nabla_B f \cdot \nabla_S x_B + \nabla_S f,\ \nabla_B f \cdot \nabla_N x_B + \nabla_N f\big) \\ &= \big(\nabla_B f \cdot (-B^{-1}S) + \nabla_S f,\ \nabla_B f \cdot (-B^{-1}N) + \nabla_N f\big) \ .\end{aligned}$$

Aquesta serà la direcció de màxim creixement sobre la varietat

$$x_B = B^{-1}d - B^{-1}Sx_S - B^{-1}Nx_N \ .$$

La direcció de màxim decreixement serà la de signe oposat. Si a més volem deixar fixes les variables no-bàsiques, només cal projectar sobre les components de S, i obtenim finalment la direcció en què ens mourem:

$$\begin{aligned} p_S &:= \nabla_B f \cdot B^{-1}S - \nabla_S f \ . \\ p_N &:= 0 \ . \end{aligned}$$

De l'expressió de x_B en termes de les altres variables, veiem que les variables bàsiques es mouran en la direcció

$$p_B := -B^1 S \cdot p_S \ .$$

El vector $p = (p_B, p_S, p_N)$ s'anomena *gradient reduït.* □

2.4.3 Algorisme

Partim d'un punt factible x i una subdivisió de les variables en bàsiques, superbàsiques i no-bàsiques: $x = (x_B, x_S, x_N)$.

1. Calculem el gradient reduït $p = (p_B, p_S, 0)$.

2. Si $p_S \equiv 0$, passem una variable no-bàsica a superbàsica.
Tornem al punt 1.

3. Fem exploració lineal en la direcció del gradient reduït:
Determinem α^* com el punt on s'assoleix

$$\min_{0 \leqslant \alpha \leqslant \bar{\alpha}} g(\alpha) = f(x + \alpha p) \ ,$$

essent

$$\bar{\alpha} = \max\{\alpha \geqslant 0 :\ 0 \leqslant x + \alpha p \leqslant u\} \ .$$

Si definim

$$\begin{aligned} \alpha_1 &:= \max\{\alpha \geqslant 0 :\ 0 \leqslant x_B + \alpha p_B \leqslant u_B\} \\ \alpha_2 &:= \max\{\alpha \geqslant 0 :\ 0 \leqslant x_S + \alpha p_S \leqslant u_S\} \end{aligned}$$

aleshores

$$\bar{\alpha} = \min\{\alpha_1, \alpha_2\} \ .$$

4. El nou punt factible serà $x + \alpha^* p$.

a) Si $\alpha^* = \alpha_1$, hi ha una variable bàsica que arriba a una de les seves cotes. La intercanviem per una de no-bàsica. Tornem al punt 1.

b) Si $\alpha^* = \alpha_2$, hi ha una variable superbàsica que arriba a una de les seves cotes i passa a ser no-bàsica. Tornem al punt 1.

c) Si $\alpha^* \neq \alpha_1$, $\alpha^* \neq \alpha_2$, tornem al punt 1. □

2.4.4 Detalls de l'algorisme

En els punts 2 i 4, cal seleccionar una variable no-bàsica per passar a superbàsica o bàsica. Com? Suposem que estem en $x = (x_B, x_S, x_N)$ i ens volem moure en una certa direcció factible $v = (v_B, v_S, v_N)$. Per α prou petit, tindrem que

$$g(x_S + \alpha v_S, x_N + \alpha v_N) - g(x_S, x_N) \approx (-\nabla_B f \cdot B^{-1} S + \nabla_S f,\ -\nabla_B f \cdot B^{-1} N + \nabla_N f) \begin{pmatrix} \alpha v_S \\ \alpha v_N \end{pmatrix} .$$

Sigui $j \in N$. Si $x_j = 0$, per força $v_j \geqslant 0$ (ja que v és direcció factible). Si deixem igual totes les altres variables (que és el que es fa), per tal que la funció disminueixi cal que

$$(-\nabla_B f \cdot B^{-1} N + \nabla_N f)_j < 0 .$$

Si $x_j = u_j$, per força $v_j \leqslant 0$, i caldrà

$$(-\nabla_B f \cdot B^{-1} N + \nabla_N f)_j > 0 .$$

Si no hi ha cap variable no-bàsica que compleixi les condicions anteriors, és que estem en un punt optimal.

Típicament, es comença a aplicar l'algorisme sense variables superbàsiques preseleccionades, i el primer que es fa és seleccionar una variable no bàsica per esdevenir superbàsica. □

3. Programació entera

3.1 Introducció

Quan plantegem problemes de programació lineal o no lineal en els quals les variables representen objectes que volem produir o comprar, podem distingir tres situacions típiques:

- Les quantitats dels objectes es mesuren per volum, pes, o altre unitat de mesura contínua. Llavors, la formulació com a problema de PL o PNL que hem vist als capítols anteriors és, amb tota seguretat, l'adequada.
- Els objectes a produir es compten per unitats enteres, però un resultat fraccionari es pot interpretar com un "ritme" o "taxa" de producció. Per exemple, si estem intentant esbrinar les unitats òptimes a produir setmanalment de certs objectes, el resultat $x_1 = 3.25$ sovint es pot interpretar com un ritme de producció que també es podria expressar com "13 objectes cada quatre setmanes". En aquesta situació també són adequats els models que hem vist fins ara.
- Els objectes es compten per unitats enteres i no té sentit considerar-ho com a ritme o taxa. Per exemple, en un problema en què calgui comprar una flota d'avions de diferents tipus, una sola vegada, una quantitat fraccionària d'avions no tindria cap sentit. En aquest cas, els models de PL i PNL no representen adequadament el problema real.

En aquest capítol veurem models per a l'última d'aquestes tres situacions. És a dir, problemes d'investigació operativa formulables, en principi, com a problemes de PL o PNL, però als quals hem d'afegir la restricció que algunes o totes les variables han de ser enteres.

Després d'aquest apartat introductori, estudiarem el mètode anomenat "Branch & Bound" per resoldre aquests models. A l'últim apartat del capítol veurem que les "variables binàries" (variables restringides a valer 0 o 1) són útils per modelar certes situacions especials.

3.1.1 Exemple

Considereu el senzill problema següent, en què s'exigeix que les dues variables siguin enteres:

$$
\begin{aligned}
&\max\ z = 10x_1 + 9x_2 \\
&\text{subj. a:} \\
&\left.\begin{array}{r}
x_1 + 2x_2 \leqslant 15 \\
6x_1 + 5x_2 \leqslant 60 \\
x_1 \qquad\ \leqslant\ \ 8 \\
x_1 \geqslant 0, \quad x_2 \geqslant 0 \\
x_1, x_2 \in \mathbb{Z}
\end{array}\right\}
\end{aligned}
$$

Per resoldre'l, se'ns pot acudir, en primer lloc, deixar de banda les restriccions d'integritat, resoldre el problema de PL resultant (s'obté $x_1 = 6 + \frac{3}{7}$, $x_2 = 4 + \frac{2}{7}$), i arrodonir cada component de la solució a l'enter més proper:

$$x_1 = 6 \ , \quad x_2 = 4 \ .$$

És una solució factible (podria no ser-ho!) però no sabem si és òptima entre totes les solucions enteres. De fet, no ho és:

$$z(6,4) = 96 \ , \quad z(7,3) = 97 \ , \quad z(8,2) = 98 \ .$$

Aquest últim valor resulta ser l'òptim. Podem observar que el vertader punt optimal pot estar lluny del que s'obté per arrodoniment, però que el valor de la funció objectiu no és molt diferent. Per tant, tot i que arrodonint no s'obté en general el millor punt, sí que s'obté habitualment una bona aproximació, que pot ser útil si hi ha pressa en obtenir quelcom.

En aquest exemple, com que només hi ha dues variables, està clar que l'òptim es pot trobar gràficament, dibuixant els conjunts de nivell de z i els punts de coordenades enteres que hi ha dins del polígon delimitat per les altres restriccions. També està clar que això només funciona per a dues variables.

Una altra idea que només funciona per a problemes molt petits és avaluar z en tots els punts de coordenades enteres i escollir entre ells el valor òptim.

Si no es poden aplicar aquestes idees elementals, aleshores el problema és més difícil i llarg de tractar que en el cas continu.

3.1.2 Definicions

Un problema de *Programació Entera* (PE) és aquell que es pot escriure com:

$$\begin{array}{l} \quad \min \ f(x) \\ \text{subj. a:} \\ \quad \left.\begin{array}{l} h_i(x) = 0, \ i = 1, \dots, p \\ g_i(x) \leqslant 0, \ i = p+1, \dots, m \\ \forall j \in J, \ x_j \in \mathbb{Z} \end{array}\right\} \end{array}$$

on $J \subset \{1, \dots, n\}$.

Diem, més concretament, que és un problema de PE *lineal* si totes les funcions involucrades són afins. Si hi ha alguna funció no afí, diem que és un problema de PE *no lineal*.

(De vegades és parla de *programació entera mixta* i es reserva el nom de programació entera pel cas particular en què *totes* les variables han de ser enteres. No farem aquesta distinció.) □

Ens restringirem, per simplicitat, al cas lineal,

$$\begin{array}{l} \quad \min \ z = c \cdot x + c_0 \\ \text{subj. a:} \\ \quad \left.\begin{array}{r} Ax = d \\ x \geqslant 0 \\ \forall j \in J, \ x_j \in \mathbb{Z} \end{array}\right\} \end{array}$$

encara que tot el que segueix es podria generalitzar al cas no lineal. Veurem un mètode general per resoldre tots els problemes d'aquesta mena. Altres mètodes poden ser més eficients si el problema té alguna forma particular.

3.2 Mètode "Branch & Bound"

El mètode "Branch & Bound" per resoldre problemes de PE es pot esquematitzar de la manera següent:

- Resolem primer el problema de programació entera com un problema de programació lineal, prescindint de les restriccions d'integritat.
  Si la solució satisfà casualment les restriccions d'integritat, ja hem acabat. Si no, continuem.
- Suposem que, en el punt òptim del problema de programació lineal, una variable x_j que hauria de tenir un valor enter, no el té. Suposem $x_j = s$.
  Construïm dos nous problemes de programació lineal, afegint a un la restricció $x_j \leqslant \lfloor s \rfloor$, i a l'altre $x_j \geqslant \lfloor s \rfloor + 1$. (La notació $\lfloor s \rfloor$ vol dir "part entera de s".)
- Resolem els dos nous problemes de programació lineal, i apliquem la mateixa ramificació.
- Es va construint d'aquesta manera un arbre de problemes de PL. Cadascun dels nodes de l'arbre pot ser:
  - ◯ Un problema que cal subdividir.
  - □ Un problema amb solució òptima complint les restriccions d'integritat. No ramifiquem.
  - ⊗ Un problema impossible. No ramifiquem.
- Òptim: Entre tots els nodes □ de l'arbre, el que doni millor valor de z.
- "Acotació": El valor de z en un problema de l'arbre no és millor que el valor de z del seu pare. Aquest fet estalvia feina si s'examinen els nodes en *preordre*, o sigui si s'examina tota la descendència de cada node abans d'examinar la del seu germà.

3.2.1 Exemple

Considereu el problema

$$
\begin{aligned}
&\max\ z = 5x_1 + 8x_2 \\
&\text{subj. a:} \\
&\left.\begin{aligned}
x_1 + \ \ x_2 &\leqslant \ 6 \\
5x_1 + 9x_2 &\leqslant 45 \\
x_j \geqslant 0, \quad j &= 1,2,3 \\
x_1, x_2 &\in \mathbb{Z}
\end{aligned}\right\}
\end{aligned}
$$

que té les característiques següents:

	Òpt. continu	Arrodoniment	Factible més proper	Òpt. enter
x_1	2.25	2	2	0
x_2	3.75	4	3	5
z	41.25	No factible	34	40

Anem a trobar l'òptim enter usant el mètode Branch & Bound. En primer lloc, resolem el problema de PL associat, traient les restriccions d'integritat, i obtenim

$$x_1 = 2.25\ , \quad x_2 = 3.75\ , \quad z = 41.25\ .$$

Podem assegurar que l'òptim serà pitjor que aquest valor (o sigui $\leqslant 41.25$). (I en aquest cas particular, com que els coeficients de la funció objectiu són enters, podem afirmar que serà $\leqslant 41$.)

Observem que el punt òptim no té coordenades enteres. Per tant, l'arrel és de tipus ◯ i hem de crear dos fills. Per exemple, escollim x_2 per ramificar. Posem la restricció addicional $x_2 \leqslant 3$ a un fill i $x_2 \geqslant 4$ a l'altre. Obtenim dos nous problemes amb regions factibles que anomenarem S_1 i S_2, respectivament.

El germà gran té solució

$$x_1 = 1.8\ , \quad x_2 = 4\ , \quad z = 41\ .$$

Com que x_1 no és enter, el node és $\bigcirc$ i fem dos fills, amb restriccions addicionals $x_1 \leqslant 1$ i $x_1 \geqslant 2$, donant lloc a les regions factibles S_3 i S_4, respectivament. El problema sobre S_4 és impossible. Per tant, és un node $\bigotimes$ i ja no tindrà fills. Per a S_3, l'òptim és

$$x_1 = 1 \ , \quad x_2 = 4 + \tfrac{4}{9} \ , \quad z = 40 + \tfrac{5}{9} \ .$$

Subdividim afegint $x_2 \leqslant 4$ i $x_2 \geqslant 5$, obtenint noves regions S_5 i S_6. S_5 és un segment, i l'òptim és

$$x_1 = 1 \ , \quad x_2 = 4 \ , \quad z = 37 \ ,$$

que és factible del problema original, i per tant tenim un node $\square$. A més, ens dóna una cota inferior del valor òptim: Ara sabem que el valor de z òptim està en $[37, 41]$. Si només ens interessa una primera aproximació del problema, podem parar aquí i sabem l'error que cometem.

Analitzem ara el seu germà (regió S_6). Només hi ha un punt factible:

$$x_1 = 0 \ , \quad x_2 = 5 \ , \quad z = 40 \ .$$

És un altre node $\square$, i la cota inferior és millor que la d'abans. Ara sabem que l'òptim és en $[40, 41]$.

Ens queda per investigar la regió S_2, el fill petit de l'arrel. La solució és

$$x_1 = 3 \ , \quad x_2 = 3 \ , \quad z = 39 \ .$$

És un node $\square$ i l'anàlisi de l'arbre s'ha acabat. Com que aquest node $\square$ no és millor que l'altre, la solució definitiva és

$$x_1 = 0, \quad x_2 = 5, \quad z = 40 \ .$$

És important notar que, encara que sobre S_2 el problema no hagués tingut solució entera, no caldria subdividir-lo, gràcies a l'acotació (part "bound" de l'algorisme). En efecte, cap descendent d'aquest node podrà millorar el valor 39, i per tant és inútil continuar. □

3.3 Variables binàries

En aquest apartat veurem la utilitat de les *variables binàries* (variables que només poden valer 0 o 1) en la modelització d'alguns problemes d'investigació operativa. No hi haurà cap mètode nou de resolució. El mètode Branch & Bound, per exemple, és en principi aplicable a tots els exemples ($x \in \{0, 1\}$ és equivalent a $x \in \mathbb{Z}$ i $0 \leqslant x \leqslant 1$.) Evidentment, cada situació concreta pot tenir mètodes més eficients adaptats a l'estructura del problema.

3.3.1 Problema de la motxilla

Situació: Anem d'excursió i volem omplir la motxilla de coses que ens han de ser útils. Però el volum d de la motxilla és limitat i no hi cap tot el que voldríem. Suposem que cada objecte o_j té un valor (utilitat) c_j, i un volum a_j. Volem fer màxima la utilitat total. Quins objectes agafem? □

Model: Considerem una variable binària x_j per a cada objecte o_j. Agafem l'objecte si $x_j = 1$; no l'agafem si $x_j = 0$. El problema es pot formular així:

$$\begin{array}{l} \max \ z = c_1 x_1 + \cdots + c_n x_n \\ \text{subj. a:} \\ \left.\begin{array}{r} a_1 x_1 + \cdots + a_n x_n \leqslant d \\ x_j \in \{0,1\}, \ \forall j \end{array}\right\} \end{array}$$ □

Exemple: Suposem que tenim cinc objectes a posar en una motxilla de volum 15, amb valors i volums següents:

Objecte	Valor	Volum
o_1	5	5
o_2	3	4
o_3	6	7
o_4	6	6
o_5	2	2

Ordenem els objectes en ordre decreixent de quocients c_j/a_j (valor/volum). Els empats es resolen arbitràriament:

Objecte	c_j/a_j
o_1	1
o_4	1
o_5	1
o_3	6/7
o_2	3/4

Es pot demostrar que la solució òptima del problema de PL associat (traient les restriccions $x_j \in \{0,1\}$) consisteix en anar posant els objectes a la motxilla en aquest ordre, fins que es desbordi, i aleshores posar només la fracció justa de l'últim objecte. A l'exemple, posarem

$$o_1 \text{ (vol= 5)} + o_4 \text{ (vol= 6)} + o_5 \text{ (vol= 2)} + \tfrac{2}{7} o_3 \text{ (vol= 2)}.$$

La funció objectiu val $z = 14 + \frac{5}{7}$.

Fem "branch" introduint les restriccions $x_3 = 0$ (l'anomenarem problema [1]) i $x_3 = 1$ (serà el problema [2]). Fem el problema [1]:

Objecte	c_j/a_j
o_1	1
o_4	1
o_5	1
o_2	3/4

Posem a la motxilla

$$o_1 \ (5) + o_4 \ (6) + o_5 \ (2) + \tfrac{1}{2} o_2 \ (2) \ . \qquad z = 14 + \tfrac{1}{2} \ .$$

Fem "branch" introduint les restriccions $x_2 = 0$ (problema [1.1]) i $x_2 = 1$ (problema [1.2]). Comencem pel problema [1.1]:

Posem a la motxilla

$$o_1 \ (5) + o_4 \ (6) + o_5 \ (2) \ . \qquad z = 13 \ . \qquad \text{(solució entera)}$$

Fem el problema [1.2]:

Objecte	c_j/a_j
o_1	1
o_4	1
o_5	1

Posem a la motxilla

$$o_2 \ (4) + o_1 \ (5) + o_4 \ (6) \ . \qquad z = 14 \ . \qquad \text{(solució entera)}$$

Fem el problema [2]:

Objecte	c_j/a_j
o_1	1
o_4	1
o_5	1
o_2	3/4

Posem a la motxilla

$$o_3\ (7) + o_1\ (5) + \tfrac{1}{2} o_4\ (3)\ . \qquad z = 14\ .$$

No cal continuar subdividint aquest problema, perquè cap fill no donarà un valor millor de 14, i ja tenim aquest valor en una solució entera. Podria ser però que hi hagués alguna altra solució amb valor 14 en aquesta branca. En efecte, treballant una mica més, es troba que el problema té dues solucions optimals:

$$x_1 = x_2 = x_4 = 1\ , \quad x_3 = x_5 = 0\ .$$
$$x_3 = x_4 = x_5 = 1\ , \quad x_1 = x_2 = 0\ .$$

3.3.2 Problema del viatjant

Situació: Un viatjant de comerç surt de la seva ciutat i vol visitar $n-1$ altres ciutats i tornar a casa amb un cost mínim. Ha de passar per cada ciutat exactament una vegada. El cost de viatjar de la ciutat i a la ciutat k és c_{ik}. □

Model (temptatiu): Considerem una variable binària x_{ik} per a cada parella de ciutats (i,k). Es fa el trajecte de la ciutat i a la k si $x_{ik}=1$; no es fa si $x_{ik}=0$.

$$\min\ z = \sum_{i=1}^{n}\sum_{k=1}^{n} c_{ik}x_{ik}$$

$$\text{subj. a:} \quad \left.\begin{aligned} \sum_{i=1}^{n} x_{ik} = 1,\ k = 1,\dots,n \\ \sum_{k=1}^{n} x_{ik} = 1,\ i = 1,\dots,n \\ \forall i,k,\ x_{ik} \in \{0,1\} \end{aligned}\right\}$$

El primer grup de restriccions diu que a la ciutat k s'hi arriba exactament una vegada; el segon diu que de la ciutat i només es pot sortir una vegada. □

Model correcte: És correcte el model anterior? No. Per exemple, si $n=6$, una ruta inconnexa unint les ciutats 1,2,3 per una banda i les ciutats 4,5,6 per l'altra seria un punt factible.

Hem d'evitar cicles que no incloguin totes les ciutats. Per exemple,

$$x_{12} + x_{23} + x_{31} + x_{21} + x_{32} + x_{13} \leqslant 2$$

evita que es formi un cicle entre les ciutats 1,2,3. Convé doncs afegir, per a cada subconjunt del conjunt de ciutats C, diferent del buit i del propi C, la restricció

$$\sum_{i,k\in C} x_{ik} \leqslant (\text{cardinal de } C) - 1\ .$$

3.3.3 Penalitzacions

Situació: Suposem que no volem imposar de manera categòrica una restricció

$$g(x_1,\dots,x_n) \leqslant 0\ ,$$

però que violar-la ha de tenir una penalització (un cost). □

Model: Suposem que la funció g està acotada superiorment per una constant D sobre el conjunt deterinat per les demés restriccions. Introduïm una variable binària y. Posem la restricció

$$g(x_1, \dots, x_n) - Dy \leqslant 0$$

i afegim un terme a la funció objectiu:

$$f(x_1, \dots, x_n) + cy \ ,$$

on $c > 0$ és la penalització per violar la restricció.

Aleshores, si en un punt factible $(x_1, \dots, x_n, y)$ es té $y = 0$, es compleix que $g(x_1, \dots, x_n) \leqslant 0$ i no hi ha penalització; si en canvi $y = 1$, aleshores hi ha penalització.

3.3.4 Disjunció de restriccions

Situació: Suposem que volem que es compleixin

$$g_1(x_1, \dots, x_n) \leqslant 0 \qquad \text{o bé} \qquad g_2(x_1, \dots, x_n) \leqslant 0$$

però no necessàriament les dues restriccions a la vegada. □

Model: Sigui D tal que $g_1(x_1, \dots, x_n) \leqslant D$, $g_2(x_1, \dots, x_n) \leqslant D$ sobre el conjunt determinat per les demés restriccions.

Introduïm dues variables binàries y_1, y_2 i les restriccions

$$\left.\begin{array}{r} g_1(x_1, \dots, x_n) - Dy_1 \leqslant 0 \\ g_2(x_1, \dots, x_n) - Dy_2 \leqslant 0 \\ y_1 + y_2 \leqslant 1 \end{array}\right\}$$

Gràcies a la tercera restricció, almenys una de les variables y_1 o y_2 ha de ser zero, i la restricció associada es complirà.

3.3.5 Restriccions condicionals

Situació: Suposem que volem imposar

$$g_1(x_1, \dots, x_n) > 0 \Rightarrow g_2(x_1, \dots, x_n) \leqslant 0 \ .$$

Model: Aquesta situació és idèntica a la de la disjunció 3.3.4, i per tant es modela igual.

3.3.6 Alternatives compostes

Situació: Suposem que volem que $(x_1, \dots, x_n)$ estigui almenys en una de les regions

$$\left.\begin{array}{r} g_1(x_1, \dots, x_n) \leqslant 0 \\ g_2(x_1, \dots, x_n) \leqslant 0 \end{array}\right\} \qquad \left.\begin{array}{r} g_3(x_1, \dots, x_n) \leqslant 0 \\ g_4(x_1, \dots, x_n) \leqslant 0 \end{array}\right\} \qquad \left.\begin{array}{r} g_5(x_1, \dots, x_n) \leqslant 0 \\ g_6(x_1, \dots, x_n) \leqslant 0 \\ g_7(x_1, \dots, x_n) \leqslant 0 \end{array}\right\}$$

Model: Sigui D una cota superior de $g_1, \dots, g_7$, com en els apartats anteriors. Introduïm 3 variables binàries y_1, y_2, y_3 i les restriccions

$$\left.\begin{array}{r} g_1(x_1, \dots, x_n) - Dy_1 \leqslant 0 \\ g_2(x_1, \dots, x_n) - Dy_1 \leqslant 0 \\ g_3(x_1, \dots, x_n) - Dy_2 \leqslant 0 \\ g_4(x_1, \dots, x_n) - Dy_2 \leqslant 0 \\ g_5(x_1, \dots, x_n) - Dy_3 \leqslant 0 \\ g_6(x_1, \dots, x_n) - Dy_3 \leqslant 0 \\ g_7(x_1, \dots, x_n) - Dy_3 \leqslant 0 \\ y_1 + y_2 + y_3 \leqslant 2 \end{array}\right\}$$

3.3.7 Costos fixos

Situació: Suposem que en un problema de minimització de costos la variable x_n té un *cost fix*. Això vol dir que intervé a la funció objectiu aportant un cost 0 si $x_n = 0$, i un cost $k + c_n x_n$ si $x_n > 0$. La constant k és el cost fix: Per poc que x_n sigui positiva, cal pagar-lo. □

Model: Sigui D una cota de x_n en la regió factible. Introduïm una variable binària y. Afegim la restricció

$$x_n - Dy \leqslant 0$$

i posem a la funció objectiu

$$f(x_1, \ldots, x_{n-1}) + c_n x_n + ky \ .$$

Aleshores, si en un punt factible $(x_1, \ldots, x_n, y)$ es té $y = 0$, es complirà també $x_n = 0$ i no hi ha cost fix; si en canvi $y = 1$, aleshores se suma el cost fix.

3.3.8 Objectius lineals a trossos

Situació: Suposem que en un problema de minimització de costos la variable x_n intervé a la funció objectiu de manera "lineal a trossos". Per exemple, la funció objectiu és de la forma

$$f(x_1, \ldots, x_{n-1}) + g(x_n) \ ,$$

on la gràfica de g consisteix en segments que tenen pendent 5 en $[0, 4]$, pendent 1 en $[4, 10]$ i pendent 3 en $[10, 15]$. □

Model: Expressarem x_n com la suma de tres variables: $x_n = \delta_1 + \delta_2 + \delta_3$, amb

$$0 \leqslant \delta_1 \leqslant 4 \ , \quad 0 \leqslant \delta_2 \leqslant 6 \ , \quad 0 \leqslant \delta_3 \leqslant 5 \ ,$$

i de manera que

$$\delta_2 > 0 \Rightarrow \delta_1 = 4 \ , \quad \delta_3 > 0 \Rightarrow \delta_2 = 6 \ .$$

El cost degut a x_n vindrà donat per $5\delta_1 + \delta_2 + 3\delta_3$.

Introduïm dues variables binàries y_1 i y_2, i afegim les restriccions

$$\left.\begin{array}{r} 4y_1 \leqslant \delta_1 \leqslant 4 \\ 6y_2 \leqslant \delta_2 \leqslant 6y_1 \\ 0 \leqslant \delta_3 \leqslant 5y_2 \end{array}\right\}$$

Aleshores:

a) Quan $y_1 = 0$, la segona restricció implica $y_2 = 0$, i estem imposant de fet

$$\left.\begin{array}{r} 0 \leqslant \delta_1 \leqslant 4 \\ \delta_2 = 0 \\ \delta_3 = 0 \end{array}\right\}$$

(és el cas $x_n \in [0, 4]$).

b) Quan $y_1 = 1$ i $y_2 = 0$, estem imposant

$$\left.\begin{array}{r} \delta_1 = 4 \\ 0 \leqslant \delta_2 \leqslant 6 \\ \delta_3 = 0 \end{array}\right\}$$

(és el cas $x_n \in [4, 10]$).

c) Quan $y_1 = 1$ i $y_2 = 1$, estem imposant

$$\left.\begin{array}{r} \delta_1 = 4 \\ \delta_2 = 6 \\ 0 \leqslant \delta_3 \leqslant 5 \end{array}\right\}$$

(és el cas $x_n \in [10, 15]$).

4. Fluxos lineals sobre xarxes

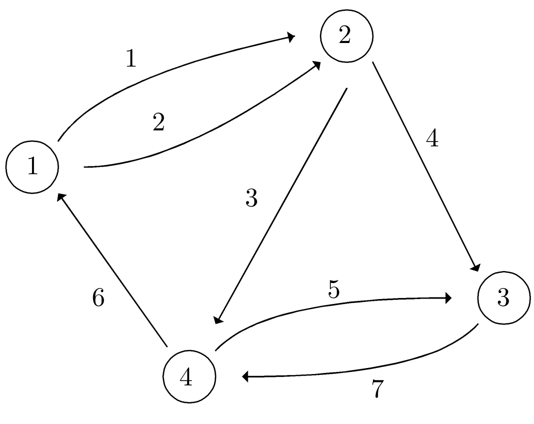
Hi ha problemes d'optimització formulables com a problemes de programació lineal, no-lineal, o entera però que tenen una estructura especial que fa que, per a ells concretament, es puguin aplicar altres mètodes millors que els mètodes generals que hem vist, o que el mètode del símplex, per exemple, sigui més eficient. Un cas especialment important són els problemes anomenats de *Fluxos sobre Xarxes.*

4.1 El problema del cost minimal

Una *xarxa* està composta per dos tipus d'entitats: *nodes* i *arcs dirigits* (fletxes que apunten d'un node a un altre). La interpretació gràfica és evident.

4.1.1 Exemple

El gràfic següent representa una xarxa:



□

Suposarem sempre que:

- No hi ha arcs que uneixen un node amb ell mateix.
- La xarxa és connexa, és a dir, que dos nodes qualsevols estan units per algun camí d'arcs (ignorant el sentit de les fletxes).

En canvi, com es veu al dibuix, entre els mateixos dos nodes poden haver-hi diversos arcs.

Una segona manera de representar una xarxa, a més del dibuix anterior, és mitjançant la matriu d'incidències.

4.1.2 Definició

Si hi ha n arcs i m nodes, la *matriu d'indidències* és $A=(a_{ij})$, de dimensions $m\times n$, definida per

$$a_{ij}=\begin{cases} +1, & \text{si l'arc } j \text{ té el node } i \text{ per origen} \\ -1, & \text{si l'arc } j \text{ té el node } i \text{ per destí} \\ 0, & \text{altrament} \end{cases}$$

4.1.3 Exemple

Per a la xarxa de l'Exemple 4.1.1, la matriu d'incidències és

$$A=\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & -1 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 & -1 & 1 \end{pmatrix}$$

□

Ara suposem que tenim el problema de moure a través de la xarxa un cert producte per complir uns determinats requeriments de cada node, que cada unitat del producte té un cost fer-lo circular per cada arc, i que volem fer mínim aquest cost.

Notem:

x_j: Quantitat de producte que viatja per l'arc j. Aquestes seran les variables del model, i s'acostumen a anomenar *fluxos*.

c_j: *Cost* unitari de circular per l'arc j.

d_i: *Requeriment* del node i, definit com

$$(\text{quantitat que surt de } i) - (\text{quantitat que arriba a } i)\ .$$

$d_i>0 \Rightarrow$ És un node *subministrador*.
$d_i<0 \Rightarrow$ És un node *receptor*.
$d_i=0 \Rightarrow$ És un node *de pas*.

S'ha de complir $\sum_{i=1}^m d_i=0$ (*conservació del flux*).

El problema es pot formular com

$$\begin{aligned} &\min\ c\cdot x \\ &\text{subj. a:} \\ &\left.\begin{aligned} Ax&=d \\ x&\geqslant 0 \end{aligned}\right\} \end{aligned}$$

Per tant, el nostre és un problema de PL, i ja el sabem resoldre. Però resulta que la matriu A té una forma molt especial. A cada columna hi ha exactament un $+1$, un -1, i tota la resta són zeros. Gràcies a aquesta forma especial, el problema es pot resoldre molt més ràpidament i senzilla que aplicant el mètode del símplex de manera rutinària.

El problema que hem descrit s'anomena *problema del cost minimal*. A l'apartat 1.3 veurem altres problemes d'optimització que es poden plantejar sobre un xarxa.

4.1.4 Exemple

S'han d'assignar n persones a n feines (una per a cadascú). Sigui c_{kj} una mesura de l'habilitat de la persona k per fer la feina j.

Volem fer l'assignació òptima: La suma de totes les habilitats de les assignacions que es facin ha de ser la màxima possible.

Posem

$$x_{kj} = \begin{cases} 1, & \text{si la persona } k \text{ fa la feina } j \\ 0, & \text{en cas contrari.} \end{cases}$$

Hem d'imposar com a restriccions que cada persona faci només un treball, i que cada treball només el faci una persona. Obtenim el planteig següent:

$$\max\ z = \sum_{k=1}^{n} \sum_{j=1}^{n} c_{kj} x_{kj}$$

subj. a:

$$\left.\begin{array}{rr} \displaystyle\sum_{j=1}^{n} x_{kj} = 1, & k = 1, \ldots, n \\ \displaystyle\sum_{k=1}^{n} x_{kj} = 1, & j = 1, \ldots, n \\ & x_{kj} \in \{0, 1\} \end{array}\right\}$$

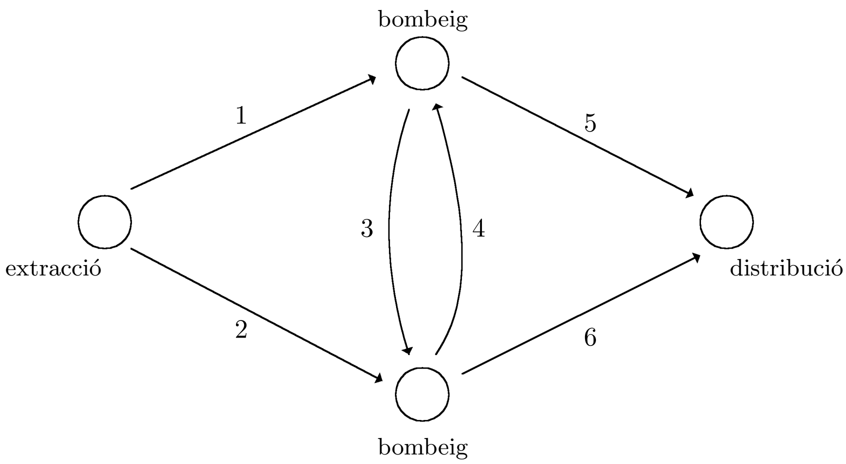
Si ara multipliquem per -1 el segon grup de restriccions, obtenim una matriu d'incidències. Per altra banda, podem substituir les condicions $x_{kj} \in \{0, 1\}$ per $x_{kj} \geqslant 0$, perquè:

- x_{kj} no podran valer > 1, degut a les altres restriccions.
- Pel fet de ser els requeriments enters, es pot demostrar que totes les solucions bàsiques són automàticament enteres.

Per tant, es tracta d'un problema de cost minimal, amb $2n$ nodes i n^2 arcs.

4.1.5 Exemple

Una companyia de gas ha de transportar gas des del lloc d'extracció al centre de distribució, a través de gasoductes i dues estacions de bombeig, distribuïts segons la xarxa següent:



La quantitat a transportar són 1200 milions de m^3 (Mm^3) mensuals. Cada gasoducte té una capacitat màxima i un cost unitari:

Gasoducte	1	2	3	4	5	6
Capacitat en Mm^3/mes	500	900	700	400	600	1000
Cost en $/$Mm^3$	20	25	10	15	20	40

El model serà:

Variables: x_j serà el volum que viatja per l'arc j.

Objectiu: $\min z = 20x_1 + 25x_2 + 10x_3 + 15x_4 + 20x_5 + 40x_6$

Restriccions:

$$\left.\begin{array}{r} x_1 + x_2 = 1200 \\ -x_1 + x_3 - x_4 + x_5 = 0 \\ -x_2 - x_3 + x_4 + x_6 = 0 \\ -x_5 - x_6 = -1200 \\ 0 \leqslant x_1 \leqslant 500 \\ 0 \leqslant x_2 \leqslant 900 \\ 0 \leqslant x_3 \leqslant 700 \\ 0 \leqslant x_4 \leqslant 400 \\ 0 \leqslant x_5 \leqslant 600 \\ 0 \leqslant x_6 \leqslant 1000 \end{array}\right\}$$

Es tracta d'un problema de cost minimal. Hi ha cotes superiors, però això no és cap inconvenient.

4.2 Mètode del Símplex per xarxes

4.2.1 Definicions

Un *cicle* és una seqüència d'arcs que uneixen un node amb ell mateix (sense tenir en compte la direcció dels arcs).

Un *arbre* és una xarxa connexa sense cicles.

Un *arbre generador de la xarxa* és un arbre que inclou tots els nodes de la xarxa.

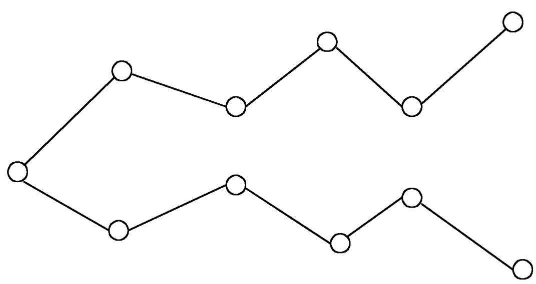
Un node és *extrem* si hi ha exactament un arc incidint en ell. □

4.2.2 Proposició

Tot arbre té com a mínim dos nodes extrems.

Demostració: Comencem per un node qualsevol (per exemple, el de més a l'esquerra del dibuix següent) i ens anem movent per arcs. Com que no hi ha cicles, en un cert moment "se'ns acabarà el camí" (haurem arribat a un extrem).

Si havíem començat en un extrem, ja en tenim dos. Si no, agafem un altre camí i acabarem en un altre extrem, que no podrà ser el mateix d'abans perquè no hi ha cicles.



□

4.2.3 Proposició

En un arbre,

$$\text{(nombre d'arcs)} = \text{(nombre de nodes)} - 1\ .$$

Demostració: És evident per a arbres amb un sol node o dos nodes. Suposem-ho cert fins per a arbres fins a $m-1$ nodes.

Sigui ara un arbre amb m nodes. Traiem un node extrem i el seu arc. Segueix essent un arbre, però amb $m-1$ nodes i per tant amb $m-2$ arcs. Tornant a posar l'arc que hem tret, fan $m-1$ arcs. □

4.2.4 Proposició

En una xarxa, si

$$(\text{nombre d'arcs}) = (\text{nombre de nodes}) - 1\ ,$$

aleshores és un arbre.

Demostració: Suposarem que no és arbre i veurem que la igualtat no es pot complir:

Suposem que hi ha cicles. Traient un arc de cada cicle, seguim tenint una xarxa connexa, i ens queda un arbre. Aquest arbre satisfà la igualtat. Tornant a posar els arcs que hem tret, la igualtat deixarà de ser certa.

4.2.5 Corol·lari

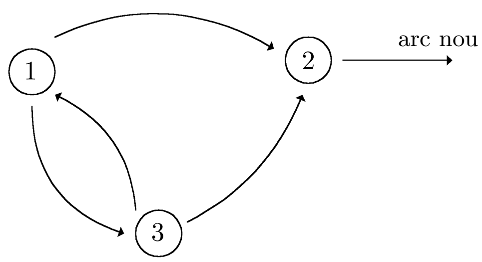
Una "subxarxa" (tros d'una xarxa) sense cicles és un arbre generador si i només si és connexa i conté $m-1$ arcs exactament. □

Ara considerem el problema

$$\begin{array}{l} \min\ c \cdot x \\ \text{subj. a:} \\ \left.\begin{array}{r} Ax = d \\ 0 \leqslant x \leqslant u \end{array}\right\} \end{array}$$

on A és una matriu d'incidències $m \times n$. Aquest problema no està en forma estàndard: Per una banda, hi haurà termes independents d_i negatius. A més, el rang de A no és m. Això és evident, perquè sumant totes les files obtenim una fila de zeros. Es pot demostrar que el rang és exactament $m-1$.

Per tenir una matriu de rang màxim, en lloc d'eliminar una fila, hi afegirem una columna que només tingui un $+1$, i la resta zeros. Gràficament, això vol dir que afegim a la xarxa un arc que surt d'un cert node i no apunta enlloc:

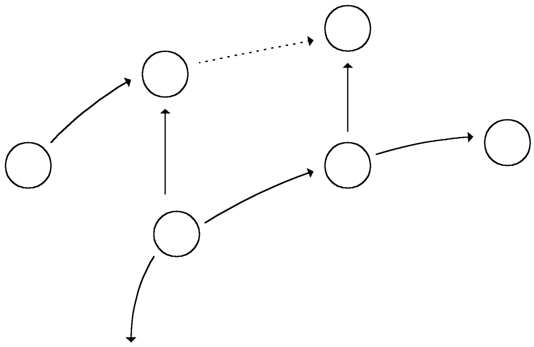


La nova columna representa una variable afegida al problema. Aquesta nova variable la posem amb cota superior igual a zero, de manera que el flux a través del seu arc serà sempre zero i no influirà en el resultat final.

Es pot demostrar que:

- La matriu A ampliada d'aquesta manera té rang m.
- En una base B formada per columnes de la matriu A ampliada, sempre hi és la nova columna.
- Els arcs d'un arbre generador de la xarxa representada per la matriu A ampliada corresponen a columnes que formen base; i viceversa, columnes que forme base corresponen a arcs que formen un arbre generador.

Per tant, tenir una base és tenir un arbre generador. I què és aleshores un canvi de base en termes d'arbres generadors? Entrar una variable a la base és afegir un arc al arbre generador (per exemple, l'arc puntejat del dibuix següent). Això forma un cicle, que caldrà desfer per tal de seguir tenint un arbre. Algun dels altres arcs del cicle haurà de sortir.



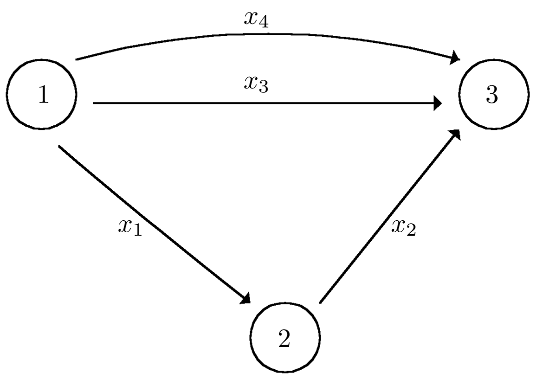
I resulta que trobar els arcs que poden entrar a l'arbre, i l'arc que ha de sortir, es pot fer gràficament, mirant la xarxa, sense necessitat d'escriure la taula del símplex.

4.2.6 Exemple

Considerem el problema següent:

$$
\begin{aligned}
&\min\ z = x_1 + x_2 + 3x_3 + 10x_4 \\
&\text{subj. a:} \\
&\left.\begin{array}{r}
x_1 \qquad + x_3 + x_4 = 5 \\
-x_1 + x_2 \qquad\qquad = 0 \\
- x_2 - x_3 - x_4 = -5 \\
0 \leqslant x_1 \leqslant 4 \\
0 \leqslant x_2 \leqslant 2 \\
0 \leqslant x_3 \leqslant 4 \\
0 \leqslant x_4 \leqslant 10
\end{array}\right\}
\end{aligned}
$$

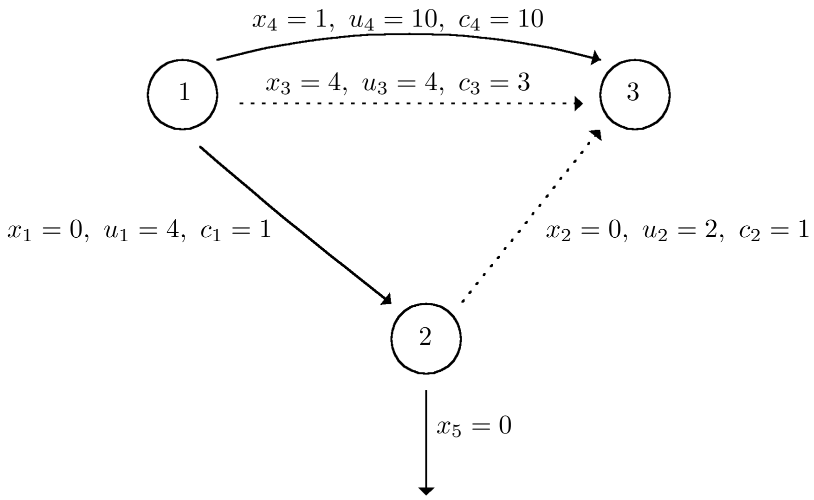
La xarxa que representa és:



Afegint l'arc nou al node 2, per exemple, queda la matriu

$$
A = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 & 0 \end{pmatrix}
$$

Un possible arbre generador per començar seria



Tenim com a matriu bàsica B (variables x_1, x_4, x_5) i matriu de variables no-bàsiques N (variables x_2, x_3).

$$B = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \qquad N = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & -1 \end{pmatrix}$$

Segons el mètode del símplex, hem de calcular el vector $c_N - c_B B^{-1} N$ per buscar una variable no-bàsica candidata a entrar a la base (aquest vector és la fila inferior de la taula del símplex, vegeu el Capítol 1). Si no hi haguessin cotes superiors, el criteri per poder entrar a la base és tenir un coeficient negatiu en aquest vector. Com que en aquest cas tenim cotes superiors, s'aplica el mateix que vam veure en el mètode del gradient reduït (vegeu el Capítol 2). És a dir: Buscarem $k \in N$ tal que

$$x_k = 0 \quad \text{i} \quad (c_N - c_B B^{-1} N)_k < 0$$

o bé

$$x_k = u_k \quad \text{i} \quad (c_N - c_B B^{-1} N)_k > 0 \ .$$

La manera més ràpida de calcular aquest vector, en principi, seria fer primer $\lambda := c_B B^{-1}$ (resolent el sistema $\lambda B = c_B$), obtenint-se en el nostre cas

$$\lambda_1 = c_5 + c_1 \ , \quad \lambda_4 = c_5 \ , \quad \lambda_5 = c_5 + c_1 - c_4 \ ,$$

i després fer

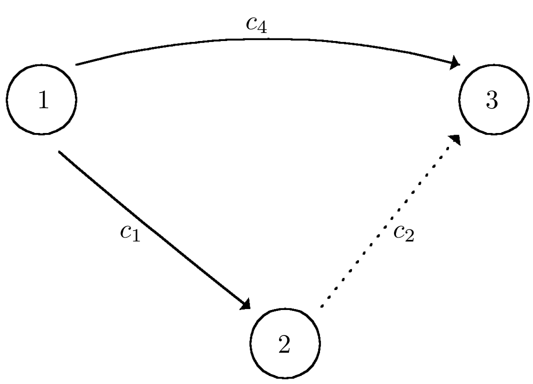
$$\lambda N = (-c_1 + c_4, c_4)$$

i, finalment,

$$c_N - \lambda N = (c_2 + c_1 - c_4, c_3 - c_4) \ .$$

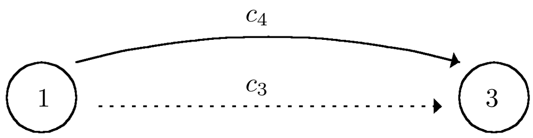
La primera component correspon a la variable no bàsica x_2 i la segona a la variable no-bàsica x_3.

Però de fet no cal fer-ho així. El vector es pot calcular fàcilment de manera gràfica sobre el dibuix de la xarxa: Si entrés x_2 a la base es formaria el cicle



Recorrem ara el cicle en el sentit de l'arc que entra, sumant els costos dels arcs que trobem en la nostre direcció, i restant els dels que van en sentit contrari. Obtenim $c_2 - c_4 + c_1$.

Si entrés x_3,



obtenim amb el mateix procediment $c_3 - c_4$.

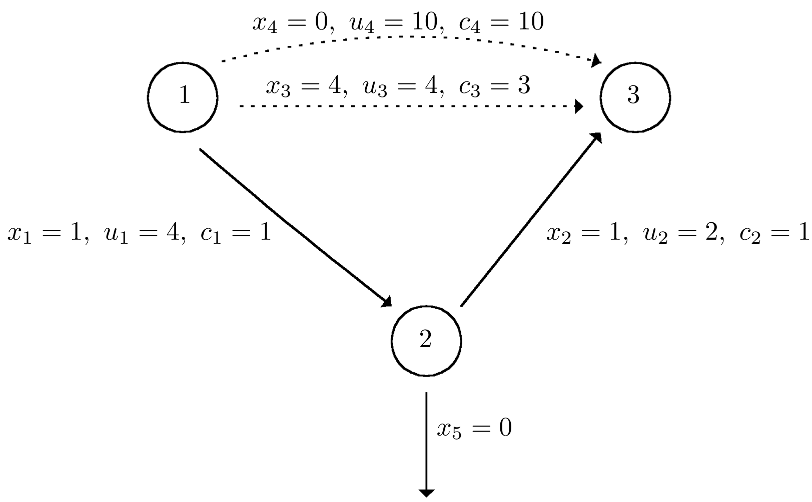
En el nostre cas tenim concretament

$$c_N - c_B B^{-1} N = (c_2 - c_4 + c_1, c_3 - c_4) = (-8, -7) \ .$$

La variable x_2 està a cota inferior i té la component negativa -8; per tant, pot entrar a la base. La variable x_3 està a cota superior i té la component negativa -7; aquesta no pot entrar a la base.

Anem ara a actualitzar l'arbre (pivotació): Imaginem que comença a passar flux per l'arc x_2, que estava a zero. Per mantenir l'equilibri entre els nodes, l'arc x_1 també ha d'augmentar de flux, en la mateixa quantitat que x_2, mentre que el x_4 ha de disminuir. Anem fent passar més flux fins que un dels arcs del cicle se satura (arriba a la seva cota superior, i no pot seguir augmentant, o arriba a zero, i no pot seguir disminuint). En el nostre cas és x_4 el primer que atura el moviment, quan passa de tenir flux 1 a tenir flux 0.

Surt doncs x_4 de la base i obtenim el nou arbre generador

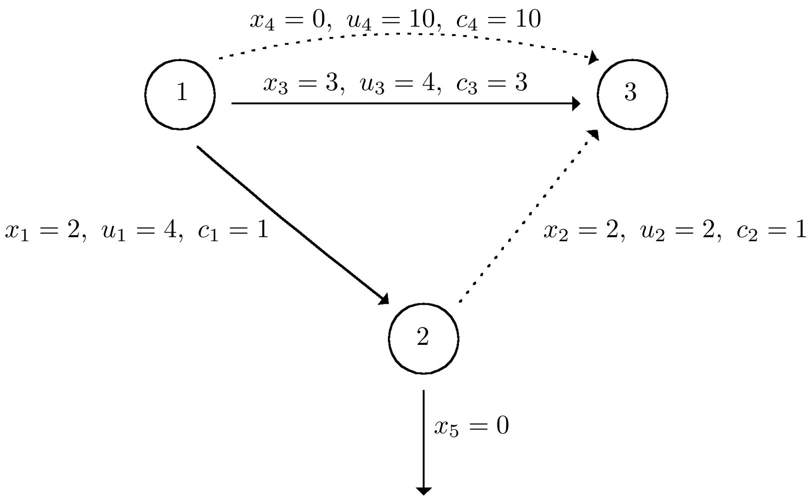


Procedim a fer una nova iteració. Ara obtenim:

$$c_N - c_B B^{-1} N = (c_3 - c_2 - c_1, c_4 - c_2 - c_1) = (+1, +8) \ ,$$

on la primera component $+1$ correspon a x_3, que està a cota superior i per tant pot entrar a la base, i la segona $+8$ a x_4, que està a cota inferior i per tant no pot entrar a la base.

Formem el cicle, comencem a disminuir el flux que passa per x_3, augmentant per tant el de x_1 i x_2, fins que l'arc x_2 arriba a la seva cota superior 2 i surt de la base. Obtenim el nou arbre



Tornem a valorar les dues variables no-bàsiques x_2, x_4:

$$c_N - c_B B^{-1} N = (c_2 - c_3 - c_1, c_4 - c_3) = (-1, +7)\ .$$

Ara la variable que està a cota superior té un coeficient negatiu i la que està a cota inferior el té positiu. O sigui que cap de les dues és apte per entrar a la base millorant la funció objectiu. Aquest és el punt òptim:

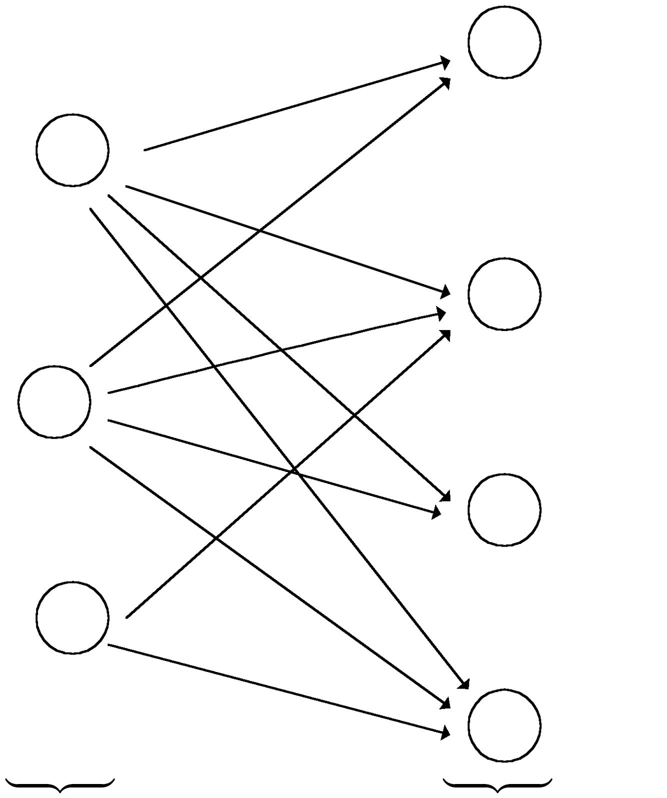
$$x_1 = 2,\ x_2 = 2,\ x_3 = 3,\ x_4 = 0\ .$$

4.3 Altres problemes amb estructura de xarxa

El problema que hem formulat era l'anomenat *problema del cost minimal* (PCM). Hi ha altres tipus de problemes d'optimització que es poden formular sobre una xarxa, però tots es poden transformar en un PCM o bé es poden pensar directament com a casos particulars del PCM. Vegem-ne uns quants.

4.3.1 El problema del transport

L'exemple prototípic és el de transportar un producte que es fabrica en uns certs centres de producció a uns certs centres de distribució.



És un cas especial del PCM en què tot node és subministrador o receptor i tots els arcs van d'un node subministrador a un node receptor. Hi ha mètodes especials, però el símplex que hem vist és prou eficient.

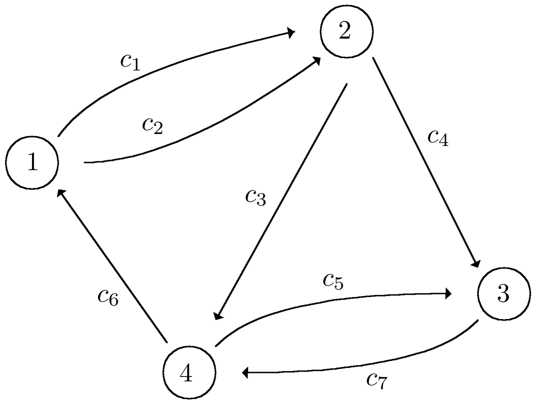
4.3.2 El problema de l'assignació

És l'Exemple 4.1.4, on havíem d'assignar n persones a n feines. És un cas especial del problema del transport en què $|d_i| = 1$ ($+1$ per als nodes-persona i -1 per als nodes-feina).

Existeix un mètode per resoldre'l, anomenat "mètode hungarès", que és més eficient que el símplex.

4.3.3 El problema del camí més curt

És el problema de trobar el camí més curt (o de cost mínim) entre dos nodes fixats d'una xarxa. Per exemple considerem el problema de trobar el camí més econòmic del node 1 al node 3 a la xarxa següent:

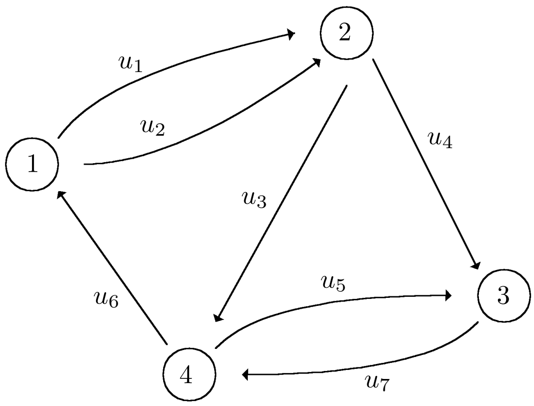


Els arcs tenen un cost, però no els posem cap limitació de capacitat. Posem els requeriments $d_1 = 1$, $d_3 = -1$, $d_2 = d_4 = 0$. Això farà anar una unitat de producte del node 1 al node 3. Veurem fàcilment quin és el camí més curt, seguint els arcs amb flux 1, i la seva longitud o cost total.

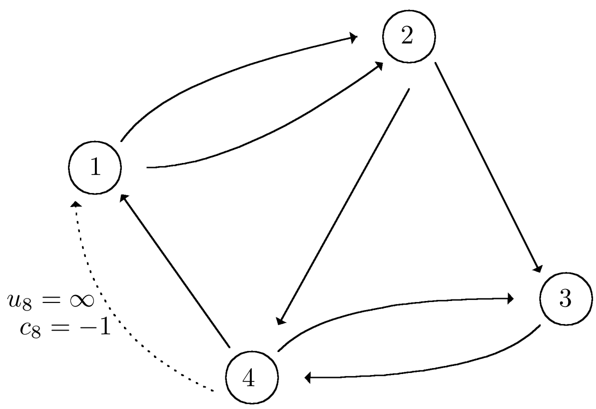
Per tant el problema del camí més curt és un cas especial del PCM amb $d_{i_1} = 1$, $d_{i_2} = -1$, per certs nodes i_1, i_2, i $d_i = 0$, $\forall i \neq i_1, i_2$.

4.3.4 El problema del flux maximal

Es tracta d'enviar tanta quantitat de producte com sigui possible de un node fixat a un altre node fixat. Els arcs tenen una capacitat, però no els associem cap cost. Per exemple:



EL transformen en un PCM amb el truc següent: Agefim un arc del node 4 al node 1 amb capacitat infinita i cost -1. Posem tots els altres arcs a cost 0. Els requeriments de tots els nodes es posen també a zero.



El PCM trobarà els fluxos que fan passar la quantitat més gran possible de flux de 4 a 1. Com tots els requeriments són zero, haurem enviat la quantitat més gran possible de 1 a 4 pels arcs de veritat, que és el que interessava.

Per tant, el problema de flux maximal és un cas particular del PCM en què $d_i = 0$, $\forall i$; $c_{j_0} = -1$ per un cert arc j_0, i $c_j = 0$, $\forall j \neq j_0$.

Hi ha diversos mètodes per resoldre'l, però el símplex és prou bo.