Versión muy parecida a la publicada.

OPTIMIZACIÓN NO LINEAL Y DINÁMICA

Héctor Manuel Mora Escobar
Departamento de Matemáticas
Universidad Nacional de Colombia
Bogotá

ISBN

A Hélène, Nicolás y Sylvie

ÍNDICE GENERAL

PRÓLOGO

El nombre de este libro difiere un poco del utilizado en la primera edición: Programación No Lineal. Actualmente hay una tendencia a utilizar la palabra *optimización* para designar los temas donde, perdón por la redundancia, se optimiza (se minimiza o se maximiza). La palabra clásica, tradicional y muy conocida, *programación*, induce cierta confusión, pues también se utiliza para programación de computadores y lenguajes de programación. La tradición de todas maneras pesa y en este libro se utilizan indistintamente las dos palabras.

Esta segunda edición tiene dos capítulos adicionales, uno sobre métodos de punto interior y otro sobre Optimización Dinámica. También tiene una sección nueva sobre métodos de región de confianza. El tema de matrices definidas positivas en un subespacio está ahora en el tercer capítulo. Finalmente algunos pequeños errores fueron corregidos.

El propósito de este libro introductorio es presentar algunos de los resultados y de los métodos más importantes y útiles de la Optimización No Lineal y una introducción a la Optimización Dinámica. Estos temas corresponden al curso Programación No Lineal y Dinámica de las carreras de Ingeniería de Sistemas y de Matemáticas. En tiempo, esto corresponde a cuatro horas semanales durante quince semanas aproximadamente. Para los estudiantes de Matemáticas puede ser conveniente dejar de lado algunos métodos y, a cambio, estudiar las demostraciones de algunos de los teoremas fundamentales.

Además de los resultados que permiten la caracterización de los mínimos de funciones de varias variables, también aparecen en el capítulo 2 algunos resultados que no se emplean directamente para el estudio de los mínimos, pero que sí hacen parte de la teoría clásica de convexidad, o se utilizan para la demostración de algunos teoremas, o sirven para el estudio de la Optimización Lineal. Creo que para poder seguir sin mucha rapidez este libro en las quince semanas, se puede omitir el capítulo 2, del cual sólo se

"necesita" la definición de conjunto convexo.

También por "razones de tiempo", este libro no contiene nada sobre dualidad, no hay resultados sobre calificación de restricciones, ni están "todos" los métodos importantes. Faltarían algunos métodos espécificos, por ejemplo, para programación cuadrática, complementariedad lineal y algoritmo de Lemke, métodos de programación cuadrática secuencial, método de conjunto activo como tal. Obviamente hay otros métodos importantes que definitivamente quedan por fuera del alcance de este libro, por ejemplo, métodos para mínimos cuadrados (sin o con restricciones), métodos para optimización global, métodos para optimización no suave (no diferenciable), métodos para problemas grandes de PNL (large-scale) y métodos paralelos.

Este libro tiene un enfoque utilitario en el sentido de que lo importante de cada resultado es su aplicación e interpretación y no su justificación. Por esta razón casi todas las proposiciones están presentadas sin demostración, éstas se pueden encontrar en los libros citados en la bibliografía. En cambio hay bastantes ejemplos que permiten, eso espero, comprender el alcance, las condiciones de aplicación y las implicaciones de cada resultado.

Mis primeros conocimientos de PNL los tuve del profesor Jean Legras en la Universidad de Nancy en Francia. Posteriormente seguí aprendiendo al dictar el curso de PNL, para lo cual usé bastante el libro de Bazaraa y Shetty (Bazaraa, Sherali y Shetty en la segunda edición). El lector notará la gran influencia del libro de Bazaraa en la redacción de estas notas.

Quiero agradecer especialmente a los profesores Lucimar Nova, Jaime Malpica, Luis G. Moreno, Argemiro Echeverry, Felix Soriano y Jorge Mauricio Ruiz, quienes tuvieron la amabilidad y la paciencia de leer la versión inicial de este libro. También quiero agradecer a los estudiantes del curso Programación No Lineal y Dinámica de las carreras de Ingeniería de Sistemas y de Matemáticas, en especial a Sandra Toro, Héctor López y David Báez. Las sugerencias, comentarios y correcciones de todos ellos, fueron muy útiles.

Deseo agradecer a la Universidad Nacional por haberme permitido destinar parte de mi tiempo de trabajo para dedicarlo a esta obra, este tiempo fue una parte importante del necesario para la realización del libro.

El texto fue escrito en Latex. Quiero también agradecer al profesor Rodrigo De Castro quien amablemente me ayudó a resolver las inquietudes y los problemas presentados.

Finalmente, y de manera muy especial, agradezco a Hélène, Nicolás y Syl-

vie. Sin su apoyo, comprensión y paciencia no hubiera sido posible escribir este libro.

NOTACIÓN

$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_j \in \mathbb{R}, \forall j\}$
$\mathbb{R}^n :=$ conjunto universal
x es un vector o punto $\Leftrightarrow x \in \mathbb{R}^n$
S es un conjunto $\Leftrightarrow S \subseteq \mathbb{R}^n$
$\mathcal{M}(m,n) = \mathbb{R}^{m\times n} =$ conjunto de matrices reales de tamaño $m \times n$
si $A \in \mathcal{M}(m,n)$, entonces A es de la forma:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

$\mathcal{M}(n,1) = \mathbb{R}^{n\times 1} =$ conjunto de matrices columna de n componentes.
$\mathcal{M}(1,n) = \mathbb{R}^{1\times n} =$ conjunto de matrices fila de n componentes.
$A^{\mathrm{T}} =$ la transpuesta de la matriz A.
$\mathbb{R}^n := \mathcal{M}(n,1) = \mathbb{R}^{n\times 1}$

$$x = (x_1, x_2, \ldots, x_n) := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$x^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}$
$A_{i\cdot} =$ fila i-ésima de la matriz $A = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix}$

$$A_{\cdot j} = \text{columna } j\text{-ésima de la matriz } A = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$

$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$

$d_p(x,y) = \|x-y\|_p, \quad p = 1, 2, \infty$

$B(c,r) = \{x : ||x-c|| < r\}$

$B_p(c,r) = \{x : ||x-c||_p < r\}$

$B[c,r] = \{x : ||x-c|| \leq r\}$

$B_p[c,r] = \{x : ||x-c||_p \leq r\}$

$x^k = x^{(k)}$ =vector x en la iteracioón k, $k = 0, 1, 2 \ldots$

$Q^k = Q^{(k)}$ =matriz Q en la iteracioón k ,$k = 0, 1, 2 \ldots$

$(Q)^k = k$ veces el producto de la matriz Q por sí misma.

Q^{-1} = inversa de la matriz Q

$R(\bar{x}, d) = \{\bar{x} + \mu d : \mu \in \mathbb{R}\}$ recta que pasa por $\bar{x}$ y es paralela a $d \neq 0$

$S(\bar{x}, d) = \{\bar{x} + \mu d : \mu \geq 0\}$ semirrecta que empieza en $\bar{x}$ y va en la dirección de $d \neq 0$

$f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_l$ son funciones de variable vectorial y valor real, es decir:

$$f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_l \ : \mathbb{R}^n \longrightarrow \mathbb{R}$$

$\mathcal{A}$: conjunto admisible = $\{x : x$ cumple todas las restricciones $\}$

$\mathcal{I} = \{1 \leq i \leq m : g_i(\bar{x}) = 0\}$

$f'(\bar{x}) = \nabla f(\bar{x}) = \text{grad}_f(\bar{x}) =$ gradiente de f calculado en $\bar{x}$

$$f'(\bar{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\bar{x}) \\ \dfrac{\partial f}{\partial x_2}(\bar{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\bar{x}) \end{bmatrix}$$

$f''(\bar{x}) = \nabla^2 f(\bar{x}) = H_f(\bar{x}) = H(\bar{x}) =$ Hessiano o matriz hessiana de f en $\bar{x}$

$$f''(\bar{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(\bar{x}) & \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}) & \dots & \dfrac{\partial^2 f}{\partial x_n \partial x_1}(\bar{x}) \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}) & \dfrac{\partial^2 f}{\partial x_2^2}(\bar{x}) & \dots & \dfrac{\partial^2 f}{\partial x_n \partial x_2}(\bar{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(\bar{x}) & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(\bar{x}) & \dots & \dfrac{\partial^2 f}{\partial x_n^2}(\bar{x}) \end{bmatrix}$$

f es acotada $\Leftrightarrow$ f es acotada inferiormente.
$\min f(x) \Leftrightarrow$ minimizar $f(x)$
$x \geq y \Leftrightarrow x_i \geq y_i$ para todo i
$x \geq 0 \Leftrightarrow x_i \geq 0$ para todo i
$x \geq 0, x \ngeq 0 \Leftrightarrow x$ no tiene componentes negativas y $x_i = 0$ para algún i
$\mathbb{R}^n_+ = \{(x_1, x_2, ..., x_n) : x_i \geq 0, \forall i\} =$ el ortante no negativo de $\mathbb{R}^n$.

$$\begin{array}{c} \min f(x) \\ x \in \mathcal{A} \end{array} \quad \Leftrightarrow \quad \begin{array}{c} \min f(x) \\ \text{sujeto a} \quad x \in \mathcal{A} \end{array}$$

$\mathcal{A}^* = \mathcal{A}^*_f = \{x^* : f(x^*) \leq f(x)$ para todo $x \in \mathcal{A}\}$
$\underset{x \in S}{\text{Argmin}}\, f(x) = \{x^* : f(x^*) \leq f(x)$ para todo $x \in S\}$
$w = \underset{x \in S}{\text{argmin}}\, f(x)$ si el conjunto $\underset{x \in S}{\text{Argmin}}\, f(x)$ tiene un solo elemento w
$\overset{\circ}{S}=$ interior de S.
$\text{co}\,(S) =$ envolvente convexa de S.
$\text{pe}\,(C) =$ conjunto de puntos extremos del convexo C.

$\bar{\lambda} =$ argmej $f(x + \lambda d)$ indica un valor de λ tal que $x + \bar{\lambda} d$ es mejor que x, o sea, $f(x + \bar{\lambda} d) < f(x)$ y cumple condiciones como la de Armijo o la de Goldstein, o ...

$e^j = j$-ésima columna de la matriz identidad

$$\text{dom}\{a, b\} = \begin{cases} a & \text{si} \quad |a| \geq |b| \\ b & \text{si} \quad |a| < |b| \end{cases} \quad : \text{valor dominante}$$

$\rho(A) = \max\{|\lambda_i|_{\mathbb{C}} : \lambda_i$ es valor propio de $A\} =$ radio espectral de A.
$\text{nf}\,(A) =$ número de filas de la matriz A.
$\log x =$ logaritmo de x en base e.
$\bar{m} =$ número de desigualdades activas.
$\Diamond$: fin del ejemplo.

$\lfloor x \rfloor$ = parte entera de x = parte entera inferior de $x = \max\{n \in \mathbb{Z} : n \leq x\}$.
$\lceil x \rceil$ = parte entera superior de $x = \min\{n \in \mathbb{Z} : n \geq x\}$.
$c = \text{mejor}\{a, b\} \Leftrightarrow c \in \{a, b\}, \varphi(c) = \min\{\varphi(a), \varphi(b)\}$para una función φ dada.

En la escritura de numeros decimales, los enteros están separados de los decimales por medio de un punto. No se usa la notación española (los enteros están separados de los decimales por una coma). No se utiliza un símbolo para separar las unidades de mil de las centenas.

Capítulo 1

INTRODUCCIÓN

1.1 GENERALIDADES

Se puede decir que un problema de programación matemática (optimización), consiste en encontrar un punto que minimice (o que maximice) el valor de una función $f(x)$, con la condición de que x esté en un conjunto $\mathcal{A}$. Este problema se puede denotar así:

$$\begin{aligned} &\min \ f(x) \\ \text{sujeto a} \quad &x \in \mathcal{A}, \end{aligned}$$

y de manera más sencilla, eliminando la expresión *sujeto a*,

$$\begin{aligned} &\min \ f(x) \\ &x \in \mathcal{A}. \end{aligned}$$

Usualmente el conjunto $\mathcal{A}$ está definido por medio de igualdades y desigualdades matemáticas.

El objetivo de la programación lineal es la minimización de una función lineal, con restricciones (igualdades o desigualdades) también lineales. De manera general, un problema de programación lineal es de la forma:

$$\begin{array}{lcl} \min & f(x) & \\ g_1(x) & \leq & 0 \\ g_2(x) & \leq & 0 \end{array}$$

$$\begin{array}{rcl} & \cdots & \\ g_m(x) & \leq & 0 \\ h_1(x) & = & 0 \\ h_2(x) & = & 0 \\ & \cdots & \\ h_l(x) & = & 0, \end{array}$$

donde f, g_i, h_j son funciones de $\mathbb{R}^n$ en $\mathbb{R}$, f es una función lineal, y las restricciones (desigualdades e igualdades) son lineales, es decir, las funciones g_i, h_j son afines. Una función es afín si se puede expresar como una función lineal más una constante. Si f es lineal se puede expresar de la forma

$$c_1\, x_1 \,+\, c_2\, x_2 \,+ \cdots + \, c_n\, x_n,$$

donde c_1, c_2, ... , c_n son constantes. Si g_i es una función afín, se puede expresar de la forma

$$a_{i1} x_1 \,+\, a_{i2} x_2 \,+ \cdots + \, a_{in} x_n \,+\, b_i,$$

donde a_{i1}, a_{i2}, ... , a_{in}, b_i son constantes.

Si f no es lineal, o si alguna restricción no es lineal (la función correspondiente no es afín), se tiene un problema de **programación no lineal**, PNL.

Minimizar una función lineal no nula sin restricciones, no tiene interés ya que el óptimo no es acotado, o sea, el valor de f no es acotado inferiormente (tampoco superiormente). Mientras no se diga lo contrario, para subconjuntos de $\mathbb{R}$ o para funciones de valor real, acotado significará acotado inferiormente.

En el caso no lineal, minimizar una función de varias variables sin restricciones, sí es un problema usual e importante. Obviamente, si además hay restricciones el problema se hace más difícil.

Vale la pena recordar y presentar algunos resultados usuales, algunas suposiciones y notaciones usadas en este libro. El conjunto de puntos que cumplen todas las restricciones se llama **conjunto admisible**, **factible** o **realizable** y se denotará por $\mathcal{A}$. Si no hay ninguna restricción se sobreentiende que $\mathcal{A} = \mathbb{R}^n$.

Todo problema de maximización se puede convertir en un problema de minimización multiplicando por -1 la función objetivo (función objetivo o

función económica es la función que se desea minimizar o maximizar), y se obtiene un problema equivalente.

$$\begin{aligned} &\max \varphi(x) \\ &x \in \mathcal{A}, \end{aligned} \tag{1.1}$$

es equivalente a

$$\begin{aligned} &\min \ f(x) = -\varphi(x) \\ &x \in \mathcal{A}, \end{aligned} \tag{1.2}$$

o sea, $\xi^* = x^*$ y $\varphi^* = -f^*$, donde ξ^* es el punto óptimo del problema (1.1), φ^* es el valor máximo de (1.1), x^* es el punto óptimo del problema (1.2), f^* es el valor mínimo de (1.2).

Si la función objetivo se multiplica o se divide por una constante positiva o si se le suma o se resta cualquier constante, entonces se obtiene un problema equivalente. Es decir, sean $\alpha > 0$, $\beta \in \mathbb{R}$,

$$\begin{aligned} &\min \ \varphi(x) = \alpha f(x) + \beta \\ &x \in \mathcal{A}, \end{aligned} \tag{1.3}$$

es equivalente a

$$\begin{aligned} &\min \ f(x) \\ &x \in \mathcal{A}, \end{aligned} \tag{1.4}$$

o sea, $\xi^* = x^*$ y $\varphi^* = \alpha f^* + \beta$, donde ξ^* es el punto óptimo del problema (1.3), φ^* es el valor mínimo de (1.3), x^* es el punto óptimo del problema (1.4), f^* es el valor mínimo de (1.4).

Una restricción de la forma $\geq$ se puede convertir en una desigualdad de la forma $\leq$ simplemente multiplicando ambos lados de la desigualdad por -1, o sea,

$$p(x) \ \geq \ 0,$$

es equivalente a

$$-p(x) \leq 0.$$

En programación lineal y en programación no lineal, por razones prácticas, no se consideran desigualdades estrictas. Una de las razones es que los números utilizados en los computadores son sólo aproximaciones de los números reales, y a veces la aproximación para dos números reales diferentes es la misma, por ejemplo, los números 1 y $1 + 10^{-20}$, en la mayoría de las representaciones usuales en computador, no se diferencian. Por otro lado, si tomamos el intervalo $[0, 1]$, el conjunto de números con que trabaja el computador es finito, y es mucho más pequeño que el conjunto de racionales en $[0, 1]$, a su vez subconjunto propio de $[0, 1]$.

Además las desigualdades estrictas generalmente dan lugar a conjuntos admisibles abiertos, en los cuales es más difícil o a veces imposible garantizar la existencia de un punto óptimo, por ejemplo, el siguiente problema no tiene solución

$$\begin{aligned} \min\ & x_1 + x_2 \\ & x_1 > 0 \\ & x_2 > 0. \end{aligned}$$

Una manera de tratar una desigualdad estricta consiste en escoger un número positivo, suficientemente pequeño ε, y suponer que hay equivalencia "práctica" entre las dos desigualdades siguientes

$$\begin{aligned} q(x) &< 0, \\ q(x) &\leq -\varepsilon. \end{aligned}$$

Si las variables x_1, ... , x_n -además de cumplir con las igualdades y desigualdades- deben ser enteras, entonces no se tiene un problema de PNL usual, sino un problema de **programación no lineal entera**. Si las variables únicamente pueden tomar los valores 0 y 1, entonces se habla de **programación no lineal binaria**.

A continuación hay algunos ejemplos sencillos de planteamiento de problemas de PNL, algunos sin restricciones, otros con restricciones, algunos con pocas variables, otros con la posibilidad de muchas variables.

1.2 APROXIMACIÓN POR MÍNIMOS CUADRADOS: UNA RECTA

Conocidos $m \geq 2$ puntos (x_1, y_1), (x_2, y_2), ...,(x_m, y_m), se desea encontrar una línea recta que pase lo "más cerca" posible de ellos. Supóngase que no hay dos puntos que tengan la misma coordenada x_i.

La manera usual de medir la cercanía de la recta a los puntos es sumando el cuadrado de la distancia vertical de cada punto a la recta. De ahí el nombre de mínimos cuadrados. Una recta cualquiera, no vertical, se puede representar de la forma $y = ax + b$. Para nuestro caso, se desea determinar los coeficientes a y b. La distancia vertical entre el punto (x_i, y_i) y la recta es

$$|y_i - (ax_i + b)|,$$

luego se desea encontrar los valores a, b que minimicen

$$f(a, b) = \sum_{i=1}^{m} (ax_i + b - y_i)^2.$$

Por ejemplo, si hay cuatro puntos $(0,\ 0.1)$, $(1,\ 0.9)$, $(2,\ 4.1)$, $(2.9,\ 8.3)$ se desea minimizar

$$f(a, b) = (b - 0.1)^2 + (a + b - 0.9)^2 + (2a + b - 4.1)^2 + (2.9a + b - 8.3)^2.$$

En una calculadora científica de bolsillo se puede resolver este problema, es decir, se obtienen los valores óptimos

$$\begin{aligned} a^* &= 2.8476\ , \\ b^* &= -0.8502\ , \\ f^* = f(a^*, b^*) &= 3.4581\ , \\ y &= 2.8476\,x - 0.8502\ . \end{aligned}$$

Si en lugar de una recta se escogiera una parábola $y = c + bx + ax^2$, la "mejor" sería

$$\begin{aligned} y &= 0.0612 - 0.0425x + 0.9997x^2 \\ f^* &= 0.0332\ .. \end{aligned}$$

1.3 APROXIMACIÓN POR MÍNIMOS CUADRADOS: CASO GENERAL

Conocidos m puntos (x_1, y_1), (x_2, y_2), ..., (x_m, y_m) y un conjunto de $n \leq m$ funciones linealmente independientes $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$, se desea encontrar coeficientes α_1, α_2, ..., α_n tales que la función

$$\varphi(x) = \alpha_1\varphi_1(x) + \alpha_2\varphi_2(x) + \cdots + \alpha_n\varphi_n(x),$$

pase lo más cerca posible, en el sentido de mínimos cuadrados, de los m puntos.

Algunos ejemplos de conjuntos de funciones pueden ser :

1, x, x^2, x^{m-1}
1, e^x, e^{2x}, ..., $e^{(m\text{-}1)x}$
1, $\cos(x)$, $\operatorname{sen}(x)$, $\cos(2x)$, ...
$\vdots$

Para encontrar la función $\varphi(x)$ que aproxima por mínimos cuadrados los m puntos se necesita minimizar

$$\begin{aligned} f(\alpha_1, ..., \alpha_n) &= \sum_{i=1}^{m} (\varphi(x_i) - y_i)^2 \\ f(\alpha) &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j\varphi_j(x_i) - y_i \right)^2 . \end{aligned}$$

Sea Φ la matriz de tamaño $m \times n$

$$\Phi = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \dots & \varphi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_m) & \varphi_2(x_m) & \dots & \varphi_n(x_m) \end{bmatrix}.$$

Entonces

$$\sum_{j=1}^{n} \alpha_j\varphi_j(x_i) = \begin{bmatrix} \varphi_1(x_i) & \varphi_2(x_i) & \dots & \varphi_n(x_i) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

$$= \quad \Phi_{i.}\,\alpha,$$

o sea, se trata de resolver el problema

$$\begin{aligned}
\min\ f(\alpha) &= \sum_{i=1}^{m} (\Phi_{i.}\,\alpha - y_i)^2 \\
&= \sum_{i=1}^{m} \big((\Phi\,\alpha)_i - y_i\big)^2 \\
&= \sum_{i=1}^{m} \big(\Phi\,\alpha - y\big)_i^2 \\
\min\ f(\alpha) &= ||\,\Phi\,\alpha - y\,||_2^2.
\end{aligned}$$

1.4 SOLUCIÓN POR MÍNIMOS CUADRADOS DE UN SISTEMA LINEAL

Un sistema de m ecuaciones lineales con n incógnitas se puede escribir en la forma

$$\begin{aligned}
A\,x &= b \\
A\,x - b &= 0 \\
||\,A\,x - b\,|| &= 0 \\
||\,A\,x - b\,||_2 &= 0 \\
||\,A\,x - b\,||_2^2 &= 0.
\end{aligned}$$

Si el número de ecuaciones es mayor que el número de incógnitas, $m \geq n$, es muy probable que el sistema no tenga solución, es decir, no se podrá encontrar un x tal que la norma de $A\,x - b$ sea nula. Sin embargo, se puede tratar de encontrar un x^* que minimice el cuadrado de la norma:

$$\min\ f(x) = ||\,A\,x - b\,||_2^2.$$

Este x^* se llama **solución por mínimos cuadrados** o **seudosolución**. Obviamente si existe solución verdadera, ésta también será seudosolución.

Se ve claramente que el problema general de aproximación por mínimos cuadrados es un problema de solución de un sistema de ecuaciones lineales (más ecuaciones que incógnitas), por mínimos cuadrados.

La manera más usual de resolver el problema anterior, cuando las columnas de A son linealmente independientes y A no es de tamaño muy grande, es mediante la solución de las ecuaciones normales ([Bj´o96], [Sto93], [Gol89]):

$$A^{\mathrm{T}} A\, x = A^{\mathrm{T}} b.$$

1.5 DISEÑO DE UNA COLUMNA TUBULAR DE PESO MÍNIMO

En este problema, adaptación del propuesto en [Aro89], se tiene una columna (vertical) en forma de tubo de altura h, empotrada en la base y libre en la parte superior. Se desea determinar el radio promedio R y el espesor e de la columna de manera que soporte una carga vertical P y que el peso de la columna sea mínimo. En esta sección e no indica el valor 2.71828... . Por razones de construcción, es necesario cumplir con ciertas medidas mínimas y máximas para el radio y para el espesor:

$$R_{\min} \leq R \leq R_{\max} \ , \quad e_{\min} \leq e \leq e_{\max}.$$

Para que la columna se pueda considerar como tubular se necesita que el radio promedio sea mayor que 20 veces el espesor. La carga máxima que puede soportar este tipo de columna, sin que falle por pandeo, está dada por $\pi^2 EI/4h^2$, donde I es el momento de inercia de la sección transversal de la columna y E es el módulo de elasticidad del material. Se conoce, además, la densidad ϱ y el máximo esfuerzo de compresión σ_a que puede soportar el material.

Si $R >> e$, entonces el área transversal y el momento de inercia están dados por:

$$A = 2\pi Re \ , \qquad I \approx \pi R^3 e.$$

Al minimizar la masa de la columna se tiene

$$\min \ f(R, e) = 2\varrho h \pi R e,$$

y las restricciones son:

$$P \ \leq \ \frac{\pi^3 R^3 E e}{4h^2}$$

$$\begin{aligned} \frac{P}{2\pi Re} &\leq \sigma_a \\ R_{\min} &\leq R \leq R_{\max} \\ e_{\min} &\leq e \leq e_{\max} \\ 0 &\leq 20e \leq R. \end{aligned}$$

Uitilizando la forma general:

$$\begin{aligned} \min \quad f(R,e) &= 2\varrho h\pi Re \\ 4Ph^2 - \pi^3 R^3 Ee &\leq 0 \\ P - 2\pi Re\sigma_a &\leq 0 \\ R_{\min} - R &\leq 0 \\ R - R_{\max} &\leq 0 \\ e_{\min} - e &\leq 0 \\ e - e_{\max} &\leq 0 \\ 20e - R &\leq 0 \\ -e &\leq 0. \end{aligned}$$

Este problema se puede enfocar de otra manera, considerando como variables el radio interno R_1 y el radio externo R_2,

$$\begin{aligned} A &= \pi(R_2^2 - R_1^2), \\ I &= \frac{\pi}{4}(R_2^4 - R_1^4). \end{aligned}$$

Entonces el problema quedaría de la siguiente forma:

$$\begin{aligned} \min\ f(R_1,R_2) &= \pi\varrho h(R_2^2 - R_1^2) \\ \frac{P}{\pi(R_2^2 - R_1^2)} &\leq \sigma_a \\ P &\leq \frac{\pi^3 E(R_2^4 - R_1^4)}{16h^2} \\ R_{\min} &\leq R_1 \\ R_2 &\leq R_{\max} \\ e_{\min} &\leq R_2 - R_1 \leq e_{\max} \\ \frac{R_1 + R_2}{2(R_2 - R_1)} &\geq 20. \end{aligned}$$

1.6 CONSTRUCCIÓN DE UNA CAJA DE COSTO MÍNIMO

Este problema es una adaptación del planteado en [Sim75]. Una compañía química necesita transportar 1000 metros cúbicos de un gas muy tóxico, desde su centro de producción a un laboratorio situado en otro departamento.

Para ello desea construir una caja muy hermética, en forma de paralelepípedo rectangular, para ser transportada en tractomula. El material para el fondo y la tapa cuesta $200000 por metro cuadrado. El de las caras laterales cuesta $100000 por metro cuadrado y su disponibilidad es de 50 metros cuadrados. Las dimensiones máximas que puede tener la caja para ser llevada en una tractomula son: 2 metros de ancho, 5 metros de largo y 3 metros de alto. Independientemente de las dimensiones de la caja, cada viaje redondo (ida y vuelta) en tractomula cuesta $800000. Asumiendo que no hay límite para el tiempo total necesario para todos los viajes, ¿qué dimensiones debe tener la caja para minimizar el costo total, es decir, el costo de transporte más el de construcción?

Sean:

$$\begin{aligned} x_1 &= \text{ancho de la caja,} \\ x_2 &= \text{largo de la caja,} \\ x_3 &= \text{alto de la caja.} \end{aligned}$$

$$\begin{aligned} \min\ z &= 800000\lceil \frac{1000}{x_1 x_2 x_3} \rceil + 200000(2x_1x_2) \\ &\quad +100000(2x_1x_3 + 2x_2x_3) \\ x_1 &\leq 2 \\ x_2 &\leq 5 \\ x_3 &\leq 3 \\ 2x_1x_3 + 2x_2x_3 &\leq 50 \\ x &\geq 0, \end{aligned}$$

donde $\lceil t \rceil$ indica la parte entera superior de t, es decir, el mínimo entero mayor o igual a t, así, por ejemplo, $\lceil 2.1 \rceil = 3$, $\lceil 2 \rceil = 2$, $\lceil -1.9 \rceil = -1$. La expresión $x \geq 0$ indica que todas las variables deben ser no negativas, o sea, para este caso, x_1, x_2, $x_3 \geq 0$.

1.7 LOCALIZACIÓN DE UNA CENTRAL

En una región hay m ciudades, de cada ciudad se conocen las coordenadas (u_i, v_i) y el número de habitantes n_i. Se desea determinar las coordenadas (x_1, x_2) de una central termoeléctrica de manera que se minimice la suma de las pérdidas entre la central y las diferentes ciudades. Se conoce una función creciente g tal que dada una distancia d_i entre la central y una ciudad, conocido $\bar{c}$ consumo promedio por habitante (igual en todas las ciudades), entonces las pérdidas entre la central y esta ciudad están dadas por $n_i \bar{c}\, g(d_i)$

La distancia entre la central y una ciudad es

$$\sqrt{(x_1 - u_i)^2 + (x_2 - v_i)^2},$$

luego las pérdidas están dadas por

$$n_i \bar{c}\, g\left(\sqrt{(x_1 - u_i)^2 + (x_2 - v_i)^2}\right).$$

Entonces se desea minimizar la suma total de pérdidas

$$\min \ f(x_1, x_2) = \sum_{i=1}^{m} n_i \bar{c}\, g\left(\sqrt{(x_1 - u_i)^2 + (x_2 - v_i)^2}\right).$$

Como $\bar{c}$ es constante,

$$\min \ f(x_1, x_2) = \sum_{i=1}^{m} n_i g\left(\sqrt{(x_1 - u_i)^2 + (x_2 - v_i)^2}\right).$$

1.8 CONVERGENCIA Y ORDEN DE CONVERGENCIA

Sea $||\ ||$ una norma sobre $\mathbb{R}^n$. Se dice que una sucesión de vectores $\{x^k\}$ converge a x^* si la sucesión de números reales $\{\, ||x^k - x^*|| \,\}$ tiende a cero. Como todas las normas de $\mathbb{R}^n$ son equivalentes entonces no importa cual norma se use. Dicho de otra forma, si utilizando una norma hay convergencia entonces con cualquiera otra norma la sucesión de vectores será convergente hacia el mismo vector x^* y, obviamente, si con una norma no hay convergencia, entonces no es posible que haya convergencia con otra norma. Si la sucesión de vectores es convergente, se denotará

$$\lim_{k\to\infty} x^k = x^*,$$

o también,

$$x^k \underset{k\to\infty}{\longrightarrow} x^*,$$

o de manera más compacta y sencilla,

$$x^k \to x^*.$$

Ejemplo 1.1.

$$\begin{aligned} x^k = (2 - \frac{1}{k}, 3 + \frac{1}{k^2}) \quad &\to \quad (2,3) = x^*, \\ \text{ya que} \quad ||x^k - x^*||_2 = \sqrt{\frac{1}{k^2} + \frac{1}{k^4}} \quad &\to \quad 0, \\ \text{o también} \quad ||x^k - x^*||_\infty = \frac{1}{k} \quad &\to \quad 0. \ \diamond \end{aligned}$$

Definición 1.1. Una sucesión de vectores $\{x^k\}$ que converge a x^* se dice que tiene **orden de convergencia** $p > 0$, si p es el mayor valor tal que

$$\lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||^p} = \beta < \infty.$$

Si $p = 1$ y $\beta < 1$ se dice que la convergencia es **lineal** con **tasa de convergencia** β. Algunos autores no colocan restricción sobre β para llamar lineal a la convergencia de orden uno. En la práctica, para que la convergencia lineal sea buena, se requiere que $\beta \leq \frac{1}{4}$, [Fle87]. Si $p = 2$ la convergencia se llama **cuadrática**. Si

$$\lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0,$$

la convergencia se llama **superlineal** y se presenta cuando $p > 1$ o cuando $p = 1$ y $\beta = 0$.

También se puede definir el orden de convergencia en términos de cotas de los cocientes

$$\begin{aligned} \frac{||x^{k+1} - x^*||}{||x^k - x^*||^p} &\leq \beta < \infty, \\ \text{o también} \quad ||x^{k+1} - x^*|| &= O(\,||x^k - x^*||^p\,). \end{aligned}$$

Ejemplo 1.2.

$$x^k = (\,2 + \frac{1}{k},\, 3 - \frac{1}{k}\,)$$

tiene orden de convergencia uno, pero no es lineal ya que $x^k \to (2,3)$ y

$$\lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = \lim_{k\to\infty} \frac{\frac{1}{k+1}\sqrt{2}}{\frac{1}{k}\sqrt{2}} = 1. \quad \diamond$$

Ejemplo 1.3.

$$x^k = \frac{1}{k^2}$$

Aparentemente, al ver un cuadrado en el denominador, se podría pensar que el orden de convergencia es 2. Sin embargo, tiene orden de convergencia uno con $\beta = 1$. O sea, tiene el mismo orden de convergencia y el mismo valor β del ejemplo anterior. $\diamond$

Ejemplo 1.4.

$$x^{k+1} = x^k\, x^k = (x^k)^2 \qquad , \text{ con} \qquad -1 < x^1 < 1,$$

tiene orden de convergencia 2, o sea, cuadrático, con $\beta = 1$. También se puede decir que la convergencia es superlineal. Los primeros valores son: -0.9, 0.81, 0.6561, 0.430467, 0.185302, 0.034337, 0.001179, 0.000001, ... $\diamond$

Ejemplo 1.5.

$$x^k = 3 + \frac{1}{2^{2^k}}$$

tiene orden de convergencia 2, con $\beta = 1$. $\diamond$

Ejemplo 1.6.

$$x^k = \frac{1}{3^k\, k^2} \qquad , \qquad y^k = \frac{1}{3^k}$$

tienen convergencia lineal con tasa $\frac{1}{3}$. $\diamond$

Ejemplo 1.7.

$$x^k = \frac{1}{k!}$$

tiene convergencia lineal con tasa 0, o sea, es superlineal. $\diamond$

Ejemplo 1.8.

$$x^{k+1} = x^k\sqrt{x^k}\ , \quad 0 < x^1 < 1.$$

tiene convergencia de orden 1.5 . $\diamond$

EJERCICIOS

1.1. Se desea encontrar una recta que pase lo más cerca posible, en el sentido de mínimos cuadrados, de los puntos $(1, 7)$, $(2, 11)$, $(3, 18)$, $(4, 28)$. Plantee explícitamente el problema de optimización correspondiente.

1.2. Se desea encontrar una recta de pendiente no negativa que pase lo más cerca posible, en el sentido de mínimos cuadrados, de los puntos $(1, 7)$, $(2, 11)$, $(3, 18)$, $(4, 28)$. Plantee explícitamente el problema de optimización correspondiente.

1.3. Se desea encontrar una parábola que pase lo más cerca posible, en el sentido de mínimos cuadrados, de los puntos $(1, 7)$, $(2, 11)$, $(3, 18)$, $(4, 28)$. Plantee explícitamente el problema de optimización correspondiente.

1.4. Se desea encontrar una parábola convexa que pase lo más cerca posible, en el sentido de mínimos cuadrados, de los puntos $(1, 7)$, $(2, 11)$, $(3, 18)$, $(4, 28)$. Plantee explícitamente el problema de optimización correspondiente.

1.5. Se desea construir un recipiente en hojalata, en forma de cilindro circular recto, con fondo, pero sin tapa. Su volumen debe ser 1000 cm^3, el cociente entre la altura y el diámetro debe variar entre 1 y 1.5 y la altura no debe ser superior a 40 cm. Plantee el problema si se desea minimizar la hojalata usada.

1.6. Se desea encontrar una seudosolución o solución por mínimos cuadrados del sistema sobredeterminado $x_1 + 2x_2 = 2$; $3x_1 + 4x_2 = 1$; $5x_1 + 6x_2 = 1$. Plantee explícitamente el problema de optimización correspondiente.

1.7. Se desea encontrar una seudosolución o solución por mínimos cuadrados, no negativa, del sistema sobredeterminado $x_1 + 2x_2 = 2$; $3x_1 + 4x_2 = 1$; $5x_1 + 6x_2 = 1$. Plantee explícitamente el problema de optimización correspondiente.

Capítulo 2

CONJUNTOS CONVEXOS

Sea V el espacio vectorial $\mathbb{R}^n$. Mientras no se diga lo contrario, todos los conjuntos son subconjuntos de $\mathbb{R}^n$, todos los puntos o vectores son elementos de $\mathbb{R}^n$, todos los números son números reales. La mayoría de las definiciones y resultados que siguen, se pueden generalizar fácilmente a otros espacios vectoriales.

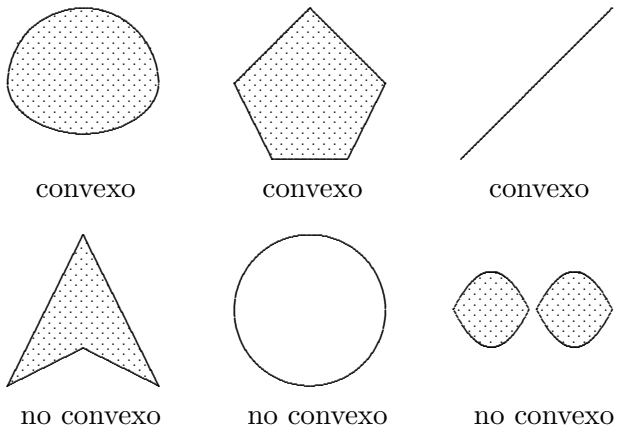


Figura 2.1

Definición 2.1. Sea C un subconjunto de V. Se dice que C es **convexo** si dados x, y en C, λ un escalar en el intervalo $[0, 1]$, entonces $z = z_{xy\lambda} =$ (1-λ)$x + \lambda y$ también está en C. Gráficamente, un conjunto C es convexo si dados dos puntos x, y en C , cualquier punto del segmento de recta que los une, también está en C.

Ejemplos triviales de conjuntos convexos son: V, $\emptyset$, $\{\bar{x}\}$.

Ejemplo 2.1. $\{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$ es convexo. $\diamond$

Ejemplo 2.2. La bola (o esfera) cerrada de $\mathbb{R}^n$, con centro en c y radio r denotada por $\bar{B}(c, r) = B[c, r] = \{x : ||x - c|| \leq r\}$ y la bola abierta $B(c, r) = \{x : ||x - c|| < r\}$, son conjuntos convexos. $\diamond$

Ejemplo 2.3. Dados $\bar{x}$, $d \neq 0$ elementos de $\mathbb{R}^n$, la recta que pasa por $\bar{x}$ y es paralela a d, o sea, $R(\bar{x}, d) = \{\bar{x} + \mu d : \mu \in \mathbb{R}\}$ y la semirrecta que empieza en $\bar{x}$ y va en la dirección de d, o sea , $S(\bar{x}, d) = \{\bar{x} + \mu d : \mu \geq 0\}$, son ejemplos de conjuntos convexos.$\diamond$

Ejemplo 2.4. $C = \{(x_1, x_2) : x_2 = x_1^2\}$ no es convexo ya que $(0, 1) = \frac{1}{2}(1, 1) + \frac{1}{2}(-1, 1)$ no está en el conjunto, aunque $(1, 1)$ y $(-1, 1)$ sí están en C. $\diamond$

Ejemplo 2.5. $\{(x_1, x_2) : x_2 \geq x_1^2\}$ sí es convexo. $\diamond$

Definición 2.2. Dados $c \in \mathbb{R}^n$, $c \neq 0$, $\alpha \in \mathbb{R}$, se llama **hiperplano** al siguiente conjunto:

$$H = H_{c,\alpha} = \{x \in \mathbb{R}^n : c^{\mathrm{T}} x = \alpha\}.$$

Este hiperplano genera dos semiespacios cerrados:

$$H^+ = \{x \in \mathbb{R}^n : c^{\mathrm{T}} x \geq \alpha\},$$
$$H^- = \{x \in \mathbb{R}^n : c^{\mathrm{T}} x \leq \alpha\},$$

y dos semiespacios abiertos:

$$\overset{\circ}{H}{}^+ = \{x \in \mathbb{R}^n : c^{\mathrm{T}} x > \alpha\},$$
$$\overset{\circ}{H}{}^- = \{x \in \mathbb{R}^n : c^{\mathrm{T}} x < \alpha\}.$$

Ejemplo 2.6. El conjunto $\{(x_1, x_2, x_3) : 2x_1 - 3x_2 + 4x_3 = 5\}$ es un hiperplano de $\mathbb{R}^3$. El conjunto $\{(x_1, x_2, x_3) : 2x_1 - 3x_2 + 4x_3 > 5\}$ es un semiespacio abierto de $\mathbb{R}^3$. $\diamond$

En $\mathbb{R}$ un hiperplano es un punto y los semiespacios semirrectas. En $\mathbb{R}^2$ los hiperplanos son las rectas y los semiespacios los semiplanos. En $\mathbb{R}^3$ los hiperplanos son los planos.

Los conjuntos H, H^+, H^-, $\overset{\circ}{H}{}^+$, $\overset{\circ}{H}{}^-$ son convexos. Veamos que H es convexo. Sean: $x, y \in H$, $\lambda \in [0,1]$, $z = (1\text{-}\lambda)x + \lambda y$. El punto z está en H si y solamente si $c^{\mathrm{T}}z = \alpha$; efectuando el cálculo:

$$c^{\mathrm{T}}z = c^{\mathrm{T}}((1\text{-}\lambda)x + \lambda y) = (1\text{-}\lambda)c^{\mathrm{T}}x + \lambda c^{\mathrm{T}}y = (1\text{-}\lambda)\alpha + \lambda\alpha = \alpha,$$

luego z está en H, luego H es convexo. En esta demostración no se utilizó que $\lambda \in [0,1]$, entonces no sólo los puntos del segmento de recta están en H, sino que todos los puntos de la recta que pasa por x y y también están en H, es decir, H es una variedad lineal. Un conjunto L es una variedad lineal o variedad afín si dados x, y en L, λ un escalar, entonces $z = z_{xy\lambda} = (1\text{-}\lambda)x + \lambda y$ también está en L.

Veámos ahora que H^+ es convexo. Sean: $x, y \in H^+$, $\lambda \in [0,1]$, $z = (1\text{-}\lambda)x + \lambda y$. Entonces $c^{\mathrm{T}}x \geq \alpha$, $c^{\mathrm{T}}y \geq \alpha$, λ, $1\text{-}\lambda \geq 0$.

$$c^{\mathrm{T}}z = c^{\mathrm{T}}((1\text{-}\lambda)x + \lambda y) = (1\text{-}\lambda)c^{\mathrm{T}}x + \lambda c^{\mathrm{T}}y \geq (1\text{-}\lambda)\alpha + \lambda\alpha = \alpha,$$

entonces H^+ también es convexo, y de manera semejante se comprueba que H^-, $\overset{\circ}{H}{}^+$, $\overset{\circ}{H}{}^-$ son convexos.

Proposición 2.1. *H es un hiperplano si y solamente si H es una translación de un subespacio vectorial de dimensión n-1. Dicho de otra forma, H es un hiperplano si y solamente si, para todo $x \in H$, el conjunto $H - x = H - \{x\} = \{y - x : y \in H\}$ es un subespacio vectorial de dimensión $n - 1$.*

Proposición 2.2. *La intersección de dos conjuntos convexos es un convexo.*

La demostración es muy sencilla. Sean: C, D convexos, $x, y \in C \cap D$, $\lambda \in [0,1]$, $z = (1\text{-}\lambda)x + \lambda y$. Como C es convexo entonces $z \in C$. Como D es convexo $z \in D$. Luego $z \in C \cap D$.

Proposición 2.3. *La intersección de cualquier familia de conjuntos convexos es un convexo, independientemente de que la familia sea finita, infinita, enumerable o no enumerable. Dicho de otra forma, sea $\{C_i\}_{i \in I}$ una familia de conjuntos convexos, entonces*

$$\bigcap_{i \in I} C_i \text{ es un conjunto convexo.}$$

En cambio, no se puede afirmar que la unión de dos convexos sea siempre un convexo. Por ejemplo, en $\mathbb{R}^2$ las bolas $B((0,0),1)$, $B((2,0),1)$ son conjuntos convexos, pero su unión no es un convexo.

Ejemplo 2.7. Las restricciones de un problema de programación lineal son igualdades, es decir representan hiperplanos, o bien, son desigualdades y en este caso representan semiespacios. Así cualquier conjunto admisible de un problema de programación lineal es simplemente la intersección de hiperplanos y semiespacios, luego es un conjunto convexo. En particular, dada una matriz real A de tamaño $m \times n$, los siguientes conjuntos son convexos.

$$\{x \in \mathbb{R}^n : Ax = b\},$$
$$\{x \in \mathbb{R}^n : Ax \geq b\},$$
$$\{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\},$$
$$\{x \in \mathbb{R}^n : Ax \geq b,\ x \geq 0\}. \quad \diamond$$

Ejemplo 2.8.

$$\bigcap_{0 \leq \theta \leq 2\pi} \{(x_1, x_2) : x_1 \cos\theta + x_2 \text{ sen } \theta \leq 1\},$$

es decir, la bola (para la norma $\| \ \|_2$) con centro en $\mathbf{0} = (0,0)$ y radio 1, denotada $B_2[\mathbf{0}, 1]$, es un conjunto convexo, puesto que es intersección de semiespacios. $\diamond$

Definición 2.3. Se llama un **polítopo** a cualquier conjunto que se pueda expresar como la intersección de un número finito de semiespacios cerrados (o de semiespacios cerrados e hiperplanos).

Directamente de la definición se puede concluir que un polítopo es un conjunto convexo y cerrado.

Definición 2.4. Un **poliedro** es un polítopo acotado.

Ejemplo 2.9. El conjunto admisible de cualquier problema de programación lineal es un polítopo. $\diamond$

Ejemplo 2.10. El subconjunto de $\mathbb{R}^2$ definido por las siguientes restricciones es un poliedro.

$$-x_1 + x_2 \ \geq \ 2$$

$$\begin{aligned} x_2 &\geq 3 \\ x_2 &\leq 5 \\ x &\geq 0. \ \diamond \end{aligned}$$

Ejemplo 2.11. $B_2[\mathbf{0},1] = \bigcap_{0\leq\theta\leq 2\pi} \{(x_1,x_2): x_1\cos\theta + x_2 \ \operatorname{sen}\theta \leq 1\}$ no es un polítopo. $\diamond$

Proposición 2.4. *Si C es un conjunto convexo y α un número, entonces*

$$\alpha C = \{\alpha x : x \in C\}$$

es un conjunto convexo.

Proposición 2.5. *Si C y D son conjuntos convexos, entonces*

$$C + D = \{x + y : x \in C, y \in D\}$$

es un conjunto convexo.

Corolario 2.1. *Si C, D son conjuntos convexos y α, β son números, entonces*

$$\alpha C + \beta D = \{\alpha x + \beta y : x \in C, y \in D\}$$

es un conjunto convexo. En particular $C - D = \{x - y : x \in C, y \in D\}$ también es convexo.

Ejemplo 2.12. Partiendo de que $B((0,0),1)$ es un conjunto convexo, se puede afirmar que

$$B((2,-3),4) = \{(2,-3)\} + 4\ B((0,0),1)$$

es también un conjunto convexo. $\diamond$

Proposición 2.6. *Si $C \subseteq \mathbb{R}^m$, $D \subseteq \mathbb{R}^p$ son convexos, entonces $C \times D \subseteq \mathbb{R}^{m+p}$, también es convexo.*

Ejemplo 2.13. El intervalo $[2,3]$ es convexo. También son convexos $B[0,2] \subseteq \mathbb{R}^2$ y $\mathbb{R}^2$. Luego el conjunto $\{(x_1,x_2,x_3): 2 \leq x_1 \leq 3,\ x_2^2 + x_3^2 \leq 4\}$ y el conjunto $\{(x_1,x_2,x_3): 2 \leq x_1 \leq 3\}$ también son convexos. $\diamond$

Definición 2.5. Se llama **combinación convexa** de x^1, x^2, ... , x^m elementos de V a una combinación lineal en la que todos los escalares son no negativos y además su suma es uno, es decir:

$$x = \lambda_1 x^1 + \lambda_2 x^2 + \ldots + \lambda_m x^m,\ \lambda_i \geq 0\ \forall i,\ \sum_{i=1}^{m} \lambda_i = 1.$$

Si todos los escalares son positivos la combinación convexa se llama **estricta**. Se denotará por cc(A) el **conjunto de todas las combinaciones convexas** de elementos de A, es decir, el conjunto de todas las combinaciones convexas de subconjuntos finitos de A.

La combinación convexa es la generalización de la expresión (1-λ)$x + \lambda y$ con λ en el intervalo $[0, 1]$.

Ejemplo 2.14. Dados $(1, 0)$, $(0, 0)$ y $(0, 1)$, son ejemplos de combinaciones convexas:

$$\begin{aligned} (\frac{1}{2}, \frac{1}{4}) &= \frac{1}{2}(1,0) + \frac{1}{4}(0,0) + \frac{1}{4}(0,1) \\ (0,1) &= 0(1,0) + 0(0,0) + 1(0,1). \quad \diamond \end{aligned}$$

Definición 2.6. Sea A un subconjunto de V. Se llama **envolvente convexa** de A, o **convexo generado** por A, o **casco convexo** de A, denotado co(A), al conjunto convexo más pequeño que contenga a A. Esto quiere decir que si C es un conjunto convexo que contiene a A , entonces necesariamente co(A) está contenido en C.

La anterior definición es descriptiva, pero no constructiva.

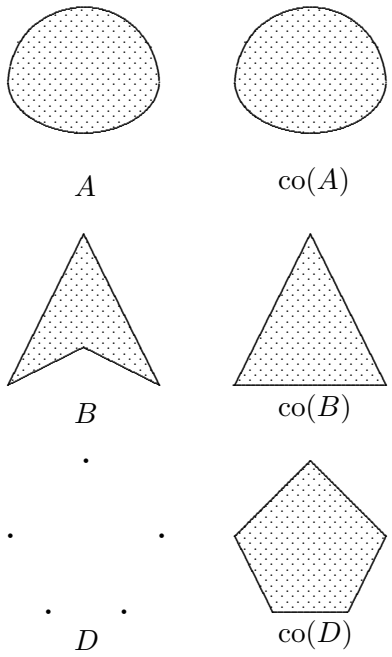


Figura 2.2

Proposición 2.7. *El convexo generado por A se puede caracterizar "constructivamente" como la intersección de todos los convexos que contienen a A,*

$$co\,(A) = \bigcap_{\substack{C \text{ convexo,} \\ A \subseteq C}} C.$$

Esta intersección está bien definida ya que por lo menos existe un conjunto convexo que contiene a A: el espacio completo $\mathbb{R}^n$.

Proposición 2.8. $co\,(A) = cc(A)$.

Ejemplo 2.15. $\mathrm{co}\,(\{(1,0),(0,1),(0,0)\}) = \{(x_1,x_2)\colon x_1+x_2 \le 1,\ x \ge 0\}$,

$\mathrm{co}\,(\{(1,0),(0,1),(0,0),(0.1,0.2)\}) = \{(x_1,x_2) : x_1+x_2 \le 1, x \ge 0\}$,

$\mathrm{co}\,(\{(x_1,x_2)\colon x_1x_2 = 0,\ x_1^2+x_2^2 \le 1,\ x \ge 0\}) = \{(x_1,x_2)\colon x_1+x_2 \le 1,\ x \ge 0\}$,

$\mathrm{co}\,(\{(x_1,x_2)\colon x_2 = x_1^2\}) = \{(x_1,x_2)\colon x_2 \ge x_1^2\}$. $\diamond$

Proposición 2.9. ***Teorema de Caratheodory.*** *Todo elemento de* $\mathrm{co}(A)$ *se puede expresar como combinación convexa de a lo más* $n+1$ *puntos de* A *(recuérdese que* $A \subseteq \mathbb{R}^n$*). Es decir, si* $x \in co(A)$ *existen* x^1, x^2, ... , x^m *en* A *y escalares* λ_1, λ_2, ... , λ_m *tales que*

$$x = \lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_m x^m,\ \lambda_i \geq 0 \quad \forall i, \sum_{i=1}^{m} \lambda_i = 1, \quad m \leq n+1.$$

Sería erróneo pensar que toda envolvente convexa se puede expresar como la envolvente de a lo más $n+1$ puntos, puesto que la existencia de $n+1$ puntos está garantizada para un punto, pero a medida que éste varía también varían los $n+1$ puntos.

Proposición 2.10. *Un conjunto* C *es un poliedro si y solamente si se puede expresar como la envolvente convexa de un número finito de puntos.*

Definición 2.7. Sea C convexo, x en C. Se dice que x es **punto extremo** de C, si no es posible expresar x como combinación convexa estricta de dos puntos distintos de C, es decir:

$$\begin{array}{l} x = (1\text{-}\lambda)u + \lambda v \\ \quad u, v \in C \\ \quad \lambda \in]0,1[\end{array} \qquad \Longrightarrow \qquad u = v = x.$$

El conjunto de puntos extremos de C se denota $\mathrm{pe}(C)$.

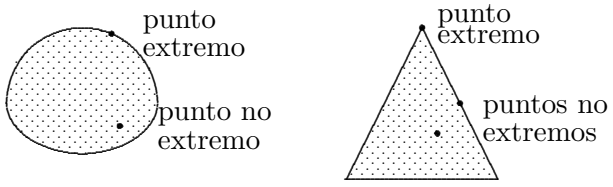


Figura 2.3

Ejemplo 2.16. Dado el conjunto convexo

$$\{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$$

$(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ es punto extremo. El punto $(0.2, 0.4)$ no es punto extremo. $\diamond$

Ejemplo 2.17. El punto $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ no es punto extremo de $\{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$. $\diamond$

Ejemplo 2.18. $(1,0)$, $(0,1)$, $(0,0)$ son puntos extremos de $\{(x_1,x_2)\colon x_1 + x_2 \leq 1, x \geq 0\}$. $\diamond$

Proposición 2.11. *Sean: C convexo, x en C. x es punto extremo de C si y solamente si al quitar x de C se tiene un convexo, es decir, si y solamente si*

$$C \setminus \{x\} = \{y \in C\colon y \neq x\} \quad \textit{es convexo.}$$

Proposición 2.12. *Todo conjunto convexo, cerrado y acotado se puede expresar como la envolvente convexa de sus puntos extremos, o sea*

$$C = \ co\big(pe(C)\big).$$

Ejemplo 2.19. $B[\mathbf{0},1] = \ \mathrm{co}\big(\{x\colon \|x\| = 1\}\big)$. $\diamond$

Ejemplo 2.20. $B(\mathbf{0},1) \neq \ \mathrm{co}\,(\emptyset)$. $\diamond$

Ejemplo 2.21. $\{x\colon x \geq 0\} \neq \ \mathrm{co}\big(\{(0,0)\}\big)$. $\diamond$

Proposición 2.13. *Un conjunto C es un poliedro si y solamente si se puede expresar como la envolvente convexa de sus puntos extremos y el número de puntos extremos es finito.*

$$C = \ co\big(pe(C)\big), \ \ \#\big(pe(C)\big) < \infty.$$

Definición 2.8. Sean: C un convexo, $d \in V$, $d \neq 0$. Se dice que d es una **dirección** de C si para todo $x \in C$ y para todo μ positivo , $x + \mu d$ también está en C. Dicho de otra forma, $d \neq 0$ es dirección de C si para todo $x \in C$ la semirrecta $S(x,d)$ es subconjunto de C.

Es claro que un conjunto convexo acotado no tiene direcciones y que un convexo con direcciones no es acotado.

Definición 2.9. Dos direcciones d^1, d^2 de un conjunto convexo C son **equivalentes** si una es múltiplo positivo de la otra, es decir, si existe $\mu > 0$ tal que $d^1 = \mu d^2$.

Definición 2.10. Una dirección d de un convexo C, se llama **dirección extrema** si no existen dos direcciones de C: d^1 y d^2, no equivalentes, tales que d sea combinación lineal positiva de d^1, d^2. Dicho de otra manera:

$$\begin{array}{c} d = \mu_1 d^1 + \mu_2 d^2 \\ d^1, d^2 \text{ direcciones de } C \\ \mu_1, \mu_2 > 0 \end{array} \qquad \Longrightarrow \qquad d, d^1, d^2 \text{ son equivalentes.}$$

Ejemplo 2.22. Consideremos el conjunto definido por las siguientes restricciones:

$$\begin{aligned} -x_1 + x_2 &\geq 2 \\ x_2 &\geq 3 \\ x &\geq 0 \end{aligned}$$

Este conjunto es convexo, ya que es intersección de cuatro semiplanos.

Los puntos extremos de este conjunto son: $(0, 3)$ y $(1, 3)$.

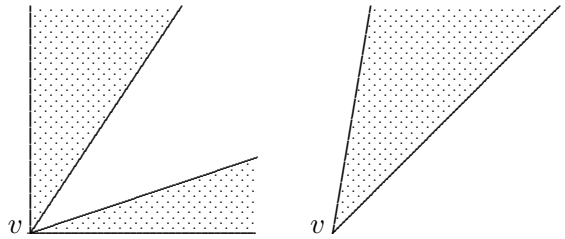
Para el mismo conjunto $(1, 4)$, $(0, 0.001)$, $(345, 345)$ son direcciones, y no lo son $(0, 0)$, $(-0.01, 100)$, $(-3, -4)$.

Son direcciones extremas $(0, 0.001)$, $(10, 10)$ o cualquier dirección equivalente a una de las dos. $\diamond$

Definición 2.11. Un conjunto K se llama **cono** con **vértice** en v , si para todo $x \in K$ y para todo $\mu > 0$ el punto $v + \mu(x - v)$ también está en K. Geométricamente, K es un cono con vértice en v si para todo x en K, la semirrecta que parte de v, sin incluirlo, y pasa por x está contenida en K.

Según la definición, el vértice v no está necesariamente en el cono. Si no se habla explícitamente del vértice se supone que éste es el origen.

Definición 2.12. Un conjunto K se llama **cono**, si para todo $x \in K$ y para todo $\mu > 0$ el punto μx también está en K.



cono no convexo cono convexo

Figura 2.4

Ejemplo 2.23. $\{(x_1, x_2) : x_1 x_2 = 0\}$, es decir, el conjunto formado por los puntos sobre los ejes, es un cono no convexo. $\diamond$

Ejemplo 2.24. $\{(x_1, x_2) : x \geq 0\}$, es decir, el primer cuadrante, es un cono convexo. $\diamond$

Ejemplo 2.25. El conjunto de direcciones del conjunto definido por las siguientes restricciones es un cono convexo.

$$\begin{aligned} -x_1 + x_2 &\geq 2 \\ x_2 &\geq 3 \\ x &\geq 0. \quad \diamond \end{aligned}$$

En general, sean C un convexo y D_C el conjunto de direcciones de C. Si $D_C \neq \emptyset$, entonces D_C es un cono convexo

Definición 2.13. Sean: A un conjunto, x un punto cualquiera. Se define la distancia del punto x al conjunto A, según la norma $\| \ \|$, al valor

$$d(x, A) = \underset{y \in A}{\rightarrow} \inf \|x - y\|$$

Obviamente si $x \in A$, entonces $d(x, A) = 0$.

Proposición 2.14. *Sean: A un conjunto cerrado, $x \notin A$. Entonces existe $\tilde{x} \in A$ tal que*

$$\|x - \tilde{x}\| = \min_{y \in A} \|x - y\| = d(x, A) > 0.$$

Como corolario se puede afirmar que para conjuntos cerrados, x está en A si y solamente si $d(x, A) = 0$. Cuando se utiliza la norma euclideana $\|\|_2$, en conjuntos convexos y cerrados, hay un único punto de A más cercano a x.

Proposición 2.15. *Sean: C un conjunto convexo cerrado, $x \notin C$. Entonces existe un* ***único*** *$x_C \in C$, el punto de C más cercano a x, tal que*

$$\|x - x_C\|_2 = \min_{y \in C} \|x - y\|_2 = d_2(x, C) > 0.$$

Si el punto x estuviera en C, el punto x_C sería el mismo x.

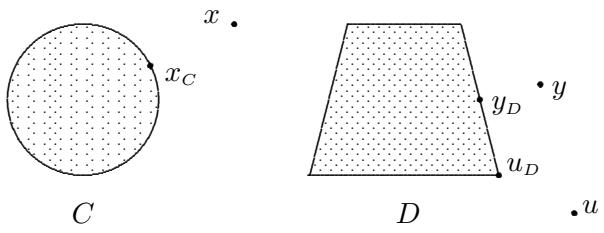


Figura 2.5

Ejemplo 2.26. Sea $A = B(\mathbf{0}, 1) = \{x : \|x\| < 1\}$. El punto $x = (-1, 0)$ no está en A, sin embargo $d(x, A) = 0$. $\diamond$

Ejemplo 2.27. Sea $A = \{x : x_1^2 + x_2^2 \leq 4, x_1^2 + x_2^2 \geq 1, x_2 \geq 0\}$ cerrado. El punto $x = (0, 0)$ no está en A, $d_2(x, A) = 1$, sin embargo, $d_2(x, (1, 0)) = 1$, $d_2(x, (0, 1)) = 1$ y los puntos $(1, 0)$, $(0, 1)$ están en A. $\diamond$

Ejemplo 2.28. Sean: $C = \{x : x_1 + x_2 \geq 2\}$, $x = (0, 0) \notin C$.

$$\begin{aligned} d_1(x, C) &= 2, \\ d_2(x, C) &= \sqrt{2}, \\ d_\infty(x, C) &= 1, \\ d_1(x, (1, 1)) &= 2, \\ d_1(x, (2, 0)) &= 2, \\ d_2(x, (1, 1)) &= \sqrt{2} \quad , \quad x_C = (1, 1). \ \diamond \end{aligned}$$

Ejemplo 2.29. Sean: $C = \{(x_1, x_2) : x_2 \geq 2\}$, $x = (0, 0) \notin C$.

$$\begin{aligned} d_2(x, C) &= 2, \\ d_\infty(x, C) &= 2, \\ d_\infty(x, (0, 2)) &= 2, \\ d_\infty(x, (1, 2)) &= 2, \\ d_2(x, (0, 2)) &= 2 \quad , \quad x_C = (0, 2). \ \diamond \end{aligned}$$

Definición 2.14. Sean A y B dos conjuntos. Un hiperplano H los **separa**si A está contenido en uno de sus semiespacios cerrados y B está contenido en el otro. Un hiperplano H los **separa estrictamente** si A está contenido en uno de sus semiespacios abiertos y B está contenido en el otro. Se dice que

A y B están **separados fuertemente** si existe $\varepsilon > 0$, $c \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$ tales que $c^{\mathrm{T}}x \leq \alpha$ para todo x en A y $c^{\mathrm{T}}x \geq \alpha+\varepsilon$ para todo x en B. Dicho de otra forma, un hiperplano H los separa fuertemente si los separa estrictamente y además la distancia de H a por lo menos uno de los conjuntos es positiva.

Proposición 2.16. *Sean: C un convexo cerrado, $x \notin C$. Entonces existe un hiperplano $H_{v\alpha} = H = \{y : v^T y = \alpha\}$ tal que contiene a x y uno de sus semiespacios abiertos contiene a C.*

En general, puede haber varios hiperplanos, pero uno que siempre sirve es el perpendicular al vector $x - x_C$ y que pasa por x, es decir:

$$\begin{array}{rrcll} & v & = & x - x_C, & \\ & \alpha & = & v^{\mathrm{T}}x, & \\ \text{entonces} & x & \in & H = \{y : v^{\mathrm{T}}y = \alpha\}, & \\ & C \subseteq \overset{\circ}{H^-} & = & \{y : v^{\mathrm{T}}y < \alpha\}, & \\ \text{o sea,} & v^{\mathrm{T}}y & < & \alpha \qquad \text{para todo } y \in C. & \end{array}$$

Si con el mismo vector normal v, se hace pasar el hiperplano por el punto $0.5x + 0.5x_C$ entonces C está en un semiespacio abierto y x en el otro, es decir, el hiperplano separa estrictamente al punto x y al conjunto C. Más aún, la separación es fuerte.

Ejemplo 2.30. Sean: $A = \{x : x_1^2 + x_2^2 \leq 4, x_1^2 + x_2^2 \geq 1, x_2 \geq 0\}$, $x = (0, 1/2) \notin A$. No existe ningún hiperplano que separe el punto x del conjunto A. $\diamond$

Ejemplo 2.31. Sean: $C = \{x : x_2 \geq x_1^2\}$, $x = (3, 0) \notin C$. El punto de C más cercano a x es $x_C = (1, 1)$. El hiperplano $H = \{x : 2x_1 - x_2 = 6\}$ pasa por x y uno de sus semiespacios abiertos contiene a C. Otro hiperplano con las mismas características puede ser $H = \{x : 2x_1 - 1.1x_2 = 6\}$. El hiperplano $H = \{x : 2x_1 - x_2 = 3.5\}$ separa estrictamente al punto x y al conjunto C. $\diamond$

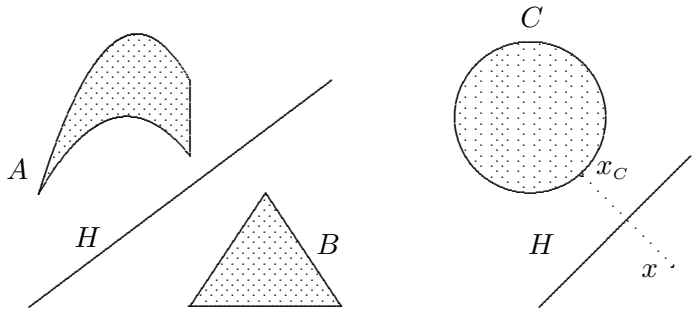


Figura 2.6

Ejemplo 2.32. Sean: $C = \{x : x_1 + x_2 \geq 4\}$, $x = (1, 1) \notin C$. El punto de C más cercano a x es $x_C = (2, 2)$. El hiperplano $H = \{x : x_1 + x_2 = 2\}$ pasa por x, uno de sus semiespacios abiertos contiene a C y es el único con estas características. $\diamond$

La última proposición se puede generalizar a dos conjuntos convexos, cerrados, uno de los cuales debe ser acotado.

Proposición 2.17. *Si C, D son dos conjuntos convexos, cerrados y por lo menos uno de ellos es acotado, entonces existe un hiperplano que los separa estrictamente.*

Ejemplo 2.33. Sean: $C = \{x : x_2 \geq x_1^2\}$, $D = \{x : (x_1 - 3)^2 + x_2^2 \leq 4\}$. El hiperplano $H = \{x : 2x_1 - x_2 = 1.5\}$ separa estrictamente los dos conjuntos. $\diamond$

Ejemplo 2.34. Sean: $C = \{x : x_2 \leq 0\}$, $D = \{x : x_1 x_2 \geq 1, x_1 \geq 0\}$. No existe ningún hiperplano que los separe estrictamente. $\diamond$

Definición 2.15. Sean: A un conjunto, H un hiperplano. H se llama **hiperplano de apoyo** de A, si H y la frontera de A (denotada ∂A) no son disyuntos, y además A está contenido en uno de los semiespacios cerrados de H.

Proposición 2.18. *Sea C un conjunto convexo. Si $x \in \partial C$, entonces existe un hiperplano de apoyo de C que pasa por x.*

Ejemplo 2.35. Sean: $A = \{x : x_1^2 + x_2^2 \leq 4, x_1^2 + x_2^2 \geq 1, x_2 \geq 0\}$, $x = (0, 1)$. $x \in \partial A$, sin embargo no existe un hiperplano de apoyo de A que pase por x. $\diamond$

Ejemplo 2.36. Sean: $C = B_\infty[\mathbf{0}, 1] = \{x : |x_1| \leq 1,\ |x_2| \leq 1\}$, $x = (1, 0)$, $y = (1, 1)$ -puntos de la frontera de C-. Existe un único hiperplano de apoyo de C que pase por x, a saber $H = \{(x_1, x_2) : x_1 = 1\} = \{(1, x_2)\}$. Este mismo hiperplano es hiperplano de apoyo de C en y. Sin embargo, en y hay muchos más hiperplanos de apoyo, por ejemplo, $\{(x_1, x_2) : x_1 + x_2 = 2\}$.$\diamond$

En matemáticas, varias definiciones se enuncian primero en un punto y luego se consideran en todo el conjunto, por ejemplo, la definición de continuidad se hace primero en un punto y después en todo un conjunto. La convexidad de un conjunto también se puede definir en un punto.

Definición 2.16. Sea C un conjunto, $\bar{x} \in C$. Se dice que C es convexo con respecto a $\bar{x}$, o también "estrellado" (forma de estrella) con respecto a $\bar{x}$, si para todo y en C, y para todo real λ en el intervalo $[0, 1]$, entonces $z = z_{\bar{x}y\lambda} = (1\text{-}\lambda)\bar{x} + \lambda y$ también está en C.

Ejemplo 2.37. El conjunto de puntos, en $\mathbb{R}^2$, que están sobre los ejes coordenados, no es convexo, pero sí es convexo con respecto al origen. $\diamond$

Ejemplo 2.38. Una estrella regular de cinco puntas no es un conjunto convexo, pero sí es un conjunto convexo con respecto a cada uno de los puntos del pentágono regular formado por los cinco vértices interiores.

Usando la definición de convexidad en un punto, se puede decir que un conjunto es convexo si es convexo con respecto a cada uno de sus puntos.

EJERCICIOS

2.1. Sea $A = \{(x_1, x_2) : x_1^2 \leq x_2\}$. Halle $\text{co}\,(A)$.

2.2. Sea $A = \{(x_1, x_2) : x_1^3 = x_2\}$. Halle $\text{co}\,(A)$.

2.3. Sea $A = \{(x_1, x_2) : x_1^3 \leq x_2\}$. Halle $\text{co}\,(A)$.

2.4. Sea $A = \{(x_1, x_2) : x_1 > 0,\ x_1|x_2| \geq 1\}$. ¿$A$ es cerrado? Halle $\text{co}\,(A)$. ¿$\text{co}\,(A)$ es cerrado? Dé condiciones suficientes para que si A es cerrado, entonces $\text{co}\,(A)$ sea cerrado.

2.5. Sea $C = \{(x_1, x_2) : |x_1| \leq x_2\}$. Halle los puntos extremos, las direcciones y las direcciones extremas de C.

2.6. Sea $C = \{(x_1, x_2) : x_1^2 \leq x_2\}$. Halle los puntos extremos, las direcciones y las direcciones extremas de C.

2.7. Sea $C = \{(x_1, x_2, x_3) : x_1^2 \leq x_2,\ x_1 + x_2 + x_3 \leq 1\}$. Halle los puntos extremos, las direcciones y las direcciones extremas de C.

2.8. Sean: C un convexo, A una matriz $m \times n$, $D = \{y : y = Ax, x \in C\}$. Muestre que D es un convexo.

2.9. Sean: A una matriz $m \times n$, $(m \leq n)$ de rango m, $D = \{d : Ad = 0, d \geq 0, d \neq 0\} \neq \emptyset$. Muestre que D es un cono convexo.

2.10. Sean x^1, ..., x^m elementos no nulos de $\mathbb{R}^n$. Una combinación lineal de estos elementos se llama no negativa si todos los escalares son no negativos, y se llama positiva si todos los escalares son positivos. Sea $\text{cnn}(x^1, ..., x^m)$ el conjunto de todas las combinaciones lineales no negativas y $\text{cp}(x^1, ..., x^m)$ el conjunto de todas las combinaciones lineales positivas. Muestre que los conjuntos $\text{cnn}(x^1, ..., x^m)$ y $\text{cp}(x^1, ..., x^m)$ son conos convexos.

Capítulo 3

MATRICES DEFINIDAS Y SEMIDEFINIDAS POSITIVAS

3.1 FACTORIZACIÓN DE CHOLESKY

Sea A una matriz simétrica. Bajo ciertas condiciones (como se verá posteriormente, si y solamente si es definida positiva), existe una matriz U triangular superior invertible tal que $U^{\mathrm{T}}U = A$. El cálculo se puede hacer por filas, es decir, primero se obtienen los elementos de la primera fila de U, en seguida los de la segunda, etc.

Supongamos conocidos los elementos de las filas $1, 2, ..., k$-1 de la matriz U. O sea, se conocen los elementos u_{ij} para $i = 1, 2, ..., k$-1 y para $j = i, i+1, ..., n$. Como U es triangular superior, se sabe también que $u_{ij} = 0$ para $i > j$. Al multiplicar la fila k de la matriz U^{T} por la columna k de la matriz U se tiene:

$$\begin{aligned} a_{kk} &= (U^{\mathrm{T}})_{k\cdot} U_{\cdot k} \\ &= (U_{\cdot k})^{\mathrm{T}} U_{\cdot k} \\ &= \sum_{i=1}^{n} u_{ik}^2 \\ &= \sum_{i=1}^{k} u_{ik}^2 + \sum_{i=k+1}^{n} u_{ik}^2 \end{aligned}$$

$$\begin{aligned} a_{kk} &= \sum_{i=1}^{k} u_{ik}^2 + \sum_{i>k} u_{ik}^2 \\ &= \sum_{i=1}^{k} u_{ik}^2 \\ &= \sum_{i=1}^{k\text{-}1} u_{ik}^2 + u_{kk}^2. \end{aligned}$$

De la última igualdad todo se conoce salvo u_{kk}, entonces

$$u_{kk} = \sqrt{a_{kk} - \sum_{i=1}^{k\text{-}1} u_{ik}^2}\ , \quad k = 1, ..., n. \tag{3.1}$$

Para que tenga sentido la raíz cuadrada se necesita que la cantidad bajo el radical sea no negativa. Como además, U es invertible (y su determinante es igual al producto de los elementos diagonales), entonces $u_{kk} \neq 0$. Luego para poder obtener U de manera adecuada se necesita que

$$a_{kk} - \sum_{i=1}^{k\text{-}1} u_{ik}^2 > 0\ , \quad k = 1, ..., n. \tag{3.2}$$

Al multiplicar la fila k de la matriz U^{T} por la columna j de la matriz U, con $k < j$, se tiene:

$$\begin{aligned} a_{kj} &= (U^{\mathrm{T}})_{k\cdot} U_{\cdot j} \\ &= (U_{\cdot k})^{\mathrm{T}} U_{\cdot j} \\ &= \textstyle\sum_{i=1}^{n} u_{ik} u_{ij} \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^{k} u_{ik} u_{ij} + \sum_{i=k+1}^{n} u_{ik} u_{ij} \\ &= \sum_{i=1}^{k} u_{ik} u_{ij} + \sum_{i>k} u_{ik} u_{ij} \end{aligned}$$

$$= \sum_{i=1}^{k} u_{ik}u_{ij}$$

$$a_{kj} = \sum_{i=1}^{k\text{-}1} u_{ik}u_{ij} + u_{kk}u_{kj}.$$

De la última igualdad todo se conoce salvo u_{kj}, entonces

$$u_{kj} = \big(a_{kj} - \sum_{i=1}^{k\text{-}1} u_{ik}u_{ij}\big)\frac{1}{u_{kk}}\ , \qquad k = 1, ..., n\text{-}1 \qquad (3.3)$$

$$j = k+1, ..., n.$$

En resumen, si siempre se cumple la condición (3.2), mediante las fórmulas (3.1) y (3.3) se puede obtener la matriz U.

Ejemplo 3.1.

$$A = \begin{bmatrix} 16 & \text{-}12 & 8 & \text{-}16 \\ \text{-}12 & 18 & \text{-}6 & 9 \\ 8 & \text{-}6 & 5 & \text{-}10 \\ \text{-}16 & 9 & \text{-}10 & 46 \end{bmatrix}$$

$$\begin{aligned} u_{11} &= \sqrt{16} = 4 \\ u_{12} &= \frac{\text{-}12}{4} = \text{-}3 \\ u_{13} &= \frac{8}{4} = 2 \\ u_{14} &= \frac{\text{-}16}{4} = \text{-}4 \end{aligned}$$

$$\begin{aligned} u_{22} &= \sqrt{18 - (\text{-}3)^2} = 3 \\ u_{23} &= \frac{\text{-}6 - (\text{-}3)(2)}{3} = 0 \\ u_{24} &= \frac{9 - (\text{-}3)(\text{-}4)}{3} = \text{-}1 \end{aligned}$$

$$u_{33} = \sqrt{5 - (2)^2 - (0)^2} = 1$$

$$u_{34} = \frac{-10-(2)(-4)-(0)(-1)}{1} = -2$$

$$u_{44} = \sqrt{46-(-4)^2-(-1)^2-(-2)^2} = 5,$$

entonces,

$$U = \begin{bmatrix} 4 & -3 & 2 & -4 \\ 0 & 3 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 5 \end{bmatrix} \quad \diamond$$

Ejemplo 3.2.

$$A = \begin{bmatrix} 16 & -12 & 8 \\ -12 & 7 & -6 \\ 8 & -6 & 5 \end{bmatrix}$$

$$\begin{aligned} u_{11} &= 4 \\ u_{12} &= -3 \\ u_{13} &= 2 \end{aligned}$$

$$u_{22} = \sqrt{7-(-3)^2} = \sqrt{-2},$$

luego no existe la factorización de Cholesky para esta matriz A.$\diamond$

Ejemplo 3.3.

$$A = \begin{bmatrix} 16 & -12 & 8 \\ -12 & 18 & -6 \\ 8 & -6 & 4 \end{bmatrix}$$

$$\begin{aligned} u_{11} &= 4 \\ u_{12} &= -3 \\ u_{13} &= 2 \\ u_{22} &= 3 \\ u_{23} &= 0 \\ u_{33} &= \sqrt{0} = 0, \end{aligned}$$

luego, aunque con

$$U = \begin{bmatrix} 4 & -3 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se tiene la igualdad $A = U^{\mathrm{T}}U$, no existe la factorización de Cholesky para esta matriz A puesto que U no es invertible. $\diamond$

3.2 ALGORITMO DE LA FACTORIZACIÓN DE CHOLESKY

La factorización de Cholesky es muy empleada en la solución de sistemas de ecuaciones lineales cuando la matriz de coeficientes es simétrica y definida positiva, lo cual es bastante frecuente. Por ejemplo, las matrices de rigidez de los problemas estructurales son definidas positivas. Dado el sistema $Ax = b$, si se conoce la factorización, se tiene entonces $U^{\mathrm{T}}Ux = b$. Si se toma $y = Ux$ el problema se convierte en $U^{\mathrm{T}}y = b$. Este sistema de ecuaciones lineales es muy fácil de resolver, puesto que es triangular inferior. Una vez obtenido y, se resuelve $Ux = y$, también fácil de efectuar puesto que es triangular superior, obteniéndose así x solución del sistema inicial. El método de Cholesky es más rápido que el método de Gauss.

Otra de las ventajas importantísimas del método de la factorización de Cholesky, cuando se hace utilizando el computador, es que la matriz U se puede almacenar exactamente donde estaba la matriz A, o sea, si el programa se elabora de forma adecuada, a medida que se obtiene un elemento u_{ij}, éste se puede almacenar justamente donde estaba el elemento a_{ij}. Obviamente al finalizar el proceso los valores de la matriz A se han perdido, y en su lugar está la matriz U. En la mayoría de los casos esto no es ningún problema, por el contrario, es una gran ventaja: únicamente se requiere espacio para almacenar una matriz y no para dos.

En el algoritmo presentado a continuación se hace uso de esta ventaja, así al comienzo los valores a_{ij} corresponden, en realidad a los elementos de A, pero al final estarán los valores de la matriz U. Si al final la variable `error` vale 0, se está indicando que se pudo efectuar la factorización de Cholesky, si vale 1 no hay factorización de Cholesky para esta matriz, o sea, A no es definida positiva. La variable de entrada ε indica la tolerancia o precisión en el siguiente sentido: los valores positivos menores que ε se pueden considerar como nulos.

datos: $A,\ n,\ \varepsilon$
resultados: A, `error`
`error` $= 0$
para $k = 1, ..., n$ **mientras** `error`$= 0$
    $s = a_{kk}$
    para $i = 1, ..., k-1$
        $s = s - a_{ik}^2$
    fin-para i
    si $s < \epsilon$ **ent**
        `error`$= 1$
        parar
    fin-ent
    sino
        $a_{kk} = \sqrt{s}$
        para $j = k+1, ..., n$
            $t = a_{kj}$
            para $i = 1, ..., k$-1
                $t = t - a_{ik}a_{ij}$
            fin-para i
            $a_{kj} = t/a_{kk}$
        fin-para j
    fin-sino
fin-para k

3.3 MATRICES DEFINIDAS POSITIVAS

Definición 3.1. Una matriz A real, simétrica (obviamente cuadrada) es **definida positiva** (o positivamente definida) si:

$$x^{\mathrm{T}}Ax > 0 \qquad \text{para todo } x \neq 0.$$

Ejemplo 3.4. La matriz identidad de orden n.

$$\begin{aligned} x^{\mathrm{T}}Ix &= x^{\mathrm{T}}x \\ &= \|x\|_2^2 \\ &\geq 0 \qquad \text{para todo } x \\ &> 0 \qquad \text{para todo } x \neq 0. \end{aligned}$$

Luego la matriz identidad es definida positiva. $\diamond$

Ejemplo 3.5. La matriz nula de orden n.

$$x^{\mathrm{T}}\mathbf{0}x = 0.$$

luego la matriz nula no es definida positiva. $\diamond$

Para una matriz simétrica A

$$x^{\mathrm{T}}Ax = \sum_{i=1}^{n} a_{ii}x_i^2 \ + 2\sum_{i=1}^{n\text{-}1}\sum_{j=i+1}^{n} a_{ij}x_ix_j.$$

Ejemplo 3.6.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$

$$\begin{array}{rcll} x^{\mathrm{T}}Ax & = & x_1^2 + 5x_2^2 + 4x_1x_2 & \\ & = & (x_1 + 2x_2)^2 + x_2^2 & \\ & \geq & 0 \qquad \text{para todo } x & \\ & = & 0 \qquad \text{sssi} \quad (x_1 + 2x_2)^2 = 0, \ \ x_2^2 = 0 & \\ & = & 0 \qquad \text{sssi} \quad x_1 + 2x_2 = 0, \ \ x_2 = 0 & \\ & = & 0 \qquad \text{sssi} \quad x_1 = 0, \ \ x_2 = 0 & \\ & = & 0 \qquad \text{sssi} \quad x = 0, & \end{array}$$

luego A es definida positiva. $\diamond$

Ejemplo 3.7.

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\begin{array}{rcl} x^{\mathrm{T}}Bx & = & x_1^2 + 4x_2^2 + 4x_1x_2 \\ & = & (x_1 + 2x_2)^2 \\ & \geq & 0 \quad \text{para todo } x, \\ & = & 0 \ , \ \text{por ejemplo para } x_1 = 2, x_2 = -1, \end{array}$$

luego B no es definida positiva. $\diamond$

Ejemplo 3.8.

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

$$\begin{aligned} x^{\mathrm{T}}Cx &= x_1^2 + 3x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 - x_2^2, \\ &= -1 \ , \text{ por ejemplo para } x_1 = 2, x_2 = -1, \end{aligned}$$

luego C no es definida positiva. $\diamond$

Definición 3.2. Una **submatriz principal** de A es la obtenida al quitar de A algunas (o ninguna) filas y exactamente esas columnas correspondientes. Una **submatriz estrictamente principal** A_k es la obtenida al quitar de A las filas y columnas $k+1$, $k+2$, ..., n con $1 \leq k \leq n$, es decir, las n submatrices estrictamente principales de A son:

$$A_1 = \begin{bmatrix} a_{11} \end{bmatrix} \ , \quad A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \ldots, \quad A_n = A.$$

Definición 3.3. El determinante de una submatriz principal se llama un **subdeterminante principal**; el determinante de una submatriz estrictamente principal se llama un **subdeterminante estrictamente principal**. Los n subteterminantes estrictamente principales son:

$$\delta_1 = \det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11}, \quad \delta_2 = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \ldots, \quad \delta_n = \det(A).$$

En ejemplos pequeños, es relativamente fácil aplicar directamente la definición para saber si una matriz simétrica es definida positiva. Para ejemplos grandes, no solo se vuelve difícil, sino casi imposible. La siguiente proposición da condiciones necesarias y suficientes para la caracterización de matrices definidas positivas.

Proposición 3.1. (Condiciones necesarias y suficientes.) *Dada una matriz simétrica las siguientes siete afirmaciones son equivalentes:*

(1) A es definida positiva;

(2) todos los λ_i, valores propios de A (reales por ser simétrica), son positivos;

(3) todos los δ_i, subdeterminantes estrictamente principales, son positivos;

(4) todos los pivotes a_{kk}^k, en el método de eliminación de Gauss sin permutación, son positivos;

(5) existe una matriz U triangular superior invertible tal que $A = U^T U$ (ésta es la factorización de Cholesky);

(6) existe una matriz W invertible tal que $A = W^T W$;

(7) todos los subdeterminantes principales son positivos.

Observación: la factorización de Cholesky, en general, no es única, pero hay únicamente una matriz U con los elementos diagonales positivos.

Para matrices grandes, en casos no triviales, la caracterización más usada es la factorización de Cholesky. Obviamente, si se tiene una matriz U triangular superior e invertible, se tiene una matriz W invertible. Tratar de averiguar si una matriz es definida positiva mediante la sexta caracterización casi nunca se utiliza, salvo en el caso, en que por anticipado, se sabe que la matriz $A = W^{\mathrm{T}} W$, con W invertible.

Ejemplo 3.9. Los valores propios de la matriz identidad son $1, 1, ..., 1$; $\delta_i = 1$ para todo i ; todos los pivotes, en la eliminación gaussiana, tienen el valor 1 ; $I = I^{\mathrm{T}} I$. Como ejemplo de la sexta caracterización:

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

Por cualquiera de los criterios se ve que la matriz identidad es definida positiva. $\diamond$

Ejemplo 3.10. Los valores propios de la matriz nula son 0, 0, ..., 0 ; $\delta_i = 0$ para todo i ; la eliminación gaussiana no se puede efectuar; como la matriz nula no es invertible no se puede expresar como producto de matrices invertibles. Por cualquiera de los criterios se ve que la matriz nula no es definida positiva. $\diamond$

Ejemplo 3.11.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad \det(A - \lambda I) = (1-\lambda)(5-\lambda) - 4 = \lambda^2 - 6\lambda + 1,$$

sus raíces (los valores propios) son: $\lambda_1 = 3 + \sqrt{8}, \quad \lambda_2 = 3 - \sqrt{8}$; $\delta_1 = 1$, $\delta_2 = 1$; en la eliminación de Gauss sin permutaciones

$$A^{(1)} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix},$$

luego $a_{11}^{(1)} = 1$, $a_{22}^{(2)} = 1$.

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

Luego, por cualquiera de los criterios, la matriz A es definida positiva. $\diamond$

Ejemplo 3.12.

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad \det(B - \lambda I) = (1-\lambda)(4-\lambda) - 4 = \lambda^2 - 5\lambda,$$

sus raíces (los valores propios) son: $\lambda_1 = 5$, $\lambda_2 = 0$; $\delta_1 = 1$, $\delta_2 = 0$; en la eliminación de Gauss sin permutaciones

$$B^{(1)} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad B^{(2)} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix},$$

luego $b_{11}^{(1)} = 1$, $b_{22}^{(2)} = 0$. Al tratar de encontrar la factorización de Cholesky:

$$B = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}.$$

Como B no es invertible no se puede encontrar W invertible, tal que $B = W^{\mathrm{T}}W$. Luego, por cualquiera de los criterios, la matriz B no es definida positiva. $\diamond$

Ejemplo 3.13.

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad \det(C - \lambda I) = (1-\lambda)(3-\lambda) - 4 = \lambda^2 - 4\lambda - 1,$$

sus raíces (los valores propios) son: $\lambda_1 = 2 + \sqrt{5}$, $\lambda_2 = 2 - \sqrt{5}$; $\delta_1 = 1$, $\delta_2 = -1$; en la eliminación de Gauss sin permutaciones

$$C^{(1)} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad C^{(2)} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix},$$

luego $c_{11}^{(1)} = 1$, $c_{22}^{(2)} = -1$. Al tratar de encontrar la factorización de Cholesky ésta no se puede realizar, ya que aparece la raíz de un número negativo. Si $C = W^{\mathrm{T}}W$ entonces $\det(C) = \det(W^{\mathrm{T}})\det(W) = \big(\det(W)\big)^2$, pero como $\det(C) = -1$, entonces no existe tal matriz W. Luego, por cualquiera de los criterios, la matriz C no es definida positiva. $\diamond$

Proposición 3.2. (Condiciones necesarias.) *Si A es una matriz simétrica definida positiva entonces:*

(1) $a_{ii} > 0$ *para todo* i *;*

(2) $a_{ij}^2 < a_{ii}a_{jj}$ *para todo* $i \neq j$ *;*

(3) $\max\limits_{i} a_{ii} = \max\limits_{i,j} |a_{ij}|$*, es decir,* $\max\limits_{i,j} |a_{ij}|$ *es un elemento diagonal;*

(4) $2|a_{ij}| < a_{ii} + a_{jj}$ *para todo* $i \neq j$*.*

Ejemplo 3.14. Sean:

$$A = \begin{bmatrix} 5 & 6 \\ 6 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 \\ 4 & 7 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 3 \\ 3 & 3 \end{bmatrix},$$

$$C = \begin{bmatrix} 10 & -3 & 5 \\ -3 & 1 & -1 \\ 5 & -1 & 3 \end{bmatrix}.$$

Para cada criterio se presenta un ejemplo de una matriz que no lo cumple, luego no es definida positiva y un segundo ejemplo de una matriz que cumple el criterio y, sin embargo, no es definida positiva.

(1) A no cumple el criterio, luego no es MDP.

B cumple el criterio y, sin embargo no es MDP.

(2) B no cumple el criterio, luego no es MDP.

C cumple el criterio y, sin embargo, no es MDP.

(3) A no cumple el criterio, luego no es MDP.

B cumple el criterio y, sin embargo, no es MDP.

(4) D no cumple el criterio, luego no es MDP.

B cumple el criterio y, sin embargo, no es MDP. $\diamond$

Definición 3.4. Una matriz es de **diagonal estrictamente dominante por filas** si:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| < |a_{ii}| \ , \quad \text{para todo } i.$$

Una matriz es de **diagonal estrictamente dominante por columnas** si:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| < |a_{jj}| \ , \quad \text{para todo } j.$$

Una matriz es de **diagonal estrictamente dominante** si es de diagonal estrictamente dominante por columnas y de diagonal estrictamente dominante por filas.

Es obvio que para matrices simétricas las tres definiciones anteriores coinciden.

Proposición 3.3. (Condiciones suficientes.) *Si A es una matriz simétrica, de diagonal estrictamente dominante y positiva, entonces es definida positiva.*

Ejemplo 3.15.

$$A = \begin{bmatrix} 4 & 1 & -2 \\ 1 & 5 & 3 \\ -2 & 3 & 6 \end{bmatrix}$$

cumple el criterio, luego es definida positiva. $\diamond$

Ejemplo 3.16.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$

no cumple el criterio y, sin embargo, es definida positiva. $\diamond$

3.4 MATRICES SEMIDEFINIDAS POSITIVAS

Definición 3.5. Una matriz A real, simétrica (obviamente cuadrada) es **semidefinida positiva** o definida no negativa si:

$$x^{\mathrm{T}}Ax \geq 0 \qquad \text{para todo } x.$$

Es evidente que toda matriz definida positiva es semidefinida positiva.

Algunas veces, no en estas notas, se agrega a la definición de matrices semidefinidas positivas la existencia de un x no nulo tal que $x^{\mathrm{T}}Ax = 0$. Según esta otra definición una matriz definida positiva no sería semidefinida positiva.

Ejemplo 3.17. Los ejemplos anteriores de matrices definidas positivas son matrices semidefinidas positivas. $\diamond$

Ejemplo 3.18. La matriz nula $\mathbf{0}$. Aplicando la definición $x^{\mathrm{T}}\mathbf{0}x = 0$, luego la matriz nula es semidefinida positiva. $\diamond$

Ejemplo 3.19.

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix},$$

$$\begin{aligned} x^{\mathrm{T}}Bx &= x_1^2 + 4x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 \\ &\geq 0 \qquad \text{para todo } x, \end{aligned}$$

luego B es semidefinida positiva. $\diamond$

Ejemplo 3.20.

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix},$$

$$\begin{aligned} x^{\mathrm{T}}Cx &= x_1^2 + 3x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 - x_2^2 \\ &= -1 \qquad \text{por ejemplo para} \quad x_1 = 2,\ x_2 = -1, \end{aligned}$$

luego C no es semidefinida positiva. $\diamond$

Ejemplo 3.21.

$$E = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\begin{aligned} x^{\mathrm{T}}Ex &= -x_2^2 \\ &= -1 \qquad \text{por ejemplo para} \quad x_1 = 0, x_2 = 1, \end{aligned}$$

luego E no es semidefinida positiva. $\diamond$

Proposición 3.4. *Dada una matriz A simétrica, las siguientes siete afirmaciones son equivalentes:*

(1) A es semidefinida positiva;

(2) todos los λ_i valores propios de A son no negativos;

(3) $\delta_i \geq 0$ para todo i y para cualquier reordenamiento simétrico de filas y de columnas, es decir, para cualquier permutación de las filas y de las columnas correspondientes;

(4) los pivotes a_{kk}^k, en el método de eliminación de Gauss posiblemente con permutaciones simétricas, son todos no negativos;

(5) existe una matriz U triangular superior, posiblemente no invertible, tal que $A = U^T U$;

(6) existe una matriz W, posiblemente no invertible, tal que $A = W^T W$;

(7) todos los subdeterminantes principales son no negativos.

Observación: El séptimo criterio es un caso particular del tercer criterio pero más fácil de aplicar.

Es claro que las caracterizaciones para las matrices definidas positivas son más fuertes que las caracterizaciones para las matrices semidefinidas positivas. En los ejemplos de matrices definidas positivas se observa claramente que éstas cumplen con las caracterizaciones de matrices semidefinidas positivas.

Ejemplo 3.22. Los valores propios de la matriz nula son 0, 0, ..., 0 ; $\delta_i = 0$ para todo i y para cualquier reordenamiento simétrico de filas y columnas; la eliminación de Gauss (con posibilidad de permutación) conduce a pivotes nulos; $\mathbf{0} = \mathbf{0}^{\mathrm{T}}\mathbf{0}$; todos los subdeterminantes principales son nulos. Por cualquiera de los criterios se ve que la matriz nula es semidefinida positiva. $\diamond$

Ejemplo 3.23.

$$B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad \det(B - \lambda I) = (1-\lambda)(4-\lambda) - 4 = \lambda^2 - 5\lambda,$$

sus raíces (los valores propios) son: $\lambda_1 = 5$, $\lambda_2 = 0$; sin reordenamientos $\delta_1 = 1$, $\delta_2 = 0$; únicamente hay un solo reordenamiento de B:

$$\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

$\delta_1 = 4$, $\delta_2 = 0$; en la eliminación de Gauss

$$B^{(1)} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad B^{(2)} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix},$$

luego $b_{11}^{(1)} = 1$, $b_{22}^{(2)} = 0$. Al tratar de encontrar la factorización de Cholesky:

$$B = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix};$$

los subdeterminantes principales son 0 (sin quitar filas ni columnas), 1 (quitando la segunda fila y la segunda columna) y 4 (quitando la primera fila y la primera columna). Luego, por cualquiera de los criterios, la matriz B es semidefinida positiva. $\diamond$

Ejemplo 3.24.

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \det(C - \lambda I) = (1-\lambda)(3-\lambda) - 4 = \lambda^2 - 4\lambda - 1,$$

sus raíces (los valores propios) son: $\lambda_1 = 2 + \sqrt{5}$, $\lambda_2 = 2 - \sqrt{5}$; sin reordenamientos $\delta_1 = 1$, $\delta_2 = -1$; en la eliminación de Gauss

$$C^{(1)} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad C^{(2)} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix},$$

luego $c_{11}^{(1)} = 1$, $c_{22}^{(2)} = -1$; al tratar de encontrar la factorización de Cholesky ésta no se puede realizar, ya que aparece la raíz de un número negativo. Si $C = W^{\mathrm{T}}W$ entonces $\det(C) = \det(W^{\mathrm{T}})\det(W) = \big(\det(W)\big)^2$, pero como $\det(C) = -1$, entonces no existe tal matriz W. Los subdeterminantes principales son -1 (sin quitar filas ni columnas), 1 (quitando la segunda fila y la segunda columna) y 3 (quitando la primera fila y la primera columna). Luego, por cualquiera de los criterios, la matriz C no es semidefinida positiva. $\diamond$

Ejemplo 3.25.

$$E = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \det(E - \lambda I) = (0-\lambda)(-1-\lambda) = \lambda^2 + \lambda,$$

sus raíces (los valores propios) son: $\lambda_1 = 0$, $\lambda_2 = -1$; sin reordenamientos $\delta_1 = 0$, $\delta_2 = 0$; el único reordenamiento simétrico (de filas y de columnas) da como resultado:

$$\begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix},$$

$\delta_1 = -1$, $\delta_2 = 0$; al tratar de encontrar la factorización de Cholesky ésta no se puede realizar, ya que aparece la raíz de un número negativo. Tampoco se puede encontrar W tal que $E = W^{\mathrm{T}}W$. Los subdeterminantes principales son 0 (sin quitar filas ni columnas), 0 (quitando la segunda fila y la segunda columna) y -1 (quitando la primera fila y la primera columna). Luego la matriz E no es semidefinida positiva. $\diamond$

Proposición 3.5. (Condiciones necesarias.) *Si A es una matriz simétrica semidefinida positiva entonces:*

(1) $a_{ii} \geq 0$ *para todo* i *;*

(2) $a_{ij}^2 \leq a_{ii}a_{jj}$ *para todo* i, j *;*

(3) $\max a_{ii} = \max |a_{ij}|$*, es decir,* $\max |a_{ij}|$ *es un elemento diagonal;*

(4) $2|a_{ij}| \leq a_{ii} + a_{jj}$ *para todo* i, j *;*

(5) $a_{ii} = 0 \Rightarrow A_{i\cdot} = 0, A_{\cdot i} = 0$ *;*

(6) $\delta_i \geq 0$ *para todo* i*.*

Ejemplo 3.26.

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 8 & 5 & 8 \\ 5 & 4 & 6 \\ 8 & 6 & 9 \end{bmatrix}.$$

Para cada criterio se presenta un ejemplo de una matriz que no lo cumple, luego no es semidefinida positiva y un segundo ejemplo de una matriz que cumple el criterio y, sin embargo, no es semidefinida positiva.

(1) A no cumple el criterio, luego no es MSDP.
C cumple el criterio y, sin embargo, no es MSDP.

(2) B no cumple el criterio, luego no es MSDP.
E cumple el criterio y, sin embargo, no es MSDP.

(3) C no cumple el criterio, luego no es MSDP.
E cumple el criterio y, sin embargo, no es MSDP.

(4) C no cumple el criterio, luego no es MSDP.
E cumple el criterio y, sin embargo, no es MSDP.

(5) D no cumple el criterio, luego no es MSDP.
A cumple el criterio y, sin embargo, no es MSDP.

(6) B no cumple el criterio, luego no es MSDP.
A cumple el criterio y, sin embargo, no es MSDP. $\diamond$

Definición 3.6. Una matriz A simétrica es **definida negativa** si $-A$ es definida positiva. Una matriz A simétrica es **semidefinida negativa** si $-A$ es semidefinida positiva. Una matriz A simétrica es **indefinida** si no es ni semidefinida positiva ni semidefinida negativa, o lo que es lo mismo, si existen $x,\ y$ tales que $(x^{\mathrm{T}}Ax)(y^{\mathrm{T}}Ay) < 0$, o también, si tiene por lo menos un valor propio positivo y por lo menos un valor propio negativo.

Ejemplo 3.27.

$$\begin{bmatrix} -1 & -2 \\ -2 & -5 \end{bmatrix} \qquad \text{es definida negativa.} \ \diamond$$

Ejemplo 3.28.

$$\begin{aligned} A &= \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \qquad \text{no es semidefinida positiva,} \\ -A &= \begin{bmatrix} \text{-1} & \text{-2} \\ \text{-2} & \text{-3} \end{bmatrix} \qquad \text{tampoco es semidefinida positiva,} \end{aligned}$$

luego A es indefinida. Sean: $x = (2, -1),\ \ y = (1, 0)$. Entonces $(x^{\mathrm{T}}Ax)$ $(y^{\mathrm{T}}Ay) = (-1)(1) = -1$, lo que comprueba que A es indefinida. $\diamond$

3.5 MATRICES DEFINIDAS POSITIVAS EN UN SUBESPACIO VECTORIAL

Es posible que una matriz no sea definida positiva pero cumpla la propiedad $x^{\mathrm{T}}Ax > 0$ para todos los vectores x en algún conjunto especial. Para subespcios vectoriales se tiene la siguiente definición.

Definición 3.7. Sea $\mathcal{E}$ un subespacio vectorial de $\mathbb{R}^n$. Una matriz simétrica A es definida positiva en $\mathcal{E}$ si

$$x^{\mathrm{T}}Ax > 0, \ \ \forall x \in \mathcal{E}, \ \ x \neq 0.$$

A es semidefinida positiva en $\mathcal{E}$ si

$$x^{\mathrm{T}}Ax \geq 0 \ , \ \ \forall x \in \mathcal{E}.$$

Obviamente si una matriz es definida positiva es definida positiva en cualquier subespacio. Si $\mathcal{E} = \mathbb{R}^n$, se tienen las definiciones usuales de matriz definida positiva y matriz semidefinida positiva. Si $\mathcal{E} = \{\mathbf{0}\}$ cualquier matriz simétrica es semidefinida positiva y definida positiva en $\mathcal{E}$.

La definición de matriz definida positiva en un subespacio vectorial es útil sobre todo cuando la matriz no es definida positiva, y sí es definida positiva en algún subespacio vectorial de $\mathbb{R}^n$.

Ejemplo 3.29. La matriz

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

no es definida positiva ni semidefinida positiva. Tampoco es semidefinida negativa. Sus valores propios son $1, -1$. Sea $x \in \mathcal{E} = \{(x_1, x_2) : 3x_1 - x_2 = 0\}$.

$$\begin{aligned} x^{\mathrm{T}}Ax &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1x_2 \\ &= 2x_1(3x_1) \\ &= 6x_1^2 \\ &> 0 \quad \text{si } x_1 \neq 0. \end{aligned}$$

Luego A es definida positiva en $\mathcal{E}$. $\diamond$

Saber si una matriz es definida positiva en un subespacio vectorial aplicando directamente la definición puede ser fácil para matrices pequeñas, sin embargo, es muy útil tener procedimientos precisos para matrices de cualquier tamaño.

Sean: $\mathcal{E}$ un subespacio vectorial de $\mathbb{R}^n$, v^1, v^2, ..., v^k una base de $\mathcal{E}$, E la matriz $n \times k$ cuyas columnas son los vectores v^1, v^2, ..., v^k. Los elementos de la forma $E\xi$ caracterizan completamente a $\mathcal{E}$, es decir, x es un elemento de $\mathcal{E}$ si y solamente si existe $\xi \in \mathbb{R}^k$ tal que $x = E\xi$. Una matriz simétrica A es definida positiva en $\mathcal{E}$ si $x^{\mathrm{T}}Ax > 0$ para todo $x \in \mathcal{E}, x \neq 0$, o sea, si $(E\xi)^{\mathrm{T}}AE\xi = \xi^{\mathrm{T}}E^{\mathrm{T}}AE\xi > 0$ para todo $\xi \in \mathbb{R}^k, \xi \neq 0$.

Proposición 3.6. *Sea A una matriz simétrica. A es definida positiva en $\mathcal{E}$ si y solamente si E^TAE es definida positiva. A es semidefinida positiva en $\mathcal{E}$ si y solamente si E^TAE es semidefinida positiva.*

En esta proposición el criterio no depende de la base escogida. Para saber si la matriz $E^{\mathrm{T}}AE$, de tamaño $k \times k$, es definida positiva se utiliza cualquiera de los criterios vistos anteriormente.

Ejemplo 3.30. Sean:

$$\begin{aligned} A &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ \mathcal{E} &= \{(x_1, x_2) : 3x_1 - x_2 = 0\}. \end{aligned}$$

La dimensión de $\mathcal{E}$ es 1. Una base de $\mathcal{E}$ es $v^1 = [1 \;\; 3]^{\mathrm{T}}$, entonces

$$E = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

$$E^{\mathrm{T}}AE = [6] = 6 > 0,$$

luego A es definida positiva en $\mathcal{E}$. $\diamond$

Ejemplo 3.31. Sean:

$$\begin{aligned} A &= \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}, \\ \mathcal{E} &= \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 0\}. \end{aligned}$$

La dimensión de $\mathcal{E}$ es 2. Una base de $\mathcal{E}$ es $v^1 = [1 \;\; 0 \;\; \text{-}1]^{\mathrm{T}}$, $v^2 = [0 \;\; 1 \;\; \text{-}1]^{\mathrm{T}}$, entonces

$$\begin{aligned} E &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \text{-}1 & \text{-}1 \end{bmatrix}, \\ E^{\mathrm{T}}AE &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Luego A es semidefinida positiva en $\mathcal{E}$. Si se escoge otra base la conclusión será la misma. Otra base de $\mathcal{E}$ es $v^1 = [1 \;\; 0 \;\; \text{-}1]^{\mathrm{T}}$, $v^2 = [1 \;\; \text{-}2 \;\; 1]^{\mathrm{T}}$, entonces

$$\begin{aligned} E &= \begin{bmatrix} 1 & 1 \\ 0 & \text{-}2 \\ \text{-}1 & 1 \end{bmatrix}, \\ E^{\mathrm{T}}AE &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \end{aligned}$$

luego A es semidefinida positiva en $\mathcal{E}$. $\diamond$

3.5.1 En el espacio nulo de una matriz

Con frecuencia un subespacio vectorial está definido como el espacio nulo de una matriz. En este caso hay un procedimiento explícito para encontrar una base.

Definición 3.8. Sea M una matriz $p \times n$. El espacion nulo de M es el conjunto

$$\mathcal{N}(M) = \{x \in \mathbb{R}^n : Mx = 0\}.$$

Este conjunto es un subespacio vectorial de $\mathbb{R}^n$. Si M no es la matriz nula existe una matriz $\bar{M}$ cuyas filas son linealmente independientes y además

$$\mathcal{N}(M) = \mathcal{N}(\bar{M}).$$

En lo que sigue se supone que **las filas de M son linealmente independientes**. Esto implica que $1 \leq p \leq n$ y que el rango de M es p. Si $p = n$, entonces $\mathcal{N}(M) = \{\mathbf{0}\}$. Como generalización se puede decir que si M no tiene filas ($p = 0$), entonces $\mathcal{N}(M) = \mathbb{R}^n$.

El problema que estudiaremos a continuación es el siguiente. Dada A matriz simétrica $n \times n$ y M matriz $p \times n$, rango$(M) = p$, se desea saber si A es definida positiva (o semidefinida positiva) en $\mathcal{N}(M)$.

Si $p = n$, entonces A es definida positiva y semidefinida positiva en $\mathcal{N}(M) = \{\mathbf{0}\}$. Para la generalización $p = 0$, se trata simplemente de averiguar si A es definida positiva en $\mathbb{R}^n$. En consecuencia podemos suponer que $1 \leq p < n$. El espacio $\mathcal{N}(M)$ tiene dimensión $q = n - p$, o sea, cualquier base de $\mathcal{N}(M)$ tiene q elementos.

Para poder aplicar la proposición 3.6 se requiere conocer una matriz E cuyas columnas son los vectores de una base de $\mathcal{N}(M)$. Como rango$(M) = p$, entonces existe B de tamaño $p \times p$, submatriz de M, invertible. Entonces el sistema

$$Mx = 0,$$

es equivalente a

$$B^{-1}Mx = M'x = 0.$$

La matriz M' tiene una característica importante, tomando adecuadamente p columnas se obtiene la matriz identidad de orden p. La matriz M' se puede obtener por multiplicación explícita o también llevando M a la forma escalonada reducida por filas.

Sean: L la submatriz $p \times q$ formada por las demás columnas de M', x_B el vector columna $p \times 1$ construido con las componentes del vector x correspondientes a las columnas de M' que forman la matriz identidad, x_L el vector columna $q \times 1$ construido con las demás componentes del vector x. Las variables x_B se llaman variables básicas. Las variables x_L se llaman variables libres.

El sistema $M'x = 0$ es equivalente a

$$\begin{bmatrix} I_p & L \end{bmatrix} \begin{bmatrix} x_B \\ x_L \end{bmatrix} = 0,$$
$$\text{luego} \qquad x_B = -Lx_L.$$

Una manera de obtener una base de $\mathcal{N}(M)$ es mediante el siguiente procedimiento: dar a la primera variable libre (componente de x_L) el valor 1 y a las demás variables libres el valor 0. De esa manera $x_B = -L_{\cdot 1}$ y se

obtiene el primer elemento de la base. Enseguida, se le da a la segunda variable libre el valor 1 y a las demás variables libres el valor 0. De esa manera $x_B = -L_{\cdot 2}$ y se obtiene el segundo elemento de la base. Y así sucesivamente hasta obtener los q elementos de la base. Entonces la matriz E tiene la forma

$$E = \begin{bmatrix} -L \\ I_q \end{bmatrix},$$

o la matriz obtenida por permutación de filas para concordar con el orden de las variables. De manera más precisa: si las variables básicas son las p primeras (y las variables libres las q últimas), entonces no se requiere hacer ninguna permutación de filas, en caso contrario se reordenan las filas.

Dicho en palabras, en las p filas de E correspondientes a las variables básicas se colocan las p filas de $-L$, y en las q filas de E correspondientes a las variables libres se colocan las q filas de I_q.

Ejemplo 3.32. Encontrar una matriz E para $\mathcal{N}(M)$ subespacio nulo de la matriz $M = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$.

La matriz M ya está en la forma escalonada reducida. Consideremos, por ejemplo, que la primera columna de M constituye I y que las dos últimas forman L

$$\begin{aligned} I &= \begin{bmatrix} 1 \end{bmatrix} \\ L &= \begin{bmatrix} 1 & 1 \end{bmatrix} \\ x_B &= \begin{bmatrix} x_1 \end{bmatrix} \\ x_L &= \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}, \end{aligned}$$

entonces

$$E = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad \diamond$$

Ejemplo 3.33. Encontrar una matriz E para $\mathcal{N}(M)$ subespacio nulo de la matriz

$$M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}.$$

Llevándola a la forma escalonada reducida resulta

$$\begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

Las dos primeras columnas de M constituyen I y las dos últimas forman L

$$\begin{aligned} L &= \begin{bmatrix} -1 & -2 \\ 2 & 3 \end{bmatrix} \\ x_B &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ x_L &= \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}, \end{aligned}$$

entonces

$$E = \begin{bmatrix} 1 & 2 \\ -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad \diamond$$

Ejemplo 3.34. Averiguar si A es semidefinida positiva en $\mathcal{N}(M)$ subespacio nulo de la matriz M,

$$\begin{aligned} A &= \begin{bmatrix} -2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ M &= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 6 & 7 & 8 \\ 3 & 6 & 9 & 10 & 12 \end{bmatrix}. \end{aligned}$$

Llevando M a la forma escalonada reducida se obtiene

$$\begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Las columnas 1, 4, 5 constituyen I, la segunda y tercera forman L,

$$
\begin{aligned}
L &= \begin{bmatrix} 2 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \\
x_B &= \begin{bmatrix} x_1 \\ x_4 \\ x_5 \end{bmatrix} \\
x_L &= \begin{bmatrix} x_2 \\ x_3 \end{bmatrix},
\end{aligned}
$$

entonces

$$
\begin{bmatrix} -L \\ I_q \end{bmatrix} = \begin{bmatrix} -2 & -3 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} x_1 \\ x_4 \\ x_5 \\ x_2 \\ x_3 \end{matrix}
$$

En este caso es indispensable reordenar las filas para obtener la matriz

$$
E = \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.
$$

Aplicando la proposición 3.6:

$$
E^{\mathrm{T}}AE = \begin{bmatrix} -11 & -16 \\ -16 & -23 \end{bmatrix}.
$$

Los valores propios de $E^{\mathrm{T}}AE$ son 0.0880, -34.0880, luego A no es semidefinida positiva en $\mathcal{N}(M)$. $\diamond$

EJERCICIOS

3.1 - 3.12 Diga si las siguientes matrices son definidas positivas, semidefinidas positivas, definidas negativas, semidefinidas negativas, o indefinidas.

$$A^1 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 6 & 6 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}, \quad A^3 = \begin{bmatrix} -4 & 5 \\ 5 & -7 \end{bmatrix},$$

$$A^4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix}, \quad A^5 = \begin{bmatrix} 9 & -18 & -15 \\ -18 & 40 & 22 \\ -15 & 22 & 42 \end{bmatrix}, \quad A^6 = \begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix},$$

$$A^7 = \begin{bmatrix} 10 & 1 & -1 & 2 \\ 1 & 11 & -2 & 3 \\ -1 & -2 & 12 & -3 \\ 2 & 3 & -3 & 13 \end{bmatrix}, \quad A^8 = \begin{bmatrix} 10 & 1 & 1 & 2 \\ 1 & 20 & -2 & 3 \\ 1 & -2 & 30 & 25 \\ 2 & 3 & 25 & 20 \end{bmatrix},$$

$$A^9 = \begin{bmatrix} 10 & 1 & -1 & 2 \\ 1 & 11 & -2 & 3 \\ -1 & -2 & 12 & -14 \\ 2 & 3 & -14 & 13 \end{bmatrix}, \quad A^{10} = \begin{bmatrix} 10 & 1 & 1 & 2 \\ 1 & 20 & -2 & 3 \\ 1 & -2 & 30 & 2 \\ 2 & 3 & 2 & -20 \end{bmatrix},$$

$$A^{11} = \begin{bmatrix} 10 & 1 & 1 & 2 \\ 1 & 11 & 2 & 3 \\ 1 & 2 & 12 & 13 \\ 2 & 3 & 13 & 11 \end{bmatrix}, \quad A^{12} = \begin{bmatrix} 10 & 1 & 1 & 2 \\ 1 & 11 & 2 & 3 \\ 1 & 2 & 12 & 13 \\ 2 & 3 & 13 & 14 \end{bmatrix}.$$

3.13 Diga si las siguientes matrices A son definidas positivas o semidefinidas positivas en el espacio nulo de la matriz M.

$$A^{13} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \quad M^{13} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}.$$

$$A^{14} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad M^{14} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}.$$

$$A^{15} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}, \quad M^{15} = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix}.$$

$$A^{16} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}, \quad M^{16} = \begin{bmatrix} 2 & 4 & 6 \\ 3 & 6 & 10 \end{bmatrix}.$$

$$A^{17} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}, \quad M^{17} = \begin{bmatrix} 2 & 4 & 6 \\ 3 & 6 & 10 \\ 10 & 10 & 10 \end{bmatrix}.$$

Capítulo 4

FUNCIONES CONVEXAS Y GENERALIZACIONES

4.1 FUNCIONES CONVEXAS

Definición 4.1. Una función f con valores reales, definida en un convexo no vacío C, es **convexa**, si para todo x, y en C y para todo λ en $[0,1]$

$$f(z_{xy\lambda}) = f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) = f_{xy\lambda}.$$

La función f es **estrictamente convexa** si para todo x, y en C, $x \neq y$ y para todo λ en $]0,1[$

$$f(z_{xy\lambda}) = f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) = f_{xy\lambda}.$$

La noción de convexidad se puede interpretar geométricamente de la siguiente manera: una función es convexa si al tomar dos puntos $(x, f(x))$, $(y, f(y))$ (de la "gráfica" de f) y unirlos por medio de un segmento de recta, éste nunca queda por debajo de la gráfica.

Definición 4.2. Una función f definida en un convexo no vacío es **cóncava** (**estrictamente cóncava**) si $-f$ es convexa (estrictamente convexa).

Es claro, según la definición, que toda función estrictamente convexa también es convexa. Como para $\lambda = 0$ y para $\lambda = 1$ siempre se tiene $f(x) \leq f(x)$ y $f(y) \leq f(y)$, entonces en la definición de convexidad se puede hacer variar λ únicamente en el intervalo abierto $]0,1[$.

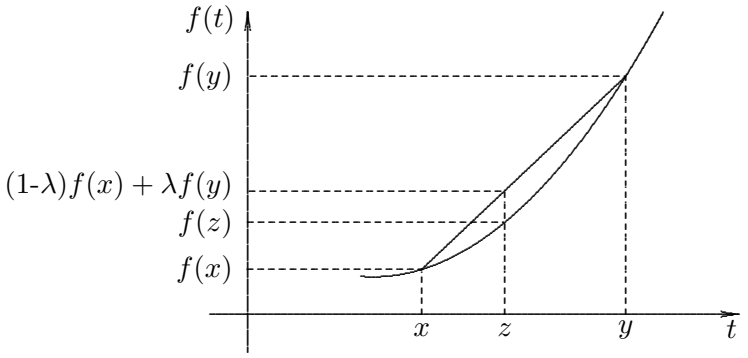


Figura 4.1

Ejemplo 4.1. $f : C \subseteq \mathbb{R} \longrightarrow \mathbb{R}, f(x) = x^2$. Gráficamente se "ve" que f es convexa. Por otro lado, sean x, y en C, λ en el intervalo $[0, 1]$,

$$\begin{aligned} f_{xy\lambda} - f(z_{xy\lambda}) &= (1-\lambda)x^2 + \lambda y^2 - \big((1-\lambda)x + \lambda y\big)^2 \\ &= \lambda x^2 - \lambda^2 x^2 + \lambda y^2 - \lambda^2 y^2 - 2\lambda xy + 2\lambda^2 xy \\ &= x^2\lambda(1-\lambda) + y^2\lambda(1-\lambda) - 2xy\lambda(1-\lambda) \\ &= (x-y)^2\lambda(1-\lambda) \\ &\geq 0. \end{aligned}$$

Luego $f(z_{xy\lambda}) \leq f_{xy\lambda}$, es decir, f es convexa. Además si $x \neq y$, entonces $(x-y)^2 > 0$, y si λ está en el intervalo abierto $]0, 1[$ se tiene que $\lambda(1-\lambda) > 0$, luego $f_{xy\lambda} - f(z_{xy\lambda}) > 0$, o sea, f es estrictamente convexa. $\diamond$

En la figura 4.2 las funciones en a), b) y c) son convexas; la función en a) es estrictamente convexa; las funciones en c) y d) son cóncavas; la función en d) es estrictamente cóncava; la función en e) no es ni convexa ni cóncava.

Ejemplo 4.2. Sean: $c \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $f(x) = c^{\mathrm{T}}x + \alpha$. Si $n = 2$ la gráfica corresponde a una parte de un plano.

$$\begin{aligned} f_{xy\lambda} - f(z_{xy\lambda}) &= (1-\lambda)(c^{\mathrm{T}}x + \alpha) + \lambda(c^{\mathrm{T}}y + \alpha) \\ &\quad -c^{\mathrm{T}}\big((1-\lambda)x + \lambda y\big) - \alpha = 0 \end{aligned}$$

Esta función es convexa, pero no estrictamente convexa, también es cóncava y no es estrictamente cóncava. $\diamond$

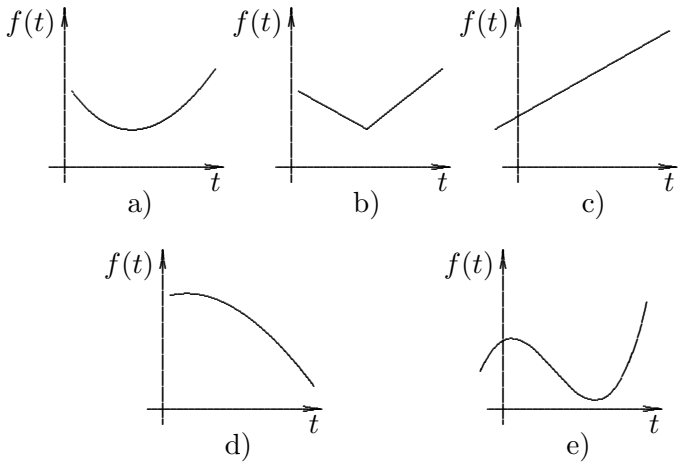


Figura 4.2

Ejemplo 4.3. $f : C \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}, f(x) = \|x\|$.

$$\begin{aligned} f(z_{xy\lambda}) = f\big((1-\lambda)x + \lambda y\big) &= \|(1-\lambda)x + \lambda y\| \\ &\leq \|(1-\lambda)x\| + \|\lambda y\| \\ &= |1-\lambda|\|x\| + |\lambda|\|y\| \\ &= (1-\lambda)f(x) + \lambda f(y). \end{aligned}$$

Entonces la norma es una función convexa. $\diamond$

Ejemplo 4.4. Sean: $x = (0,0)$, $y = (0,1)$, $\lambda = 1/2$, $z = z_{xy\lambda} = (0,1/2)$. Entonces $\|z\| = \|0.5y\| = 0.5\|y\|$. Por otro lado (1-$\lambda$)$\|x\| + \lambda\|y\| = 0.5\|y\|$. Luego ninguna norma $\|\|$ es una función estrictamente convexa. Esto no contradice el resultado de Análisis Funcional, según el cual la norma euclidiana es una norma estrictamente convexa. $\diamond$

Ejemplo 4.5. $f : [-2,3] \longrightarrow \mathbb{R}, f(x) = x^3$. Si $x = -1$, $y = 1$, $\lambda = 1/4$, entonces

$$\begin{aligned} f_{xy\lambda} - f(z_{xy\lambda}) &= \frac{3}{4}(-1) + \frac{1}{4}(1) - \big(\frac{3}{4}(-1) + \frac{1}{4}(1)\big)^3 \\ &= -\frac{1}{2} - \frac{-1}{8} \end{aligned}$$

$$= \quad -\frac{3}{8}.$$

Luego f no es convexa. Tampoco es cóncava. Sin embargo, para esta función, al cambiar el conjunto de definición, se puede tener convexidad. $\diamond$

Ejemplo 4.6. $f : [0,3] \longrightarrow \mathbb{R}, f(x) = x^3$ es convexa, más aún, es estrictamente convexa. $\diamond$

Definición 4.3. Sean: $f : A \longrightarrow \mathbb{R}$ una función, α un número real. Se llama un **conjunto de nivel** al subconjunto de A:

$$\Gamma_\alpha = \{x \in A : f(x) \leq \alpha\}.$$

Proposición 4.1. *Sea C un convexo. Si $f : C \longrightarrow \mathbb{R}$ es convexa, entonces todos los conjuntos de nivel son convexos. En particular $\{x : f(x) \leq 0\}$ es un convexo.*

Proposición 4.2. *Sea C un conjunto convexo. Si $f : C \longrightarrow \mathbb{R}$, $g : C \longrightarrow \mathbb{R}$ son funciones convexas y α y β son números no negativos, entonces*

$$\alpha f + \beta g : C \longrightarrow \mathbb{R} \quad \textit{es convexa.}$$

En particular, αf, $f + g$ son convexas.

Obviamente también se tiene la generalización para más de dos funciones. Si $f_1, ..., f_m$ son funciones convexas definidas en un convexo C, y $\alpha_1, ...\alpha_m$ son escalares no negativos, entonces

$$\alpha_1 f_1 + ... + \alpha_m f_m : C \longrightarrow \mathbb{R} \qquad \text{es convexa.}$$

Uno esperaría que el producto de dos funciones convexas fuera una función convexa, pero no es cierto en general. Basta con considerar x, x^2 que son funciones convexas en $\mathbb{R}$, pero $f(x) = x^3$ no es convexa en $\mathbb{R}$.

Proposición 4.3. *Sea C un conjunto convexo. Si $f : C \longrightarrow \mathbb{R}$ es convexa, entonces las siguientes "ampliaciones" de f son funciones convexas:*

$$\begin{array}{lll} f_1 & : \mathbb{R}^m \times C \times \mathbb{R}^p \longrightarrow \mathbb{R}, & f_1(u,x,v) = f(x) \\ f_2 & : \mathbb{R}^m \times C \longrightarrow \mathbb{R}, & f_2(u,x) = f(x) \\ f_3 & : C \times \mathbb{R}^p \longrightarrow \mathbb{R}, & f_3(x,v) = f(x). \end{array}$$

Ejemplo 4.7. Como la función $f(x) = x^2$ es convexa, entonces $f_1(x_1, x_2) = x_1^2$, $f_2(x_1, x_2) = x_2^2$ son convexas, luego $g(x_1, x_2) = x_1^2 + x_2^2$ es una función convexa. $\diamond$

Proposición 4.4. *Sean: C un convexo no vacío, $f : C \longrightarrow \mathbb{R}$. Si f es convexa, entonces es continua en el interior de C, o sea, la continuidad en el interior de C es una condición necesaria para la convexidad.*

Ejemplo 4.8. $f : [0,1] \longrightarrow \mathbb{R}$ definida por $f(x) = [x]$ (parte entera) no es continua en $[0,1]$, pero es convexa en $[0,1]$ y continua en $]0,1[$. $\diamond$

Ejemplo 4.9.

$$f(x) = \begin{cases} x^2 & \text{si} \quad x \neq 0 \\ 1 & \text{si} \quad x = 0, \end{cases}$$

Como f no es continua en $x = 0$, entonces no es convexa. $\diamond$

Definición 4.4. Sean: $C \subseteq \mathbb{R}^n$ un convexo no vacío, $f : C \longrightarrow \mathbb{R}$. Se llama **epígrafo** de f al subconjunto de $\mathbb{R}^{n+1}$:

$$\text{epi}(f) = \{(x, \lambda) : f(x) \leq \lambda\}.$$

Proposición 4.5. *Sea $C \subseteq \mathbb{R}^n$ un convexo no vacío. $f : C \longrightarrow \mathbb{R}$ es convexa si y solamente si su epígrafo es un conjunto convexo.*

Ejemplo 4.10. Sean: $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = |x|$.

$$\begin{aligned} \text{epi}(f) &= \{(x_1, x_2) : |x_1| \leq x_2\} \\ &= \{(x_1, x_2) : x_1 \leq x_2 \text{ si } x_1 \geq 0, -x_1 \leq x_2 \text{ si } x_1 \leq 0\} \\ &= \{(x_1, x_2) : x_1 - x_2 \leq 0 \text{ si } x_1 \geq 0, -x_1 - x_2 \leq 0 \text{ si } x_1 \leq 0\} \\ &= \{(x_1, x_2) : x_1 - x_2 \leq 0, -x_1 - x_2 \leq 0\} \end{aligned}$$

es convexo por ser intersección de dos semiespacios, luego f es convexa. $\diamond$

Ejemplo 4.11. Sean:

$$\begin{aligned} x &= (0,0,0) \in \text{epi}(f), \\ y &= (0,\text{-}1,\text{-}1) \in \text{epi}(f), \\ \lambda &= 1/2, \\ z &= \frac{1}{2}\,x + \frac{1}{2}\,y = (0,\text{-}1/2,\text{-}1/2) \notin \text{epi}(f), \\ &\quad \text{ya que } f(0,\text{-}1/2) = \text{-}1/8 \not\leq \text{-}1/2. \end{aligned}$$

Entonces epi(f)no es convexo, luego f no es convexa. $\diamond$

Proposición 4.6. *Sean: C un convexo no vacío, $f : C \longrightarrow \mathbb{R}$ convexa, $\varphi : f(C) \longrightarrow \mathbb{R}$ convexa y creciente ($u \leq v \Rightarrow \varphi(u) \leq \varphi(v)$). Entonces $g = \varphi \circ f : C \longrightarrow \mathbb{R}$, definida por $g(x) = \varphi(f(x))$, es convexa.*

Ejemplo 4.12. La función $g(x_1, x_2) = (2x_1 + 3x_2 + 4)^2$ definida para $x \geq 0$ es convexa, pues $f(x_1, x_2) = 2x_1 + 3x_2 + 4$ es convexa y $\varphi(t) = t^2$ es convexa y creciente en $[16, \infty]$. La función g es convexa en todo $\mathbb{R}^2$, pero esto no se deduce de la proposición anterior. $\diamond$

Ejemplo 4.13. La función $g(x_1, x_2) = \exp(2x_1 + 3x_2 + 4)$ es convexa pues $f(x_1, x_2) = 2x_1 + 3x_2 + 4$ es convexa y $\varphi(t) = \exp(t)$ es convexa y creciente en $\mathbb{R}$. $\diamond$

Proposición 4.7. *Sean: $C \subseteq \mathbb{R}^n$ un convexo abierto y no vacío, $f : C \longrightarrow \mathbb{R}$ diferenciable. Entonces f es convexa si y solamente si para todo $x, \bar{x} \in C$*

$$f(x) - f(\bar{x}) \geq f'(\bar{x})^T (x - \bar{x}).$$

La anterior desigualdad se puede presentar así:

$$f(x) \geq f(\bar{x}) + f'(\bar{x})^{\mathrm{T}} (x - \bar{x}).$$

Si $n = 1$ la expresión del lado derecho representa la recta tangente a la curva en el punto $(\bar{x}, f(\bar{x}))$, o sea, una función derivable es convexa si y solamente si la curva queda por encima de todas las rectas tangentes.

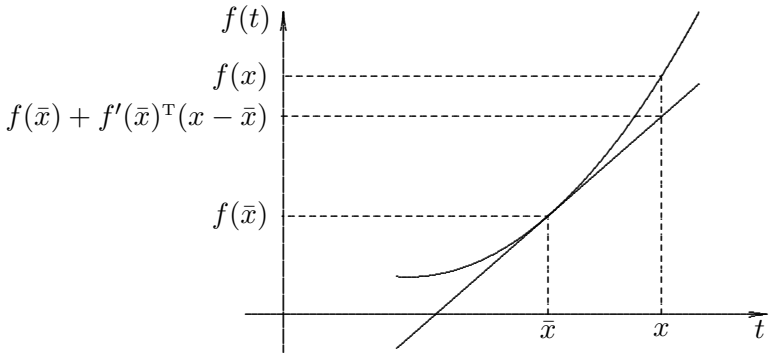


Figura 4.3

En general, la expresión del lado derecho representa la aproximación lineal de la función en $\bar{x}$, o sea, una función diferenciable es convexa si y solamente si cualquier aproximación lineal subevalúa la función.

Ejemplo 4.14. Sea $f(x_1, x_2) = x_1^2 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es diferenciable y además

$$\begin{aligned} f(x) - f(\bar{x}) &- f'(\bar{x})^{\mathrm{T}}(x - \bar{x}) = \\ &= x_1^2 + x_2^2 - \bar{x}_1^2 - \bar{x}_2^2 - \begin{bmatrix} 2\bar{x}_1 & 2\bar{x}_2 \end{bmatrix} \begin{bmatrix} x_1\text{-}\bar{x}_1 \\ x_2\text{-}\bar{x}_2 \end{bmatrix} \\ &= x_1^2 + x_2^2 - \bar{x}_1^2 - \bar{x}_2^2 - 2\bar{x}_1 x_1 + 2\bar{x}_1^2 - 2\bar{x}_2 x_2 + 2\bar{x}_2^2 \\ &= x_1^2 + x_2^2 + \bar{x}_1^2 + \bar{x}_2^2 - 2\bar{x}_1 x_1 - 2\bar{x}_2 x_2 \\ &= (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2 \\ &\geq 0. \end{aligned}$$

Entonces f es convexa. $\diamond$

Proposición 4.8. *Sean: $C \subseteq \mathbb{R}^n$ un convexo abierto y no vacío, $f : C \longrightarrow \mathbb{R}$ diferenciable. Entonces f es convexa si y solamente si*

$$\big(f'(y) - f'(x)\big)^{T}(y - x) \geq 0 \quad \textit{para todo } x, y \in C.$$

Si $n = 1$ el gradiente es simplemente la derivada, y así la anterior desigualdad está indicando que $f'(x)$ es creciente.

Ejemplo 4.15. Sea $f(x_1, x_2) = x_1^2 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es diferenciable y además

$$\begin{aligned} \big(f'(y) - f'(x)\big)^{\mathrm{T}}(y - x) &= \big(f'(y)^{\mathrm{T}} - f'(x)^{\mathrm{T}}\big)(y - x) \\ &= \big([2y_1 \quad 2y_2] - [2x_1 \quad 2x_2]\big) \begin{bmatrix} y_1\text{-}x_1 \\ y_2\text{-}x_2 \end{bmatrix} \\ &= [2(y_1 - x_1) \quad 2(y_2 - x_2)] \begin{bmatrix} y_1\text{-}x_1 \\ y_2\text{-}x_2 \end{bmatrix} \\ &= 2(y_1 - x_1)^2 + 2(y_2 - x_2)^2 \\ &\geq 0. \ \diamond \end{aligned}$$

Entonces f es convexa.

Ejemplo 4.16. Sea $f(x_1, x_2) = x_1^3 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es diferenciable y además

$$\big(f'(y) - f'(x)\big)^{\mathrm{T}}(y - x) = \big([3y_1^2 \quad 2y_2] - [3x_1^2 \quad 2x_2]\big) \begin{bmatrix} y_1\text{-}x_1 \\ y_2\text{-}x_2 \end{bmatrix}$$

$$\begin{aligned} &= [3(y_1^2 - x_1^2) \quad 2(y_2 - x_2)] \begin{bmatrix} y_1\text{-}x_1 \\ y_2\text{-}x_2 \end{bmatrix} \\ &= 3(y_1 + x_1)(y_1 - x_1)^2 + 2(y_2 - x_2)^2 \\ &= -3, \quad \text{si } x = (-1, 0), \ \ y = (0, 0). \end{aligned}$$

Entonces f no es convexa. $\diamond$

Ejemplo 4.17. Sea $f(x_1, x_2) = x_1^3 + x_2^2$ definida en $C = \{(x_1, x_2) : x_1 \geq 0\}$. Como f es diferenciable y además

$$\begin{aligned} \big(f'(y) - f'(x)\big)^{\mathrm{T}}(y - x) &= 3(y_1 + x_1)(y_1 - x_1)^2 + 2(y_2 - x_2)^2 \\ &\geq 0, \end{aligned}$$

entonces, según la última proposición, f es convexa en el interior de C. Faltaría por ver que también es convexa en todo C. $\diamond$

Definición 4.5. Sea $A \subseteq \mathbb{R}^n$ un conjunto abierto. $F : A \longrightarrow \mathbb{R}^n$ se llama **monótona** si

$$\big(F(y) - F(x)\big)^{\mathrm{T}}(y - x) \geq 0 \quad \text{para todo } x, y \in A.$$

La última proposición se puede enunciar así: una función diferenciable f, definida en un convexo abierto, es convexa si y solamente si su gradiente es monótono.

Una función de una sola variable (definida en un intervalo de $\mathbb{R}$) es **monótona** si es creciente en todo el conjunto de definición o si es decreciente en todo el conjunto de definición.

Proposición 4.9. *Sean: C un convexo abierto y no vacío, $f : C \to \mathbb{R}$ doblemente diferenciable. Entonces f es convexa si y solamente si la matriz hessiana $H_f(x) = f''(x) = \left[\dfrac{\partial^2 f}{\partial x_j \partial x_i}(x)\right]$ es semidefinida positiva en todo punto x de C.*

Recuérdese que para funciones doblemente diferenciables se cumple que $\dfrac{\partial^2 f}{\partial x_j \partial x_i} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$, luego la matriz hessiana es simétrica.

Ejemplo 4.18. Sea $f(x_1, x_2) = x_1^2 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es doblemente diferenciable y además

$$f''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

es definida positiva, luego semidefinida positiva, entonces f es convexa. $\diamond$

Una característica de las funciones cuadráticas, de las afines (lineal más una constante) y de las constantes es que su matriz hessiana es constante, es decir, no depende del punto donde se evalúe.

Proposición 4.10. *Sean: C un convexo abierto no vacío, $f : C \longrightarrow \mathbb{R}$ doblemente diferenciable. Si la matriz hessiana $f''(x)$ es definida positiva en todo punto x de C, entonces f es estrictamente convexa.*

Ejemplo 4.19. Sea $f(x_1, x_2) = x_1^2 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es doblemente diferenciable y además

$$f''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

es definida positiva, entonces f es estrictamente convexa. $\diamond$

Ejemplo 4.20. Sea $f(x_1, x_2) = x_1^4 + x_2^2$ definida en todo $\mathbb{R}^2$. Como f es doblemente diferenciable y además

$$f''(x) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}$$

es semidefinida positiva para todo x, entonces f es convexa. La matriz hessiana no siempre es definida positiva, por ejemplo para $x_1 = 0$, luego a partir de la última proposición no se puede afirmar que f sea estrictamente convexa, sin embargo, sí lo es. $\diamond$

Ejemplo 4.21. Sea $f : B[(3, 2), 1] \to \mathbb{R}$ definida por $f(x_1, x_2) = x_1^3 + x_2^2$. La función f es doblemente diferenciable, además

$$f''(x) = \begin{bmatrix} 6x_1 & 0 \\ 0 & 2 \end{bmatrix}$$

es semidefinida positiva siempre que $x_1 \geq 0$, en particular, en el conjunto abierto definido por la restricción $x_1 > 0$, que contiene la bola $B[(3, 2), 1]$, entonces f es convexa, más aún, es estrictamente convexa. $\diamond$

Ejemplo 4.22. Sea $f(x_1, x_2) = x_1^2 x_2^2$ definida en todo $\mathbb{R}^2$. La función f es doblemente diferenciable,

$$f''(x) = \begin{bmatrix} 2x_2^2 & 4x_1x_2 \\ 4x_1x_2 & 2x_1^2 \end{bmatrix}$$

no es semidefinida positiva ya que $\delta_2 = -12x_1^2x_2^2$, entonces f no es convexa. $\diamond$

Ejemplo 4.23. Sea $f(x_1, x_2) = 5(3x_1 - 4x_2 - 2)^2 + 6(x_1 - 2x_2)^2$ definida en todo $\mathbb{R}^2$. Como f es doblemente diferenciable y además

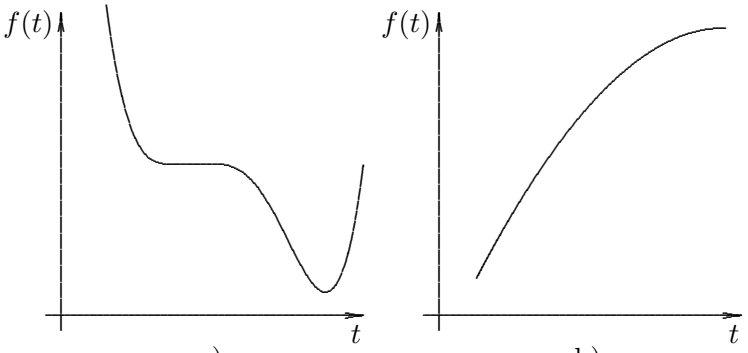
$$f''(x) = \begin{bmatrix} 102 & -144 \\ -144 & 208 \end{bmatrix}, \quad \delta_1 = 102, \quad \delta_2 = 480.$$

Entonces la matriz hessiana es definida positiva y la función f es estrictamente convexa. $\diamond$

4.2 GENERALIZACIONES DE FUNCIONES CONVEXAS

Definición 4.6. Una función f con valores reales, definida en un convexo no vacío C, es **cuasiconvexa**, si para todo $x, y \in C$ y para todo λ en el intervalo $[0, 1]$

$$f(z_{xy\lambda}) = f((1 - \lambda)x + \lambda y) \leq \max\{f(x), f(y)\}.$$



a) b)

Figura 4.4

Definición 4.7. Una función f con valores reales, definida en un convexo no vacío C, es **cuasicóncava**, si $-f$ es cuasiconvexa.

Proposición 4.11. *Toda función convexa es cuasiconvexa.*

Ejemplo 4.24. $f(x_1, x_2) = x_1^2 + x_2^2$ es convexa, luego es cuasiconvexa. $\diamond$

Proposición 4.12. *Sea $C \subseteq \mathbb{R}$ convexo, $f : C \longrightarrow \mathbb{R}$. Si f es monótona, entonces es cuasiconvexa.*

Ejemplo 4.25. $f(x) = x^3$ siempre es creciente, luego es cuasiconvexa. $\diamond$

Ejemplo 4.26. $f(x) = x^2$ definida en todo $\mathbb{R}$ no es monótona, sin embargo, sí es cuasiconvexa. $\diamond$

Proposición 4.13. *Sean: $C \subseteq \mathbb{R}^n$ convexo, $f : C \longrightarrow \mathbb{R}$. f es cuasiconvexa si y solamente si para todo real α, el conjunto de nivel Γ_α es convexo.*

Obviamente la anterior caracterización de funciones cuasiconvexas es una condición necesaria para funciones convexas

Ejemplo 4.27. $f(x_1, x_2) = (x_1 + 2x_2 - 3)^3 + 4$ no es convexa en $\mathbb{R}^2$, ya que

$$f''(x) = \begin{bmatrix} 6(x_1 + 2x_2 - 3) & 12(x_1 + 2x_2 - 3) \\ 12(x_1 + 2x_2 - 3) & 24(x_1 + 2x_2 - 3) \end{bmatrix}$$

y para $x = (0, 0)$, $\delta_1 = -18$, luego la matriz hessiana no es semidefinida positiva. Sin embargo,

$$\begin{aligned} \Gamma_\alpha &= \{(x_1, x_2) : (x_1 + 2x_2 - 3)^3 + 4 \leq \alpha\} \\ &= \{(x_1, x_2) : (x_1 + 2x_2 - 3)^3 \leq \alpha - 4\} \\ &= \{(x_1, x_2) : x_1 + 2x_2 - 3 \leq \sqrt[3]{\alpha - 4}\} \\ &= \{(x_1, x_2) : x_1 + 2x_2 \leq \sqrt[3]{\alpha - 4} + 3\} \end{aligned}$$

es convexo (es un semiespacio) para todo α, luego f es cuasiconvexa en todo $\mathbb{R}^2$. $\diamond$

Ejemplo 4.28. $f : (2, \infty) \longrightarrow \mathbb{R}$, $f(x) = \sqrt{x}$ no es convexa, sin embargo,

$$\Gamma_\alpha = \begin{cases} \emptyset & \text{si } \alpha < \sqrt{2} \\ [2, \alpha^2] & \text{si } \alpha \geq \sqrt{2}, \end{cases}$$

es convexo para todo α, entonces f es cuasiconvexa. Por otro lado, la función es estrictamente creciente, luego es creciente y, por lo tanto, es cuasiconvexa. $\diamond$

Ejemplo 4.29. $f(x_1, x_2) = x_1^2 x_2^2$.

$$\begin{aligned}\Gamma_0 &= \{(x_1, x_2) : x_1^2 x_2^2 \leq 0\} \\ &= \{(x_1, x_2) : x_1^2 x_2^2 = 0\} \\ &= \{(x_1, x_2) : x_1 x_2 = 0\} \\ &= \{(x_1, x_2) : x_1 = 0, \text{ o }, x_2 = 0\} \\ &= \{ \text{ puntos sobre los ejes } \}\end{aligned}$$

no es un conjunto convexo, entonces f no es cuasiconvexa y mucho menos convexa. $\diamond$

Ejemplo 4.30. Sea $f : [-2, 5[\longrightarrow \mathbb{R}$ definida por $f(x) = -|x|$. f no es convexa. Además, si $x = -2, \quad y = 3, \quad \lambda = 0.2$, entonces

$$\begin{aligned}\max\{f(x), f(y)\}\text{-}f(z_{xy\lambda}) &= \max\{f(x), f(y)\}\text{-}f((1-\lambda)x + \lambda y) \\ &= \max\{\text{-}2, \text{-}3\} - f(\text{-}1) \\ &= -2 - -1 \\ &= -1.\end{aligned}$$

Luego f no es cuasiconvexa. $\diamond$

Proposición 4.14. *Sean: C un convexo abierto y no vacío, $f : C \longrightarrow \mathbb{R}$ una función diferenciable. Entonces f es cuasiconvexa si y solamente si para todo $x, y \in C$:*

$$f(y) \leq f(x) \Rightarrow f'(x)^T(y - x) \leq 0.$$

La anterior implicación es obviamente equivalente a $f'(x)^{\mathrm{T}}(y - x) > 0 \Rightarrow f(y) > f(x)$.

Ejemplo 4.31. $f(x) = x^3$ es cuasiconvexa ya que: $f(y) \leq f(x)$ implica que $y \leq x$, o sea, $y - x \leq 0$ entonces $\alpha(y - x) \leq 0$ para $\alpha \geq 0$. En particular $3x^2(y - x) \leq 0$. $\diamond$

Ejemplo 4.32. $f(x_1, x_2) = (x_1 + x_2)^3$ es cuasiconvexa ya que: $f'(x)^{\mathrm{T}}(y - x) > 0$ es equivalente a $3(x_1 + x_2)^2((y_1 + y_2) - (x_1 + x_2)) > 0$, lo que implica que $y_1 + y_2 > x_1 + x_2$, luego $f(y) > f(x)$. $\diamond$

Ejemplo 4.33. $f(x_1, x_2) = x_1 x_2$ no es cuasiconvexa en $\mathbb{R}^2$ ya que para $x = (1, 0)$, $y = (-1, 1)$, se cumple $f(y) = -1 \leq f(x) = 0$, pero $f'(x)^{\mathrm{T}}(y - x) = [01][-21]^{\mathrm{T}} = 1 \not\leq 0$. $\diamond$

Definición 4.8. Sean: C un conjunto convexo, $f : C \longrightarrow \mathbb{R}$. Se dice que f es **estrictamente cuasiconvexa** si para todo $x, y \in C$ con $f(x) \neq f(y)$ y para todo λ en el intervalo $]0, 1[$

$$f(z_{xy\lambda}) = f((1-\lambda)x + \lambda y) < \max\{f(x), f(y)\}.$$

Un poco en contra de lo que se espera, no se tiene que cuasiconvexidad estricta implica cuasiconvexidad.

Ejemplo 4.34. Sea f definida en todos los reales por

$$f(x) = \begin{cases} 1 & \text{si} \quad x = 0 \\ 0 & \text{si} \quad x \neq 0. \end{cases}$$

f es estrictamente cuasiconvexa, pero no es cuasiconvexa. $\diamond$

Ejemplo 4.35. Sea f definida en todos los reales por

$$f(x) = \begin{cases} 1 & \text{si} \quad x \neq 0 \\ 0 & \text{si} \quad x = 0. \end{cases}$$

f es cuasiconvexa, pero no es estrictamente cuasiconvexa. $\diamond$

Definición 4.9. Sean: C un conjunto convexo, $f : C \longrightarrow \mathbb{R}$. Se dice que f es **fuertemente cuasiconvexa** si para todo $x \neq y \in C$ y para todo λ en el intervalo $]0, 1[$

$$f(z_{xy\lambda}) = f((1-\lambda)x + \lambda y) < \max\{f(x), f(y)\}.$$

Ejemplo 4.36. $f(x) = 1 - \exp(\text{-}x^2)$ no es convexa, es cuasiconvexa, fuertemente y estrictamente cuasiconvexa. $\diamond$

Ejemplo 4.37. $f(x) = \max\{0, x^2\text{-}1\}$ es convexa, no es fuertemente cuasiconvexa, sí es cuasiconvexa y estrictamente cuasiconvexa. $\diamond$

Proposición 4.15.

convexidad estricta	$\Longrightarrow$	*cuasiconvexidad fuerte.*
cuasiconvexidad fuerte	$\Longrightarrow$	*cuasiconvexidad estricta*
cuasiconvexidad fuerte	$\Longrightarrow$	*cuasiconvexidad .*

Definición 4.10. Sea $f : A \longrightarrow \mathbb{R}$ doblemente diferenciable en un punto $\bar{x} \in A$. Se llama **hessiano orlado (bordered)** de f en el punto $\bar{x}$ a la matriz de tamaño $(n+1) \times (n+1)$

$$B = B(\bar{x}) = B(f, \bar{x}) = \widehat{f''(\bar{x})} = \begin{bmatrix} f''(\bar{x}) & f'(\bar{x}) \\ f'(\bar{x})^{\mathrm{T}} & 0 \end{bmatrix}.$$

Definición 4.11. Se llama **submatriz principal del hessiano orlado**, a toda submatriz de B obtenida quitando algunas de las primeras n filas de B y quitando exactamente esas mismas columnas. De manera más precisa, sean: $k \in \{1,\ldots,n\}$, $\gamma = \{i_1,\ldots,i_k\}$ con $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. La matriz $B_{k\gamma}(\bar{x})$ de tamaño $(k+1)\times(k+1)$, obtenida al dejar únicamente los elementos b_{ij} tales que $i,j \in \{i_1,\ldots,i_k,n+1\}$, es una submatriz principal del hessiano orlado. Si $\gamma = \{1,2,\ldots,k\}$ la matriz se denota simplemente $B_k(\bar{x})$ y se llama **submatriz estrictamente principal del hessiano orlado**.

Ejemplo 4.38. $B_n(\bar{x}) = B(\bar{x})$. $\diamond$

Ejemplo 4.39. Sean: $n = 2$, $k = 1$, $\gamma = \{2\}$. Entonces

$$B_{1\gamma} = \begin{bmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{bmatrix}. \ \diamond$$

Ejemplo 4.40. Sean: $n = 2$, $k = 1$. Entonces

$$B_1 = \begin{bmatrix} b_{11} & b_{13} \\ b_{31} & b_{33} \end{bmatrix}. \ \diamond$$

Proposición 4.16. *Sean: C un convexo sólido (de interior no vacío), $f : C \longrightarrow \mathbb{R}$ doblemente diferenciable. Si f es cuasiconvexa, entonces*

$$\det\big(B_k(x)\big) \leq 0\ , \quad \forall k = 1,\ldots,n, \quad \forall x \in C.$$

Ejemplo 4.41. Sea $f(x_1,x_2) = x_1^3 + x_2^2$ defininida en todo $\mathbb{R}^2$.

$$\begin{aligned} B(x) &= \begin{bmatrix} 6x_1 & 0 & 3x_1^2 \\ 0 & 2 & 2x_2 \\ 3x_1^2 & 2x_2 & 0 \end{bmatrix}, \\ \det(B_1(x)) &= \det \begin{bmatrix} 6x_1 & 3x_1^2 \\ 3x_1^2 & 0 \end{bmatrix} = -9x_1^4, \\ \det(B_2(x)) &= \det(B(x)) = -18x_1^4 - 24x_1x_2^2. \end{aligned}$$

Si $x_1 = -1$ y $x_2 = 1$, entonces $\det(B_2(x)) = 6$, luego f no es cuasiconvexa. Lo anterior también se puede ver aplicando directamente la definición. Si $x = (-5,5)$, $y = (1,0)$, $\lambda = 0.6$, entonces $z = (-1.4,2)$, $f(x) = -100$, $f(y) = 1$, $f(z) = 1.256$. $\diamond$

Para funciones de una variable, la proposición anterior no sirve para nada pues para toda función doblemente diferenciable

$$B(x) = \begin{bmatrix} f''(x) & f'(x) \\ f'(x) & 0 \end{bmatrix}$$

y $\det(B_1(x)) = \det(B(x)) = -(f'(x))^2 \leq 0$.

Ejemplo 4.42. Sea $f(x) = x^3$ definida en todo $\mathbb{R}$.

$$\begin{aligned} B(x) &= \begin{bmatrix} 6x & 3x^2 \\ 3x^2 & 0 \end{bmatrix}, \\ \det(B_1(x)) &= \det(B(x)) = -9x^4. \end{aligned}$$

Entonces la proposición apenas permite decir que posiblemente f es cuasiconvexa. Recuérdese que x^3 sí es cuasiconvexa. Por otro lado, si se considera $f(x) = -x^2$ que no es cuasiconvexa, la proposición anterior dice que $-x^2$ cumple esta condición necesaria para cuasiconvexidad. $\diamond$

Proposición 4.17. *Sean: C un convexo sólido, $f : C \longrightarrow \mathbb{R}$ doblemente diferenciable. Si*

$$\det\big(B_k(x)\big) < 0 \quad , \forall k = 1, \ldots, n, \quad \forall x \in C,$$

entonces f es cuasiconvexa.

Ejemplo 4.43. Sea $f(x_1, x_2) = x_1^2 - x_2^2$ definida en $\mathbb{R}^2$.

$$\begin{aligned} B(x) &= \begin{bmatrix} 2 & 0 & 2x_1 \\ 0 & \text{-}2 & \text{-}2x_2 \\ 2x_1 & \text{-}2x_2 & 0 \end{bmatrix}, \\ \det(B_2(x)) &= 8(x_1^2 - x_2^2) \\ &= 8 \quad \text{para } x = (1, 0). \end{aligned}$$

Luego la función no es cuasiconvexa. $\diamond$

Ejemplo 4.44. Sea $f(x) = \log x$ con $x \in C =]0, \infty[$. En este libro log indica el logaritmo en base e.

$$\begin{aligned} B(x) &= \begin{bmatrix} -\frac{1}{x^2} & \frac{1}{x} \\ \frac{1}{x} & 0 \end{bmatrix} \\ \det(B_1(x)) &= -\frac{1}{x^2} < 0 \quad \forall x \in C. \end{aligned}$$

Luego $f(x) = \log x$ es cuasiconvexa. $\diamond$

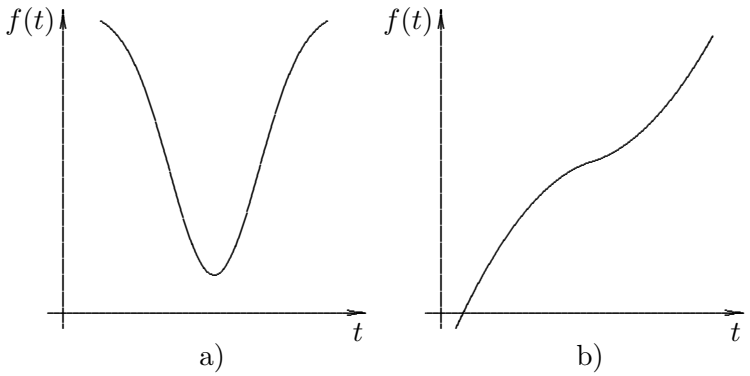
Ejemplo 4.45. Sea $f(x) = x^2$ en $\mathbb{R}$.

$$\begin{aligned} B(x) &= \begin{bmatrix} 2 & 2x \\ 2x & 0 \end{bmatrix}, \\ \det(B_1(x)) &= -4x^2. \end{aligned}$$

Como $\det(B_k(x))$ no es siempre negativo, entonces la proposición anterior no permite garantizar que $f(x) = x^2$ sea cuasiconvexa. Sin embargo, f sí es cuasiconvexa, más aún es estrictamente convexa. $\diamond$

Definición 4.12. Sean: C un convexo abierto y no vacío, $f : C \longrightarrow \mathbb{R}$ diferenciable. f es **seudoconvexa** si para todo $x, y \in C$

$$f'(x)^{\mathrm{T}}(y - x) \geq 0 \Rightarrow f(y) - f(x) \geq 0.$$



a) b)

Figura 4.5

Estas dos funciones de la Figura 4.5 son seudoconvexas. En la Figura 4.4 , la primera función no es seudoconvexa, la segunda sí lo es.

La definición de función seudoconvexa es muy parecida a la caracterización de funciones diferenciables cuasiconvexas, pero es más fuerte. En la seudoconvexidad la desigualdad no estricta implica que $f(y) > f(x)$, en cambio en la cuasiconvexidad se requiere la desigualdad estricta para implicar el mismo resultado.

Proposición 4.18. *Toda función seudoconvexa es cuasiconvexa.*

Si una función de una variable, derivable, es seudoconvexa, entonces, cuando se anula la derivada se tiene un mínimo. Además si en un punto la derivada es positiva, entonces, a partir de ese punto la función es creciente. Si en un punto la derivada es negativa, entonces antes de ese punto la función es decreciente.

Proposición 4.19. *Sea C un convexo, abierto y no vacío. Si $f : C \longrightarrow \mathbb{R}$ es diferenciable y convexa, entonces f es seudoconvexa.*

Ejemplo 4.46. $f(x_1, x_2) = x_1^2 + x_2^2$ es convexa y diferenciable, luego es seudoconvexa. $\diamond$

Ejemplo 4.47. $f :]-4, -2[\longrightarrow \mathbb{R}$, $f(x) = x^3$ no es convexa. Veámos que sí es seudoconvexa. Ante todo, f es diferenciable. Sean x, $y \in]-4, -2[$.

$$\begin{aligned} f'(x)^{\mathrm{T}}(y-x) \geq 0 &\Leftrightarrow 3x^2(y-x) \geq 0 \\ &\Leftrightarrow (y-x) \geq 0 \\ &\Leftrightarrow y \geq x \\ &\Leftrightarrow y^3 \geq x^3 \\ &\Leftrightarrow y^3 - x^3 \geq 0 \\ &\Leftrightarrow f(y) - f(x) \geq 0. \end{aligned}$$

Luego f es seudoconvexa. $\diamond$

Ejemplo 4.48. $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = x^3$ no es convexa, pero sí es cuasiconvexa. Veámos que no es seudoconvexa. Sean: $x = 0$, $y =$ -1.

$$\begin{aligned} f'(x)^{\mathrm{T}}(y-x) &= 0(-1-0) \\ &= 0 \\ &\geq 0, \end{aligned}$$

sin embargo,

$$f(y) - f(x) = (\text{-}1)^3 - (0)^3 = -1 \ngeq 0. \quad \diamond$$

Ejemplo 4.49. Sea $f(x) = 1 - e^{-x^2}$ definida en toda la recta real. ¿Es convexa? ¿Es cuasiconvexa? ¿Es seudoconvexa?

$$\begin{aligned} f'(x) &= 2xe^{-x^2}, \\ f''(x) &= e^{-x^2}(2 - 4x^2), \\ f''(1) &= -2e^{-1} < 0. \end{aligned}$$

Como hay puntos donde la "matriz" hessiana no es semidefinida positiva, entonces se puede afirmar que f no es convexa. Obsérvese que x^2 varía entre cero e infinito, entonces $\exp(-x^2)$ varía entre cero y uno, luego $f(x)$ toma valores entre cero y uno. En particular toma el valor cero, valor mínimo, para $x = 0$. Por medio de los conjuntos de nivel se puede saber si es cuasiconvexa.

$$\begin{aligned} 1 - e^{-x^2} &\leq \alpha \\ e^{-x^2} &\geq 1 - \alpha \\ -x^2 &\geq \log(1 - \alpha) \\ x^2 &\leq -\log(1 - \alpha) \\ |x| &\leq \sqrt{-\log(1\text{-}\alpha)} = t. \end{aligned}$$

Entonces

$$\Gamma_\alpha = \begin{cases} \emptyset & \text{si } \alpha < 0, \\ [-t, t] & \text{si } 0 \leq \alpha < 1, \\ \mathbb{R} & \text{si } \alpha \geq 1. \end{cases}$$

Luego f es cuasiconvexa. Para averiguar si f es seudoconvexa, estudiemos tres casos diferentes: x nulo, x positivo, x negativo.

i) Sea $x = 0$.

$$\begin{aligned} f'(x)^{\mathrm{T}}(y - x) &\geq 0 \\ f'(0)(y - 0) &\geq 0 \\ 0 &\geq 0. \end{aligned}$$

Como lo anterior siempre es cierto, se debería cumplir que

$$f(y) - f(x) \geq 0$$

$$\begin{aligned} f(y) - f(0) &\geq 0 \\ f(y) &\geq 0, \end{aligned}$$

y siempre se cumple ya que cero es el valor mínimo de la función.

ii) Sea $x > 0$. Entonces $f'(x) > 0$.

$$\begin{aligned} f'(x)^{\mathrm{T}}(y-x) &\geq 0 \\ f'(x)(y-x) &\geq 0 \\ (y-x) &\geq 0 \\ y &\geq x > 0 \\ f(y) &\geq f(x) \qquad (\text{ por ser } f \text{ creciente en }]0, \infty[) \\ f(y) - f(x) &\geq 0. \end{aligned}$$

iii) Sea $x < 0$. Entonces $f'(x) < 0$.

$$\begin{aligned} f'(x)^{\mathrm{T}}(y-x) &\geq 0 \\ f'(x)(y-x) &\geq 0 \\ (y-x) &\leq 0 \\ y &\leq x < 0 \\ f(y) &\geq f(x) \qquad (\text{ por ser } f \text{ decreciente en }]-\infty, 0\,[\ ) \\ f(y) - f(x) &\geq 0. \end{aligned}$$

Luego en los tres casos se cumple la implicación, entonces la función f es seudoconvexa. $\diamond$

Proposición 4.20. *Sean: C un convexo abierto, $f : C \longrightarrow \mathbb{R}$ doblemente diferenciable. Si*

$$\det\big(B_k(x)\big) < 0, \ \forall k = 1, \ldots, n, \quad \forall x \in C,$$

entonces f es seudoconvexa.

Estas condiciones suficientes son las mismas de la cuasiconvexidad. Más adelante hay un resultado que relaciona cuasiconvexidad y seudoconvexidad para funciones doblemente diferenciables definidas sobre conjuntos sólidos. Por las mismas razones de dos ejemplos anteriores, $f(x) = \log x$ es seudoconvexa para $x > 0$. Por otro lado, usando la proposición anterior, no se puede afirmar que $f(x) = x^2$ sea seudoconvexa.

Proposición 4.21. *Sean: C un convexo sólido, $f : C \longrightarrow \mathbb{R}$ cuadrática. Entonces f es cuasiconvexa en C si y solamente si f es seudoconvexa en $\overset{\circ}{C}$.*

En la literatura sobre métodos de minimización en una variable es muy frecuente encontrar métodos para funciones unimodales. Las definiciones de unimodalidad pueden variar ligeramente.

Definición 4.13. [Avr76] Una función φ es **unimodal** en el intervalo $I = [a, b]$ si existe λ^* minimizador de φ en I tal que φ es estrictamente decreciente en $[a, \lambda^*]$ y estrictamente creciente en $[\lambda^*, b]$.

Esta definición de unimodalidad es equivalente, para funciones de una variable definidas en un intervalo, a la cuasiconvexidad fuerte con existencia de un minimizador global.

Definición 4.14. [Lue89] Una función φ es unimodal en el intervalo $I = [a, b]$ si existe un único λ^* minimizador local de φ en I.

Ejemplo 4.50. Sea φ definida en el intervalo $[-1, 1]$ por

$$\varphi(x) = \begin{cases} 0 & \text{si } x = 0, \\ 1 & \text{si } x \neq 0 \text{ y es racional}, \\ 2 & \text{si } x \text{ es irracional}. \end{cases}$$

Este ejemplo muestra que las dos definiciones no son equivalentes ya que $x^* = 0$ es el único minimizador local, pero φ no es estrictamente creciente en $[0, 1]$. $\diamond$

Proposición 4.22. *Toda función de una variable, estrictamente convexa es unimodal.*

4.3 CONVEXIDAD Y GENERALIZACIONES EN UN PUNTO

Las definiciones de convexidad, cuasiconvexidad y seudoconvexidad se han hecho para todos los puntos de un conjunto convexo C, pero se pueden modificar para aplicarlas a un solo punto.

Definición 4.15. Sean: C un conjunto convexo, $f : C \longrightarrow \mathbb{R}$, $\bar{x}$ un elemento de C. Se dice que f es **convexa en** $\bar{x}$ si para todo $y \in C$ y para todo $\lambda \in [0, 1]$,

$$f(z_{\bar{x}y\lambda}) = f((1-\lambda)\bar{x} + \lambda y) \leq (1-\lambda)f(\bar{x}) + \lambda f(y) = f_{\bar{x}y\lambda}.$$

Ejemplo 4.51.

$$f(x) = \begin{cases} x^2 & \text{si} \quad x \neq 0 \\ 1 & \text{si} \quad x = 0, \end{cases}$$

no es convexa (no es continua), sin embargo, f sí es convexa en $\bar{x} = 0$. $\diamond$

Definición 4.16. Sean: C un conjunto convexo, $f : C \longrightarrow \mathbb{R}$, $\bar{x}$ un elemento de C. Se dice que f es **cuasiconvexa en** $\bar{x}$ si para todo $y \in C$ y para todo $\lambda \in [0, 1]$,

$$f(z_{\bar{x}y\lambda}) = f((1 - \lambda)\bar{x} + \lambda y) \leq \max\{f(\bar{x}), f(y)\}.$$

Ejemplo 4.52.

$$f(x) = \begin{cases} \cos x & \text{si} \quad x \neq 1 \\ 2 & \text{si} \quad x = 1 \end{cases}$$

no es cuasiconvexa (basta con tomar x = -$\pi, y = \pi, \lambda = 1/2, z = 0$), sin embargo, f sí es cuasiconvexa en $\bar{x} = 1$. $\diamond$

Definición 4.17. Sean: C un convexo abierto y no vacío, $f : C \longrightarrow \mathbb{R}$ diferenciable, $\bar{x} \in C$. Se dice que f es **seudoconvexa en** $\bar{x}$ si para todo $y \in C$,

$$f'(\bar{x})^{\mathrm{T}}(y - \bar{x}) \geq 0 \Longrightarrow f(y) - f(\bar{x}) \geq 0.$$

Ejemplo 4.53. $f(x) = x^3$ es seudoconvexa en cualquier punto diferente de cero. $\diamond$

Si se hubiera presentado primero la definición de convexidad en un punto, la definición de convexidad (global) se podría presentar de la siguiente manera. Sea C un convexo; se dice que $f : C \longrightarrow \mathbb{R}$ es convexa en C, si es convexa para todo $x \in C$.

Hasta ahora las definiciones de convexidad, cuasiconvexidad y seudoconvexidad se aplican a funciones definidas en conjuntos convexos. Mangasarian [Man69] presenta definiciones que pueden aplicarse a funciones definidas en conjuntos no convexos, por ejemplo:

Definición 4.18. Una función $f : A \longrightarrow \mathbb{R}$ se llama convexa en un punto $\bar{x} \in A$ si para todo $y \in A$ y para todo $\lambda \in [0, 1]$ tales que $z_{\bar{x}y\lambda} = (1 - \lambda)\bar{x} + \lambda y \in A$ se cumple que

$$f(z_{\bar{x}y\lambda}) = f((1 - \lambda)\bar{x} + \lambda y) \leq (1 - \lambda)f(\bar{x}) + \lambda f(y) = f_{\bar{x}y\lambda}.$$

Ejemplo 4.54. Sean: a, b dos números diferentes, $f : \{a, b\} \to \mathbb{R}$. Según la definición de Mangasarian, f es convexa en a y también en b, luego es convexa en todo el conjunto $\{a, b\}$. $\diamond$

La convexidad y sus generalizaciones se conservan cuando se restringe la función a un conjunto más pequeño. Por ejemplo, si f es convexa en C, entonces es convexa en cualquier subconjunto convexo de C. En particular, si f es convexa en C, entonces, para todo $x \in C$ y para todo $d \in \mathbb{R}^n$, $d \neq 0$, la restricción de f a $R(x, d) \cap C$ también es convexa. El conjunto $R(x, d)$ es la recta que pasa por x y es paralela a d, entonces la restricción de f a $R(x, d) \cap C$ es una función de una sola variable.

EJERCICIOS

4.1. Determine si $f(x_1, x_2) = x_1^2 + 2x_1x_2 - 10x_1 + 5x_2$ es convexa, o cóncava, o ninguna de las dos. Determine en qué conjunto f es convexa.

4.2. Determine si $f(x_1, x_2) = x_1 e^{-x_1 - x_2}$ es convexa, o cóncava, o ninguna de las dos. Determine en qué conjunto f es convexa.

4.3. Determine si $f(x_1, x_2) = -x_1^2 - 5x_2^2 + 2x_1x_2 + 10x_1 - 10x_2$ es convexa, o cóncava, o ninguna de las dos. Determine en qué conjunto f es convexa.

4.4. Sea $C = \{(x_1, x_2) : |x_i| \leq 1, \forall i\}$. Determine si $f(x_1, x_2) = 2(x_2 - x_1^2)^2$ es convexa en C, o cóncava, o ninguna de las dos. Determine en qué conjunto f es convexa.

4.5. Sea $f(x_1, x_2) = x_1^2 + x_2^3$. ¿Es f convexa? ¿Es f convexa en $(0, 0)$? ¿Es f seudoconvexa? ¿Es f seudoconvexa en $(0, 0)$? Determine en qué conjunto f es convexa.

4.6. Sea $f(x_1, x_2) = x_1^4 + x_2^3$. ¿Es f convexa? ¿Es f convexa en $(0, 0)$? ¿Es f seudoconvexa? ¿Es f seudoconvexa en $(0, 0)$? Determine en qué conjunto f es convexa.

4.7. Sean: C un convexo, $g : C \longrightarrow \mathbb{R}$ convexa. Muestre que $\{x \in C : g(x) \leq 0\}$ es un convexo.

4.8. Sea

$$f(x) = \begin{cases} x^2 & \text{si } x < 0, \\ x^3 & \text{si } x \geq 0. \end{cases}$$

¿Es f continua? ¿Es f diferenciable? ¿Es f doblemente diferenciable? ¿Es f convexa? ¿Es f seudoconvexa?

4.9. Sean: C un convexo, $g : C \longrightarrow \mathbb{R}$ cóncava. Muestre que $\{x \in C : g(x) > 0\}$ es un convexo.

4.10. Sean: C un convexo, $g : C \longrightarrow \mathbb{R}$ cóncava. Muestre que $f(x) = 1/g(x)$ es convexa en $\{x \in C : g(x) > 0\}$.

4.11. Sean: $c \in \mathbb{R}^n$, $f(x) = (c^{\mathrm{T}}x)^2$. Determine si f es convexa. Determine en qué conjunto f es convexa.

4.12. Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ continua. Determine las condiciones que debe cumplir f para que $\{x : f(x) = 0\}$ sea convexo.

4.13. Sean: C convexo, $f : C \longrightarrow \mathbb{R}$. Muestre que f es convexa si y solamente si

$$f(\frac{1}{2}x + \frac{1}{2}y) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) \quad \forall x, y \in C$$

4.14. Sean: f_1, ..., f_m funciones convexas definidas en un convexo C; α_1, ..., $\alpha_m \geq 0$. Muestre que $f(x) = y\alpha_1 f_1(x) + ... + \alpha_m f_m(x)$ es convexa.

4.15. Sean f_1, ..., f_m funciones convexas definidas en un convexo C. Muestre que $f(x) = \max\{f_1(x), ..., f_m(x)\}$ es convexa.

4.16. Sean: $c, d \in \mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}$, $C = \{x : d^{\mathrm{T}}x + \beta > 0\}$,

$$f(x) = \frac{c^{\mathrm{T}}x + \alpha}{d^{\mathrm{T}}x + \beta}.$$

¿Es f seudoconvexa en C? ¿Es f seudocóncava en C?

Capítulo 5

CONDICIONES DE OPTIMALIDAD EN PUNTOS INTERIORES

Sean: $\mathcal{B}$ un subconjunto de $\mathbb{R}^n$, $f : \mathcal{B} \longrightarrow \mathbb{R}$, $\mathcal{A}$ un conjunto abierto contenido en $\mathcal{B}$. En este capítulo se estudiarán condiciones de optimalidad para el problema de minimizar $f(x)$ cuando x varía en $\mathcal{A}$. Este problema de minimización se denotará simplemente

$$\begin{aligned} &\min f(x) \\ &x \in \mathcal{A}. \end{aligned} \tag{PM}$$

Si $\mathcal{A}$ no es abierto, las condiciones de optimalidad, de este capítulo, se aplican a los puntos interiores de $\mathcal{A}$. En este caso, si f y $\mathcal{A}$ cumplen ciertas propiedades adicionales, aunque el estudio se haga en el interior de $\mathcal{A}$ los resultados pueden ser válidos en todo $\mathcal{A}$.

Con frecuencia $\mathcal{A} = \mathcal{B} = \mathbb{R}^n$ y así se tiene un problema de minimización sin restricciones (irrestricto), que se denotará simplemente por

$$\min f(x),$$

sobreentendiendo que el conjunto $\mathcal{A}$ es todo $\mathbb{R}^n$.

Definición 5.1. Un punto $x^* \in \mathcal{A}$ se llama **minimizador global** o absoluto (o también punto de mínimo global) y $f^* = f(x^*)$ se llama **mínimo global** o absoluto (o valor mínimo global) de PM, si

$$f^* = f(x^*) \leq f(x) \qquad \forall x \in \mathcal{A}.$$

Un punto $x^* \in \mathcal{A}$ se llama **minimizador global estricto** y $f^* = f(x^*)$ se llama **mínimo global estricto** de PM, si

$$f^* = f(x^*) < f(x) \qquad \forall x \in \mathcal{A},\ x \neq x^*.$$

Un punto $x^* \in \mathcal{A}$ se llama **minimizador local** o relativo y $f^* = f(x^*)$ se llama **mínimo local** o relativo de PM, si existe $r > 0$ tal que

$$f^* = f(x^*) \leq f(x) \qquad \forall x \in \mathcal{A} \cap B(x^*, r).$$

Un punto $x^* \in \mathcal{A}$ se llama **minimizador local estricto** y $f^* = f(x^*)$ se llama **mínimo local estricto** de PM, si existe $r > 0$ tal que

$$f^* = f(x^*) < f(x) \qquad \forall x \in \mathcal{A} \cap B(x^*, r), x \neq x^*.$$

Un punto $x^* \in \mathcal{A}$ se llama **minimizador global aislado** si existe una vecindad de x^* donde no hay otros minimizadores globales.

Un punto $x^* \in \mathcal{A}$ se llama **minimizador local aislado** si existe una vecindad de x^* donde no hay otros minimizadores locales.

Obsérvese que todo minimizador global también es minimizador local. Para minimizadores globales, estricto es equivalente a único, y cualquiera de estas dos características implica aislamiento. Es posible que no haya minimizadores locales o que no haya minimizadores globales.

Ejemplo 5.1. Sea $f(x) = \cos x$ en el intervalo $[-20, 20]$. El punto $x = \pi$ es minimizador local estricto y aislado, también es minimizador global aislado. El valor -1 es un mínimo local y global. $\diamond$

Ejemplo 5.2. Sea $f(x) = x^2$ en todos los reales. El punto $x = 0$ es minimizador local y global estricto y aislado. El valor 0 es un mínimo local y global. $\diamond$

Ejemplo 5.3. Sea $f(x) = x^2$ en el intervalo $[3, 9]$. El punto $x = 3$ es minimizador local y global estricto y aislado. $\diamond$

Ejemplo 5.4. Sea $f(x) = \lfloor x \rfloor$ (parte entera, también llamada parte entera inferior). El conjunto de minimizadores locales es $\mathbb{R} \setminus \mathbb{Z}$ (el conjunto de todos los reales no enteros). $\diamond$

Ejemplo 5.5. Sea $f(x) = \lceil x \rceil$ (parte entera superior). Todos los puntos son minimizadores locales. $\diamond$

Ejemplo 5.6. Sea $f(x) = x^2$ en el intervalo $]3, 9]$. Para este problema no hay minimizadores locales ni globales. $\diamond$

Ejemplo 5.7. Sea $f(x) = x^3$ en toda la recta real. Para este problema no hay minimizadores locales ni globales. $\diamond$

Ejemplo 5.8. Sea

$$f(x) = \begin{cases} x^2\big(1 + x^2 + \sin(1/x)\big) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0. \end{cases}$$

El punto $x = 0$ es un minimizador local y global estricto, pero no es minimizador local aislado. $\diamond$

Si el problema planteado es de maximización, basta con multiplicar por menos uno a la función f y así un minimizador de un problema es maximizador del otro. Obviamente el valor mínimo global de un problema corresponde al valor máximo global del otro problema multiplicado por menos uno.

$$\begin{aligned} &\min f(x) \\ &x \in \mathcal{A}. \end{aligned} \tag{PM}$$

$$\begin{aligned} &\max g(x) = -f(x) \\ &x \in \mathcal{A}. \end{aligned} \tag{PMAX}$$

Sean:

$$\begin{array}{rcl} \mathcal{A}^* & : & \text{conjunto de minimizadores globales del PM,} \\ f^* & : & \text{mínimo global del PM,} \\ \overset{*}{\mathcal{A}} & : & \text{conjunto de maximizadores globales del PMAX,} \\ g^* & : & \text{máximo global del PMAX.} \end{array}$$

Entonces

$$\mathcal{A}^* \;=\; \overset{*}{\mathcal{A}},$$

$$f^* \;=\; -g^*.$$

Los resultados anteriores también son ciertos si se cambia la palabra global por la palabra local. El siguiente enunciado es simplemente una manera de decir algo obvio con otras palabras.

Proposición 5.1. *El PM tiene minimizador global si y solamente si el conjunto de imágenes $f(\mathcal{A})$ tiene mínimo.*

La "proposición" anterior, utilizada junto con condiciones suficientes para que $f(\mathcal{A})$ tenga mínimo, permite garantizar la existencia de un minimizador global para el PM.

Proposición 5.2. *Las siguientes son algunas de las condiciones suficientes para que exista minimizador global:*

- *$f(\mathcal{A})$ es cerrado y acotado.*
- *$f(\mathcal{A})$ es cerrado y acotado inferiormente.*
- *f es continua, $\mathcal{A}$ es cerrado y acotado.*

Ejemplo 5.9. Sea $f(x_1, x_2) = \operatorname{sen} x_1 + \cos x_2$ en todo $\mathbb{R}^2$. El conjunto imagen es simplemente $f(\mathcal{A}) = [-2, 2]$ cerrado y acotado, luego existe por lo menos un minimizador global. $\diamond$

Ejemplo 5.10. Sea $f(x_1, x_2) = x_1^2 + x_2^3$ en $\mathcal{A} = \{x : x \geq 0\}$. El conjunto imagen es simplemente $f(\mathcal{A}) = [0, \infty[$ cerrado y acotado inferiormente, luego existe por lo menos un minimizador global. $\diamond$

Ejemplo 5.11. Sean: $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 = 2, x_1 \geq 0, x_2 \geq 0\}$, $f(x) = x_1^3 + x_2^2$. Como $\mathcal{A}$ es cerrado y acotado y f es continua, entonces existe un minimizador global de f en $\mathcal{A}$. $\diamond$

Ejemplo 5.12. Sea $f(x) = 1/(1 + x^2)$ en todo $\mathbb{R}$. Esta función es continua en un cerrado, $f(\mathcal{A})$ es acotado inferiormente ya que $f(x) \geq 0$ para todo x, sin embargo, no existe minimizador. $\diamond$

Definición 5.2. Una función f definida en S subconjunto no acotado de $\mathbb{R}^n$ es **coercitiva** si

$$\lim_{||x|| \to \infty} f(x) = +\infty.$$

Esto quiere decir que dado cualquier $M > 0$ existe $R > 0$ tal que si $x \in S$ y $||x|| > R$, entonces $f(x) > M$.

Ejemplo 5.13. $S = \mathbb{R}^n, f(x) = ||x||$ es coercitiva.
$S = \mathbb{R}^n, f(x) = ||x||^2$ es coercitiva.
$S = \mathbb{R}^n, f(x) = ||x - \bar{x}||^2$ (para un $\bar{x}$ fijo) es coercitiva.
$S = \mathbb{R}^n_+, f(x) = ||x - \bar{x}||^2$ (para un $\bar{x}$ fijo) es coercitiva.
$S = \mathbb{R}^2, f(x) = e^{x_1^2} + e^{x_2^2} - x_1^{100} - x_2^{100}$ es coercitiva.
$S = \mathbb{R}^2, f(x) = a_1x_1 + a_2x_2 + a_3$ no es coercitiva. $\diamond$

Ejemplo 5.14. Sean: $S = \mathbb{R}^2$, $f(x) = (x_1 - x_2)^2$. f no es coercitiva, aunque es coercitiva con respecto a x_1 y con respecto a x_2. $\diamond$

Proposición 5.3. *Sea $\mathcal{A}$ cerrado y no acotado. Si $f : \mathcal{A} \longrightarrow \mathbb{R}$ es continua y coercitiva, entonces existe x^* minimizador global de f en $\mathcal{A}$.*

Ejemplo 5.15. Sean: $\mathcal{A} = \mathbb{R}^2_+$, $f(x_1, x_2) = (x_1 - 4)^3 + (x_2 - 5)^2$. Como $\mathcal{A}$ es cerrado y f es continua y coercitiva, entonces existe por lo menos un minimizador global. $\diamond$

Ejemplo 5.16. Sean: $\mathcal{A}$ el interior de $\mathbb{R}^2_+$, $f(x_1, x_2) = (x_1 - 4)^2 + (x_2 - 5)^2$. Como $\mathcal{A}$ no es cerrado no se puede aplicar la proposición anterior. Sin embargo, sí existe un minimizador global (el punto $x = (4, 5)$). $\diamond$

De la definición se deduce fácilmente que, si una función es coercitiva en un conjunto no acotado, entonces es coercitiva en cada uno de sus subconjuntos no acotados. En particular si una función es coercitiva en $\mathbb{R}^n$, entonces es coercitiva en cualquier conjunto no acotado.

Proposición 5.4. *Sean: $\bar{x}$ un punto interior de $\mathcal{A}$, f diferenciable en $\bar{x}$. Si $\bar{x}$ es un minimizador local del PM, entonces $f'(\bar{x}) = 0$.*

Los puntos donde se anula el gradiente se llaman **puntos críticos**.

Ejemplo 5.17. Sean: $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = (x_1 + x_2 - 5)^2 + (3x_1 - 2x_2)^2$. Entonces f es continua, $\mathcal{A}$ es cerrado y f es coercitiva, luego existe por lo menos un minimizador global. Al obtener el gradiente $f'(x) = (20x_1 - 10x_2 - 10, -10x_1 + 10x_2 - 10)$ se deduce que el único punto que cumple la condición necesaria de anular el gradiente es $x = (2, 3)$, luego necesariamente es el minimizador global ya que todos los puntos son interiores. $\diamond$

Ejemplo 5.18. Sea $f(x) = x^3$ en el intervalo $[2, \infty[$. Como $\mathcal{A}$ es cerrado, f es continua y coercitiva, entonces existe un minimizador global. Si el minimizador global x^* fuera punto interior, entonces $f'(x^*) = 0$, pero esto no es posible, luego el minimizador tiene que ser el único punto no interior, es decir, $x^* = 2$. $\diamond$

Ejemplo 5.19. Sean: $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = x_1^2 + x_2^3$. En este caso no se puede garantizar que $f(\mathcal{A})$ tenga mínimo. Como $f'(x) = (2x_1, 3x_2^2)$, se deduce que el único punto que cumple la condición necesaria de anular el gradiente es $\bar{x} = (0, 0)$, pero no se puede afirmar que es minimizador local o minimizador global. Más aún, $\bar{x} = (0, 0)$ no es minimizador local -y tampoco global- pues los puntos de la forma $(0, -\varepsilon)$, con $\varepsilon > 0$ y pequeño, son mejores que $\bar{x}$ y están muy cerca de $\bar{x}$. $\diamond$

Ejemplo 5.20. Sean: $\mathcal{A} = \{(x_1, x_2) \colon x_1 + x_2 \geq 1,\ x_1 \geq 0,\ x_2 \geq 0\}$, $f(x_1, x_2) = x_1^2 + x_2^3$. Entonces f es continua, $\mathcal{A}$ es cerrado y $f(\mathcal{A})$ es continua y coercitiva, luego existe por lo menos un minimizador global. Al obtener el gradiente $f'(x) = (2x_1, 3x_2^2)$ se deduce que no hay puntos interiores que cumplan la condición necesaria de anular el gradiente, luego el minimizador global debe ser un punto de $\mathcal{A}$ que esté en la frontera. $\diamond$

Ejemplo 5.21. $f(x_1, x_2) = x_1^2 - x_2^2$ en todo $\mathbb{R}^2$, $\bar{x} = (0, 0)$ es el único candidato a ser minimizador local. No es minimizador global ya que, por ejemplo, $x = (0, 1)$ es mejor. Tampoco es minimizador local pues , si ε es un número positivo pequeño, los puntos $(0, \varepsilon)$ son puntos vecinos y mejores que $\bar{x}$. $\diamond$

Proposición 5.5. *Sean: $\bar{x}$ un punto interior de $\mathcal{A}$, f doblemente diferenciable en $\bar{x}$. Si $\bar{x}$ es un minimizador local del PM, entonces $f'(\bar{x}) = 0$ y $f''(\bar{x})$ es semidefinida positiva.*

Ejemplo 5.22. Sean: $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = x_1^2 - x_2^2$. Al calcular el gradiente y la matriz hessiana, se obtiene

$$f'(x) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}, \qquad f''(x) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Se deduce que el único punto que cumple la condición necesaria de anular el gradiente es $x = (0, 0)$, sin embargo, la matriz hessiana en este punto no es semidefinida positiva, luego $x = (0, 0)$ no es minimizador local, luego no existe minimizador local ni global. $\diamond$

Ejemplo 5.23. Sean: $\mathcal{A} = \{(x_1, x_2) \colon x_1 + x_2 > -4\}$, $f(x_1, x_2) = (x_1 + x_2 - 5)^2 + (3x_1 - 2x_2)^2$. Al calcular el gradiente y la matriz hessiana se obtiene

$$f'(x) = \begin{bmatrix} 20x_1\text{-}10x_2\text{-}10 \\ -10x_1 + 10x_2\text{-}10 \end{bmatrix}, \qquad f''(x) = \begin{bmatrix} 20 & \text{-}10 \\ \text{-}10 & 10 \end{bmatrix}.$$

Se ve que en $x = (2, 3)$ se anula el gradiente y la matriz hessiana es semidefinida positiva, luego $x = (2, 3)$ es candidato a ser minimizador local. $\diamond$

Proposición 5.6. *Sean: $\bar{x}$ un punto interior de $\mathcal{A}$, f doblemente diferenciable en $\bar{x}$. Si $f'(\bar{x}) = 0$ y además $f''(\bar{x})$ es definida positiva, entonces $\bar{x}$ es un minimizador local del PM.*

Ejemplo 5.24. Sean: $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 > -4\}$, $f(x_1, x_2) = (x_1 + x_2 - 5)^2 + (3x_1 - 2x_2)^2$. Considerando el ejemplo anterior se observa que en $\bar{x} = (2, 3)$ se anula el gradiente y la matriz hessiana no sólo es semidefinida positiva sino que también es definida positiva, por lo tanto se concluye que $x = (2, 3)$ es minimizador local. $\diamond$

Ejemplo 5.25. Sean: $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = x_1^4 + x_2^2$. Al calcular el gradiente y la matriz hessiana se obtiene

$$f'(x) = \begin{bmatrix} 4x_1^3 \\ 2x_2 \end{bmatrix}, \qquad f''(x) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Se observa que en $x = (0, 0)$ se anula el gradiente y la matriz hessiana es semidefinida positiva, luego es candidato a minimizador local. Como la matriz hessiana no es definida positiva en $x = (0, 0)$, entonces la proposición anterior no permite garantizar que sea un minimizador local. Sin embargo, sí lo es, ya que es el único punto donde la función vale cero. Más aún, es minimizador global. $\diamond$

Proposición 5.7. *Sean: $\bar{x}$ un punto interior de $\mathcal{A}$, f doblemente diferenciable en $\bar{x}$. Si $f'(\bar{x}) = 0$ y existe $r > 0$ tal que $f''(x)$ es semidefinida positiva para todo $x \in B(\bar{x}, r)$, entonces $\bar{x}$ es un minimizador local del PM.*

Ejemplo 5.26. Sean: $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = x_1^4 + x_2^2$. Al calcular el gradiente y la matriz hessiana se obtiene

$$f'(x) = \begin{bmatrix} 4x_1^3 \\ 2x_2 \end{bmatrix}, \qquad f''(x) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Se observa que en $x = (0, 0)$ se anula el gradiente y para cualquier $r > 0$ la matriz hessiana es semidefinida positiva en $B((0, 0), r)$, luego $x = (0, 0)$ es minimizador local del PM. $\diamond$

Proposición 5.8. *Sean: $\mathcal{A}$ un conjunto convexo, abierto, $f : \mathcal{A} \longrightarrow \mathbb{R}$ convexa y diferenciable o simplemente seudoconvexa. Un punto $\bar{x}$ elemento de $\mathcal{A}$ es minimizador global del PM si y solamente si $f'(\bar{x}) = 0$.*

Ejemplo 5.27. $\mathcal{A} = \mathbb{R}^2$, $f(x_1, x_2) = x_1^2 + x_2^2$ es diferenciable y se vio que es convexa, luego $\bar{x}$ es minimizador global si y solamente si

$$f'(\bar{x}) = \begin{bmatrix} 2\bar{x}_1 \\ 2\bar{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Luego el único minimizador global es $\bar{x} = (0, 0)$. $\diamond$

Proposición 5.9. *Sean: $\mathcal{A}$ un conjunto convexo y abierto, $f : \mathcal{A} \longrightarrow \mathbb{R}$ seudoconvexa. Si $\bar{x}$ es minimizador local, entonces es minimizador global.*

Proposición 5.10. *Sean: C convexo, $f : C \longrightarrow \mathbb{R}$ convexa o estrictamente cuasiconvexa. Si x^* es un minimizador local del* PM*, entonces x^* es un minimizador global del PM.*

Las dos últimas proposiciones se pueden resumir de la siguiente forma. Si f es seudoconvexa (en un convexo abierto) o convexa o estrictamente cuasiconvexa (en un convexo cualquiera), entonces todo minimizador local es minimizador global.

Proposición 5.11. *Sea C convexo, $f : C \longrightarrow \mathbb{R}$ fuertemente cuasiconvexa (o estrictamente convexa). Si x^* es un minimizador local del PM, entonces x^* es el único minimizador global del PM.*

Ejemplo 5.28. Sea $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 > -4\}$, $f(x_1, x_2) = (x_1 + x_2 - 5)^2 + (3x_1 - 2x_2)^2$. El punto $x = (2, 3)$ es minimizador local y como f es convexa, entonces es minimizador global. Además, como f es estrictamente convexa, entonces $x = (2, 3)$ es el único minimizador global. $\diamond$

Ejemplo 5.29. Sea $\mathcal{A} = \mathbb{R}$, $f(x) = [x]$. El punto $x = 3/2$ es minimizador local, f es cuasiconvexa, sin embargo, $x = 3/2$ no es minimizador global.

Una propiedad importante de las funciones definidas sobre un intervalo cerrado y acotado, cuando son estrictamente cuasiconvexas o cuando son unimodales, es que si se calcula la función en dos puntos interiores se puede construir un intervalo más pequeño donde tambien está el minimizador. $\diamond$

Proposición 5.12. *Sean $\varphi : [a, b] \longrightarrow \mathbb{R}$ unimodal o estrictamente cuasiconvexa, σ, τ tales que $a < \sigma < \tau < b$, λ^* minimizador de φ en $[a, b]$. Entonces*

$$\begin{aligned} \varphi(\sigma) &\leq \varphi(\tau) \quad \Rightarrow \quad \lambda^* \in [a, \tau], \\ \varphi(\sigma) &\geq \varphi(\tau) \quad \Rightarrow \quad \lambda^* \in [\sigma, b]. \end{aligned}$$

Como la unimodalidad implica la cuasiconvexidad estricta, entonces en la proposición anterior se puede quitar la hipótesis alterna de unimodalidad.

Ejemplo 5.30. Sea $f(x) = x^2$ definida en $[-1, 5]$. Como $f(1) < f(3)$, entonces el minimizador de f en $[-1, 5]$ está en $[-1, 3]$. $\diamond$

EJERCICIOS

En los ejercicios 5.1 a 5.10 estudie el problema propuesto, use condiciones necesarias, suficientes y otros argumentos. Encuentre, si es posible, un punto crítico, los puntos críticos, un minimizador, los minimizadores. ¿Son estos puntos minimizadores globales?

5.1. Minimizar $f(x_1, x_2, x_3) = x_2^2 + 4x_2x_3 + 3x_3^2 - 6x_2 - 10x_3 + 100$ en $\mathbb{R}^3$.

5.2. Minimizar $f(x_1, x_2) = \operatorname{sen}(x_1x_2)$ en $\mathbb{R}^2$.

5.3. Minimizar $f(x_1, x_2) = (x_1 - 1)^2 + e^{x_1} + x_2^2 - 2x_2 + 1$ en $\mathbb{R}^2$.

5.4. Minimizar $f(x_1, x_2) = x_1^3 + 3x_1^2 + x_2^2 + 3x_1 - 2x_2 + 2$ en $\mathbb{R}^2$.

5.5. Minimizar $f(x_1, x_2) = x_1^4 + 4x_1^3 + 6x_1^2 + x_2^2 + 4x_1 - 2x_2 + 2$ en $\mathbb{R}^2$.

5.6. Minimizar $f(x_1, x_2) = 11x_1^2 + 11x_2^2 - 18x_1x_2 + 4x_1 + 4x_2 + 142$ en $\mathbb{R}^2$.

5.7. Minimizar $f(x_1, x_2) = (x_1 + x_2)/(x_1^2 + x_2^2 + 1)$ en $\mathbb{R}^2$.

5.8. Minimizar $f(x_1, x_2) = (x_1^2 - x_2)^2$ en $\mathbb{R}^2$.

5.9. Minimizar $f(x_1, x_2, x_3) = 2x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_2x_3 - 6x_1 - 7x_2 - 8x_3 + 20$ en $\mathbb{R}^3$.

5.10. Minimizar $f(x_1, x_2) = x_1^2 + 2x_1x_2 - 10x_1 + 5x_2 - 8x_3 + 20$ en $\mathbb{R}^2$.

5.11. Sea x^* minimizador global de f en A. Sea B subconjunto de A. Dé condiciones sobre B (por ejemplo, suficientes), para que exista minimizador global de f en B.

Capítulo 6

CONDICIONES DE KARUSH-KUHN-TUCKER

Considérese inicialmente el mismo problema de minimización visto hasta ahora:

$$\begin{aligned} &\min f(x) \\ &x \in \mathcal{A}. \end{aligned} \tag{PM}$$

donde $\mathcal{A}$ es un conjunto cualquiera, no necesariamente abierto. Los resultados presentados en este capítulo permiten estudiar las condiciones de optimalidad en puntos cualesquiera (interiores o no interiores).

6.1 GENERALIDADES

Definición 6.1. Sean: $d \in \mathbb{R}^n$, $d \neq 0$, $\bar{x} \in \mathcal{A}$. d es **dirección** (local) **de descenso** de f en el punto $\bar{x}$, si existe $\varepsilon > 0$ tal que

$$f(\bar{x} + \lambda d) < f(\bar{x}) \qquad \text{para todo } \lambda \in]0, \varepsilon[.$$

Definición 6.2. Sean: $d \in 0\mathbb{R}^n$, $d \neq 0$, $\bar{x} \in \mathcal{A}$. d es **dirección** (local) **de ascenso** de f en el punto $\bar{x}$, si existe $\varepsilon > 0$ tal que

$$f(\bar{x} + \lambda d) > f(\bar{x}) \qquad \text{para todo } \lambda \in]0, \varepsilon[.$$

Ejemplo 6.1. $f(x_1, x_2) = x_1^2 + x_2^2$ en $\mathbb{R}^2$, $\bar{x} = (3, 4)$. En este ejemplo sencillo minimizar f es equivalente a minimizar la distancia de un punto al

origen. La distancia del punto $\bar{x}$ al origen es 5 y una dirección de descenso es aquella que permite, a partir de $\bar{x}$, acercarse al origen. El punto $\bar{x}$ está en la circunferencia con centro en el origen y radio 5, entonces las direcciones de descenso son las que permiten entrar al círculo a partir de $\bar{x}$, o sea, $d = (d_1, d_2)$ es dirección de descenso si $3d_1 + 4d_2 < 0$. Obsérvese que las direcciones tangentes a la circunferencia en $\bar{x}$, es decir, $3d_1 + 4d_2 = 0$ con $d \neq 0$, son direcciones de ascenso. Las otras direcciones de ascenso deben cumplir con $3d_1 + 4d_2 > 0$. $\diamond$

Proposición 6.1. *Sean: $d \in \mathbb{R}^n$, $d \neq 0$, $\bar{x} \in \mathbb{R}^n$, f diferenciable en $\bar{x}$.*

- *si $f'(\bar{x})^T d < 0$, entonces d es dirección de descenso.*
- *si $f'(\bar{x})^T d > 0$, entonces d es dirección de ascenso.*
- *si $f'(\bar{x})^T d = 0$, entonces d puede ser dirección de descenso, o de ascenso o ninguna de las dos.*

Además si $f'(\bar{x}) \neq 0$ y f es convexa, entonces $d \neq 0$ es dirección de descenso si y solamente si $f'(\bar{x})^T d < 0$

Utilicemos la siguiente notación:

$$
\begin{array}{rcl}
D_d & : & \text{conjunto de direcciones de descenso de } f \text{ en } \bar{x}, \\
D_a & : & \text{conjunto de direcciones de ascenso de } f \text{ en } \bar{x}, \\
F_- & : & \text{conjunto de direcciones tales que } f'(\bar{x})^{\mathrm{T}} d < 0, \\
F_+ & : & \text{conjunto de direcciones tales que } f'(\bar{x})^{\mathrm{T}} d > 0, \\
\bar{F} & : & \text{conjunto de direcciones no nulas tales que } f'(\bar{x})^{\mathrm{T}} d = 0, \\
\bar{F}_- & : & \text{conjunto de direcciones no nulas tales que } f'(\bar{x})^{\mathrm{T}} d \leq 0, \\
 & = & F_- \cup \bar{F}, \\
\bar{F}_+ & : & \text{conjunto de direcciones no nulas tales que } f'(\bar{x})^{\mathrm{T}} d \geq 0, \\
 & = & F_+ \cup \bar{F}.
\end{array}
$$

Entonces la proposición anterior se puede expresar así:

$$
\begin{aligned}
F_- &\subseteq D_d \subseteq \bar{F}_-, \\
F_+ &\subseteq D_a \subseteq \bar{F}_+.
\end{aligned}
$$

Si f es convexa y $f'(\bar{x}) \neq 0$, entonces

$$\begin{aligned} F_- &= D_d, \\ F_+ &= D_a. \end{aligned}$$

Ejemplo 6.2. $f(x_1, x_2) = x_1^2 + x_2^2$ en $\mathbb{R}^2$, $\bar{x} = (3, 4)$. Utilizando la notación anterior, el ejemplo 6.1 se resume así:

$$\begin{aligned} D_d &= \{(d_1, d_2) : 3d_1 + 4d_2 < 0\}, \\ D_a &= \{(d_1, d_2) : 3d_1 + 4d_2 \geq 0, d \neq 0\}. \end{aligned}$$

Utilizando el valor $f'(\bar{x})^{\mathrm{T}} d$

$$\begin{aligned} F_- &= \{(d_1, d_2) : 6d_1 + 8d_2 < 0\}, \\ F_+ &= \{(d_1, d_2) : 6d_1 + 8d_2 > 0\}, \\ \bar{F}_+ &= \{(d_1, d_2) : 6d_1 + 8d_2 \geq 0, d \neq 0\}. \end{aligned}$$

En este caso

$$\begin{aligned} D_d &= F_- \ , \\ D_a &= \bar{F}_+ \ . \ \diamond \end{aligned}$$

Ejemplo 6.3. $f(x_1, x_2) = x_1^3 + (x_2 - 3)^4$ en $\mathbb{R}^2$, $\bar{x} = (0, 2)$.

$$\begin{aligned} F_- &= \{(d_1, d_2) : 0d_1 - 4d_2 < 0\}, \\ F_- &= \{(d_1, d_2) : d_2 > 0\}, \\ \bar{F} &= \{(d_1, d_2) : d_2 = 0, d_1 \neq 0\}, \\ F_+ &= \{(d_1, d_2) : d_2 < 0\}. \end{aligned}$$

En este ejemplo, en $\bar{F}$ hay direcciones de ascenso y de descenso:

$$\begin{aligned} d^1 &= (-1, 0) \in \bar{F} \quad \text{es dirección de descenso}, \\ d^2 &= (1, 0) \in \bar{F} \quad \text{es dirección de ascenso.} \ \diamond \end{aligned}$$

Definición 6.3. Sean: $d \in \mathbb{R}^n$, $d \neq 0$, $\bar{x} \in \mathcal{A}$. d es **dirección realizable** o **admisible** (local) o **factible** en el punto $\bar{x}$ con respecto al conjunto $\mathcal{A}$, si existe $\varepsilon > 0$ tal que

$$\bar{x} + \lambda d \in \mathcal{A} \qquad \text{para todo } \lambda \in]0, \varepsilon[.$$

Sea D_r el conjunto de direcciones realizables o admisibles en el punto $\bar{x}$ con respecto al conjunto $\mathcal{A}$. Los tres resultados siguientes son consecuencias directas de las definiciones.

Proposición 6.2. *Si $\bar{x}$ es un punto interior de $\mathcal{A}$, entonces $D_r = \mathbb{R}^n \setminus \{\mathbf{0}\}$. O sea, todo vector no nulo es dirección realizable.*

Proposición 6.3. *Si $\bar{x}$ es un minimizador local de PM, entonces*

$$D_r \cap D_d = \emptyset.$$

Proposición 6.4. *Sean: $\bar{x} \in \mathcal{A}$, f diferenciable en $\bar{x}$. Si $\bar{x}$ es un minimizador local del PM, entonces*

$$D_r \cap F_- \quad = \quad \emptyset.$$

La aplicación de este último resultado se facilita en algunos problemas de dos variables dibujando la región admisible, siempre y cuando la función objetivo tenga una interpretación sencilla. Para problemas de tres o más variables la interpretación geométrica es casi siempre muy difícil o imposible. Para problemas donde la región admisible está definida por desigualdades se verá, más adelante, un criterio para caracterizar la mayoría de las direcciones realizables mediante los gradientes de algunas desigualdades.

Proposición 6.5. *Estos conjuntos de direcciones D_d , D_r , F_- , F_+ ,... son conos.*

Ejemplo 6.4. Minimizar $f(x) = x_1^2 + x_2^2$ en $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 \geq 2, x \geq 0\}$. Considérese el punto $\bar{x} = (2, 0)$.

Siempre la primera comprobación que se debe hacer es la factibilidad del punto, en este caso, como $\bar{x}$ cumple todas las restricciones, entonces está en $\mathcal{A}$. Al dibujar el conjunto $\mathcal{A}$ se observa que éste es simplemente el primer cuadrante quitándole un pedazo de punta de forma triángular. El punto $\bar{x}$ es uno de los puntos extremos de $\mathcal{A}$. Las direcciones realizables son las comprendidas entre la dirección oriente inclusive y la dirección noroccidental inclusive,

$$D_r = \{(d_1, d_2) : d_2 \geq 0, d_1 + d_2 \geq 0, d \neq 0\}.$$

Las direcciones de descenso son simplemente aquellas que van, por lo menos parcialmente, hacia la izquierda,

$$D_d = \{(d_1, d_2) : d_1 < 0\}.$$

Como los dos conjuntos no son disyuntos, por ejemplo $d = (-1, 2)$ está en ambos, entonces $\bar{x} = (2, 0)$ no es minimizador local. $\diamond$

Ejemplo 6.5. Minimizar $f(x) = x_1^2 + x_2^2$ en $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 \geq 2, x \geq 0\}$. Considérese el punto $\bar{x} = (2, 2)$. Este punto cumple todas las restricciones, entonces es factible. Como es punto interior

$$\begin{aligned} D_r &= \mathbb{R}^n \setminus \{\mathbf{0}\}. \\ \text{Además} \quad F_- &= \{(d_1, d_2) : d_1 + d_2 < 0\}. \end{aligned}$$

Como los dos conjuntos no son disyuntos, entonces $\bar{x} = (2, 2)$ no es minimizador local. $\diamond$

Ejemplo 6.6. Minimizar $f(x) = x_1^2 + x_2^2$ en $\mathcal{A} = \{(x_1, x_2) : x_1 + x_2 \geq 2, x \geq 0\}$. Considérese el punto $\bar{x} = (1, 1)$. Este punto cumple todas las restricciones, entonces es factible. Las direcciones realizables son las comprendidas entre la dirección suroriental inclusive, pasando por la dirección oriente y la dirección norte, hasta la dirección noroccidental inclusive:

$$\begin{aligned} D_r &= \{(d_1, d_2) : d_1 + d_2 \geq 0, d \neq 0\}. \\ D_d = F_- &= \{(d_1, d_2) : d_1 + d_2 < 0\}. \end{aligned}$$

Como los dos conjuntos son disyuntos, entonces $\bar{x} = (1, 1)$ es buen candidato a minimizador local. Por otro lado, es "claro" que es el punto de $\mathcal{A}$ más cercano al origen. $\diamond$

Ejemplo 6.7. Minimizar $f(x) = \text{-}x_1$ en $\mathcal{A} = \{(x_1, x_2) : x_1^2 + x_2^2 \geq 1, x_1^2 + (x_2 + 1)^2 \leq 4, x_1 \geq 0\}$. Considérese el punto factible $\bar{x} = (0, 1)$. El conjunto admisible tiene forma de cuerno. En la punta $(0, 1)$ hay dos circunferencias tangentes. Razonamientos geométricos permiten afirmar que en este punto no hay direcciones realizables, entonces $D_r = \emptyset$, luego $D_r \cap D_d = \emptyset$. Según lo anterior $\bar{x} = (0, 1)$ es un buen candidato a minimizador local. Minimizar f

corresponde a maximizar x_1, o sea, el minimizador es el punto realizable que tenga más grande la primera coordenada. El minimizador global es (2, -1). Si $\varepsilon > 0$ es pequeño, el punto $(\varepsilon, \sqrt{1-\varepsilon^2})$ es factible, está muy cerca de $(0, 1)$ y además mejora el valor de f. Lo anterior muestra que $\bar{x} = (0, 1)$ no es minimizador local. $\diamond$

La mayoría de los problemas de programación no lineal se plantean como la minimización de una función f en un conjunto definido por desigualdades e igualdades. Para facilitar el estudio de las condiciones de Karush-Kuhn-Tucker, KKT, se presentará primero el problema únicamente con desigualdades y posteriormente con desigualdades e igualdades. Las funciones g_i, h_j están definidas en $\mathbb{R}^n$ y tienen valor real.

$$\begin{aligned} \min\ & f(x) \\ & g_1(x) \leq 0 \\ & g_2(x) \leq 0 \\ & \cdots \\ & g_m(x) \leq 0. \end{aligned} \tag{PMD}$$

$$\begin{aligned} \min\ & f(x) \\ & g_1(x) \leq 0 \\ & g_2(x) \leq 0 \\ & \cdots \\ & g_m(x) \leq 0 \\ & h_1(x) = 0 \\ & h_2(x) = 0 \\ & \cdots \\ & h_l(x) = 0. \end{aligned} \tag{PMDI}$$

En todos los casos $\mathcal{A}$ indica el conjunto de puntos admisibles o realizables, es decir, el conjunto de puntos que cumplen todas las restricciones.

Definición 6.4. Se dice que la desigualdad $g_i(x) \leq 0$ está **activa** o **saturada** en un punto $\bar{x}$ si se cumple exactamente la igualdad, es decir, si $g_i(\bar{x}) = 0$. Se dice que la desigualdad $g_i(x) \leq 0$ está **inactiva** o **no saturada** o **pasiva** en un punto $\bar{x}$ si se cumple estrictamente la desigualdad,

es decir, si $g_i(\bar{x}) < 0$. Sea $\bar{x}$ admisible. Se denotará por $\mathcal{I}$ el conjunto de índices de las desigualdades activas o saturadas:

$$\mathcal{I} = \{i : g_i(\bar{x}) = 0\}.$$

Obsérvese que una desigualdad se cumple o no se cumple en un punto $\bar{x}$. Si se cumple puede estar activa o inactiva.

Si para todo $i \in \mathcal{I}$ las funciones g_i son diferenciables, sea

$$\begin{aligned} G_- &= \{d : g_i'(\bar{x})^{\mathrm{T}} d < 0 \ \forall i \in \mathcal{I}\}, \\ \bar{G}_- &= \{d : g_i'(\bar{x})^{\mathrm{T}} d \leq 0 \ \forall i \in \mathcal{I},\ d \neq 0\}. \end{aligned}$$

Proposición 6.6. *Sea $\bar{x}$ tal que:*
$\bar{x}$ es un punto admisible del PMD,
g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$,
g_i es continua en $\bar{x}$ para $i \notin \mathcal{I}$.
Entonces:

$$G_- \subseteq D_r \subseteq \bar{G}_-.$$

Como corolario de la proposición anterior:

Proposición 6.7. *Sea $\bar{x}$ tal que:*
$\bar{x}$ es un punto admisible del PMD,
f es diferenciable en $\bar{x}$,
g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$,
g_i es continua en $\bar{x}$ para $i \notin \mathcal{I}$.
Si $\bar{x}$ es un minimizador local del PMD, entonces:

$$\begin{aligned} D_d \cap G_- &= \emptyset, \\ F_- \cap D_r &= \emptyset, \\ F_- \cap G_- &= \emptyset. \end{aligned}$$

Ejemplo 6.8. Considere $\bar{x} = (1, 3)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ x_2 &\geq x_1^2 + 2 \\ x_1 &\geq 0 \end{aligned}$$

$$x_2 \geq 5/2.$$

Es necesario plantear el problema en la forma usual:

$$\begin{aligned} \min f(x) &= (x_1-3)^2+(x_2-2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

Ante todo se estudia la factibilidad del punto $\bar{x}$ y se observa que cumple todas las restricciones. Además se deduce que $\mathcal{I} = \{1\}$ ya que la única restricción activa es la primera. También f, g_1 son diferenciables en $\bar{x}$; g_2, g_3 son continuas en $\bar{x}$; $f'(\bar{x}) = [-4 \quad 2]^{\mathrm{T}}$, $g_1'(\bar{x}) = [2 \quad -1]^{\mathrm{T}}$,

$$\begin{aligned} F_- &= \{(d_1,d_2): -4d_1 + 2d_2 < 0\}, \\ G_- &= \{(d_1,d_2): 2d_1 - d_2 < 0\}. \end{aligned}$$

Claramente los dos conjuntos son disyuntos, entones $\bar{x} = (1,3)$ es un buen candidato a minimizador local. Al dibujar la región admisible y teniendo en cuenta que f representa el cuadrado de la distancia de un punto $x = (x_1, x_2)$ a $(3,2)$, se "observa" que el minimizador global es precisamente $\bar{x} = (1,3)$. $\diamond$

Ejemplo 6.9. Considere $\bar{x} = (1,1)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1-3)^2+(x_2-2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ no es admisible. $\diamond$

Ejemplo 6.10. Considere $\bar{x} = (0, 5/2)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1-3)^2+(x_2-2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \end{aligned}$$

$$-x_2 + 5/2 \;\leq\; 0.$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{2, 3\}$; f, g_2, g_3 son diferenciables en $\bar{x}$; g_1 es continua en $\bar{x}$; $f'(\bar{x}) = [\text{-}6 \;\; 1]^{\mathrm{T}}$, $g_2'(\bar{x}) = [\text{-}1 \;\; 0]^{\mathrm{T}}$, $g_3'(\bar{x}) = [0 \;\; \text{-}1]^{\mathrm{T}}$,

$$\begin{aligned} F_- &= \{(d_1, d_2) : \, -6d_1 + d_2 < 0\}, \\ G_- &= \{(d_1, d_2) : d_1 > 0, \; d_2 > 0\}. \end{aligned}$$

Estos dos conjuntos no son disyuntos, $d = (1, 1)$ está en ambos conjuntos, luego $\bar{x} = (0, 5/2)$ no es minimizador local. $\diamond$

Ejemplo 6.11. Considere $\bar{x} = (1, 1)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 5)^2 + x_2^2 \\ x_1 + x_2 - 2 &\leq 0 \\ \text{-}x_1\text{-}x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{1, 2\}$; f, g_1, g_2 son diferenciables en $\bar{x}$; g_3, g_4 son continuas en $\bar{x}$; $f'(\bar{x}) = [\text{-}8 \;\; 2]^{\mathrm{T}}$, $g_1'(\bar{x}) = [1 \;\; 1]^{\mathrm{T}}$, $g_2'(\bar{x}) = [\text{-}1 \;\; \text{-}1]^{\mathrm{T}}$,

$$\begin{aligned} F_- &= \{(d_1, d_2) : \, -8d_1 + 2d_2 < 0\}, \\ G_- &= \{(d_1, d_2) : d_1 + d_2 < 0, \; -d_1 - d_2 < 0\} = \emptyset. \end{aligned}$$

Estos dos conjuntos son disyuntos, luego $\bar{x} = (1, 1)$ es candidato a minimizador local. Sin embargo, no lo es ya que si $\varepsilon > 0$ es pequeño, el punto $(1 + \varepsilon, 1 - \varepsilon)$ es factible, está muy cerca de $(1, 1)$ y además mejora el valor de f. Lo anterior muestra que $\bar{x} = (1, 1)$ no es minimizador local. El minimizador global es el punto $(2, 0)$. $\diamond$

6.2 PROBLEMAS CON DESIGUALDADES

El estudio sistemático de los conjuntos $F_{\text{-}}$, $G_{\text{-}}$ da lugar a las condiciones de Fritz John y a las condiciones de Karush-Kuhn-Tucker. Estas condiciones también son conocidas con el nombre de Kuhn-Tucker únicamente, sin embargo, se da también crédito al trabajo de Karush en una tesis de

maestría de la Universidad de Chicago, en 1939. Las condiciones de KKT, cuando se pueden aplicar, permiten un manejo algorítmico y preciso, aún para problemas con muchas variables.

Definición 6.5. Un punto $\bar{x}$ admisible para el PMD tal que: g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$, se llama **regular** si los gradientes $g_i'(\bar{x})$ para $i \in \mathcal{I}$ son linealmente independientes o si $\mathcal{I} = \emptyset$.

Proposición 6.8. Condiciones necesarias de KKT para el PMD. *Sea $\bar{x}$ tal que:*

$\bar{x}$ es un punto admisible del PMD,
f es diferenciable en $\bar{x}$,
g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$,
g_i es continua en $\bar{x}$ para $i \notin \mathcal{I}$, y
$\bar{x}$ es regular.

Si $\bar{x}$ es un minimizador local del PMD, entonces existen escalares u_i, $i \in \mathcal{I}$, tales que:

$$\begin{aligned} f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) = 0, \\ u_i \geq 0 \ , \quad i \in \mathcal{I}. \end{aligned} \tag{6.1}$$

Ejemplo 6.12. Considere $\bar{x} = (0, 2)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es no es factible, luego no puede ser minimizador. $\diamond$

Ejemplo 6.13. Considere $\bar{x} = (1, 3)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{1\}$; f, g_1 son diferenciables en $\bar{x}$; g_2, g_3 son continuas en $\bar{x}$; el conjunto formado por el gradiente $g_1'(\bar{x}) = [2 \ \ \text{-}1]^{\mathrm{T}}$ es linealmente independiente, luego $\bar{x}$ es regular.

$$f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) = \begin{bmatrix} \text{-}4 \\ 2 \end{bmatrix} + u_1 \begin{bmatrix} 2 \\ \text{-}1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

entonces

$$u_1 = 2 \geq 0.$$

Luego $\bar{x} = (1, 3)$ es un buen candidato a minimizador local. $\diamond$

Ejemplo 6.14. Considere $\bar{x} = (0, 5/2)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible, $\mathcal{I} = \{2, 3\}$, f, g_2, g_3 son diferenciables en $\bar{x}$, g_1 es continua en $\bar{x}$, el conjunto formado por los gradientes $g_2'(\bar{x}) = [\text{-}1 \ \ 0]^{\mathrm{T}}$, $g_3'(\bar{x}) = [0 \ \ \text{-}1]^{\mathrm{T}}$ es linealmente independiente, luego $\bar{x}$ es regular.

$$f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) = \begin{bmatrix} \text{-}6 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} \text{-}1 \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} 0 \\ \text{-}1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

entonces

$$u_2 = -6 \ngeq 0 \ , \quad u_3 = 1.$$

Luego $\bar{x} = (0, 5/2)$ no es minimizador local. $\diamond$

Ejemplo 6.15. Considere $\bar{x} = (1, 1)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= x_1^2 + x_2^2 \\ 2 - x_1 - x_2 &\leq 0 \\ 1 - x_1 &\leq 0 \end{aligned}$$

$$1 - x_2 \quad \leq \quad 0.$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{1, 2, 3\}$; f, g_1, g_2, g_3 son diferenciables en $\bar{x}$, pero el conjunto formado por los tres gradientes no es linealmente independiente, luego no se puede aplicar el teorema. Sin embargo, el punto $\bar{x} = (1, 1)$ sí es el minimizador global. En este ejemplo sencillo se puede ver que al suprimir la primera restricción el conjunto admisible no cambia. O sea, el problema anterior es exactamente equivalente a:

$$\begin{aligned} \min f(x) &= x_1^2 + x_2^2 \\ 1 - x_1 &\leq 0 \\ 1 - x_2 &\leq 0. \end{aligned}$$

El punto $\bar{x} = (1, 1)$ es factible; $\mathcal{I} = \{1, 2\}$, f, g_1, g_2 son diferenciables en $\bar{x}$; el conjunto formado por los gradientes $g_1'(\bar{x}) = [\text{-1} \;\; 0]^{\mathrm{T}}$, $g_2'(\bar{x}) = [0 \;\; \text{-1}]^{\mathrm{T}}$ es linealmente independiente, luego $\bar{x}$ es regular.

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} + u_1 \begin{bmatrix} \text{-1} \\ 0 \end{bmatrix} + u_2 \begin{bmatrix} 0 \\ \text{-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

entonces,

$$u_1 = 2 \geq 0 \ , \quad u_2 = 2 \geq 0.$$

Luego $\bar{x} = (1, 1)$ es candidato a minimizador local. $\diamond$

Ejemplo 6.16. Considere $\bar{x} = (0, 1)$ para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2) &= -x_2 \\ -x_1^2 - x_2^2 + 1 &\leq 0 \\ x_1^2 + (x_2 + 1)^2 - 4 &\leq 0 \\ -x_2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{1, 2\}$; f, g_1, g_2 son diferenciables en $\bar{x}$; g_3 es continua en $\bar{x}$; el conjunto formado por los dos gradientes $g_1'(\bar{x}) = [0 \;\; \text{-2}]^{\mathrm{T}}$, $g_2'(\bar{x}) = [0 \;\; 4]^{\mathrm{T}}$ no es linealmente independiente, luego no se puede aplicar el teorema. Sin embargo, al construir la región admisibles, se "ve" que el punto $\bar{x} = (0, 1)$ sí es el minimizador global. Acá no se puede quitar ninguna de las restricciones pues el conjunto admisible se altera. Habría necesidad de utilizar otros criterios para el estudio de este problema. $\diamond$

En los ejemplos anteriores ha sido relativamente simple saber si un punto $\bar{x}$ cumple o no cumple condiciones de KKT. En un caso general, suponiendo $\bar{x}$ admisible y regular, se trata de resolver un problema de la forma

$$\begin{aligned} Ay &= b \\ y &\geq 0, \end{aligned}$$

donde A es una matriz $n \times \bar{m}$, siendo $\bar{m}$ el número de desigualdades saturadas; $b = -f'(\bar{x})$; y el vector columna $\bar{m} \times 1$ compuesto por las variables $u_i, i \in \mathcal{I}$; las columnas de A son los gradientes de las desigualdades activas evaluados en $\bar{x}$. Como $\bar{x}$ es regular, entonces las columnas de A son linealmente independientes, luego $\bar{m} \leq n$.

El problema anterior se puede resolver de dos maneras, la primera consiste en resolver el siguiente problema de programación lineal

$$\begin{aligned} \min z &= 0^{\mathrm{T}} y \\ Ay &= b \\ y &\geq 0. \end{aligned}$$

Resolver este problema de PL equivale simplemente a encontrar un punto que cumpla las restricciones $Ay = b$, $y \geq 0$. Muy posiblemente es necesario introducir variables artificiales y efectuar únicamente la primera fase del método de las dos fases para saber si hay puntos admisibles de este problema de PL ($\bar{x}$ es punto de KKT) o no hay puntos admisibles ($\bar{x}$ no es minimizador).

La segunda forma de encontrar la solución consiste en resolver el sistema $Ay = b$ y ver si hay solución tal que $y \geq 0$. Más precisamente:

1) Construir la matriz ampliada $\hat{A} = [A \;\; b]$ de tamaño $n \times (\bar{m} + 1)$.

2) Convertirla, mediante operaciones elementales por filas, en una matriz

$$\hat{A}' = \begin{bmatrix} I & c \\ 0 & d \end{bmatrix},$$

donde I es la matriz identidad de orden $\bar{m}$, 0 es la matriz $(n - \bar{m}) \times \bar{m}$ compuesta por ceros, c es un vector columna $\bar{m} \times 1$, d es un vector columna $(n - \bar{m}) \times 1$, es decir, se obtiene el siguiente sistema equivalente

$$\begin{bmatrix} I \\ 0 \end{bmatrix} y = \begin{bmatrix} c \\ d \end{bmatrix},$$

o sea,

$$\begin{aligned} Iy &= c \\ 0y &= d. \end{aligned}$$

El paso 2) siempre es posible puesto que las columnas de A son linealmente independientes.

3) Si $d \neq 0$, el sistema $Ay = b$ no tiene solución y el punto $\bar{x}$ no es minimizador local. Si $d = 0$, el sistema $Ay = b$ tiene como única solución $y = c$.

4) Si $y = c \geq 0$, el punto $\bar{x}$ es punto de KKT. Si $y = c \ngeq 0$, entonces $\bar{x}$ no es minimizador local.

Ejemplo 6.17. Considere $\bar{x} = (1, 1, 1, 1)$ para el siguiente problema:

$$\min f(x) = x_1^2 + 3x_2^2 + x_3^2 + x_4^2$$

$$\begin{aligned} (x_1 - 4)^2 + x_2^2 + x_3^2 + x_4^2 - 15 &\leq 0 \\ 1 - x_2 &\leq 0 \\ 10 - x_1 - 2x_2 - 3x_3 - 4x_4 &\leq 0 \\ 4 - x_1 - x_2 - x_3 - x_4 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{2, 3, 4\}$; f, g_2, g_3, g_4 son diferenciables en $\bar{x}$; g_1 es continua en $\bar{x}$; el conjunto formado por los gradientes

$$\begin{aligned} g_2'(\bar{x}) &= [\ \ 0 \ \ -1 \ \ 0 \ \ 0\ \ ]^{\mathrm{T}}, \\ g_3'(\bar{x}) &= [\ -1 \ \ -2 \ \ -3 \ \ -4\ \ ]^{\mathrm{T}}, \\ g_4'(\bar{x}) &= [\ -1 \ \ -1 \ \ -1 \ \ -1\ \ ]^{\mathrm{T}} \end{aligned}$$

es linealmente independiente. El gradiente de f es $f'(\bar{x}) = [2 \ \ 6 \ \ 2 \ \ 2]^{\mathrm{T}}$. El problema que hay que resolver es el siguiente:

$$\begin{bmatrix} 0 & -1 & -1 \\ -1 & -2 & -1 \\ 0 & -3 & -1 \\ 0 & -4 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -6 \\ -2 \\ -2 \end{bmatrix}$$

$$y \geq 0,$$

donde $y_1 = u_2$, $y_2 = u_3$, $y_3 = u_4$. Si se resuelve por PL,

$$\min z = 0y_1 + 0y_2 + 0y_3$$

$$\begin{bmatrix} 0 & -1 & -1 \\ -1 & -2 & -1 \\ 0 & -3 & -1 \\ 0 & -4 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -6 \\ -2 \\ -2 \end{bmatrix}$$
$$y \geq 0.$$

Cambiando signos para obtener términos independientes no negativos, introduciendo 4 variables artificiales (podría bastar con 3 ya que y_1 sirve como segunda variable básica) y planteando la función objetivo artificial de la primera fase, se tiene

$$\min z_a = 0y_1 + 0y_2 + 0y_3 + y_4 + y_5 + y_6 + y_7$$

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 & 1 & 0 \\ 0 & 4 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_7 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 2 \\ 2 \end{bmatrix}$$
$$y \geq 0.$$

Al final de la primera fase $u_2 = y_1 = 4$, $u_3 = y_2 = 0$, $u_4 = y_3 = 2$, $y_4 = y_5 = y_6 = y_7 = 0$, $z_a^* = 0$, luego hay puntos factibles. Entonces $\bar{x}$ es punto de KKT.

Si se resuelve de la segunda manera, se construye la matriz ampliada

$$\begin{bmatrix} 0 & -1 & -1 & -2 \\ -1 & -2 & -1 & -6 \\ 0 & -3 & -1 & -2 \\ 0 & -4 & -1 & -2 \end{bmatrix},$$

y por medio de operaciones elementales sobre las filas se llega a

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

lo cual indica, primero, que el sistema sí tiene solución, y, segundo, que la solución tiene todas sus componentes no negativas, luego $\bar{x}$ es punto de KKT. $\diamond$

Si las funciones $g_i, i \notin \mathcal{I}$ son diferenciables en $\bar{x}$, y si se considera que los u_i correspondientes son nulos, la admisibilidad y las condiciones necesarias de KKT se pueden escribir así:

$$\begin{aligned} g_i(\bar{x}) &\leq 0\ , \quad i = 1, ..., m && (6.2)\\ f'(\bar{x}) + \sum_{i=1}^{m} u_i g_i'(\bar{x}) &= 0 && (6.3)\\ u_i &\geq 0\ , \quad i = 1, ..., m \\ u_i g_i(\bar{x}) &= 0\ , \quad i = 1, ..., m. && (6.4) \end{aligned}$$

La primera condición que debe cumplir un punto $\bar{x}$ para tratar de ser minimizador, es la de ser admisible, o sea, cumplir todas las restricciones $g_i(\bar{x}) \leq 0$. Estas condiciones (6.2) se conocen también con el nombre de **condiciones de factibilidad principal** o **"primal"** (con frecuencia se utiliza en español técnico la palabra "primal", aunque puede ser un anglicismo). La existencia de $u_i \geq 0$, tales que $f'(\bar{x}) + \sum u_i g_i'(\bar{x}) = 0$, se conoce con el nombre de **condiciones de factibilidad dual** (6.3). Las condiciones (6.4) $u_i g_i(\bar{x}) = 0$ se conocen con el nombre de **condiciones de holgura complementaria**.

Si $g(\bar{x})$ denota el vector columna $[g_1(\bar{x}) \quad g_2(\bar{x}) \quad ... \quad g_m(\bar{x})]^{\mathrm{T}}$, $g'(\bar{x})$ denota la matriz $n \times m$ cuyas columnas son los gradientes $g_1'(\bar{x})$, $g_2'(\bar{x})$,..., $g_m'(\bar{x})$, y teniendo en cuenta que $u^{\mathrm{T}} g(\bar{x})$ es suma de números no positivos, entonces la admisibilidad y las condiciones necesarias de KKT se pueden escribir

$$\begin{aligned} g(\bar{x}) &\leq 0\\ f'(\bar{x}) + g'(\bar{x})u &= 0\\ u &\geq 0\\ u^{\mathrm{T}} g(\bar{x}) &= 0. \end{aligned}$$

Introduciendo la función lagrangiana, o simplemente el lagrangiano, función de $n + m$ variables $x_1, ..., x_n$, $u_1, ..., u_m$

$$L(x,u) = f(x) + \sum_{i=1}^{m} u_i g_i(x) = f(x) + u^{\mathrm{T}} g(x),$$

y denotando por $L'_x(\bar{x},u)$ las componentes del gradiente $L'(\bar{x},u)$ correspondientes a las derivadas parciales de L con respecto a las variables x_j, y de manera análoga $L'_u(\bar{x},u)$, entonces la admisibilidad y las condiciones de KKT se expresan

$$\begin{aligned} L'_u(\bar{x},u) &\leq 0 \\ L'_x(\bar{x},u) &= 0 \\ u &\geq 0 \\ u^{\mathrm{T}} L'_u(\bar{x},u) &= 0. \end{aligned}$$

Los coeficientes u_i se conocen con el nombre de **coeficientes de KKT** o **coeficientes de Lagrange**. Un punto que cumple las condiciones necesarias de KKT se llama un **punto de KKT**.

Proposición 6.9. Condiciones suficientes de KKT para el PMD. *Si el punto $\bar{x}$ cumple las condiciones necesarias de KKT para el PMD y además*

f es seudoconvexa en $\bar{x}$,

g_i es cuasiconvexa y diferenciable en $\bar{x}$ para $i \in \mathcal{I}$,

entonces $\bar{x}$ es un minimizador global del PMD.

Ejemplo 6.18. Considere $\bar{x} = (1,3)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1-3)^2 + (x_2-2)^2 \\ x_1^2 - x_2 + 2 &\leq 0 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0. \end{aligned}$$

El punto $\bar{x}$, como se vio en el ejemplo 6.8, cumple las condiciones necesarias de KKT. Falta por ver si f es seudoconvexa en $\bar{x}$ y si g_1 es diferenciable y cuasiconvexa en $\bar{x}$. La matriz hessiana de f,

$$f''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

es definida positiva para todo x, luego f es convexa. Como es diferenciable, entonces es seudoconvexa y en particular es seudoconvexa en $\bar{x}$. La función g_1 es diferenciable, su matriz hessiana,

$$g_1''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},$$

es semidefinida positiva para todo x, luego g_1 es convexa, entonces es cuasiconvexa y en particular es cuasiconvexa en $\bar{x}$. Luego $\bar{x} = (1, 3)$ es minimizador global. $\diamond$

Ejemplo 6.19. Considere $\bar{x} = (1, 1)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= \cos x_1 + \exp(x_1 + x_2) \\ 2 - x_1 - x_2 &\leq 0 \\ x_1^2 + x_2^2 - 2 &\leq 0. \end{aligned}$$

Obviamente este punto es minimizador global puesto que es el único punto admisible. Sin embargo, no es un punto de KKT (no es regular), luego no cumple condiciones suficientes. $\diamond$

6.3 PROBLEMAS CON DESIGUALDADES E IGUALDADES

Para el PMDI (problema de minimización con desiguladades e igualdades), sea $p = \bar{m} + l$, es decir, p es el número de desigualdades activas más el número de igualdades.

Definición 6.6. Un punto $\bar{x}$ admisible para el PMDI tal que: g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$, h_j es diferenciable en $\bar{x}$ para todo j, se llama **regular** si los gradientes $g_i'(\bar{x})$ con $i \in \mathcal{I}$ y $h_j'(\bar{x})$, $\forall j$ son linealmente independientes, o si $p = 0$, o sea, si $\mathcal{I} = \emptyset$ y $l = 0$.

Proposición 6.10. Condiciones necesarias de Karush-Kuhn-Tucker para el PMDI. *Sea $\bar{x}$ tal que:*

$\bar{x}$ es un punto admisible del PMDI,
f es diferenciable en $\bar{x}$,
g_i es diferenciable en $\bar{x}$ para $i \in \mathcal{I}$,
h_j es diferenciable en $\bar{x}$ para todo j,
g_i es continua en $\bar{x}$ para $i \notin \mathcal{I}$, y

$\bar{x}$ es regular.
Si $\bar{x}$ es un minimizador local del PMDI, entonces existen escalares u_i , $i \in \mathcal{I}$, $v_j \forall j$ tales que:

$$f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) + \sum_{j=1}^{l} v_j h_j'(\bar{x}) = 0, \qquad (6.5)$$
$$u_i \geq 0 \ , \quad i \in \mathcal{I}$$

Ejemplo 6.20. Considere $\bar{x} = (1, 3)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0 \\ -x_1^2 + x_2 - 2 &= 0. \end{aligned}$$

El punto $\bar{x}$ es factible, $\mathcal{I} = \emptyset$, f, h_1 son diferenciables en $\bar{x}$, g_1, g_2 son continuas en $\bar{x}$, el conjunto formado por el gradiente $h_1'(\bar{x}) = [\text{-}2 \ \ 1]^{\mathrm{T}}$ es linealmente independiente, o sea, $\bar{x}$ es regular.

$$f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) + \sum_{j=1}^{l} v_j h_j'(\bar{x}) = \begin{bmatrix} \text{-}4 \\ 2 \end{bmatrix} + v_1 \begin{bmatrix} \text{-}2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

entonces

$$v_1 = -2.$$

Luego $\bar{x} = (1, 3)$ es buen candidato a minimizador local. $\diamond$

Ejemplo 6.21. Considere $\bar{x} = (\sqrt{2}/2, 5/2)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0 \\ -x_1^2 + x_2 - 2 &= 0. \end{aligned}$$

El punto $\bar{x}$ es factible, $\mathcal{I} = \{2\}$, f, g_2, h_1 son diferenciables en $\bar{x}$, g_1 es continua en $\bar{x}$, el conjunto formado por los gradientes $g_2'(\bar{x}) = [0 \ \text{-}1]^{\mathrm{T}}$, $h_1'(\bar{x}) = [\text{-}\sqrt{2} \ \ 1]^{\mathrm{T}}$ es linealmente independiente, o sea, $\bar{x}$ es regular.

$$\begin{bmatrix} \sqrt{2}\text{-}6 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} 0 \\ \text{-}1 \end{bmatrix} + v_1 \begin{bmatrix} \text{-}\sqrt{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

entonces

$$u_2 = 2 - 3\sqrt{2} \not\geq 0\ , \quad v_1 = 1 - 3\sqrt{2}.$$

Luego $\bar{x} = (\sqrt{2}/2, 5/2)$ no es minimizador local. $\diamond$

De manera análoga a los problemas con desigualdades, para saber si $\bar{x}$ admisible es punto de KKT, se debe resolver un problema de la forma

$$\begin{aligned} A \begin{bmatrix} y \\ v \end{bmatrix} &= b \\ y &\geq 0. \end{aligned}$$

donde A es una matriz $n \times p$, siendo $p = \bar{m}+l$, $\bar{m}$ el número de desigualdades saturadas; $b = -f'(\bar{x})$; y es el vector columna $\bar{m} \times 1$ compuesto por las variables $u_i, i \in \mathcal{I}$; v el vector columna $l \times 1$ compuesto por las variables v_j; las columnas de A son los gradientes de las desigualdades activas y de las igualdades evaluados en $\bar{x}$. Como $\bar{x}$ es regular, entonces las columnas de A son linealmente independientes, luego $p = \bar{m} + l \leq n$.

El problema anterior se puede resolver de dos maneras, la primera consiste en resolver el siguiente problema de programación lineal

$$\begin{aligned} \min z = 0^{\mathrm{T}} y + 0^{\mathrm{T}} v & \\ A \begin{bmatrix} y \\ v \end{bmatrix} &= b \\ y &\geq 0 \\ v &\in \mathbb{R}^l. \end{aligned}$$

Si el algoritmo que se va a usar supone que todas las variables son no negativas, entonces es necesario convertir el problema de PL anterior en uno con todas las variables no negativas. Lo usual es descomponer cada variable v_j en la diferencia de dos variables no negativas $v_j = v_j^+ - v_j^-$, con $v_j^+, v_j^- \geq 0$, y efectuar los cambios correspondientes. Además muy posiblemente es necesario introducir variables artificiales y efectuar únicamente la primera

fase del método de las dos fases para saber si hay puntos admisibles de este problema de PL ($\bar{x}$ es punto de KKT) o no hay puntos admisibles ($\bar{x}$ no es minimizador).

La segunda forma de encontrar la solución consiste en resolver el sistema

$$A \begin{bmatrix} y \\ v \end{bmatrix} = b$$

y ver si hay solución tal que $y \geq 0$. Más precisamente:

1) Construir la matriz ampliada $\hat{A} = [A \;\; b]$ de tamaño $n \times (p+1)$.

2) Convertirla, mediante operaciones elementales por filas, en una matriz

$$\hat{A}' = \begin{bmatrix} I_{\bar{m}} & 0 & c \\ 0 & I_l & e \\ 0 & 0 & d \end{bmatrix},$$

donde c es un vector columna $\bar{m} \times 1$, e es un vector columna $l \times 1$, d es un vector columna $(n - \bar{m} - l) \times 1$, es decir, se obtiene el siguiente sistema equivalente

$$\begin{bmatrix} I_{\bar{m}} & 0 \\ 0 & I_l \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ v \end{bmatrix} = \begin{bmatrix} c \\ e \\ d \end{bmatrix},$$

o sea,

$$\begin{aligned} I_{\bar{m}} y &= c \\ I_l v &= e \\ 0y + 0v &= d. \end{aligned}$$

El paso 2) siempre es posible puesto que las columnas de A son linealmente independientes.

3) Si $d \neq 0$ el sistema $A \begin{bmatrix} y \\ v \end{bmatrix} = b$ no tiene solución y el punto $\bar{x}$ no es minimizador local. Si $d = 0$ el sistema tiene como única solución $y = c$, $v = e$.

4) Si $y = c \geq 0$ el punto $\bar{x}$ es punto de KKT. Si $y = c \ngeq 0$, entonces $\bar{x}$ no es minimizador local.

Ejemplo 6.22. Considere $\bar{x} = (1, 1, 1, 1)$ para el siguiente problema:

$$\min f(x) = x_1^2 + 3x_2^2 + x_3^2 + x_4^2$$

$$\begin{aligned} (x_1-4)^2 + x_2^2 + x_3^2 + x_4^2 - 15 &\leq 0 \\ 1 - x_2 &\leq 0 \\ x_1 + 2x_2 + 3x_3 + 4x_4 - 10 &= 0 \\ x_1 + x_2 + x_3 + x_4 - 4 &= 0. \end{aligned}$$

El punto $\bar{x}$ es factible, $\mathcal{I} = \{2\}$, f, g_2, h_1, h_2 son diferenciables en $\bar{x}$, g_1 es continua en $\bar{x}$, el conjunto formado por los gradientes

$$\begin{aligned} g_2'(\bar{x}) &= [\ 0\ \ \text{-}1\ \ 0\ \ 0\ ]^{\mathrm{T}}, \\ h_1'(\bar{x}) &= [\ 1\ \ 2\ \ 3\ \ 4\ ]^{\mathrm{T}}, \\ h_2'(\bar{x}) &= [\ 1\ \ 1\ \ 1\ \ 1\ ]^{\mathrm{T}} \end{aligned}$$

es linealmente independiente. El gradiente de f es $f'(\bar{x}) = [2\ \ 6\ \ 2\ \ 2]^{\mathrm{T}}$. El problema que hay que resolver es el siguiente:

$$\begin{bmatrix} 0 & 1 & 1 \\ \text{-}1 & 2 & 1 \\ 0 & 3 & 1 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \text{-}2 \\ \text{-}6 \\ \text{-}2 \\ \text{-}2 \end{bmatrix}$$
$$y \geq 0,$$

donde $y_1 = u_2$. Si se resuelve como un problema de PL con variables no negativas, entonces se hacen los cambios $v_1 = v_1^+ - v_1^-$, $v_2 = v_2^+ - v_2^-$, y se considera la función objetivo más sencilla

$$\min z = 0y_1 + 0v_1^+ + 0v_2^+ + 0v_1^- + 0v_2^-$$

$$\begin{bmatrix} 0 & 1 & 1 & \text{-}1 & \text{-}1 \\ \text{-}1 & 2 & 1 & \text{-}2 & \text{-}1 \\ 0 & 3 & 1 & \text{-}3 & \text{-}1 \\ 0 & 4 & 1 & \text{-}4 & \text{-}1 \end{bmatrix} \begin{bmatrix} y_1 \\ v_1^+ \\ v_2^+ \\ v_1^- \\ v_2^- \end{bmatrix} = \begin{bmatrix} \text{-}2 \\ \text{-}6 \\ \text{-}2 \\ \text{-}2 \end{bmatrix}$$
$$y, v_1^+, v_2^+, v_1^-, v_2^- \geq 0.$$

Al resolver este problema de PL con variables no negativas se obtiene $u_2 = y_1 = 4$, $v_1^+ = 0$, $v_2^+ = 0$, $v_1^- = 0$, $v_2^- = 2$, luego $v_1 = 0$, $v_2 = -2$. Entonces $\bar{x}$ es un punto de KKT.

Si se resuelve de la segunda manera, se construye la matriz ampliada

$$\begin{bmatrix} 0 & 1 & 1 & -2 \\ -1 & 2 & 1 & -6 \\ 0 & 3 & 1 & -2 \\ 0 & 4 & 1 & -2 \end{bmatrix}.$$

Por medio de operaciones elementales sobre las filas se llega a

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

lo cual indica primero, que el sistema sí tiene solución, y, segundo, que y, una parte de la solución, tiene todas sus componentes no negativas: $u_2 = y_1 = 4$, luego $\bar{x}$ es punto de KKT. $\diamond$

Si las funciones $g_i, i \notin \mathcal{I}$ son diferenciables en $\bar{x}$, y si se considera que los u_i correspondientes son nulos, la admisibilidad y las condiciones necesarias de KKT se pueden escribir así:

$$\begin{aligned} g_i(\bar{x}) &\leq 0\,, \quad i = 1, ..., m \\ h_j(\bar{x}) &= 0\,, \quad j = 1, ..., l \\ f'(\bar{x}) + \sum_{i=1}^{m} u_i g_i'(\bar{x}) + \sum_{j=1}^{l} v_j h_j'(\bar{x}) &= 0 \\ u_i &\geq 0\,, \quad i = 1, ..., m \\ u_i g_i(\bar{x}) &= 0\,, \quad i = 1, ..., m. \end{aligned}$$

Si $g(\bar{x})$ denota el vector columna $[g_1(\bar{x}) \;\; g_2(\bar{x}) \;\; ... \;\; g_m(\bar{x})]^{\mathrm{T}}$, $g'(\bar{x})$ denota la matriz $n \times m$ cuyas columnas son los gradientes $g_1'(\bar{x})$, $g_2'(\bar{x})$,..., $g_m'(\bar{x})$, $h(\bar{x})$ denota el vector columna $[h_1(\bar{x}) \;\; h_2(\bar{x}) \;\; ... \;\; h_l(\bar{x})]^{\mathrm{T}}$, $h'(\bar{x})$ denota la matriz $n \times l$ cuyas columnas son los gradientes $h_1'(\bar{x})$, $h_2'(\bar{x})$,..., $h_l'(\bar{x})$, entonces la admisibilidad y las condiciones necesarias de KKT se pueden escribir

$$\begin{aligned} g(\bar{x}) &\leq 0 \\ h(\bar{x}) &= 0 \\ f'(\bar{x}) + g'(\bar{x})u + h'(\bar{x})v &= 0 \\ u &\geq 0 \\ u^{\mathrm{T}} g(\bar{x}) &= 0. \end{aligned}$$

Utilicemos la función lagrangiana, o simplemente el lagrangiano, función de $n + m + l$ variables $x_1, ..., x_n$, $u_1, ..., u_m$, $v_1, ..., v_l$, definida por:

$$\begin{aligned} L(x, u, v) &= f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x) \\ &= f(x) + u^{\mathrm{T}} g(x) + v^{\mathrm{T}} h(x). \end{aligned}$$

Denotemos por $L'_x(\bar{x}, u, v)$ las componentes del gradiente del lagrangiano $L'(\bar{x}, u, v)$ correspondientes a las derivadas parciales de L con respecto a las variables x_j, y de manera análoga $L'_u(\bar{x}, u, v)$ y $L'_v(\bar{x}, u, v)$. Entonces la admisibilidad y las condiciones de KKT se expresan así:

$$\begin{aligned} L'_u(\bar{x}, u, v) &\leq 0 \\ L'_v(\bar{x}, u, v) &= 0 \\ L'_x(\bar{x}, u, v) &= 0 \\ u &\geq 0 \\ u^{\mathrm{T}} L'_u(\bar{x}, u, v) &= 0. \end{aligned}$$

Estudiar, uno a uno, puntos admisibles puede ser un proceso inacabable. Para problemas pequeños, sin partir de un punto admisible explícito, se puede tratar de resolver completamente un problema estudiando todas las posibilidades para el conjunto $\mathcal{I}$.

Ejemplo 6.23. Resolver, utilizando condiciones de KKT, el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ -x_1 &\leq 0 \end{aligned}$$

$$\begin{aligned} -x_2 + 5/2 &\leq 0 \\ -x_1^2 + x_2 - 2 &= 0. \end{aligned}$$

En este ejemplo el conjunto admisible $\mathcal{A}$ es no acotado y cerrado (pues es intersección de cerrados), la función f es continua y coercitiva, entonces el PMDI tiene por lo menos un minimizador global. Sea $\bar{x}$ un punto admisible. Como no se sabe exactamente qué punto es, entonces es necesario estudiar todas las posibilidades. Estas se pueden agrupar en función de las posibilidades de $\mathcal{I}$: $\mathcal{I} = \emptyset$, $\mathcal{I} = \{1\}$, $\mathcal{I} = \{2\}$, $\mathcal{I} = \{1, 2\}$.

i) $\mathcal{I} = \emptyset$:

$$f'(\bar{x}) + \sum_{i \in \mathcal{I}} u_i g_i'(\bar{x}) + \sum_{j=1}^{l} v_j h_j'(\bar{x}) = \begin{bmatrix} 2(\bar{x}_1\text{-}3) \\ 2(\bar{x}_2\text{-}2) \end{bmatrix} + v_1 \begin{bmatrix} -2\bar{x}_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Como se tiene que cumplir la igualdad, entonces

$$-\bar{x}_1^2 + \bar{x}_2 - 2 = 0\ , \qquad \text{luego } \bar{x}_2 = \bar{x}_1^2 + 2,$$

$$\begin{bmatrix} 2(\bar{x}_1 - 3) \\ 2(\bar{x}_1^2 + 2 - 2) \end{bmatrix} + v_1 \begin{bmatrix} \text{-}2\bar{x}_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

De donde

$$\begin{aligned} 2(\bar{x}_1 - 3) - 2v_1\bar{x}_1 &= 0 \\ 2\bar{x}_1^2 + v_1 &= 0. \end{aligned}$$

Reemplazando $v_1 = -2\bar{x}_1^2$,

$$\begin{aligned} 2(\bar{x}_1 - 3) + 4\bar{x}_1^3 &= 0 \\ 2(2\bar{x}_1^3 + \bar{x}_1 - 3) &= 0 \\ (\bar{x}_1 - 1)(2\bar{x}_1^2 + 2\bar{x}_1 + 3) &= 0. \end{aligned}$$

El polinomio cuadrático no tiene raíces reales ya que su discriminante vale -20, entonces:

$$\bar{x}_1 = 1$$

$$\begin{aligned} \bar{x}_2 &= 3 \\ v_1 &= -2. \end{aligned}$$

Entonces la suposición $\mathcal{I} = \emptyset$ conduce al punto $\bar{x} = (1, 3)$ que cumple las condiciones necesarias de KKT.

ii) $\mathcal{I} = \{1\}$: como está activa la primera desigualdad y se cumple la igualdad, entonces:

$$\begin{aligned} -\bar{x}_1 &= 0 \\ -\bar{x}_1^2 + \bar{x}_2 - 2 &= 0. \end{aligned}$$

Luego $\bar{x}_2 = 2$, es decir, la suposición $\mathcal{I} = \{1\}$ conduce al punto $\bar{x} = (0, 2)$ que no es realizable.

iii) $\mathcal{I} = \{2\}$: como está activa la segunda desigualdad y se cumple la igualdad, entonces:

$$\begin{aligned} -\bar{x}_2 + 5/2 &= 0 \\ -\bar{x}_1^2 + \bar{x}_2 - 2 &= 0. \end{aligned}$$

Luego $\bar{x}_2 = 5/2$, $\bar{x}_1 = \pm\sqrt{2}/2$, es decir, la suposición $\mathcal{I} = \{2\}$ conduce, o bien al punto $\bar{x} = (-\sqrt{2}/2, 5/2)$ que no es admisible, o bien al punto $\bar{x} = (\sqrt{2}/2, 5/2)$ que no cumple condiciones necesarias de KKT.

iv) $\mathcal{I} = \{1, 2\}$: como están activas la primera y la segunda desigualdad y se cumple la igualdad, entonces:

$$\begin{aligned} -\bar{x}_1 &= 0 \\ -\bar{x}_2 + 5/2 &= 0 \\ -\bar{x}_1^2 + \bar{x}_2 - 2 &= 0. \end{aligned}$$

Pero no existe ningún punto que cumpla las tres igualdades. En resumen, hay un solo punto que cumple condiciones necesarias de KKT, y como existe por lo menos un minimizador, entonces este punto $\bar{x} = (1, 3)$ debe ser el minimizador global. $\diamond$

Proposición 6.11. Condiciones suficientes de KKT para el PMDI. *Si el punto $\bar{x}$ cumple las condiciones necesarias de KKT para el PMDI y además*

f es seudoconvexa en $\bar{x}$,
g_i es cuasiconvexa en $\bar{x}$ para $i \in \mathcal{I}$,
h_j es cuasiconvexa en $\bar{x}$ para $v_j > 0$,
h_j es cuasicóncava en $\bar{x}$ para $v_j < 0$,
entonces $\bar{x}$ es un minimizador global del PMDI.

Ejemplo 6.24. Considere $\bar{x} = (1, 3)$ para el siguiente problema:

$$\begin{aligned} \min f(x) &= (x_1 - 3)^2 + (x_2 - 2)^2 \\ -x_1 &\leq 0 \\ -x_2 + 5/2 &\leq 0 \\ -x_1^2 + x_2 - 2 &= 0. \end{aligned}$$

El punto $\bar{x}$ cumple condiciones necesarias de KKT, además

$$f''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

es definida positiva para todo x, luego f es convexa, como es diferenciable, entonces es seudoconvexa y en particular es seudoconvexa en $\bar{x}$. La función h_1 es diferenciable, la matriz hessiana de $-h_1$,

$$-h_1''(x) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix},$$

es semidefinida positiva para todo x, luego $-h_1$ es convexa, es decir, h_1 es cóncava, entonces es cuasicóncava y en particular es cuasicóncava en $\bar{x}$. Luego $\bar{x} = (1, 3)$ es minimizador global. $\diamond$

6.4 CONDICIONES DE SEGUNDO ORDEN

Así como para puntos interiores hay condiciones necesarias (... y hessiana semidefinida positiva) y condiciones suficientes de segundo orden (... y hessiana definida positiva), es decir, que utilizan derivadas parciales de segundo orden, también hay condiciones de segundo orden para puntos no interiores. La diferencia principal consiste en que no se exige que la hessiana sea semidefinida positiva en todo $\mathbb{R}^n$, sino en un subespacio.

Definición 6.7. Sean: $\mathcal{A}$ el conjunto admisible de PMDI, $\bar{x}$ admisible, g_i diferenciable en $\bar{x}$ para $i \in \mathcal{I}$, h_j diferenciable en $\bar{x}$ para todo j, g_i continua en $\bar{x}$ para $i \notin \mathcal{I}$, $\bar{x}$ regular. El **espacio tangente** $\mathcal{T}$ es el conjunto de

puntos de $\mathbb{R}^n$ ortogonales a los gradientes activos $\{g_i'(\bar{x}), i \in \mathcal{I},\ h_j'(\bar{x}), \forall j\}$, es decir

$$\mathcal{T} = \{y \in \mathbb{R}^n : g_i'(\bar{x})^{\mathrm{T}} y = 0, \forall i \in \mathcal{I}\,,\ \ h_j'(\bar{x})^{\mathrm{T}} y = 0, \forall j\}.$$

Si $p = \bar{m}+l = 0$, es decir, si no hay desigualdades activas y no hay igualdades

$$\mathcal{T} = \mathbb{R}^n.$$

Si $p = n$

$$\mathcal{T} = \{\mathbf{0}\}.$$

Denotando por M la matriz $p \times n = (\bar{m} + l) \times n$, cuyas filas son los gradientes $g_i'(\bar{x})^{\mathrm{T}}, i \in \mathcal{I},\ \ h_j'(\bar{x})^{\mathrm{T}}, j = 1, ..., l$, (exactamente la transpuesta de la matriz A utilizada en el cálculo de los coeficientes u_i, v_j para un $\bar{x}$ dado), el espacio tangente se define sencillamente como

$$\mathcal{T} = \{y \in \mathbb{R}^n : My = 0\}.$$

Dicho de otra forma, $\mathcal{T}$ es simplemente el espacio nulo de la matriz M, y claro está, es un subespacio vectorial de $\mathbb{R}^n$. El significado de $\mathcal{T}$ puede tener la siguiente interpretación geométrica en $\mathbb{R}^3$ (o en $\mathbb{R}^2$): Si $\bar{x}$ está en la frontera de $\mathcal{A}$, y si se puede construir un plano tangente a la superficie en $\bar{x}$ o una única recta tangente, el espacio tangente es un plano (o una recta) que pasa por el origen y es paralelo al plano tangente (o a la recta tangente).

Ejemplo 6.25. Sean: $\mathcal{A}$ en $\mathbb{R}^2$ definido por dos desigualdades

$$\begin{aligned} (x_1 + 1)^2 + x_2^2 - 25 &\leq 0 \\ -x_1 - x_2 &\leq 0, \end{aligned}$$

$\bar{x} = (2, 4)$. El conjunto $\mathcal{A}$ es un pedazo de círculo. La recta tangente a $\mathcal{A}$ en el punto $\bar{x} = (2, 4)$ es

$$R = \{(y_1, y_2) : 3y_1 + 4y_2 = 22\}.$$

El espacio tangente es la recta paralela a R que pasa por el origen:

$$\mathcal{T} = \{(y_1, y_2) : 3y_1 + 4y_2 = 0\}.$$

Si se considera $\bar{x} = (1, 3)$, el espacio tangente es todo $\mathbb{R}^2$. Si se considera $\bar{x} = (-4, 4)$, no se puede construir "la recta tangente" a $\mathcal{A}$ en $(-4, 4)$. El espacio tangente es $\mathcal{T} = \{(0, 0)\}$. $\diamond$

Ejemplo 6.26. Sean: $\mathcal{A}$ en $\mathbb{R}^3$ definido por

$$x_1^2 + x_2^2 + x_3^2 - 20 \leq 0,$$

$\bar{x} = (2, 0, 4)$. El plano tangente a $\mathcal{A}$ en $\bar{x}$ es

$$P = \{(y_1, y_2, y_3) : y_1 + 2y_3 = 10\}.$$

El espacio tangente es el plano paralelo a P que pasa por el origen:

$$\mathcal{T} = \{(y_1, y_2, y_3) : y_1 + 2y_3 = 0\}. \ \diamond$$

Ejemplo 6.27. Sean: $\mathcal{A}$ en $\mathbb{R}^3$ definido por

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 - 20 &\leq 0, \\ x_1 + 2x_2 + 3x_3 - 14 &= 0, \end{aligned}$$

$\bar{x} = (2, 0, 4)$. La recta tangente a $\mathcal{A}$, en $\bar{x}$ es

$$R = \{(2, 0, 4) + t(\text{-}4, \text{-}1, 2) : t \in \mathbb{R}\}.$$

El espacio tangente es la recta paralela a R que pasa por el origen:

$$\begin{aligned} \mathcal{T} &= \{(y_1, y_2, y_3) : y_1 + 2y_3 = 0, y_1 + 2y_2 + 3y_3 = 0\} \\ &= \{t(\text{-}4, \text{-}1, 2) : t \in \mathbb{R}\} \ \diamond. \end{aligned}$$

Sean: $\bar{x}$ un punto de KKT para el PMDI, $\bar{u}$, $\bar{v}$ sus coeficientes de Lagrange, $L''_x(\bar{x}, \bar{u}, \bar{v})$ la submatriz de la matriz hessiana del lagrangiano correspondiente a las derivadas parciales de segundo orden con respecto a las variables x_i, o sea,

$$L''_x = L''_x(\bar{x}, \bar{u}, \bar{v}) = f''(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i g_i''(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j h_j''(\bar{x}).$$

Como para las desigualdades inactivas $u_i = 0$, entonces la definición de L''_x se puede dar como expresión de los hessianos de las desigualdades activas y de las igualdades:

$$L''_x = L''_x(\bar{x}, \bar{u}, \bar{v}) = f''(\bar{x}) + \sum_{i \in \mathcal{I}} \bar{u}_i g_i''(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j h_j''(\bar{x}).$$

Sean: $\mathcal{I}^+ = \{i : \bar{u}_i > 0\}$, $\mathcal{I}^o = \{i : \bar{u}_i = 0\}$. Si $i \in \mathcal{I}^+$ se dice que la restricción $g_i(x) \leq 0$ está **fuertemente activa**. Si $i \in \mathcal{I}^o$ se dice que la restricción $g_i(x) \leq 0$ está **débilmente activa**. Sean: $\bar{m}^+$ el número de desigualdades fuertemente activas, $\bar{m}^o$ el número de desigualdades débilmente activas. Sea $p^+ = \bar{m}^+ + l$, o sea, el número de desigualdades fuertemente activas más el número de igualdades. Obviamente $\mathcal{I}$ es unión disyunta de $\mathcal{I}^+$ e $\mathcal{I}^0$ y además $\bar{m} = \bar{m}^+ + \bar{m}^o$.

Sea $\tilde{M}$ la matriz de tamaño $p^+ \times n$ cuyas filas son los gradientes de las desigualdades fuertemente activas y los gradientes de las igualdades calculados en $\bar{x}$. Dicho de otra forma, $\tilde{M}$ se obtiene quitando de M las filas correspondientes a los gradientes de las desigualdades débilmente activas. Sea $\tilde{\mathcal{T}}$ el espacio nulo de $\tilde{M}$, o sea, $\tilde{\mathcal{T}} = \{y \in \mathbb{R}^n : \tilde{M}y = 0\}$. Si p^+, (el número de filas de $\tilde{M}$) es cero se considera que $\tilde{\mathcal{T}} = \mathbb{R}^n$.

Proposición 6.12. Condiciones necesarias de segundo orden. *Sean: $\bar{x}$ un punto de KKT del PMDI, f, $g_i, i \in \mathcal{I}$, $h_j, j = 1, ..., l$ doblemente diferenciables en $\bar{x}$. Si $\bar{x}$ es un minimizador local, entonces L''_x es semidefinida positiva en $\mathcal{T}$.*

Proposición 6.13. Condiciones suficientes de segundo orden. *Sean: $\bar{x}$ un punto de KKT del PMDI, f, $g_i, i \in \mathcal{I}$, $h_j, j = 1, ..., l$ doblemente diferenciables en $\bar{x}$. Si L''_x es definida positiva en $\tilde{\mathcal{T}}$, entonces $\bar{x}$ es un minimizador local estricto.*

Ejemplo 6.28. Considere $\bar{x} = (0, 0)$ para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 - x_2^2 & \\ -x_1 - x_2 \quad &\leq \quad 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{1\}$; f, g_1 son diferenciables en $\bar{x}$; el conjunto formado por el gradiente $g_1'(\bar{x}) = [-1 \quad -1]^{\mathrm{T}}$ es linealmente independiente, o sea, $\bar{x}$ es regular.

$$\begin{aligned} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + u_1 \begin{bmatrix} -1 \\ -1 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\ u_1 &= 0, \end{aligned}$$

luego $\bar{x} = (0, 0)$ es un punto de KKT.

$$\begin{aligned} L''_x &= \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} + 0 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \\ \mathcal{T} &= \{(y_1, y_2) : -y_1 - y_2 = 0\}. \end{aligned}$$

Una base de este subespacio puede ser el conjunto formado por el vector $[1 \quad -1]^{\mathrm{T}}$, luego

$$\begin{aligned} E &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ E^{\mathrm{T}} L''_x E &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= 0, \text{ matriz semidefinida positiva.} \end{aligned}$$

Luego $\bar{x}$ cumple condiciones necesarias de segundo orden. La desigualdad activa es débilmente activa, luego $\bar{m}^+ = 0$, $p^+ = 0$, $\tilde{M}$ no tiene filas, $\tilde{\mathcal{T}} = \mathbb{R}^2$. Como L''_x no es definida positiva en $\mathbb{R}^2$, entonces $\bar{x}$ no cumple condiciones suficientes de segundo orden. En resumen, utilizando condiciones de KKT y de segundo orden se puede decir que $\bar{x}$ es un buen candidato a minimizador local. Sin embargo, no es minimizador local ya que un punto de la forma $(0, \varepsilon)$ es admisible, mejor que $\bar{x}$ y está muy cerca de él. $\diamond$

Ejemplo 6.29. Considere $\bar{x} = (0, 0, 0)$ para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2, x_3) = x_1^2 - x_2^2 + \frac{1}{3}(x_3 - 1)^3 & \\ -6 - x_1 - x_2 - x_3 &\leq 0 \\ -x_3 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{2\}$; f, g_2 son diferenciables en $\bar{x}$; g_1 es continua en $\bar{x}$; el conjunto formado por el gradiente $g'_2(\bar{x}) = [0 \quad 0 \quad -1]^{\mathrm{T}}$ es linealmente independiente, o sea, $\bar{x}$ es regular.

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$u_2 = 1,$$

luego $\bar{x} = (0, 0, 0)$ es un punto de KKT.

$$\begin{aligned} L''_x &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} + 1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \\ \mathcal{T} &= \{(y_1, y_2, y_3) : -y_3 = 0\}. \end{aligned}$$

Una base de este subespacio puede ser el conjunto formado por los vectores $[1 \;\; 0 \;\; 0]^{\mathrm{T}}$, $[0 \;\; 1 \;\; 0]^{\mathrm{T}}$, luego

$$\begin{aligned} E &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \\ E^{\mathrm{T}} L''_x E &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \quad \text{no es semidefinida positiva,} \end{aligned}$$

luego $\bar{x}$ no cumple condiciones necesarias de segundo orden, entonces no es minimizador local ni global. $\diamond$

Ejemplo 6.30. Considere $\bar{x} = (0, 3, 0)$ para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2, x_3) = x_1^2 - x_2^2 + x_3^2 & \\ -x_2 &\leq 0 \\ x_2 - 3 &\leq 0. \end{aligned}$$

El punto $\bar{x}$ es factible; $\mathcal{I} = \{2\}$; f, g_2 son diferenciables en $\bar{x}$; g_1 es continua en $\bar{x}$; el conjunto formado por el gradiente $g'_2(\bar{x}) = [0 \;\; 1 \;\; 0]^{\mathrm{T}}$ es linealmente independiente, o sea, $\bar{x}$ es regular.

$$\begin{bmatrix} 0 \\ -6 \\ 0 \end{bmatrix} + u_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$
$$u_2 = 6.$$

Luego $\bar{x} = (0, 3, 0)$ es un punto de KKT.

$$\begin{aligned} L''_x &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix} + 6 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \\ \mathcal{T} &= \{(y_1, y_2, y_3) : y_2 = 0\}. \end{aligned}$$

Una base de este subespacio puede ser el conjunto formado por los vectores $[1 \;\; 0 \;\; 0]^{\mathrm{T}}$, $[0 \;\; 0 \;\; 1]^{\mathrm{T}}$, luego

$$\begin{aligned} E &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \\ E^{\mathrm{T}} L''_x E &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{es definida positiva.} \end{aligned}$$

Entonces $\bar{x}$ cumple condiciones necesarias de segundo orden. Más aún, no hay desigualdades débilmente activas, entonces $\tilde{M} = M$, $\tilde{\mathcal{T}} = \mathcal{T}$. En consecuencia, L''_x es definida positiva en $\tilde{\mathcal{T}}$, es decir, $\bar{x} = (0, 3, 0)$ cumple condiciones suficientes y es minimizador local estricto. $\diamond$

EJERCICIOS

En los ejercicios 6.1 a 6.15 estudie el problema propuesto, use condiciones necesarias, suficientes, condiciones de segundo orden y otros argumentos. Encuentre, si es posible, puntos de KKT. ¿Son estos puntos minimizadores? ¿Son minimizadores globales?

6.1. Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$, sujeto a $3x_1 + 4x_2 = 12$.

6.2. Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$, sujeto a $3x_1 + 4x_2 \geq 12$.

6.3. Minimizar $f(x_1, x_2) = 2x_1^2 + x_2^2$, sujeto a $3x_1 + 4x_2 = 12$.

6.4. Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$, sujeto a $3x_1 + 4x_2 = 12$, $x_1 \leq 0$.

6.5. Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$, sujeto a $3x_1 + 4x_2 = 12$, $x_1 \leq 0$, $x_2 \geq 3$.

6.6. Minimizar $f(x_1, x_2) = x_1^2 - x_2^2$, sujeto a $3x_1 + 4x_2 = 12$.

6.7. Minimizar $f(x_1, x_2) = x_1^2 - x_2^2$, sujeto a $3x_1 + 4x_2 \geq 12$.

6.8. Minimizar $f(x_1, x_2) = x_1^2 - x_2^2$, sujeto a $3x_1 + 4x_2 \leq 12$, $x \geq 0$.

6.9. Minimizar $f(x_1, x_2) = x_1^2 - x_2^2$, sujeto a $3x_1 + 4x_2 \geq 12$, $x \geq 0$.

6.10. Minimizar $f(x) = c^{\mathrm{T}}x$, sujeto a $||x - a||_2 = r$, con $c \neq 0$, $r > 0$.

6.11. Minimizar $f(x) = c^{\mathrm{T}}x$, sujeto a $||x - a||_2 \leq r$, con $c \neq 0$, $r > 0$.

6.12. Minimizar $f(x_1, x_2) = x_1^3 + x_2^2$, sujeto a $3x_1 + 4x_2 = 12$.

6.13. Minimizar $f(x_1, x_2) = x_1^3 + x_2^2$, sujeto a $3x_1 + 4x_2 \geq 12$.

6.14. Minimizar $f(x_1, x_2) = x_1^3 + x_2^2$, sujeto a $3x_1 + 4x_2 \leq 12$, $x \geq 0$.

6.15. Minimizar $f(x_1, x_2) = x_1^3 + x_2^2$, sujeto a $3x_1 + 4x_2 \geq 12$, $x \geq 0$.

Capítulo 7

MÉTODOS DE MINIMIZACIÓN EN UNA VARIABLE

En la mayoría de métodos de minimización en varias variables sin restricciones o con restricciones, es necesario resolver problemas implícitos de minimización en una variable. Si f es una función de varias variables y valor real, y están fijos un punto $x = x^k$ y una dirección $d = d^k \neq \mathbf{0}$, entonces con mucha frecuencia se debe resolver el siguiente subproblema: minimizar una función φ que depende únicamente de λ

$$\min \varphi(\lambda) = f(x + \lambda d),$$

donde λ varía en un conjunto Λ. Los casos más corrientes son:

$$\begin{aligned} \Lambda &= \mathbb{R}, \\ \Lambda &= [0, \infty[, \\ \Lambda &= [0, \lambda_{\max}]. \end{aligned}$$

Los tres casos anteriores se pueden generalizar a:

$$\begin{aligned} \Lambda &= \mathbb{R}, \\ \Lambda &= [a, \infty[, \\ \Lambda &= [a, b]. \end{aligned}$$

Casi siempre cuando $\Lambda = [0, \infty]$ o cuando $\Lambda = [0, \lambda_{\max}]$ la dirección $d = d^k$ es una dirección de descenso, entonces existe $\varepsilon > 0$ tal que $\varphi(\lambda) < \varphi(0)$ para todo $\lambda \in [0, \varepsilon]$.

En general, cada método de minimización en una variable es adecuado para un solo caso de variación de λ, sin embargo, frecuentemente, se puede modificar el método para adaptarlo a otro conjunto Λ.

La minimización en una variable también es llamada **búsqueda lineal**. Siempre se supondrá, por lo menos, que φ es continua.

Algunos métodos de minimización en una variable necesitan usar segundas derivadas (y primeras derivadas), otros no necesitan usar segundas derivadas, pero sí necesitan primeras derivadas, y finalmente otros métodos solamente requieren valores de la función φ en varios puntos.

Dependiendo de la función φ y del método, en algunos casos se construye una sucesión $\{\lambda_k\}$ que tiende a λ^* minimizador local o global de la función. Obviamente como se trata de una sucesión infinita, no es posible calcular todos los elementos, pero como es usual, el proceso se detiene cuando

$$|\lambda_{k+1} - \lambda_k| \quad \leq \varepsilon_\lambda \text{ dado,}$$

o también,

$$\left| \frac{\lambda_{k+1} - \lambda_k}{\lambda_k} \right| \quad \leq \varepsilon'_\lambda \text{ dado,}$$

o para evitar denominadores cercanos a cero,

$$\frac{|\lambda_{k+1} - \lambda_k|}{1 + |\lambda_k|} \quad \leq \varepsilon'_\lambda \text{ dado.}$$

Este hecho indica que λ_{k+1} está suficientemente cerca de λ^*, o bien, que el proceso está avanzando muy lentamente.

Otro criterio usado para detener el proceso (puede ser usado simultáneamente) consiste en parar cuando el valor de φ varía muy poco, o sea, si

$$|\varphi(\lambda_{k+1}) - \varphi(\lambda_k)| \quad \leq \varepsilon_\varphi \text{dado ,}$$

o también,

$$\left|\frac{\varphi(\lambda_{k+1})-\varphi(\lambda_k)}{\varphi(\lambda_k)}\right| \leq \varepsilon'_\varphi \text{ dado },$$

o para evitar denominadores muy pequeños,

$$\frac{|\varphi(\lambda_{k+1})-\varphi(\lambda_k)|}{1+|\varphi(\lambda_k)|} \leq \varepsilon'_\varphi \text{ dado }.$$

Cuando se emplean primeras derivadas de φ también se puede usar simultáneamente un criterio de parada que verifica si φ' está cerca de cero, o sea, si λ_k es casi un punto crítico,

$$|\varphi'(\lambda_k)| \leq \varepsilon_{\varphi'} \text{ dado.}$$

Siempre es necesario utilizar en casi todos los métodos un número máximo de iteraciones, llamado en este libro MAXIT. Esto permite impedir que un método gaste demasiado tiempo, o peor aún, que resulte un ciclo sin fin y obligue a abortar el programa o a apagar el computador.

En cualquier algoritmo de minimización en una variable, cuando λ varía en un conjunto no acotado ($[0,\infty[$ ó $\mathbb{R}$) es necesario preveer la posibilidad de que φ no esté acotada inferiormente, lo que implica que no tiene minimizador.

Otras veces la sucesión tiende a un punto λ' donde simplemente se anula la primera derivada, lo cual bajo ciertas condiciones, garantiza un minimizador.

En otros métodos se parte de un intervalo de incertidumbre $[a_0, b_0]$ y el método construye intervalos $[a_k, b_k]$ encajonados en los anteriores y de menor tamaño que también contienen al minimizador λ^*. En este caso el proceso termina cuando

$$b_k - a_k < \varepsilon \text{ dado.}$$

En los métodos llamados de descenso, además de buscar convergencia, se busca que siempre la función disminuya de valor, o sea, una condición indispensable es que

$$\varphi(\lambda_{k+1}) < \varphi(\lambda_k).$$

La minimización de funciones de una variable es un subproblema que se efectúa muchas veces. Si cada vez se pide mucha precisión, cada vez se gasta

un tiempo no despreciable. En algunos métodos las primeras veces que se utiliza la minimización en una variable no se requiere mucha precisión en la aproximación del minimizador λ^*. Es una buena idea entonces detener el proceso cuando se ha obtenido una disminución suficiente en el valor de φ

$$\varphi(\lambda_0) - \varphi(\lambda_k) > \mu|(\lambda_k - \lambda_0)d^{\mathrm{T}} f'(x)|,$$

donde $0 < \mu < 1$. Obviamente si se escoge μ muy pequeño la condición se satisface muy fácilmente. Con frecuencia la dirección es tal que $d^{\mathrm{T}} f'(x) < 0$ (lo cual garantiza una dirección de descenso), $\lambda_0 = 0$, $\lambda > 0$ y se trata de un método de descenso, entonces el criterio de disminución suficiente es:

$$\varphi(\lambda_k) < \varphi(0) + \mu\lambda_k d^{\mathrm{T}} f'(x).$$

Los métodos de minimización en una variable, que se presenten en este libro, siempre estarán enfocados a minimizar $f(x + \lambda d)$, pero obviamente también se pueden aplicar para funciones que son explícitamente de una variable, por ejemplo, para minimizar $\varphi(\lambda) = \cos\lambda + \lambda/2$ en el intervalo $[-1, 1]$.

Los valores de ε los da el usuario, pero generalmente se exige que no le pidan al método más de la mitad de cifras significativas de las dadas por la máquina usada. Por ejemplo, si un computador da 16 cifras significativas, entonces el valor de ε dado por el usuario debe ser mayor que 10^{-8}. Es frecuente relacionar la anterior condición con $\varepsilon_{\mathrm{maq}}$: el ε de la máquina, definido como la menor cantidad positiva tal que el computador puede hacer la diferencia entre 1 y $1 + \varepsilon_{\mathrm{maq}}$.

$$\varepsilon_{\mathrm{maq}} = \min\{\varepsilon : 1 + \varepsilon \quad \text{“diferente de” } 1\}.$$

Una forma aproximada de encontrar $\varepsilon_{\mathrm{maq}}$ consiste en empezar con $\varepsilon = 1$ e ir dividiéndolo por 2 hasta que el computador no diferencie entre 1 y $1+\varepsilon$. Entonces se toma como $\varepsilon'_{\mathrm{maq}}$ el penúltimo valor obtenido, es decir, el doble del útimo valor obtenido.

$\varepsilon = 1$
$s = 1 + \varepsilon$
mientras $s \neq 1$
    $\varepsilon = \varepsilon/2$
    $s = 1 + \varepsilon$
fin-mientras
$\varepsilon'_{\mathrm{maq}} = 2\varepsilon$

Obviamente la aproximación del ε_{maq} depende del tipo de números usados: "flotantes", "doble precisión", ... Por ejemplo, trabajando en un microcomputador con procesador AMD 486DX40 (tiene coprocesador matemático) y en precisión sencilla, en MS-Fortran 5.1 o en Turbo C 3.0, se obtiene el valor

$$\varepsilon'_{\text{maq}} = 1.192 * 10^{-7}.$$

En este caso no hubo diferencia entre los dos lenguajes (y compiladores), pero podría haberla. Trabajando en doble precisión el valor encontrado es

$$\varepsilon'_{\text{maq}} = 1.0842 * 10^{-19}.$$

Utilizando ε_{maq}, para que el valor de ε no sea demasiado pequeño, se acostumbra exigir que

$$\varepsilon > \sqrt{\varepsilon_{\text{maq}}}.$$

Cuando se trabaja con números en doble precisión los arreglos gastan el doble de memoria que los arreglos en precisión sencilla. El tiempo para las operaciones también puede ser mayor. Sin embargo, como regla general, es preferible utilizar números en doble precisión.

7.1 CÁLCULO DE LAS DERIVADAS

La derivada de φ esta íntimamente relacionada con la derivada direccional de f en la dirección d,

$$\begin{aligned}
\varphi'(\lambda) &= \lim_{\eta \to 0} \frac{\varphi(\lambda + \eta) - \varphi(\lambda)}{\eta} \\
&= \lim_{\eta \to 0} \frac{f(x + \lambda d + \eta d) - f(x + \lambda d)}{\eta} \\
&= \lim_{\eta \to 0} \frac{f(x + \lambda d + \eta \delta d') - f(x + \lambda d)}{\eta},
\end{aligned}$$

donde

$$\delta = \|d\| \neq 0 \ , \quad d' = \frac{d}{\delta} \ , \quad \|d'\| = 1.$$

Así:

$$\begin{aligned}\varphi'(\lambda) &= \lim_{\eta\to 0}\delta\frac{f(x+\lambda d+\eta\delta d')-f(x+\lambda d)}{\eta\delta}\\ &= \delta f'(x+\lambda d;d'),\end{aligned}$$

o sea, δ veces la derivada direccional de f en el punto $x+\lambda d$ con respecto a la dirección (de norma uno) d'.

$$\begin{aligned}\varphi'(\lambda) &= \delta f'(x+\lambda d)^{\mathrm{T}}d'\\ \varphi'(\lambda) &= f'(x+\lambda d)^{\mathrm{T}}d, \qquad (7.1)\end{aligned}$$

en particular,

$$\varphi'(0) = f'(x)^{\mathrm{T}}d.$$

De manera semejante se puede expresar la segunda derivada de φ usando el hessiano:

$$\varphi''(\lambda) = d^{\mathrm{T}}f''(x+\lambda d)d, \qquad (7.2)$$

en particular,

$$\varphi''(0) = d^{\mathrm{T}}f''(x)d.$$

Los métodos de minimización en una variable que utilizan la primera o segunda derivada requerirían entonces el conocimiento del gradiente y de la matriz hessiana. O sea, en una parte del programa de computador, debe haber subprogramas, procedimientos, subrutinas o funciones para el cálculo del gradiente y del hessiano. Este cálculo puede ser exacto, es decir, mediante las expresiones analíticas de las derivadas parciales de la función f, o bien mediante aproximaciones numéricas.

Cuando en el programa se tiene la expresión analítica del gradiente y del hessiano, sí se puede pensar en calcular φ' y φ'' mediante (7.1) y (7.2). Sin embargo, se debe evaluar la necesidad de precisión y el tiempo necesario para el cálculo exacto del gradiente y del hessiano, y para el producto matricial.

Si por el contrario, se tiene simplemente una fórmula aproximada para calcular el gradiente y el hessiano de f, no es adecuado utilizar las fórmulas

(7.1) y (7.2) para calcular φ' y φ''. En este caso es mucho mejor aproximar directa y numéricamente φ' y φ''. Para tal efecto algunas de las fórmulas más utilizadas son:

$$\varphi'(\lambda) \approx \frac{\varphi(\lambda+\eta)-\varphi(\lambda)}{\eta}, \tag{7.3}$$

$$|\text{error}| \leq \frac{\eta}{2} \max_{[\lambda,\lambda+\eta]} |\varphi''|.$$

$$\varphi'(\lambda) \approx \frac{\varphi(\lambda+\eta)-\varphi(\lambda-\eta)}{2\eta}, \tag{7.4}$$

$$|\text{error}| \leq \frac{\eta^2}{6} \max_{[\lambda-\eta,\lambda+\eta]} |\varphi'''|.$$

$$\varphi''(\lambda) \approx \frac{\varphi(\lambda+\eta)-2\varphi(\lambda)+\varphi(\lambda-\eta)}{\eta^2}, \tag{7.5}$$

$$|\text{error}| \leq \frac{\eta^2}{12} \max_{[\lambda-\eta,\lambda+\eta]} |\varphi''''|.$$

Estas fórmulas, para aproximar numéricamente derivadas de funciones de una variable, son teóricamente más exactas cuando η es más pequeño. Sin embargo, en la práctica esto no es necesariamente cierto. Dependiendo de φ y de λ, si η es demasiado pequeño el numerador o el denominador pueden ser cero (numéricamente). En las tablas 7.1, 7.2, 7.3 hay ejemplos de utilización de la fórmula (7.4) con números en doble precisión (8 bytes) en un microcomputador sin coprocesador matemático.

7.2 MÉTODO DE NEWTON

Este método busca un punto $\bar{\lambda}$ donde la primera derivada de φ se anula. Si $\varphi''(\bar{\lambda}) > 0$, entonces $\bar{\lambda}$ es un minimizador local. Si φ es convexa, entonces $\bar{\lambda}$ es un minimizador global. La fórmula se puede deducir de dos formas equivalentes: aplicando el método de Newton para hallar una raíz de φ', o bien buscando un punto crítico de la aproximación cuadrática de φ alrededor de λ_k.

La fórmula del método de Newton para resolver $g(\lambda) = 0$ es

$$\lambda_{k+1} = \lambda_k - \frac{g(\lambda_k)}{g'(\lambda_k)},$$

$\varphi(\lambda) = \lambda^4,\ \lambda = 2,\ \varphi'(2) = 32$	
η	$\widetilde{\varphi'}$
1.0E+00	4.00000000E+01
1.0E-01	3.20800000E+01
1.0E-02	3.20008000E+01
1.0E-03	3.20000080E+01
1.0E-04	3.20000001E+01
1.0E-05	3.20000000E+01
1.0E-06	3.20000000E+01
1.0E-07	3.20000000E+01
1.0E-08	3.19999999E+01
1.0E-09	3.20000026E+01
1.0E-10	3.20000026E+01
1.0E-11	3.20000026E+01
1.0E-12	3.20028448E+01
1.0E-13	3.19744231E+01
1.0E-14	3.23296945E+01
1.0E-15	3.19744231E+01
1.0E-16	0.00000000E+00
⋮	⋮
1.0E-20	0.00000000E+00

Tabla 7.1

entonces la fórmula del método de Newton para "buscar" un minimizador es:

$$\lambda_{k+1} = \lambda_k - \frac{\varphi'(\lambda_k)}{\varphi''(\lambda_k)}. \tag{7.6}$$

La aproximación cuadrática de φ alrededor de λ_k es

$$\varphi(\lambda) \approx p_2(\lambda) = \varphi(\lambda_k) + \varphi'(\lambda_k)(\lambda - \lambda_k) + \frac{1}{2}\varphi''(\lambda_k)(\lambda - \lambda_k)^2.$$

Al buscar un punto $\lambda = \lambda_{k+1}$ donde se anule la derivada de p_2 se obtiene exactamente la fórmula (7.6).

Proposición 7.1. [Lue89] *Sean: φ con tercera derivada continua, $\varphi'(\bar{\lambda}) = 0$, $\varphi''(\bar{\lambda}) \neq 0$. Si λ_0 está suficientemente cerca de $\bar{\lambda}$, entonces el método de Newton para hallar una raíz de φ' converge a $\bar{\lambda}$ con orden de convergencia superior o igual a dos.*

$\varphi(\lambda) = \lambda^4, \lambda = 2E+6, \varphi'(2) = 3.2E+19$	
η	$\widetilde{\varphi}'$
1.0E+00	3.20000000E+19
1.0E-01	3.20000000E+19
1.0E-02	3.20000000E+19
1.0E-03	3.19999974E+19
1.0E-04	3.20000189E+19
1.0E-05	3.20002981E+19
1.0E-06	3.19996538E+19
1.0E-07	3.19652941E+19
1.0E-08	3.19975064E+19
1.0E-09	3.00647711E+19
1.0E-10	0.00000000E+00
⋮	⋮
1.0E-20	0.00000000E+00

Tabla 7.2

En [Baz93] hay condiciones más específicas para garantizar la convergencia.

Generalmente, las condiciones son favorables y el método de Newton converge muy rápidamente ya que su convergencia es cuadrática. Sin embargo, en otros casos sus inconvenientes son:

- puede no converger,
- puede converger a un maximizador,
- se requiere verificar que $\varphi''(\lambda_k) \neq 0$,
- es necesario calcular primeras y segundas derivadas de φ.

Ejemplo 7.1. Aplicar el método de Newton para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$.

k	λ_k	$\varphi(\lambda_k)$	$\varphi'(\lambda_k)$	$\varphi''(\lambda_k)$	λ_{k+1}
0	0.0000	17.0000	38.0000	66.0000	-0.5758
1	-0.5758	4.6436	7.1925	42.3416	-0.7456
2	-0.7456	4.0056	0.4736	36.8815	-0.7585
3	-0.7585	4.0026	0.0025	36.4969	-0.7585
4	-0.7585	4.0026	0.0000	36.4949	-0.7585

$\varphi(\lambda) = 5120\lambda^{1/10}, \lambda = 1024, \varphi'(1024) = 1$	
η	$\widetilde{\varphi}'$
1.0E+00	1.00000027E+00
1.0E-01	1.00000000E+00
1.0E-02	1.00000000E+00
1.0E-03	9.99999999E-01
1.0E-04	1.00000000E+00
1.0E-05	9.99999975E-01
1.0E-06	1.00000034E+00
1.0E-07	9.99998520E-01
1.0E-08	9.99989425E-01
1.0E-09	1.00044417E+00
1.0E-10	1.00044417E+00
1.0E-11	1.00044417E+00
1.0E-12	9.09494702E-01
1.0E-13	0.00000000E+00
⋮	⋮
1.0E-307	0.00000000E+00
1.0E-308	división por 0.

Tabla 7.3

En la tabla de resultados aparece la columna de $\varphi(\lambda_k)$ simplemente para ver que el valor de la función φ de este ejemplo va disminuyendo.

Conociendo el valor de $\bar{\lambda} = -0.758533053$, se puede comprobar la convergencia cuadrática del método de Newton.

k	λ_k	$\lvert\lambda_k\text{-}\bar{\lambda}\rvert/\lvert\lambda_{k\text{-}1}\text{-}\bar{\lambda}\rvert^2$	
0	0.000000000		
1	-0.575757559	2.31900354	
2	-0.745625404	0.84487552	
3	-0.758465288	0.67049867	
4	-0.758533051	0.65922594	
5	-0.758533053	0.65916705	
6	-0.758533053	0.65916705	$\diamond$

7.3 MÉTODO DE LA SECANTE

Es una aproximación al método de Newton. También se busca un punto $\bar{\lambda}$ donde la primera derivada de φ se anule. No utiliza segundas derivadas $\varphi''(\lambda_k)$, utiliza únicamente valores de la primera derivada de φ. La fórmula del método de la secante se puede obtener simplemente reemplazando, en la fórmula (7.6) del método de Newton, $\varphi''(\lambda_k)$ por una aproximación

$$\varphi''(\lambda_k) \approx \frac{\varphi'(\lambda_k) - \varphi'(\lambda_{k\text{-}1})}{\lambda_k - \lambda_{k\text{-}1}}.$$

Entonces la fórmula del método de la secante para "buscar" un minimizador es:

$$\begin{aligned} \lambda_{k+1} &= \lambda_k - \frac{\varphi'(\lambda_k)}{\dfrac{\varphi'(\lambda_k) - \varphi'(\lambda_{k\text{-}1})}{\lambda_k - \lambda_{k\text{-}1}}} \\ &= \frac{\lambda_{k\text{-}1}\varphi'(\lambda_k) - \lambda_k\varphi'(\lambda_{k\text{-}1})}{\varphi'(\lambda_k) - \varphi'(\lambda_{k\text{-}1})}. \end{aligned} \tag{7.7}$$

Se puede mostrar, [Ham91] o [Atk78], que la convergencia del método de la secante es de orden $(1+\sqrt{5})/2 \approx 1.618$, o sea, la convergencia es bastante buena, pero no tanto como en el método de Newton. Las desventajas del método de la secante son:

- puede no converger,
- puede converger a un maximizador,
- se requiere verificar que $\varphi'(\lambda_k) - \varphi'(\lambda_{k\text{-}1}) \neq 0$,
- es necesario calcular primeras derivadas de φ.

Ejemplo 7.2. Aplicar el método de la secante para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$.

k	λ_k	$\varphi'(\lambda_k)$	λ_{k+1}	
0	0.0000	38.0000		
1	0.1000	44.8440	-0.5552	
2	-0.5552	8.0688	-0.6990	
3	-0.6990	2.2265	-0.7538	
4	-0.7538	0.1737	-0.7584	
5	-0.7584	0.0042	-0.7585	
6	-0.7585	0.0000	-0.7585	
7	-0.7585	0.0000	-0.7585	$\diamond$

La única ventaja del método de la secante con respecto al de Newton es que no necesita evaluar segundas derivadas. Como en ambos métodos se necesita el valor de las primeras derivadas, estos dos métodos no son muy utilizados (salvo casos específicos) para minimizar funciones $\varphi(\lambda) = f(x + \lambda d)$.

7.4 MÉTODO DE NEWTON CON DERIVACIÓN NUMÉRICA

El método de Newton se puede utilizar, utilizando simplemente valores de la función para aproximar numéricamente las dos derivadas:

$$\begin{aligned}
\lambda_{k+1} &= \lambda_k - \frac{\dfrac{\varphi(\lambda_k+\eta) - \varphi(\lambda_k-\eta)}{2\eta}}{\dfrac{\varphi(\lambda_k+\eta) - 2\varphi(\lambda_k) + \varphi(\lambda_k-\eta)}{\eta^2}} \\
\lambda_{k+1} &= \lambda_k - \frac{\eta}{2}\,\frac{\varphi(\lambda_k+\eta) - \varphi(\lambda_k-\eta)}{\varphi(\lambda_k+\eta) - 2\varphi(\lambda_k) + \varphi(\lambda_k-\eta)}.
\end{aligned}$$

Como este método es casi el método de Newton, entonces cuando se dan las condiciones adecuadas converge rápidamente. También hereda la mayoría de la desventajas:

- puede no converger,
- puede converger a un máximo,
- problemas con un denominador nulo o casi nulo,
- se necesita escoger adecuadamente el valor de η.

Ejemplo 7.3. Aplicar el método de Newton con derivación numérica para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $\lambda_0 = 0$, $\eta = 0.1$, $\varepsilon = 0.00005$.

k	λ_k	$\varphi(\lambda_k)$	$\varphi(\lambda_k + \eta)$	$\varphi(\lambda_k\text{-}\eta)$
0	0.0000	17.0000	21.1381	13.5221
1	-0.5768	4.6362	5.5683	4.1272
2	-0.7470	4.0050	4.2365	4.1421
3	-0.7598	4.0026	4.1852	4.1848
4	-0.7599	4.0026	4.1850	4.1850

Es interesante notar que hubo convergencia hacia -0.7599 y no hacia -0.7585 como en el método de Newton exacto o en el método de la secante. Si se escoge $\varepsilon = 10^{-10}$ se hace una iteración más, pero de todas maneras la convergencia se da hacia -0.7599. La razón es que para este ejemplo $\eta = 0.1$ es demasiado grande y las aproximaciones de las derivadas no son suficientemente buenas. Con $\eta = 0.01$ la convergencia (con 4 decimales) sí se da hacia -0.7585. Es bueno recordar que valores de η muy pequeños pueden dar lugar a numeradores nulos o casi nulos (entonces $\lambda_{k+1} \approx \lambda_k$) o a denominadores nulos o casi nulos. $\diamond$

Ejemplo 7.4. Aplicar el método de Newton con derivación numérica para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $\lambda_0 = 0$, $\eta = 0.01$, $\varepsilon = 0.00005$.

k	λ_k	$\varphi(\lambda_k)$	$\varphi(\lambda_k + \eta)$	$\varphi(\lambda_k\text{-}\eta)$	
0	0.0000	17.0000	17.3833	16.6233	
1	-0.5758	4.6435	4.7176	4.5737	
2	-0.7456	4.0056	4.0122	4.0027	
3	-0.7585	4.0026	4.0044	4.0044	
4	-0.7585	4.0026	4.0044	4.0044	$\diamond$

Aquí como en los demás ejemplos, los cálculos intermedios se hacen en doble precisión y los resultados tabulados son aproximaciones con un número fijo de cifras decimales. O sea, $\lambda_2 = -0.7456$ no se obtiene con los valores tabulados de la fila $k = 1$, sino con los valores internos. Con los valores tabulados se obtendría -0.7431. $\diamond$

7.5 MÉTODOS DE ENCAJONAMIENTO (BRACKETING)

Cuando se desea minimizar $\varphi(\lambda)$ en $[a, b]$, se le da al intervalo $[a, b]$ el nombre de **intervalo de incertidumbre**. La mayoría de los métodos van construyendo intervalos de incertidumbre anidados, $[a_{k+1}, b_{k+1}] \subseteq [a_k, b_k]$, y de tamaño cada vez menor, hasta obtener un intervalo tan pequeño como se desee. Para que el método sea adecuado se necesitan dos cosas: primero que el tamaño de los intervalos tienda a cero, y segundo, que en cada subintervalo haya por lo menos un minimizador global. Si la función φ es cuasiconvexa, entonces se puede cumplir la segunda condición mediante la adecuada aplicación del siguiente resultado:

Proposición 7.2. *Sean: φ^* el valor mínimo de φ en el intervalo $\Lambda = [a, b]$, Λ^* el conjunto de minimizadores (globales), $a < \sigma < \tau < b$, $\Lambda' = [a, \tau]$ si $\varphi(\sigma) < \varphi(\tau)$, $\Lambda' = [\sigma, b]$ si $\varphi(\sigma) \geq \varphi(\tau)$. Si φ es cuasiconvexa en λ, entonces $\Lambda^* \cap \Lambda' \neq \emptyset$. Dicho de otra forma, el valor mínimo de φ en el intervalo Λ' también es φ^*.*

Si hay un solo minimizador λ^* en Λ, entonces necesariamente λ^* está en Λ'. La proposición anterior se puede generalizar a $m \geq 2$ puntos intermedios.

Proposición 7.3. *Sean: φ^* el valor mínimo de φ en el intervalo $\Lambda = [a, b]$, Λ^* el conjunto de minimizadores (globales), $a = \mu_0 < \mu_1 < ... < \mu_m < \mu_{m+1} = b$, μ_j el mejor punto (menor valor de φ) entre $\mu_1, ..., \mu_m$, $\Lambda' = [\mu_{j-1}, \mu_{j+1}]$. Si φ es cuasiconvexa en Λ, entonces $\Lambda^* \cap \Lambda' \neq \emptyset$.*

Los dos métodos siguientes son aplicaciones de las dos proposiciones anteriores donde además se garantiza que el tamaño de los intervalos tiende a cero. En el método de la sección dorada se disminuye el número de evaluaciones de la función en cada iteración.

7.6 BÚSQUEDA SECUENCIAL

Este método sirve para minimizar funciones estrictamente cuasiconvexas en un intervalo $[a, b]$. Dado un intervalo de incertidumbre $[a_k, b_k]$ se evalúa la función en $m \geq 2$ puntos intermedios igualmente espaciados $\mu_1, \mu_2, .., \mu_m$, se escoge el mejor punto, o sea, el μ_j donde la función φ toma el menor valor, y así el nuevo intervalo estará centrado en él.

$$\delta = \frac{b_k - a_k}{m+1},$$

$$\begin{aligned} \mu_j &= a_k + j\delta, j = 1, ..., m, \\ \mu^* &= \text{ mejor } \{\mu_1, ..., \mu_m\}, \\ a_{k+1} &= \mu^* - \delta, \\ b_{k+1} &= \mu^* + \delta. \end{aligned}$$

El esquema de algoritmo sería el siguiente:

datos: $a,\ b,\ \varepsilon,\ m$
$l = b - a$
mientras $l \geq \varepsilon$
    $\delta = l/(m+1)$
    $\mu_j = a + j,\ j = 1, ..., m$
    $\mu^* = \text{mejor}\{\mu_1, ..., \mu_m\}$
    $a = \mu^* - \delta$
    $b = \mu^* + \delta$
    $l = b - a$
fin-mientras
$\lambda^* = \text{mejor}\{a, \mu^*, b\}$

Ejemplo 7.5. Aplicar el método de búsqueda secuencial para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $[a, b] = [-1, 0]$, $m = 2$, $\varepsilon = 0.1$.

k	a_k	b_k	μ_1	$\varphi(\mu_1)$	μ_2	$\varphi(\mu_2)$	μ^*
0	-1.000	0.000	-0.667	4.160	-0.333	7.716	-0.667
1	-1.000	-0.333	-0.778	4.009	-0.556	4.798	-0.778
2	-1.000	-0.556	-0.852	4.158	-0.704	4.058	-0.704
3	-0.852	-0.556	-0.753	4.003	-0.654	4.206	-0.753
4	-0.852	-0.654	-0.786	4.016	-0.720	4.030	-0.786
5	-0.852	-0.720	-0.808	4.047	-0.764	4.003	-0.764
6	-0.808	-0.720					

$$\lambda^* \approx 0.764$$

Ejemplo 7.5

El método de búsqueda secuencial puede ser usado con $m = 2$ o con $m = 100$. Para escoger el valor de m se necesita evaluar el tiempo total que dura el algoritmo para hallar un intervalo de incertidumbre suficientemente pequeño. La mayoría del tiempo se gasta en evaluaciones de la función y en comparaciones para hallar μ^*. El número total de evaluaciones es

igual al número de iteraciones multiplicado por el número de evaluaciones de la función en cada iteración. Obviamente con m grande el número de iteraciones es menor, pero el número de evaluaciones por iteración es mayor. Sea k el número de iteraciones.

$$\begin{aligned} b_k - a_k &= \frac{2}{m+1}(b_{k\text{-}1} - a_{k\text{-}1}) \\ &= \left(\frac{2}{m+1}\right)^k (b_0 - a_0). \end{aligned}$$

Como se desea que $b_k - a_k \leq \varepsilon$,

$$\begin{aligned} \left(\frac{2}{m+1}\right)^k &\leq \frac{\varepsilon}{b_0 - a_0} \\ k \log \frac{2}{m+1} &\leq \log \frac{\varepsilon}{b_0 - a_0}. \end{aligned}$$

Como $\log(2/(m+1)) < 0$,

$$\begin{aligned} k &\geq \frac{\log \dfrac{\varepsilon}{b_0 - a_0}}{\log \dfrac{2}{m+1}} \\ &= \frac{\log \dfrac{b_0 - a_0}{\varepsilon}}{\log \dfrac{m+1}{2}}. \end{aligned}$$

Sea

$$L_0 = \frac{b_0 - a_0}{\varepsilon},$$

entonces

$$k = \lceil \frac{L_0}{\log \dfrac{m+1}{2}} \rceil. \tag{7.8}$$

$$\text{núm. total evaluaciones} = mk,$$

$$\text{núm. total evaluaciones} \;\approx\; \frac{m}{\log \dfrac{m+1}{2}} L_0 \;=\; c(m)\, L_0. \qquad (7.9)$$

Si la evaluación de la función φ es demorada con respecto a las demás operaciones y comparaciones, se debe escoger el valor de m que minimice el número total de evaluaciones. Si no se tiene en cuenta la parte entera superior en (7.8), o sea, si se minimiza la aproximación del número total de evaluaciones (7.9), el valor óptimo de m sería $= 3.311$. Como obviamente m debe ser entero se puede aproximar por $m = 3$. Para $m = 2$ o $m = 4$ el número total de evaluaciones no se altera mucho. De todas maneras es claro, según la fórmula, que para valores grandes de m el número total de evaluaciones también se hace grande.

m	c(m)
2	4.93
3	4.33
4	4.37
5	4.55
6	4.79
7	5.05

Afinando un poquito se puede observar que cuando m es impar el μ_j que queda en el punto medio del intervalo $[a_k, b_k]$ coincide con el μ^* del intervalo anterior, luego en ese caso se hacen $m - 1$ evaluaciones por iteración.

m	c(m)
2	4.93
3	2.89
4	4.37
5	3.64
6	4.79
7	4.33

En ambos casos se ve que el valor $m = 3$ es la mejor opción, cuando lo más costoso es evaluar la función. Esto sucede en la mayoría de los casos. Si no hay seguridad sobre la cuasiconvexidad de la función, puede ser mejor utilizar un valor grande de m, para evitar cambios bruscos no previstos en un subintervalo y tener así más posibilidad de obtener un minimizador

7.7 SECCIÓN DORADA O ÁUREA

Este método sirve para minimizar funciones estrictamente cuasiconvexas o unimodales en un intervalo $[a, b]$. Dado un intervalo de incertidumbre $[a_k, b_k]$ se evalúa la función en dos puntos intermedios σ_k, τ_k y de manera análoga a la búsqueda secuencial, el nuevo intervalo estará dado por los vecinos del mejor punto. Hay dos diferencias importantes: la primera consiste en que los puntos no están igualmente espaciados, y la segunda, la más importante, los puntos σ_k, τ_k se escogen de tal manera que se haga una sola evaluación de la función por iteración.

En la deducción del método se hacen suposiciones muy razonables:

$$\begin{aligned} & a_k < \sigma_k < \tau_k < b_k, \\ & \sigma_k - a_k = b_k - \tau_k \quad \text{(simetría)}. \end{aligned}$$

Es necesario definir que tan cerca o que tan lejos están los puntos intermedios de los extremos.

$$\begin{aligned} \sigma_k &= a_k + \alpha(b_k - a_k)\ , \quad \text{con} \quad 0 < \alpha < \frac{1}{2}, \\ \tau_k &= b_k - \alpha(b_k - a_k), \\ &= a_k + (1-\alpha)(b_k - a_k). \end{aligned}$$

Para deducir el valor de α se puede suponer que

$$\varphi(\sigma_k) \leq \varphi(\tau_k),$$

entonces el nuevo intervalo es

$$[a_{k+1}, b_{k+1}] = [a_k, \tau_k]$$

En este nuevo intervalo está σ_k y se busca que uno de los puntos σ_{k+1}, τ_{k+1} coincida con σ_k, así basta con evaluar la función una sola vez por iteración, en el punto que no coincidió. En resumen hay dos posibilidades: $\sigma_{k+1} = \sigma_k$, o bien $\tau_{k+1} = \sigma_k$.

Si $\sigma_{k+1} = \sigma_k$

$$a_{k+1} + \alpha(b_{k+1} - a_{k+1}) \quad = \quad a_k + \alpha(b_k - a_k)$$

$$\begin{aligned} a_k + \alpha(\tau_k - a_k) &= a_k + \alpha(b_k - a_k) \\ a_k + \alpha(1-\alpha)(b_k - a_k) &= a_k + \alpha(b_k - a_k), \end{aligned}$$

entonces

$$\begin{aligned} \alpha(1-\alpha) &= \alpha \\ \alpha &= 0 \quad \text{valor inadecuado.} \end{aligned}$$

Si $\tau_{k+1} = \sigma_k$

$$\begin{aligned} a_{k+1} + (1-\alpha)(b_{k+1} - a_{k+1}) &= a_k + \alpha(b_k - a_k) \\ a_k + (1-\alpha)(\tau_k - a_k) &= a_k + \alpha(b_k - a_k) \\ a_k + (1-\alpha)(1-\alpha)(b_k - a_k) &= a_k + \alpha(b_k - a_k), \end{aligned}$$

entonces

$$\begin{aligned} (1-\alpha)^2 &= \alpha \\ \alpha &= \begin{cases} \dfrac{3+\sqrt{5}}{2} & \text{valor inadecuado} \\ \dfrac{3-\sqrt{5}}{2} \approx 0.381966011. \quad \surd \end{cases} \end{aligned}$$

De manera análoga, si

$$\varphi(\sigma_k) > \varphi(\tau_k),$$

el nuevo intervalo es

$$\begin{aligned} [a_{k+1}, b_{k+1}] &= [\sigma_k, b_k], \\ \sigma_{k+1} &= \tau_k. \end{aligned}$$

El esquema del algoritmo sería el siguiente:

Ejemplo 7.6. Aplicar el método de la sección áurea para minimizar la función $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $[a, b] = [-1, 0]$, $\varepsilon = 0.1$.

datos: $a,\ b,\ \varepsilon$
$l = b - a, \quad \delta = \alpha l$
$\sigma = a + \delta, \quad \tau = b - \delta$
$\texttt{fs} = \varphi(\sigma), \quad \texttt{ft} = \varphi(\tau)$
mientras $l \geq \varepsilon$
  $l = (1 - \alpha)l, \quad \delta = \alpha l$
  si $\texttt{fs} \leq \texttt{ft}$ **ent**
    $b = \tau, \quad \tau = \sigma, \quad \sigma = a + \delta$
    $\texttt{ft} = \texttt{fs}, \quad \texttt{fs} = \varphi(\sigma)$
  fin-ent
  sino
    $a = \sigma, \quad \sigma = \tau, \quad \tau = b - \delta$
    $\texttt{fs} = \texttt{ft}, \quad \texttt{ft} = \varphi(\tau)$
  fin-sino
fin-mientras
$\lambda^* = \text{mejor}\{a, \sigma, \tau, b\}$

k	a_k	b_k	σ_k	$\varphi(\sigma_k)$	τ_k	$\varphi(\tau_k)$
0	-1.000	0.000	-0.618	4.377	-0.382	6.875
1	-1.000	-0.382	-0.764	4.003	-0.618	4.377
2	-1.000	-0.618	-0.854	4.165	-0.764	4.003
3	-0.854	-0.618	-0.764	4.003	-0.708	4.049
4	-0.854	-0.708	-0.798	4.031	-0.764	4.003
5	-0.798	-0.708	-0.764	4.003	-0.743	4.073

El número total de evaluaciones de la función en el método de la sección dorada se calcula de manera semejante al método de búsqueda secuencial.

$$\begin{aligned} b_k - a_k &= (1 - \alpha)(b_{k\text{-}1} - a_{k\text{-}1}) \\ &= (1 - \alpha)^k (b_0 - a_0). \end{aligned}$$

Como se desea que $b_k - a_k \leq \varepsilon$, entonces

$$(1 - \alpha)^k \leq \frac{\varepsilon}{b_0 - a_0}.$$

Tomando logaritmo en ambos lados de la desigualdad

$$
\begin{aligned}
k\log(1-\alpha) &\leq \log\frac{\varepsilon}{b_0-a_0},\\
k &\geq \frac{\log\dfrac{\varepsilon}{b_0-a_0}}{\log(1-\alpha)}\\
k &\geq \frac{\log\dfrac{b_0-a_0}{\varepsilon}}{-\log(1-\alpha)} = \frac{L_0}{-\log(1-\alpha)},\\
k &= \lceil 2.0781 L_0 \rceil,
\end{aligned}
$$

núm. total evaluaciones $\approx 2.0871\, L_0$.

7.8 MINIMIZACIÓN POR INTERPOLACIÓN CUADRÁTICA

Para hallar λ^* minimizador de φ en $\mathbb{R}$ se puede construir una parábola que pase por tres puntos conocidos y aproximar λ^* por el punto donde se anula la derivada de la parábola. Obviamente si la parábola es estrictamente convexa (hacia arriba) el punto donde se anula la derivada es exactamente el minimizador de la parábola.

Partiendo de un λ_0 (generalmente $\lambda_0 = 0$), se obtienen tres puntos t_1, t_2, t_3, se construye la parábola $p(t)$ que pase por $(t_1, \varphi(t_1))$, $(t_2, \varphi(t_2))$, $(t_3, \varphi(t_3))$ y se aproxima λ^* por t' donde se anula $p'(t)$. Si φ es estrictamente convexa, entonces p es estrictamente convexa. Pero puede suceder que φ no sea estrictamente convexa y que p sí lo sea. En el peor de los casos, en que p no es estrictamente convexa, tratar de minimizar por interpolación cuadrática, puede traer, sin embargo, un resultado aceptable si por lo menos uno de los puntos t_1, t_2, t_3, t' es mejor que λ_0.

Si se necesita bastante precisión, aun a costa de gastar mucho más tiempo, se puede aplicar varias veces la interpolación cuadrática de la siguiente manera: partiendo de λ_0 obtener t_1, t_2, t_3, t' y el mejor de estos cuatro puntos será λ_1; partiendo de λ_1 obtener otros t_1, t_2, t_3, t' y el mejor de estos cuatro puntos será λ_2 y así sucesivamente.

Cuando φ es cuadrática, entonces $p(t)$ y $\varphi(t)$ coinciden. Si además φ es estrictamente convexa, el minimizador de p es exactamente λ^*.

7.8.1 Interpolación cuadrática

Sea φ una función de variable y valor real. Sean: t_1, t_2, t_3 tres valores diferentes, y $\varphi_1, \varphi_2, \varphi_3$ los valores de φ en estos tres puntos. O sea,

$$\varphi_i = \varphi(t_i), \quad i = 1, 2, 3.$$

El polinomio de grado no superior a dos (muy posiblemente una parábola) que pasa por los puntos (t_i, φ_i) es el siguiente:

$$\begin{aligned} p(t) &= \varphi_1 \frac{(t-t_2)(t-t_3)}{(t_1-t_2)(t_1-t_3)} + \varphi_2 \frac{(t-t_1)(t-t_3)}{(t_2-t_1)(t_2-t_3)} + \varphi_3 \frac{(t-t_1)(t-t_2)}{(t_3-t_1)(t_3-t_2)} \\ &= \frac{1}{d}\big[t^2(\varphi_1(t_3-t_2) - \varphi_2(t_3-t_1) + \varphi_3(t_2-t_1)) \\ &\quad + t(-\varphi_1(t_3^2-t_2^2) + \varphi_2(t_3^2-t_1^2) - \varphi_3(t_2^2-t_1^2)) \\ &\quad + \varphi_1(t_3-t_2)t_3t_2 - \varphi_2(t_3-t_1)t_3t_1 + \varphi_3(t_2-t_1)t_2t_1\big], \\ &= \frac{1}{d}\,[at^2 + bt + c] \end{aligned}$$

donde

$$\begin{aligned} d &= (t_2-t_1)(t_3-t_1)(t_3-t_2), \\ a &= \varphi_1(t_3-t_2) - \varphi_2(t_3-t_1) + \varphi_3(t_2-t_1), \\ b &= -\varphi_1(t_3^2-t_2^2) + \varphi_2(t_3^2-t_1^2) - \varphi_3(t_2^2-t_1^2), \\ c &= \varphi_1(t_3-t_2)t_3t_2 - \varphi_2(t_3-t_1)t_3t_1 + \varphi_3(t_2-t_1)t_2t_1. \end{aligned}$$

Además de necesitar que los tres valores t_1, t_2, t_3 sean diferentes se puede suponer que están ordenados, es decir: $t_1 < t_2 < t_3$. En este caso t_2 se puede expresar como combinación convexa estricta de t_1, t_3:

$$t_2 = (1-\nu)t_1 + \nu t_3\ , \quad 0 < \nu < 1.$$

Para saber si $p(t)$ es una parábola estrictamente convexa (tiene mínimo), una parábola degenerada (una recta) o una parábola estrictamente cóncava, basta con estudiar el signo de a, numerador del coeficiente de t^2, ya que $d > 0$.

$$\begin{aligned} a &= \varphi_1(t_3-t_1)(1-\nu) - \varphi_2(t_3-t_1) + \varphi_3(t_3-t_1)\nu \\ &= (t_3-t_1)\big(\varphi_1(1-\nu) - \varphi_2 + \varphi_3\nu\big). \end{aligned}$$

Es claro que:

$$\begin{aligned} a &> 0 \quad \text{si} \quad \varphi_2 < (1-\nu)\varphi_1 + \nu\varphi_3 \ , \\ a &= 0 \quad \text{si} \quad \varphi_2 = (1-\nu)\varphi_1 + \nu\varphi_3 \ , \\ a &< 0 \quad \text{si} \quad \varphi_2 > (1-\nu)\varphi_1 + \nu\varphi_3 \ . \end{aligned}$$

Estos resultados se interpretan geométricamente de la siguiente forma: $p(t)$ es una parábola esctrictamente convexa si el punto (t_2, φ_2) queda por debajo del segmento de recta que une los puntos (t_1, φ_1), (t_3, φ_3) y es una recta si el punto (t_2, φ_2) queda en ese segmento de recta. Un caso particular de parábola estrictamente convexa se tiene cuando $\varphi_2 < \varphi_1, \varphi_3$. Si $t_1 > t_2 > t_3$ y $\varphi_2 < \varphi_1, \varphi_3$ también se obtiene una parábola estrictamente convexa.

7.8.2 Cálculo del minimizador de la parábola.

Para el caso de una parábola estrictamente convexa, su minimizador $t' = t_p^*$ está dado por $-b/2a$, o sea,

$$t_p^* = \frac{1}{2} \frac{s_{23}\varphi_1 + s_{31}\varphi_2 + s_{12}\varphi_3}{t_{23}\varphi_1 + t_{31}\varphi_2 + t_{12}\varphi_3}, \tag{7.10}$$

donde

$$\begin{aligned} t_{ij} &= t_i - t_j \ , \\ s_{ij} &= t_i^2 - t_j^2. \end{aligned}$$

Conocidos los valores t_i, φ_i, para el cálculo de t_p^* es necesario efectuar 10 sumas o restas, 10 multiplicaciones y 1 división.

Otra expresión equivalente, pero con un número más pequeño de operaciones (7 sumas o restas, 5 multiplicaciones y 1 división) es la siguiente:

$$t_p^* = t_2 - \frac{(t_2 - t_1)^2(\varphi_2 - \varphi_3) - (t_2 - t_3)^2(\varphi_2 - \varphi_1)}{2[(t_2 - t_1)(\varphi_2 - \varphi_3) - (t_2 - t_3)(\varphi_2 - \varphi_1)]}. \tag{7.11}$$

Para el caso particular de puntos igualmente espaciados: $t_2 = t_1 + \eta$, $t_3 = t_2 + \eta$,

$$t_p^* = t_2 - \frac{\eta}{2} \ \frac{\varphi_3 - \varphi_1}{\varphi_1 - 2\varphi_2 + \varphi_3}. \tag{7.12}$$

Obsérvese que la fórmula anterior es la misma del método de Newton con aproximación numérica de las derivadas. Otro caso particular se tiene cuando: $t_2 = t_1 + \eta$, $t_3 = t_2 + 2\eta$,

$$t_p^* = t_2 - \frac{\eta}{2} \frac{4\varphi_1 - 3\varphi_2 - \varphi_3}{\text{-}2\varphi_1 + 3\varphi_2 - \varphi_3}. \tag{7.13}$$

7.9 MÉTODO DE LOS TRES PUNTOS PARA $\lambda \in \mathbb{R}$

La idea de este método de minimización, sobre toda la recta real, es muy sencilla. Se trata de encontrar, a partir de λ_k, tres valores adecuados t_1, t_2, t_3 que permitan obtener el minimizador t_p^* de la parábola que pasa por los tres puntos y escoger enseguida como λ_{k+1} el mejor de los cuatro valores t_1, t_2, t_3, t_p^*. Se espera así construir a partir de λ_0 una sucesión $\lambda_1, \lambda_2, \ldots$ tal que $\lambda_k \to \lambda^*$.

Los tres valores t_1, t_2, t_3 son adecuados si por lo menos uno de ellos es mejor que λ_k (para algún i $\varphi_i = \varphi(t_i) < \varphi(\lambda_k)$) y si además la parábola de interpolación es estrictamente convexa. Esto se logra, por ejemplo, si $\varphi_2 < \varphi_1, \varphi_3, \varphi(\lambda_k)$. Así t_p^* queda en el intevalo $]t_1, t_3[$ y el mejor de los cuatro puntos es simplemente el mejor entre t_2,t_p^*. Esto garantiza que $\varphi(\lambda_{k+1}) < \varphi(\lambda_k)$. Obviamente si dos valores t_i son mejores que λ_k, entonces posiblemente λ_{k+1}será mucho mejor.

El primer paso consiste simplemente en encontrar un punto t_2 mejor que λ_k. Se puede empezar con $\lambda_k + \eta$. Si éste no es mejor se averigua si λ_k-η es mejor. Si ninguno de los dos es mejor entonces se reduce el tamaño de η. El proceso continúa hasta encontrar un punto mejor que λ_k, o bien hasta que η sea casi nulo, lo que indica que λ_k es posiblemente un minimizador local.

datos: $\lambda_k,\ \varepsilon,\ \eta_0,\ 0 < \varepsilon < |\eta_0|, \alpha \in]0,1[$
`exito` $= 0,\ \eta = \eta_0$
mientras `exito` $= 0$ y $|\eta| > \varepsilon$
  si $\varphi(\lambda_k + \eta) < \varphi(\lambda_k)$ **ent** `exito` $= 1$
  sino
    $\eta = -\eta$
    si $\varphi(\lambda_k + \eta) < \varphi(\lambda_k)$ **ent** `exito` $= 1$
    sino $\eta = -\alpha\eta$
  fin-sino
fin-mientras
si `exito` $= 1$ **ent** $t_1 = \lambda_k,\ \ t_2 = \lambda_k + \eta$

El valor de α puede ser simplemente $\alpha = 1/2$. El valor de η_0 utilizado como dato para la búsqueda de un punto mejor que λ_k puede ser constante o puede estar relacionado con el cambio en λ, por ejemplo $\gamma(\lambda_k - \lambda_{k\text{-}1})$, con $\gamma = 1/2$ o cualquier otro valor mayor que cero. Para que cumpla con la condición $\varepsilon < |\eta_0|$ se puede utilizar $\eta_0 = \text{dom}\{\gamma(\lambda_k\text{-}\lambda_{k\text{-}1}), 2\varepsilon\}$.

Después de haber encontrado un punto mejor que λ_k, se trata de buscar puntos cada vez mejores hasta encontrar un punto peor que el mejor punto obtenido. Si $\lambda_k + \eta$ es mejor que λ_k se averigua si $\lambda_k + 2\eta$ es aún mejor y después se estudia $\lambda_k + 4\eta$ y así sucesivamente hasta encontrar uno peor. Esto se puede lograr mediante el siguiente algoritmo.

datos: $t_1,\ t_2,\ R$ suficientemente grande
$\eta = t_2 - t_1$
$\beta = 2$ o cualquier valor mayor que 1
`exito` $= 0$
mientras `exito` $= 0$ y $|t_2 + \eta| < R$
    $\eta = \beta\eta$
    si $\varphi(t_2 + \eta) < \varphi(t_2)$ **ent** $t_1 = t_2,\ \ t_2 = t_2 + \eta$
    sino $t_3 = t_2 + \eta,\ \ $ `exito` $= 1$
fin-mientras

Al final de este algoritmo, si `exito`=1 hay tres puntos adecuados t_1, t_2, t_3, y a continuación es necesario calcular t_p^* minimizador de la parábola y escoger el mejor entre t_2 y t_p^*. Por otro lado, `exito`=0 indica que no fue posible encontrar los tres puntos y que, para los valores utilizados para la variable independiente t, posiblemente la función no está acotada inferiormente y no tiene minimizador.

Ejemplo 7.7. Aplicar el método de los tres puntos para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $\lambda_0 = 0$.

k	λ_k $\varphi(\lambda_k)$	t mejor $\varphi(t)$	t_1 $\varphi(t_1)$	t_2 $\varphi(t_2)$	t_3 $\varphi(t_3)$	t_p^* $\varphi(t_p^*)$
0	0.0000	-0.1000	-0.3000	-0.7000	-1.5000	-0.8062
	17.0000	13.5221	8.3621	4.0661	12.3125	4.0434
1	-0.8062	-0.7558	-0.8062	-0.7558	-0.6550	-0.7592
	4.0434	4.0027	4.0434	4.0027	4.2038	4.0026
2	-0.7592	-0.7584	-0.7592	-0.7584	-0.7570	-0.7585
	4.0026	4.0026	4.0026	4.0026	4.0026	4.0026
3	-0.7585					
	4.0026					$\diamond$

7.10 MÉTODO DE LOS TRES PUNTOS PARA $\lambda \geq 0$

Se trata simplemente de adaptar el método anterior de tal forma que siempre $\lambda_k \geq 0$. Se desea entonces encontrar, a partir de λ_k, tres valores t_1, t_2, t_3, no negativos, tales que t_2 es el punto intermedio, y es mejor que los otros dos.

El primer paso consiste simplemente en encontrar un punto $t_2 \geq 0$ mejor que λ_k. Esta búsqueda se puede esquematizar así:

datos: $\lambda_k,\ \varepsilon,\ \eta_0,\ 0 < \varepsilon < |\eta_0|, \alpha = \in]0,1[$
$\texttt{exito} = 0, \quad \eta = \eta_0$
mientras $\texttt{exito} = 0$ y $|\eta| > \varepsilon$
  si $\lambda_k + \eta \geq 0$ y $\varphi(\lambda_k + \eta) < \varphi(\lambda_k)$ **ent** $\texttt{exito} = 1$
  sino
    $\eta = -\eta$
    si $\lambda_k + \eta \geq 0$ y $\varphi(\lambda_k + \eta) < \varphi(\lambda_k)$ **ent** $\texttt{exito} = 1$
    sino $\eta = -\alpha\eta$
  fin-sino
fin-mientras
si $\texttt{exito} = 1$ **ent** $t_1 = \lambda_k, \quad t_2 = \lambda_k + \eta$

Al final del algoritmo anterior, si $\texttt{exito} = 0$, entonces posiblemte λ_k es un minimizador.

Después de haber encontrado un punto, no negativo, mejor que λ_k, se trata de buscar puntos, no negativos, cada vez mejores hasta encontrar un punto peor que el mejor punto obtenido.

datos: $: t_1 \geq 0,\ t_2 \geq 0,\ \varepsilon > 0,\ R$ suf. grande
$\eta = \text{dom}\{t_2\text{-}t_1, 2\varepsilon\}$
$\beta = 2$ o cualquier valor mayor que 1
`exito` $= 0$
mientras `exito` $= 0$ y $|t_2 + \eta| < R$ y $|\eta| \geq \varepsilon$,
  $\eta = \beta\eta$
  si $t_2 + \eta \geq 0$ **ent**
    si $\varphi(t_2 + \eta) < \varphi(t_2)$ **ent** $t_1 = t_2,\ \ t_2 = t_2 + \eta$
    sino $t_3 = t_2 + \eta,\ \ $ `exito` $= 1$
  fin-ent
  sino $\eta = \eta/\beta^2$
fin-mientras
si `exito` $= 1$ **ent** $t_p^* = \ldots,\ \ \lambda_{k+1} = \text{mejor}\{t_p^*, t_2\}$
sino
  si $|\eta| > \varepsilon$ **ent** posiblemente no hay minimizador
  sino $\lambda_{k+1} = t_2$
fin-sino

Ejemplo 7.8. Aplicar el método de los tres puntos para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (1, -4)$, con $\lambda \geq 0$, $\lambda_0 = 0$.

k	λ_k	t mejor	t_1	t_2	t_3	t_p^*
	$\varphi(\lambda_k)$	$\varphi(t)$	$\varphi(t_1)$	$\varphi(t_2)$	$\varphi(t_3)$	$\varphi(t_p^*)$
0	0.0000	0.1000	0.3000	0.7000	1.5000	0.8062
	17.0000	13.5221	8.3621	4.0661	12.3125	4.0434
1	0.8062	0.7255	0.8062	0.7255	0.5643	0.7592
	4.0434	4.0226	4.0434	4.0226	4.7287	4.0026
2	0.7592	0.7585	0.7591	0.7585	0.7574	0.7585
	4.0026	4.0026	4.0026	4.0026	4.0026	4.0026
3	0.7585					
	4.0026					◇

Ejemplo 7.9. Aplicar el método de los tres puntos para minimizar $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, a partir de $x = (-2, 3)$, en la dirección $d = (-1, 4)$, con $\lambda \geq 0$, $\lambda_0 = 0$.

$\lambda^* = 0.0000$. $\diamond$

7.11 INTERPOLACIÓN CUADRÁTICA EN UN INTERVALO $[a, b]$

Dada φ una función estrictamente cuasiconvexa, se desea hallar λ^* minimizador de φ en un intervalo $[a, b]$. Se trata de combinar adecuadamente la interpolación cuadrática con el método de búsqueda secuencial con 5 puntos intermedios no equidistantes. Los puntos se escogen para favorecer la tendencia que muestre la función, pero también asegurando que el tamaño del intervalo decrece de manera adecuada.

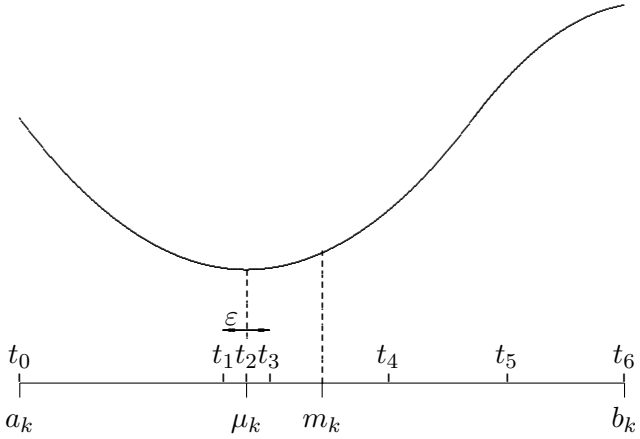


Figura 7.1

En la iteración k, dado el intervalo $[a_k, b_k]$, se calcula el punto medio m_k. Si este punto es mejor que a_k y b_k, entonces p_k, el minimizador de la parábola, queda en el interior del intervalo. Sea μ_k el mejor entre p_k y m_k. Si $\mu_k < m_k$ se hace $t_2 = \mu_k$, alrededor de t_2 se escogen 2 puntos muy cercanos, $t_1 = t_2 - \varepsilon/2$, $t_3 = t_2 + \varepsilon/2$ y, además, se escogen t_4, t_5 igualmente espaciados entre t_3 y b_k. Ver Figura 7.1 .

Si $\mu_k \geq m_k$ se hace $t_4 = \mu_k$, alrededor de t_4 se escogen 2 puntos muy cercanos, $t_3 = t_4 - \varepsilon/2$, $t_5 = t_4 + \varepsilon/2$ y, además, se escogen t_1, t_2 igualmente espaciados entre a_k y t_3. Ver Figura 7.2 .

Los extremos a_k, b_k corresponden a t_0, t_6. Como en la búsqueda secuencial, se escoge t_i el mejor de los cinco puntos interiores y el nuevo intervalo

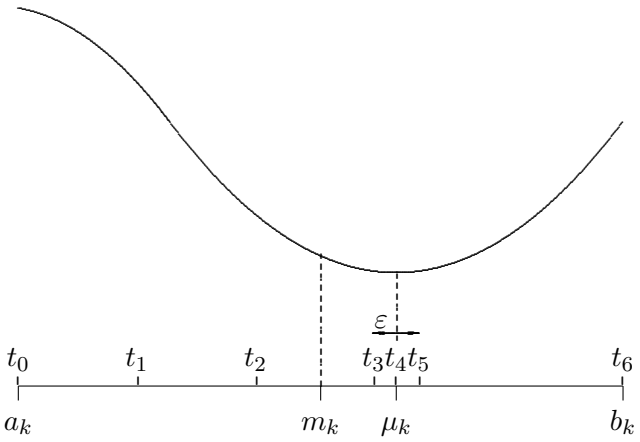


Figura 7.2

será $[t_{i-1}, t_{i+1}]$. Con buena suerte el mejor punto es μ_k y así el nuevo intervalo tiene longitud ε. En el peor de los casos, por ejemplo, μ_k está muy cerca de a_k, $t_i = t_5$ y $b_{k+1} - a_{k+1} = t_6 - t_4 = 2(b_k - t_3)/3 < 2(b_k - a_k)/3$.

Si m_k no es mejor que los extremos y si a_k es mejor que b_k, se supone cierta tendencia de que el minimizador esté cerca de a_k. Se escogen dos puntos muy cercanos a a_k: $t_1 = a_k + \varepsilon/2$, $t_2 = a_k + \varepsilon$. Los puntos t_3, t_4 y t_5 son puntos igualmente espaciados de $[a_k, b_k]$. Con buena suerte el mejor punto es t_1 y así el nuevo intervalo tiene longitud ε. En el peor de los casos $b_{k+1} - a_{k+1} = (b_k - a_k)/2$. Ver Figura 7.3 .

De manera análoga, si m_k no es mejor que los extremos y si b_k es mejor que a_k se supone cierta tendencia de que el minimizador esté cerca de b_k. Se escogen dos puntos muy cercanos a b_k: $t_5 = b_k - \varepsilon/2$, $t_4 = b_k - \varepsilon$. Los puntos t_1, t_2 y t_3 son puntos igualmente espaciados de $[a_k, b_k]$. Con buena suerte el mejor punto es t_5 y así el nuevo intervalo tiene longitud ε. En el peor de los casos $b_{k+1} - a_{k+1} = (b_k - a_k)/2$. Ver Figura 7.4 .

Ver esquema general del algoritmo al final del capítulo.

Ejemplo 7.10. Sean: $f(x_1, x_2) = 10(2x_1 + x_2)^4 + (x_1 - x_2)^2$, $x = (1, 2)$, $d = (0, -2)$. Hallar el minimizador de $\varphi(\lambda) = f(x + \lambda d)$, $\lambda \in [0, 2]$, con $\varepsilon = 0.01$.

Ver tabla de resultados al final del capítulo.

Ejemplo 7.11. Sean: $f(x_1, x_2) = 10(2x_1 + x_2)^4 + (x_1 - x_2)^2$, $x = (1, 2)$, $d = (0, -2)$. Hallar el minimizador de $\varphi(\lambda) = f(x + \lambda d)$, $\lambda \in [0, 1.5]$, con

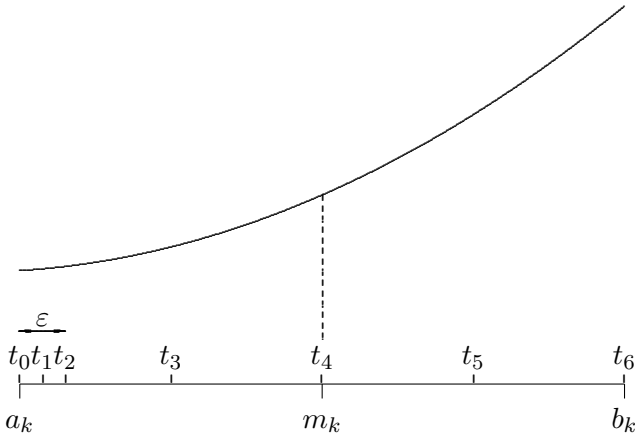


Figura 7.3

$\varepsilon = 0.01$.

Ver tabla de resultados al final del capítulo.

7.12 MINIMIZACIÓN IMPRECISA

El tiempo necesario para buscar el minimizador exacto de $\varphi(\lambda)$ no es despreciable. Si además hay que efectuar este proceso muchas veces, el tiempo total de las minimizaciones en una variable representa una parte importante del tiempo total. Algunos métodos de minimización en varias variables conservan la convergencia si en cada minimización en una variable se halla simplemente un punto mejor (con ciertas condiciones) que el anterior. Esto es lo que se denomina minimización o búsqueda imprecisa (o inexacta).

Hay varios criterios para decidir cuando se tiene un λ que aproxima adecuadamente el minimizador o cuando se ha obtenido una disminución suficiente en el valor de φ.

7.12.1 Criterio del porcentaje

Sean: λ^* el minimizador de φ en Λ, $0 < c < 1$ (por ejemplo $c = 0.10$). Se desea encontrar un $\bar{\lambda}$ tal que

$$|\bar{\lambda} - \lambda^*| \leq c|\lambda^*|.$$

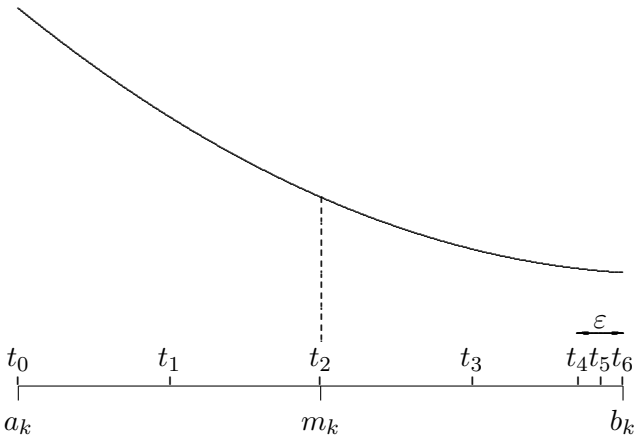


Figura 7.4

Obviamente se requiere conocer un valor aproximado o una cota de λ^*. Esto sucede, por ejemplo, cuando se minimiza en un intervalo $[0, \lambda_{\max}]$.

7.12.2 Regla de Armijo

El objetivo de esta regla es encontrar un λ que no sea ni demasiado pequeño ni demasiado grande. Sean: $0 < \beta < 1$, $\alpha > 1$ (por ejemplo $\beta = 0.2$, $\alpha = 2$). Sea d una dirección tal que $d^{\mathrm{T}} f'(x) < 0$ (entonces d es una dirección de descenso). Sea φ_1 la aproximación de orden 1 de φ alrededor de 0. Sea

$$\begin{aligned} \varphi(\lambda) &= \varphi_1(\lambda) + \lambda^2 \varphi''(t)/2 \quad , \quad t \text{ entre 0 y } \lambda, \\ \varphi_1(\lambda) &= \varphi(0) + \varphi'(0)\lambda \\ \varphi_1(\lambda) &= \varphi(0) + d^{\mathrm{T}} f'(x)\lambda. \end{aligned}$$

φ_1 es una recta que pasa por el punto $(0, \varphi(0))$ con pendiente negativa $\varphi'(0)$ y es tangente a φ. Sea $\hat{\varphi}$ un recta que pasa por el mismo punto $(0, \varphi(0))$, pero con una pendiente menor (menos negativa).

$$\hat{\varphi}(\lambda) = \varphi(0) + \beta\varphi'(0)\lambda.$$

Obviamente si $\bar{\lambda}$ es positivo y muy pequeño $\varphi(\bar{\lambda}) < \varphi(0)$. Supóngase que $\alpha = 2$. Para evitar que $\bar{\lambda}$ sea muy pequeño se le exige que sea por lo menos igual a la mitad del $\lambda > 0$ tal que $\varphi(\lambda) = \hat{\varphi}(\lambda)$, o sea,

$$\varphi(\alpha\bar{\lambda}) > \hat{\varphi}(\alpha\bar{\lambda}).$$

Por otro lado se considera adecuado el valor de $\bar{\lambda}$ si

$$\varphi(\bar{\lambda}) \leq \hat{\varphi}(\bar{\lambda}).$$

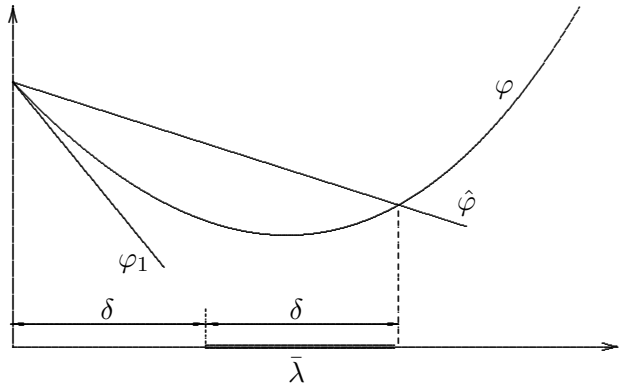


Figura 7.5

Con frecuencia, más no siempre, después de hacer **una** iteración del método de los tres puntos el valor λ obtenido cumple la regla de Armijo (y otras reglas de minimización imprecisa). Otra manera de encontrar un $\bar{\lambda}$ que cumpla esta regla es la siguiente: Empezar con un $\lambda_0 > 0$ tal que $\varphi(\lambda_0) \leq \hat{\varphi}(\lambda_0)$ (generalmente cualquier valor positivo "pequeño" sirve), e irlo multiplicando por α hasta que $\varphi(\lambda_0\alpha^k) > \hat{\varphi}(\lambda_0\alpha^k)$, entonces $\bar{\lambda} = \lambda_0\alpha^{k-1}$. Esto quiere decir que se empieza con un valor de λ mejor que $\lambda = 0$ (pero no demasiado grande), luego se incrementa varias veces el valor de λ hasta obtener un valor de λ muy grande y que no sirve, pero el penúltimo valor sí sirve.

datos: $0 < \beta < 1,\ \alpha > 1,\ R$ grande, $\varphi_0 = \varphi(0),\ \varphi'_0 = \varphi'(0),$
$\lambda_0 > 0$ tal que $\varphi(\lambda_0) < \hat{\varphi}(\lambda_0)$
`exito` $= 0,\ \bar{\lambda} = \lambda_0,\ c = \beta\varphi'_0$
mientras `exito` $= 0$ y $\bar{\lambda} < R$
$\bar{\lambda} = \alpha\bar{\lambda}$
si $\varphi(\bar{\lambda}) > \varphi_0 + c\bar{\lambda}$ **ent**
$\bar{\lambda} = \bar{\lambda}/\alpha$
`exito` $= 1$
fin-ent
fin-mientras

Si al final de este proceso `exito` $= 1$, entonces se ha obtenido un $\bar{\lambda}$ adecuado. En caso contrario, el valor de λ aumentó mucho, lo que indica que posiblemente φ no tiene minimizador.

Aunque el algoritmo funciona empezando con cualquier $\lambda_0 > 0$ tal que $\varphi(\lambda_0) < \hat{\varphi}(\lambda_0)$, es aconsejable buscar que este λ_0 no sea demasiado pequeño, para evitar así tener que aumentarlo muchas veces hasta llegar al tamaño adecuado.

El proceso completo para obtener un punto que cumpla la regla de Armijo tiene en realidad tres pasos:

- dado $\lambda''_0 > 0$, encontrar $\lambda'_0 > 0$ y mejor que 0,
- dado $\lambda'_0 > 0$ y mejor que 0 , encontrar $\lambda_0 > 0$ tal que $\varphi(\lambda_0) < \hat{\varphi}(\lambda_0)$,
- dado $\lambda_0 > 0$ tal que $\varphi(\lambda_0) < \hat{\varphi}(\lambda_0)$ encontrar $\bar{\lambda}$ que cumpla regla de Armijo.

El valor obtenido en una iteración del método de los tres puntos puede ser utilizado como λ'_0. Se puede escoger como λ''_0 el valor $\lambda_{k\text{-}1}$ (valor de λ en la iteración anterior), sin importar que corresponda a otro punto $x^{k\text{-}1}$ y a otra dirección $d^{k\text{-}1}$.

Si λ''_0 no es mejor que 0, entonces se disminuye, por ejemplo multiplicando por 0.5, y así sucesivamente hasta obtener un punto mejor que 0. Como se supone que $\varphi'(0) < 0$, entonces este proceso de buscar un λ mejor que 0 únicamente tiene dos salidas: o se encuentra un λ mejor que 0, o, menos corriente, se obtiene un λ positivo, pero demasiado pequeño para la precisión usada lo que indica que aunque $\varphi'(0) < 0$ se puede considerar que 0 es casi un minimizador.

Si λ'_0 no cumple con la condición $\varphi(\lambda'_0) < \hat{\varphi}(\lambda'_0)$, entonces se divide una o varias veces por α hasta obtener el cumplimiento de la condición.

Ejemplo 7.12. Sean: $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, $x = (-2, 3)$, $d = (1, -4)$. Hallar un $\bar{\lambda} \geq 0$ adecuado que "minimice" $\varphi(\lambda) = f(x + \lambda d)$ utilizando la regla de Armijo con $\beta = 0.2$, $\alpha = 2$.

En este caso $\varphi(0) = 17$, $\varphi'(0) = -38$. Para $\lambda = 0.1$ se tiene que $\varphi(\lambda) = 13.5221$ y $\hat{\varphi}(\lambda) = 17 + (0.1)(0.2)(-38) = 16.24$, luego 0.1 es un valor inicial aceptable. Para obtener un valor que no sea demasiado pequeño

λ	$\varphi(\lambda)$		$\hat{\varphi}(\lambda)$
0.0000	17.0000		17.0000
0.1000	13.5221	$<$	16.2400
0.2000	10.6576	$<$	15.4800
0.4000	6.5936	$<$	13.9600
0.8000	4.0336	$<$	10.9200
1.6000	14.4656	$>$	4.8400

Entonces el valor $\bar{\lambda} = 0.8$ sirve.

En este ejemplo, $\varphi(1.2329) = 7.6292$, $\hat{\varphi}(1.2329) = 7.6300$, luego cualquier valor de λ en el intervalo $[0.6165, 1.2329]$ sirve.

Si se hubiera utilizado el método de los tres puntos empezando con $t = 0.1$, mejor que $\lambda = 0$, entonces

λ	$\varphi(\lambda)$
0.0000	17.0000
0.1000	13.5221
0.3000	8.3621
0.7000	4.0661
1.5000	12.3125

$$t_p^* = 0.8062, \ \ \varphi(t_p^*) = 4.0434,$$
$$\bar{\lambda} = \text{mejor}\{0.7, \ t_p^*\} = 0.8062.$$

$$\begin{aligned} \varphi(2(0.8062)) = 14.7467 &> \hat{\varphi}(2(0.8062)) = 4.7458, \\ \varphi(0.8062) = 4.0435 &< \hat{\varphi}(0.8062) = 10.8729. \end{aligned}$$

Luego $\lambda = 0.8062$ cumple la regla de Armijo.

Si se hubiera empezado con $\lambda_0'' = 5$, valor para el cual $\varphi(5) = 277$, hubiera sido necesario multiplicar varias veces por 0.5 hasta obtener $\lambda_0' = 1.25$, $\varphi(1.25) = 7.8789$. Como $\hat{\varphi}(1.25) = 7.5$, entonces es necesario dividir por 2 hasta obtener $\lambda_0 = 0.625$ que también cumple la regla de Armijo. $\diamond$

7.12.3 Regla de Goldstein

El objetivo de esta regla también es encontrar un λ que no sea ni demasiado pequeño ni demasiado grande. Sea $0 < \varepsilon < 1/2$ (por ejemplo $\varepsilon = 0.2$). Se debe cumplir una de las condiciones de Armijo

$$\varphi(\bar{\lambda}) \leq \varphi(0) + \varepsilon\varphi'(0)\bar{\lambda} = \hat{\varphi}(\bar{\lambda}).$$

Se considera que $\bar{\lambda}$ no es muy pequeño si

$$\varphi(\bar{\lambda}) > \varphi(0) + (1-\varepsilon)\varphi'(0)\bar{\lambda} = \bar{\varphi}(\bar{\lambda}).$$

Las tres rectas $\hat{\varphi}$, $\bar{\varphi}$, φ_1, pasan por el punto $(0, \varphi(0))$, las tres tienen pendiente negativa, $\hat{\varphi}$ es la menos inclinada y φ_1 la más inclinada.

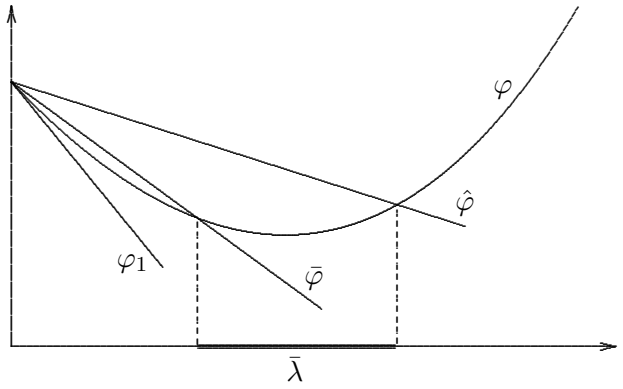


Figura 7.6

Una manera de obtener un λ que cumpla la regla de Goldstein es la siguiente:

- empezar con un valor positivo pequeño,
- incrementarlo hasta que cumpla la primera condición,
- si cumple la segunda condición se tiene un valor adecuado,
- si no la cumple, disminuir lentamente su valor hasta que cumpla la segunda condición.

Ejemplo 7.13. Sean: $f(x_1, x_2) = x_1^4 + (x_1 + x_2)^2$, $x = (-2, 3)$, $d = (1, -4)$. Hallar un $\bar{\lambda} \geq 0$ adecuado que "minimice" $\varphi(\lambda) = f(x + \lambda d)$ utilizando la regla de Goldstein con $\varepsilon = 0.2$.

Recuérdese que $\varphi(0) = 17$, $\varphi'(0) = -38$.

λ	$\varphi(\lambda)$	$\bar{\varphi}(\lambda)$	$\hat{\varphi}(\lambda)$
0.1000	13.5221	13.9600	
0.2000	10.6576	10.9200	
0.4000	6.5936	4.8400	
0.4000	6.5936		13.9600

Entonces el valor $\lambda = 0.4$ sirve.

Si se hubiera utilizado otra manera de incrementar λ, el valor de λ que cumple la primera condición puede no cumplir la segunda. Entonces es necesario que el ritmo de disminución sea más lento que el ritmo de incremento.

λ	$\varphi(\lambda)$	$\bar{\varphi}(\lambda)$	$\hat{\varphi}(\lambda)$
0.0200	16.2531	13.3920	
0.2000	10.6576	10.9200	
2.0000	25.0000	-43.8000	
2.0000	25.0000		1.8000
1.0000	5.0000		9.4000

Entonces el valor $\lambda = 1.0$ sirve.

Para este ejemplo

$$\begin{array}{rclrcl} \varphi(0.2444) & = & 9.5707\,, & \bar{\varphi}(0.2444) & = & 9.5702, \\ \varphi(1.2329) & = & 7.6292\,, & \hat{\varphi}(1.2329) & = & 7.6300. \end{array}$$

Entonces los valores de λ en el intervalo $[0.2444, 1.2329]$ cumplen la regla de Goldstein. $\diamond$

7.13 MINIMIZACIÓN DE UNA FUNCIÓN CUADRÁTICA

Sea φ una función cuadrática no degenerada (no es una recta). Dados tres puntos diferentes t_1, t_2, t_3 cualquier fórmula de interpolación cuadrática

se puede aplicar (el denominador no es nulo) y da un punto crítico. Si φ es convexa (en este caso estrictamente convexa) el punto crítico es un minimizador.

Si f es una función cuadrática se puede expresar de dos formas, la primera es

$$f(x) = \sum_{i=1}^{n} \alpha_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} c_i x_i + \kappa.$$

Como el valor de κ no altera en nada el minimizador, se puede suponer que $\kappa = 0$.

$$f(x) = \sum_{i=1}^{n} \alpha_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} c_i x_i.$$

La otra forma es una expresión matricial

$$f(x) = \frac{1}{2} x^{\mathrm{T}} H x + c^{\mathrm{T}} x,$$

donde H es una matriz $n \times n$ simétrica y c es un vector columna de n componentes. Obviamente $c = [c_1 \quad c_2 \quad ... \quad c_n]^{\mathrm{T}}$ y

$$\begin{aligned} h_{ii} &= 2\alpha_{ii}, \quad i = 1, ..., n, \\ h_{ij} &= h_{ji} = \beta_{ij}, \quad i \neq j. \end{aligned}$$

Las expresiones del gradiente y de la matriz hessiana (constante) son:

$$\begin{aligned} f'(x) &= Hx + c, \\ f''(x) &= H. \end{aligned}$$

Dejando fijos un punto x y una dirección $d \neq \mathbf{0}$, la función $\varphi(\lambda)$ es cuadrática, pero, dependiendo de d, puede ser estrictamente convexa (una parábola hacia arriba), o una recta, o estrictamente cóncava. Si H es semidefinida positiva, φ puede ser estrictamente convexa o puede ser una recta. Si H es definida positiva, entonces φ siempre es estrictamente convexa.

$$\varphi(\lambda) = \frac{1}{2}(x + \lambda d)^{\mathrm{T}} H (x + \lambda d) + c^{\mathrm{T}}(x + \lambda d)$$

$$= \frac{\lambda^2}{2} d^{\mathrm{T}} H d + \lambda d^{\mathrm{T}}(Hx + c) + f(x).$$

Para obtener un punto crítico λ',

$$\lambda' = -\frac{d^{\mathrm{T}}(Hx + c)}{d^{\mathrm{T}} H d} = -\frac{d^{\mathrm{T}} f'(x)}{d^{\mathrm{T}} H d}, \quad \text{si } d^{\mathrm{T}} H d \neq 0.$$

Este punto crítico λ' es el minimizador si $d^{\mathrm{T}} H d > 0$. Si H es definida positiva el punto crítico siempre es el minimizador. Si $d^{\mathrm{T}} f'(x) < 0$ (entonces d es dirección de descenso), el minimizador es positivo.

Ejemplo 7.14. Minimizar $f(\bar{x} + \lambda d)$, donde $f(x_1, x_2) = 2x_1^2 + (x_1 + x_2)^2$, $\bar{x} = (4, 3)$, $d = (-2, -1)$.

$$\lambda^* = -\frac{\begin{bmatrix} -2 & -1 \end{bmatrix} \begin{bmatrix} 30 \\ 14 \end{bmatrix}}{\begin{bmatrix} -2 & -1 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -1 \end{bmatrix}} = \frac{74}{34} = 2.1765. \ \diamond$$

EJERCICIOS

En los ejercicios 7.1 a 7.11 estudie el problema propuesto, averigüe cuáles métodos puede usar, use algunos de ellos, verifique si el valor obtenido cumple algunos criterios (regla de Armijo, Goldstein, ...), utilice valores iniciales diferentes.

7.1. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in \mathbb{R}$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (1, -2)$.

7.2. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in \mathbb{R}$, donde $f(x_1, x_2) = x_1^4 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (1, -2)$.

7.3. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, \infty]$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (1, -2)$.

7.4. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, 6]$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (1, -2)$.

7.5. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, 2]$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (1, -2)$.

7.6. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in \mathbb{R}$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (-1, 2)$.

7.7. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, \infty]$, donde $f(x_1, x_2) = x_1^2 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (-1, 2)$.

7.8. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in \mathbb{R}$, donde $f(x_1, x_2) = x_1^3 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (-1, 2)$.

7.9. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, \infty]$, donde $f(x_1, x_2) = x_1^3 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (-1, 2)$.

7.10. Minimizar $f(\bar{x} + \lambda d)$, con $\lambda \in [0, 10]$, donde $f(x_1, x_2) = x_1^3 - 2x_1 + x_2^2 + 2x_2 + 10$, $\bar{x} = (-3, 4)$, $d = (-1, 2)$.

7.11. Minimizar $f(\bar{x}+\lambda d)$, con $\lambda \in [0, 20]$, donde $f(x_1, x_2) = x_1 \operatorname{sen}(x_1) + x_2^2$, $\bar{x} = (3, 4)$, $d = (2, -1)$.

ALGORITMO DE LOS TRES PUNTOS PARA $\lambda \in [a, b]$

datos: $a,\ b,\ \varepsilon$
$l = b - a$
mientras $l > \varepsilon$
    $m = (b + a)/2,\ \ t_0 = a,\ \ t_6 = b$
    si $\varphi(m) < \varphi(a), \varphi(b)$ **ent**
        $p =$ minimiz. parábola: $(a, \varphi(a)), (m, \varphi(m)), (b, \varphi(b))$
        $\mu = \text{mejor}\{m, p\},\ \ \delta = \min\{\varepsilon/2,\ l/8\}$
        si $\mu \leq m$ **ent**
            $t_2 = \mu,\ \ t_1 = t_2 - \delta,\ \ t_3 = t_2 + \delta$
            $t_4 = (2t_3 + b)/3,\ \ t_5 = (t_3 + 2b)/3$
        fin-ent
        sino
            $t_4 = \mu,\ \ t_3 = t_4 - \delta,\ \ t_5 = t_4 + \delta$
            $t_1 = (2a + t_3)/3,\ \ t_2 = (a + 2t_3)/3$
        fin-sino
    fin-ent
    sino
        $\delta = \min\{\varepsilon/2,\ l/12\}$
        si $\varphi(a) \leq \varphi(b)$ **ent**
            $t_1 = a + \delta,\ \ t_2 = a + 2\delta$
            $t_3 = a + l/4,\ \ t_4 = m,\ \ t_5 = b - l/4$
        fin-ent
        sino
            $t_1 = a + l/4,\ \ t_2 = m,\ \ t_3 = b - l/4$
            $t_4 = b - 2\delta,\ \ t_5 = b - \delta$
        fin-sino
    fin-sino
    $t_i = \text{mejor}\{t_1, ..., t_5\},\ a = t_{i-1},\ b = t_{i+1},\ l = b - a$
fin-mientras
$\lambda^* = \text{mejor}\{a, t_i, b\}$

k	a	b	m	p	μ	t_1	t_2	t_3	t_4	
	$\varphi(a)$	$\varphi(b)$	$\varphi(m)$	$\varphi(p)$	$\varphi(\mu)$	$\varphi(t_1)$	$\varphi(t_2)$	$\varphi(t_3)$	$\varphi(t_4)$	φ
0	0.0000	2.0000	1.0000			0.5000	1.0000	1.5000	1.9900	1
	2561.0000	9.0000	161.0000			810.0000	161.0000	14.0000	8.8804	8
1	1.5000	1.9950	1.7475	1.8156	1.7475	1.7425	1.7475	1.7525	1.8333	1
	14.0000	8.9401	6.8754	7.1085	6.8754	6.8787	6.8754	6.8753	7.2346	8
2	1.7475	1.8333	1.7904			1.7525	1.7575	1.7690	1.7904	1
	6.8754	7.2346	6.9694			6.8753	6.8785	6.8969	6.9694	7
3	1.7475	1.7575	1.7525	1.7500	1.7500	1.7488	1.7500	1.7513	1.7533	1
	6.8754	6.8785	6.8753	6.8750	6.8750	6.8751	6.8750	6.8751	6.8757	6
	1.7488	1.7513								
	6.8751	6.8751								

$$\lambda^* \approx 1.7500$$

Ejemplo 7.10

k	a_k	b_k	m	p	μ	t_1	t_2	t_3	t_4	t_5
	$\varphi(a_k)$	$\varphi(b_k)$	$\varphi(m)$	$\varphi(p)$	$\varphi(\mu)$	$\varphi(t_1)$	$\varphi(t_2)$	$\varphi(t_3)$	$\varphi(t_4)$	$\varphi(t$
0	0.0000	1.5000	0.7500			0.3750	0.7500	1.1250	1.4900	1.4
	2561.0000	14.0000	390.8750			1115.7266	390.8750	95.3516	14.7447	14.3
1	1.4900	1.5000	1.4950			1.4925	1.4950	1.4975	1.4983	1.4
	14.7447	14.0000	14.3661			14.5539	14.3661	14.1815	14.1207	14.0
2	1.4983	1.5000								
	14.1207	14.0000								

$$\lambda^* \approx 1.7500$$

Ejemplo 7.11

Capítulo 8

MÉTODOS DE MINIMIZACIÓN EN VARIAS VARIABLES SIN RESTRICCIONES

Sea $f : \mathbb{R}^n \to \mathbb{R}$. Los métodos de este capítulo se utilizan para hallar f^* valor mínimo de f, y x^* minimizador de f cuando x varía en $\mathbb{R}^n$. Bajo ciertas condiciones de la función, algunas veces se obtiene el minimizador global único (más exactamente, una aproximación del minimizador), a veces se alcanza un minimizador global, otras veces se obtiene simplemente un minimizador local, otras veces solamente se consigue un punto crítico, otras veces apenas se consigue un punto donde el proceso se hizo muy lento (los cambios en f o en x son muy pequeños).

De todas formas los métodos presentados en este capítulo y en los dos siguientes, buscan un minimizador local. Los métodos de minimización global, que buscan mínimos globales, son mucho más complejos y están fuera del alcance de este libro. No se debe confundir minimización global con convergencia global. Un método tiene convergencia global cuando, no importando el punto inicial, converge hacia un minimizador local o global. Puede ser posible tener un método de minimización local con convergencia global.

Las condiciones casi ideales se tienen cuando f es continua, doblemente diferenciable y estrictamente convexa. Obviamente esto no se tiene o no se puede garantizar o verificar siempre. De todas formas, lo mínimo que se debe exigir a f, cuando no se utilizan derivadas, es que sea continua.

De manera semejante a la minimización en una variable, hay métodos que utilizan segundas y primeras derivadas, es decir utilizan la matriz hessiana y el gradiente, hay métodos que utilizan únicamente derivadas primeras, y hay métodos que no utilizan derivadas.

Generalmente, entre más información se utilice, es decir, entre más derivadas sean usadas, más eficiente es el método en cuanto al número de iteraciones. Pero por otra parte, conseguir bastante información cuesta y, dependiendo de la función, puede hacer lento el proceso. Por otro lado a veces no se puede conseguir cierta información, por ejemplo, no es posible calcular derivadas primeras o segundas, y entonces la única solución es utilizar métodos donde no se requiera esta información.

Dado un punto inicial x^0, aproximación (ojalá buena) de x^*, se construye una sucesión de puntos $\{x^k\}$ con la esperanza, y bajo ciertas condiciones con la garantía, de que

$$\begin{aligned} x^k &\longrightarrow x^* \quad \text{minimizador global, o} \\ x^k &\longrightarrow x^* \quad \text{minimizador local, o} \\ x^k &\longrightarrow x^c \quad \text{punto crítico.} \end{aligned}$$

En casi todos los métodos, dado un punto x^k se construye una dirección $d^k \neq \mathbf{0}$ y en seguida se obtiene λ_k^* minimizador de $\varphi(\lambda) = f(x^k + \lambda d^k)$ cuando λ varía en $\mathbb{R}$.

$$\begin{aligned} \lambda_k^* &= \operatorname{argmin} f(x^k + \lambda d^k), \lambda \in \mathbb{R}, \\ x^{k+1} &= x^k + \lambda_k^* d^k. \end{aligned}$$

Los métodos se diferencian en la construcción de la dirección d^k. Si se cumple que $d^{\mathrm{T}} f'(x^k) < 0$, entonces la dirección es de descenso y basta minimizar para valores positivos:

$$\begin{aligned} \lambda_k^* &= \operatorname{argmin} f(x^k + \lambda d^k), \lambda \geq 0, \\ x^{k+1} &= x^k + \lambda_k^* d^k. \end{aligned}$$

Hallar λ_k^* minimizador exacto o casi exacto, puede ser muy costoso en tiempo. En muchos métodos no se necesita indispensablemente encontrar el minimizador en cada iteración, sino que basta con encontrar λ_k que mejora suficientemente el valor de la función. Esto se denotará

$$\begin{aligned} \lambda_k &= \text{argmej } f(x^k + \lambda d^k), \\ x^{k+1} &= x^k + \lambda_k d^k, \end{aligned}$$

y quiere decir que

- $f(x^k + \lambda_k d^k) < f(x^k) - \delta$ con $\delta > 0$, o que
- λ_k cumple la regla de Armijo, o que
- λ_k cumple la regla de Goldstein, o que ...

Casi todos los métodos son de descenso, o sea, cumplen (o deberían cumplir) que

$$f(x^{k+1}) < f(x^k).$$

Los criterios de parada son semejantes a los vistos para una variable. Pueden ser utilizados individualmente o en conjunto mediante la conjunción ***y***, o en conjunto mediante la conjunción **o**. El primer criterio consiste en considerar, por seguridad, un número máximo de iteraciones MAXIT, aunque ojalá que éste no sea el criterio por el cual se detiene un proceso. Otro criterio de parada se tiene cuando el gradiente es nulo o casi nulo

$$||f'(x^k)|| < \varepsilon_g \quad \text{dado.}$$

También se puede detener el proceso iterativo cuando el cambio, absoluto o relativo, en x es muy pequeño

$$||x^{k+1} - x^k|| < \varepsilon_x \ \text{ dado,}$$

o también,

$$\frac{||x^{k+1} - x^k||}{||x^k||} < \varepsilon'_x \ \text{ dado,}$$

o para evitar denominadores muy pequeños,

$$\frac{||x^{k+1} - x^k||}{1 + ||x^k||} \leq \varepsilon'_x \ \text{ dado.}$$

También se puede detener el proceso iterativo cuando el cambio, absoluto o relativo, en el valor de la función es muy pequeño

$$|f(x^{k+1}) - f(x^k)| < \varepsilon_f \ \text{ dado ,}$$

o también,

$$\frac{|f(x^{k+1}) - f(x^k)|}{|f(x^k)|} < \varepsilon'_f \quad \text{dado} ,$$

o para evitar denominadores muy pequeños,

$$\frac{|f(x^{k+1}) - f(x^k)|}{|1 + f(x^k)|} \leq \varepsilon'_f \quad \text{dado.}$$

Por ejemplo, si se utiliza $\varepsilon_g = 10^{-6}$ **y** $\varepsilon_x = 10$ muy posiblemente el proceso se detendrá únicamente teniendo en cuenta el criterio del gradiente. Si se utiliza $\varepsilon_g = 10$ **y** $\varepsilon_x = 10^{-6}$ muy posiblemente el proceso se detendrá únicamente teniendo en cuenta el criterio del cambio en x.

Si se utiliza $\varepsilon_g = 10^{-6}$ **o** $\varepsilon_x = 10^{-100}$ muy posiblemente el proceso se detendrá únicamente teniendo en cuenta el criterio del gradiente. Si se utiliza $\varepsilon_g = 10^{-100}$ **o** $\varepsilon_x = 10^{-6}$ muy posiblemente el proceso se detendrá únicamente teniendo en cuenta el criterio del cambio en x.

No sobra notar que el proceso iterativo se debe detener cuando aparentemente $\varphi(\lambda) = f(x^k + \lambda d^k)$ no es acotada inferiormente y no posee minimizador, pues ésto también indica que f no es acotada inferiormente y no tiene minimizador global.

Es importantísimo evitar, antes de que el computador pare por un error, las operaciones no permitidas o no deseadas, por ejemplo, solución de un sistema singular o casi singular (con matriz no invertible), división por cero o por un número casi nulo, raíz cuadrada de un número negativo, etc.

8.1 CÁLCULO DEL GRADIENTE Y DEL HESSIANO

Si es posible obtener las expresiones analíticas de las primeras y segundas derivadas parciales y si evaluarlas no es demasiado costoso, ésta es la mejor opción para obtener el gradiente y el hessiano. También existe una técnica llamada diferenciación automática [Gri00], cada vez más eficiente y de mayor uso. Por su complejidad y nivel no está dentro del alcance de este libro. Una tercera salida consiste en la aproximación numérica.

Los programas de computador para programación no lineal, deben tener un subprograma (función, procedimiento o subrutina) donde dado un vector x se evalúa f. Posiblemente, también se tienen subprogramas para calcular mediante la expresión analítica el gradiente y el hessiano. Cuando el programa se utiliza con frecuencia para diferentes funciones, es necesario cambiar no sólo el subprograma donde se define f sino los subprogramas

para el cálculo del gradiente y del hessiano. Para n grande, el proceso de la derivación parcial analítica puede ser dispendioso y aumenta la posibilidad de los errores humanos.

Las aproximaciones numéricas de las derivadas parciales, son las mismas de las aproximaciones numéricas de las derivadas de funciones de una variable, o se deducen a partir de ellas. En una variable se utilizó un valor η; para varias variables puede ser conveniente utilizar diferentes η_j para las diferentes variables.

$$\frac{\partial f}{\partial x_i}(\bar{x}) \approx \frac{f(\bar{x}+\eta_i e^i)-f(\bar{x})}{\eta_i}, \tag{8.1}$$

$$\frac{\partial f}{\partial x_i}(\bar{x}) \approx \frac{f(\bar{x}+\eta_i e^i)-f(\bar{x}-\eta_i e^i)}{2\eta_i} \tag{8.2}$$

$$\frac{\partial^2 f}{\partial x_i^2}(\bar{x}) \approx \frac{f(\bar{x}+\eta_i e^i)-2f(\bar{x})+f(\bar{x}-\eta_i e^i)}{\eta_i^2} \tag{8.3}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}) &= \frac{\partial}{\partial x_i}\Big(\frac{\partial f}{\partial x_j}(\bar{x})\Big) \\
&\approx \frac{1}{\eta_i}\Big(\frac{\partial f}{\partial x_j}(\bar{x}+\eta_i e^i)-\frac{\partial f}{\partial x_j}(\bar{x})\Big) \\
&\approx \frac{1}{\eta_i}\Big(\frac{f(\bar{x}+\eta_i e^i+\eta_j e^j)-f(\bar{x}+\eta_i e^i)}{\eta_j} \\
&\qquad -\frac{f(\bar{x}+\eta_j e^j)-f(\bar{x})}{\eta_j}\Big),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}) \approx \frac{1}{\eta_i \eta_j}\Big(& f(\bar{x}+\eta_i e^i+\eta_j e^j)-f(\bar{x}+\eta_i e^i) \\
& -f(\bar{x}+\eta_j e^j)+f(\bar{x})\Big),
\end{aligned} \tag{8.4}$$

o también,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(\bar{x})\right)$$

$$
\begin{aligned}
&\approx \frac{1}{2\eta_i}\left(\frac{\partial f}{\partial x_j}(\bar{x}+\eta_i e^i) - \frac{\partial f}{\partial x_j}(\bar{x}-\eta_i e^i)\right) \\
&\approx \frac{1}{2\eta_i}\left(\frac{f(\bar{x}+\eta_i e^i+\eta_j e^j) - f(\bar{x}+\eta_i e^i-\eta_j e^j)}{2\eta_j}\right. \\
&\quad \left. -\frac{f(\bar{x}-\eta_i e^i+\eta_j e^j) - f(\bar{x}-\eta_i e^i-\eta_j e^j)}{2\eta_j}\right), \\
\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}) &\approx \frac{1}{4\eta_i\eta_j}\Big(f(\bar{x}+\eta_i e^i+\eta_j e^j) + f(\bar{x}-\eta_i e^i-\eta_j e^j) \\
&\quad -f(\bar{x}+\eta_i e^i-\eta_j e^j) - f(\bar{x}-\eta_i e^i+\eta_j e^j)\Big).
\end{aligned} \tag{8.5}
$$

Conocido el valor de $f(\bar{x})$, si se aproxima el gradiente usando 8.1 se necesitan n evaluaciones adicionales de la función. Si se utiliza la fórmula 8.2 se necesitan $2n$ evaluaciones de f. Para aproximar el hessiano es necesario calcular los n elementos diagonales y $n(n-1)/2$ elementos de la parte triangular estrictamente superior. Si se aproxima el hessiano usando 8.3 y 8.4 se necesitan $2n + 3n(n-1)/2$ evaluaciones adicionales. Si se almacenan los valores $f(\bar{x}+\eta_i e^i)$ el número de evaluaciones adicionales se reduce a $2n + n(n-1)/2$. Si se aproxima el hessiano utilizando 8.3 y 8.5 se necesitan $2n + 2n(n-1)$ evaluaciones adicionales.

	fórmula	núm. de eval.	orden del núm. eval.	orden error
gradiente	8.1	n	n	η_i
gradiente	8.2	$2n$	$2n$	η_i^2
hessiano	8.3, 8.4	$2n + n(n-1)/2$	$n^2/2$	
hessiano	8.3, 8.5	$2n + 2n(n-1)$	$2n^2$	

El hessiano también se puede aproximar mediante evaluaciones, o aproximaciones del gradiente. En este caso las columnas del hessiano se pueden aproximar por una de las siguientes fórmulas

$$
\begin{aligned}
H_{\cdot j} &\approx \frac{f'(\bar{x}+\eta_j e^j) - f'(\bar{x})}{\eta_j}, \\
H_{\cdot j} &\approx \frac{f'(\bar{x}+\eta_j e^j) - f'(\bar{x}-\eta_j e^j)}{2\eta_j}.
\end{aligned}
$$

La segunda expresión necesita el doble de operaciones que la primera, pero es más precisa. Por otro lado, ninguna de las dos expresiones tiene

en cuenta la simetría de la matriz hessiana, por lo tanto se deben utilizar preferiblemente cuando se tiene la expresión analítica del gradiente o cuando se dispone de una manera de calcular el gradiente, rápida y precisa.

8.2 MÉTODO DE NEWTON

Si $g : \mathbb{R} \to \mathbb{R}$, la fórmula de Newton (o Newton-Raphson) para encontrar una raíz de $g(x) = 0$ es la siguiente

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

Esta fórmula se puede reescribir como

$$x_{k+1} = x_k - (g'(x_k))^{-1} g(x_k).$$

Si $g : \mathbb{R}^n \to \mathbb{R}^n$, la fórmula anterior se generaliza inmediatamente a

$$x^{k+1} = x^k - (g'(x^k))^{-1} g(x^k),$$

donde $g'(x^k)$ es la matriz jacobiana de la función g en el punto x^k. La anterior expresión es la fórmula de Newton para hallar una raíz de la ecuación vectorial $g(x) = \mathbf{0}$. Para evitar el cálculo de la inversa, haciendo $d^k = -(g'(x^k))^{-1} g(x^k)$, o lo que es lo mismo, $g'(x^k)d^k = -g(x^k)$, la fórmula de Newton se expresa en dos partes:

$$\begin{array}{lrcl} \text{resolver} & g'(x^k)d^k & = & -g(x^k), \\ & x^{k+1} & = & x^k + d^k. \end{array}$$

Si $f : \mathbb{R}^n \to \mathbb{R}$ es doblemente diferenciable, el método de Newton se puede utilizar para hallar un punto crítico ($f'(x) = 0$), reemplazando en las fórmulas anteriores g por f'

$$\begin{array}{lrcll} \text{resolver} & f''(x^k)d^k & = & -f'(x^k), & (8.6) \\ & x^{k+1} & = & x^k + d^k. & (8.7) \end{array}$$

La fórmula (8.6) expresa que se debe resolver un sistema de ecuaciones lineales donde la matriz de coeficientes es simplemente el hessiano, el vector de términos independientes es menos el gradiente, y las incógnitas son las componentes del vector columna d^k.

Ejemplo 8.1. Utilizar el método de Newton para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$.

k	x	$f(x)$	$f'(x)$	$f''(x)_{.1}$	$f''(x)_{.2}$	d
0	3.0000	254.0000	604.0000	922.0000	-120.0000	-0.0198
	4.0000		-100.0000	-120.0000	20.0000	4.8812
1	2.9802	3.9212	4.0071	712.5421	-119.2079	-1.9648
	8.8812		-0.0078	-119.2079	20.0000	-11.7104
2	1.0154	149.0215	156.8260	238.9004	-40.6171	-0.0002
	-2.8292		-77.2065	-40.6171	20.0000	3.8599
3	1.0152	0.0002	0.0305	84.4553	-40.6092	-0.0152
	1.0307		-0.0000	-40.6092	20.0000	-0.0309
4	1.0000	0.0000	0.0093	82.0093	-40.0000	-0.0000
	0.9998		-0.0046	-40.0000	20.0000	0.0002
5	1.0000	0.0000	0.0000	82.0000	-40.0000	-0.0000
	1.0000		-0.0000	-40.0000	20.0000	-0.0000

Es interesante observar que el método de Newton no es siempre un método de descenso. En este ejemplo, como se ve en la tabla, en la iteración 2 hubo un aumento en el valor de la función. $\diamond$

Proposición 8.1. [Lue89] *Sean: f una función triple y continuamente diferenciable, $f'(\bar{x}) = 0$, $f''(\bar{x})$ invertible. Si x^0 está suficientemente cerca de $\bar{x}$, entonces el método de Newton para hallar un punto crítico de f converge a $\bar{x}$ con orden de convergencia superior o igual a dos.*

En [Baz93] hay condiciones más específicas para garantizar la convergencia. Las ventajas del método de Newton son:

- si las condiciones son favorables, converge muy rápidamente,
- no es necesario minimizar en una variable.

Las desventajas son:

- puede no converger,
- puede converger a un maximizador,
- la matriz $f''(x^k)$ puede ser singular o casi singular,
- es necesario calcular primeras y segundas derivadas parciales de f,

- es necesario resolver un sistema de ecuaciones.

Es claro que para funciones cuadráticas, con hessiano no singular, el método converge en una iteración al punto crítico.

Ejemplo 8.2. Utilizar el método de Newton para minimizar $f(x) = (x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$ partiendo de $x^0 = (3, 5)$.

k	x	$f(x)$	$f'(x)$	$f''(x)_{.1}$	$f''(x)_{.2}$	d
0	3.0000	26.0000	8.0000	4.0000	0.0000	-2.0000
	5.0000		12.0000	0.0000	4.0000	-3.0000
1	1.0000	0.0000	0.0000			
	2.0000		0.0000			

8.3 MÉTODOS DE LA REGIÓN DE CONFIANZA

Generalmente se considera que los métodos de la región de confianza, "*trust region*", son "hijos" del método de Levenberg-Marquardt usado para resolver problemas de mínimos cuadrados no lineales. Son modificaciones y adaptaciones para problemas de optimización más generales. En el fondo son modificaciones del método de Newton.

En el método de Newton, conocido x^k, en lugar minimizar directamente $f(x)$, se busca hallar el mínimo de la aproximación cuadrática de f alrededor de x^k, es decir,

$$\min\ q_k(x) = f(x^k) + f'(x^k)^{\mathrm{T}}(x - x^k) + \frac{1}{2}(x - x^k)^{\mathrm{T}} f''(x^k)(x - x^k). \quad (8.8)$$

En el método de la región de confianza se restringe la minimización de la aproximación cuadrática a una vecindad de x_k, o sea,

$$\min\ q_k(x) = f(x^k) + f'(x^k)^{\mathrm{T}}(x - x^k) + \frac{1}{2}(x - x^k)^{\mathrm{T}} f''(x^k)(x - x^k) \quad (8.9)$$
$$||x - x^k||_2 \leq \Delta_k.$$

Si se hace el cambio de variable $d = x - x^k$,

$$\min\ q_k(x^k + d) = f(x^k) + f'(x^k)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} f''(x^k) d \quad (8.10)$$

$$||d||_2 \leq \Delta_k.$$

La región de confianza es la bola $B[x^k, \Delta_k]$ o, vista después del cambio de variable, es la bola centrada en el origen $B[\mathbf{0}, \Delta_k]$. De iteración en iteración el radio Δ_k puede cambiar.

Se puede pensar en una aproximación de f más general, cambiando $f''(x^k)$, la matriz hessiana en x^k por una matriz M,

$$\begin{aligned} \min\ q_k(x^k + d) &= f(x^k) + f'(x^k)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} M d \\ ||d||_2 &\leq \Delta_k. \end{aligned} \tag{8.11}$$

El esquema simplificado del método es entonces:

$$\begin{aligned} d^k &= \text{solución de (8.10) o de (8.11)}\ , \\ x^{k+1} &= x^k + d^k. \end{aligned}$$

El valor de Δ_k puede variar en cada iteración de acuerdo a la mejoría en el valor de f. Si la mejoría es grande entonces para la iteración siguiente el valor del radio Δ aumenta. Por el contrario, si la mejoría es muy pequeña (o si no hay mejoría), esto quiere decir que el radio utilizado es muy grande y debe ser diminuido para la iteración siguiente. En este caso, además, el x no cambia, es decir, $x^{k+1} = x^k$. Más adelante se define ρ_k, una "medida" de la mejoría.

Hay varias estrategias para la implementación del método de la región de confianza. Algunas utilizan una solución exacta de (8.10), otras simplemente una solución aproximada. Los criterios para clasificar la mejoría en el valor de f pueden variar. Las estrategias exactas tienen un planteamiento más sencillo, pero generalmente son menos eficentes en la implementación.

8.3.1 Una estrategia exacta

El problema (8.10) se puede reescribir

$$\begin{aligned} \min\ q_k(x^k + d) &= f(x^k) + f'(x^k)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} f''(x^k) d \\ ||d||_2^2 - \Delta_k^2 &\leq 0. \end{aligned}$$

Al aplicar condiciones de KKT se tiene

$$\begin{aligned} f'(x^k) + f''(x^k)d + u_1 d = 0, \\ u_1 \geq 0, \\ u_1\big(||d||_2^2 - \Delta_k^2\big) = 0. \end{aligned}$$

Haciendo $\lambda = u_1$,

$$f'(x^k) + f''(x^k)d + \lambda d = 0, \tag{8.12}$$
$$\lambda \geq 0, \tag{8.13}$$
$$\lambda\big(||d||_2^2 - \Delta_k^2\big) = 0. \tag{8.14}$$

La igualdad (8.12) se puede reescribir

$$\Big(f''(x^k)d + \lambda I\Big)d = -f'(x^k).$$

Las condiciones de segundo orden exigen que la matriz $f''(x^k)d + \lambda I$ sea semidefinida positiva.

Se puede mostrar que si $f''(x^k)$ es definida positiva y si Δ_k es suficientemente grande, entonces la solución de (8.10) está dada por

$$f''(x^k)d = -f'(x^k),$$

o sea, simplemente se tiene el método de Newton. Si las suposiciones no se cumplen, entonces el método garantiza que

$$\Delta_k \geq ||d||_2 = ||\big(f''(x^k)d + \lambda I\big)^{-1} f'(x^k)||_2.$$

También se puede mostrar, que en estas condiciones si $\Delta_k \to 0$, entonces $\lambda \to \infty$ y

$$d \approx -\frac{1}{\lambda} f'(x^k).$$

Lo anterior es (aproximadamente) equivalente al método del descenso más pendiente.

Cuando λ varía entre 0 e ∞, entonces $d = d(\lambda)$ varía continuamente entre la dirección del método de Newton (si $f''(x^k)$ es definida positiva) y un múltiplo de la dirección del descenso más pendiente.

Para definir el valor de Δ_k en cada iteración se utiliza el parámetro ρ_k que es simplemente el cociente entre la reducción o disminución real en el valor de f y la reducción prevista

$$\rho_k = \frac{\text{reducción real}}{\text{reducción prevista}} = \frac{f(x^k) - f(x^k + d^k)}{f(x^k) - q_k(x^k + d^k)}. \tag{8.15}$$

Si d^k se obtiene por minimización (solución de (8.10)), entonces la reducción prevista es positiva. Luego el signo de ρ_k está dado por la reducción real.

A continuación están los detalles de la estrategia presentada en [NaSo96]. El algoritmo utiliza dos parámetros

$$0 < \eta < \mu < 1,$$

por ejemplo,

$$\eta = \frac{1}{4}, \quad \mu = \frac{3}{4}.$$

Si $\rho_k > \eta$ se considera que la iteración fue exitosa y se construye $x^{k+1} = x^k + d^k$. En caso contrario $x^{k+1} = x^k$. Además, si $\rho_k \leq \eta$, entonces el parámetro Δ se disminuye: $\Delta_{k+1} = \frac{1}{2}\Delta_k$. Si $\eta < \rho_k < \mu$, entonces el parámetro Δ no cambia: $\Delta_{k+1} = \Delta_k$. Si $\mu \leq \rho_k$, entonces el parámetro Δ se incrementa: $\Delta_{k+1} = 2\Delta_k$.

$$\begin{array}{llllll} \text{Si} & \rho_k \leq \eta & : & & x^{k+1} = x^k, & \Delta_{k+1} = \frac{1}{2}\Delta_k \\ \text{Si} & \eta < \rho_k < \mu, & : & \text{éxito}, & x^{k+1} = x^k + d^k, & \Delta_{k+1} = \Delta_k \\ \text{Si} & \mu \leq \rho_k, & : & \text{éxito}, & x^{k+1} = x^k + d^k, & \Delta_{k+1} = 2\Delta_k \end{array} \tag{8.16}$$

A continuación está un esquema general del algoritmo:

REGIÓN DE CONFIANZA: ESTRATEGIA EXACTA
datos: x^0, $\Delta_0 > 0$, $0 < \eta < \mu < 1$, ε_g, `MAXIT`
para $k = 0, 1, ..., $`MAXIT`
  si $||f'(x^k)|| < \varepsilon_g$ **ent parar**
  $d^k =$ solución de (8.10)
  calcular ρ_k según (8.15)
  construir x^{k+1}, Δ_{k+1} según (8.16)
fin-para k

Algunas veces se utiliza como radio inicial $\Delta_0 = \min(10, ||f'(x^0)||)$. El paso difícil de este algoritmo es la solución del problema (8.10). En la práctica se obtiene la solución exacta cuando es fácil hacerlo; en caso contrario se usa una solución aproximada.

En (8.10) se tiene una función continua en un compacto (cerrado y acotado), luego necesariamente hay minimizador. Éste está en el interior de la bola o en la frontera. Desde el punto de vista de condiciones de KKT hay dos posibilidades:

- La desigualdad no está activa ($\mathcal{I} = \emptyset$), entonces

$$\begin{aligned} f''(x^k)d &= -f'(x^k), \\ \lambda &= 0, \\ ||d||_2 &< \Delta_k. \end{aligned} \tag{8.17}$$

- La desigualdad está activa ($\mathcal{I} = \{1\}$), entonces

$$\begin{aligned} \big(f''(x^k) + \lambda I\big)d &= -f'(x^k), \\ \lambda &\geq 0, \\ ||d||_2 &= \Delta_k. \end{aligned} \tag{8.18}$$

Si $\mathcal{I} = \emptyset$ y $f''(x^k)$ es definida positiva y $||-f''(x^k)^{-1}f'(x^k)|| < \Delta_k$, se tiene el caso sencillo de minimización no restringida con la dirección plena de Newton

$$d_N \text{ solución de } f''(x^k)\, d = -f'(x^k), \tag{8.19}$$

o de manera teórica,

$$d_N = -f''(x^k)^{-1}f'(x^k).$$

En caso contrario, la (una) solución está en la frontera y hay que resolver una ecuación no lineal de una variable, λ, pero que involucra $n+1$ variables: d_1, d_2, ..., d_n, λ,

$$\begin{aligned} \big(f''(x^k) + \lambda I\big)d &= -f'(x^k), \\ ||d||_2 &= \Delta_k. \end{aligned} \tag{8.20}$$

Se puede expresar explícitamente, únicamente en función de λ, así:

$$\phi(\lambda) = || - \left(f''(x^k) + \lambda I\right)^{-1} f'(x^k)||_2 - \Delta_k = 0. \tag{8.21}$$

Para evitar este proceso iterativo, dentro del proceso iterativo general, en algunas estrategias para la región de confianza, simplemente se halla una solución aproximada de (8.18).

El esquema completo del algoritmo, incluyendo la solución de (8.10) es:

REGIÓN DE CONFIANZA: ESTRATEGIA APROXIMADA
datos: x^0, $\Delta_0 > 0$, $0 < \eta < \mu < 1$, ε_g, `MAXIT`
para $k = 0, 1, ..., $`MAXIT`
  si $||f'(x^k)|| < \varepsilon_g$ **ent parar**
  $d_N =$ solución de $f''(x^k)d = -f'(x^k)$
  si $||d_N||_2 \leq \Delta_k$ y $f''(x^k)$ s.d.p. **ent** $d^k = d_N$
  sino $d^k =$ solución de (8.20)
  calcular ρ_k según (8.15)
  construir x^{k+1}, Δ_{k+1} según (8.16)
fin-para k

Ejemplo 8.3. Aplicar la estrategia del método de región de confianza descrita anteriormente para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$ con $\Delta_0 = 1$.

$$f(x^0) = 254, \quad f'(x^0) = \begin{bmatrix} 604 \\ -100 \end{bmatrix}, \quad f''(x^0) = \begin{bmatrix} 922 & -120 \\ -120 & 20 \end{bmatrix}.$$

La dirección de Newton está dada por la solución de $f''(x^k)d = -f'(x^k)$,

$$d_N = \begin{bmatrix} -0.0198 \\ 4.8812 \end{bmatrix}, \quad ||d_N|| = 4.8812 \not\leq \Delta_0 = 1.$$

Entonces es necesario encontrar d, λ tales que

$$\begin{bmatrix} 922 + \lambda & -120 \\ -120 & 20 + \lambda \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = - \begin{bmatrix} 604 \\ -100 \end{bmatrix}, \quad ||d|| = 1.$$

Si se expresa todo en función de λ, como en (8.21), y se aplica un método para solución de ecuaciones en una variable, por ejemplo el método de la secante, se obtiene después de 10 iteraciones

$$\lambda = 22.7308, \quad d^0 = \begin{bmatrix} -0.5318 \\ 0.8469 \end{bmatrix}, \quad ||d^0|| = 1.$$

Cálculo de ρ_0 :

$$\rho_0 = \frac{f(x^0) - f(x^0 + d^0)}{f(x^0) - q_0(x^0 + d^0)} = \frac{254 - 17.6635}{254 - 39.6975} = \frac{236.3365}{214.3025} = 1.1028,$$

entonces

$$x^1 = x^0 + d^0 = \begin{bmatrix} 2.4682 \\ 4.8469 \end{bmatrix}, \quad \Delta_1 = 2\Delta_0 = 2.$$

$$f(x^1) = 17.6635, \quad f'(x^1) = \begin{bmatrix} 125.8844 \\ -24.9060 \end{bmatrix},$$
$$f''(x^1) = \begin{bmatrix} 539.1876 & -98.7295 \\ -98.7295 & 20.0000 \end{bmatrix}.$$

La dirección de Newton:

$$d_N = \begin{bmatrix} -0.0567 \\ 0.9655 \end{bmatrix}, \quad ||d_N|| = 0.9672 < 2.$$

Luego, esta dirección es la solución del subproblema.

$$\rho_1 = \frac{f(x^1) - f(x^1 + d^1)}{f(x^1) - q_1(x^1 + d^1)} = \frac{17.6635 - 1.9926}{17.6635 - 2.0725} = \frac{15.6709}{15.5910} = 1.0051,$$

entonces

$$x^2 = x^1 + d^1 = \begin{bmatrix} 2.4116 \\ 5.8124 \end{bmatrix}, \quad \Delta_2 = 2\Delta_1 = 4.$$

$$f(x^2) = 1.9926, \quad f'(x^2) = \begin{bmatrix} 3.1330 \\ -0.0642 \end{bmatrix},$$
$$f''(x^2) = \begin{bmatrix} 467.3788 & -96.4625 \\ -96.4625 & 20.0000 \end{bmatrix}.$$

La dirección de Newton:

$$d_N = \begin{bmatrix} -1.3264 \\ -6.3940 \end{bmatrix}, \quad ||d_N|| = 6.5301 \not< 4.$$

La dirección de Newton no sirve, es necesario encontrar $d,\ \lambda$ tales que

$$\big(f''(x^2) + \lambda I\big)d = -f'(x^2), \quad \lambda \geq 0, \quad ||d||_2 = 4.$$

En 5 iteraciones se obtiene,

$$\lambda = 0.0553, \quad d^2 = \begin{bmatrix} -0.8149 \\ -3.9161 \end{bmatrix}.$$

$$\rho_2 = \frac{f(x^2) - f(x^2 + d^2)}{f(x^2) - q_2(x^2 + d^2)} = \frac{1.9926 - 4.6224}{1.9926 - 0.3999} = \frac{-2.6298}{1.5927} = -1.6511,$$

entonces

$$x^3 = x^2 = \begin{bmatrix} 2.4116 \\ 5.8124 \end{bmatrix}, \quad \Delta_3 = \frac{1}{2}\Delta_2 = 2.$$

$$f(x^3) = 1.9926, \quad f'(x^3) = \begin{bmatrix} 3.1330 \\ -0.0642 \end{bmatrix},$$
$$f''(x^3) = \begin{bmatrix} 467.3788 & -96.4625 \\ -96.4625 & 20.0000 \end{bmatrix}.$$

La dirección de Newton:

$$d_N = \begin{bmatrix} -1.3264 \\ -6.3940 \end{bmatrix}, \quad ||d_N|| = 6.5301 \not< 4.$$

La dirección de Newton no sirve, es necesario encontrar $d,\ \lambda$ tales que

$$\big(f''(x^3) + \lambda I\big)d = -f'(x^3), \quad \lambda \geq 0, \quad ||d||_2 = 2.$$

En 6 iteraciones se obtiene,

$$\lambda = 0.1979, \quad d^3 = \begin{bmatrix} -0.4105 \\ -1.9574 \end{bmatrix}.$$

$$\rho_3 = \frac{f(x^3) - f(x^3 + d^3)}{f(x^3) - q_3(x^3 + d^3)} = \frac{1.9926 - 1.2246}{1.9926 - 1.0167} = \frac{0.7680}{0.9760} = 0.7870,$$

entonces

$$x^4 = x^3 + d^3 = \begin{bmatrix} 2.0010 \\ 3.8550 \end{bmatrix}, \quad \Delta_4 = 2\Delta_3 = 4.$$

Después de otras 7 iteraciones se obtiene

$$x^{11} = \begin{bmatrix} 1.0000 \\ 1.0000 \end{bmatrix}, \quad f'(x^{11}) = \begin{bmatrix} 0.0001 \\ 0.0000 \end{bmatrix}. \ \diamond$$

La siguiente proposición garantiza la convergencia global del algoritmo (desde cualquier punto inicial) a un punto crítico. Ver la demostración en [NaSo96],

Proposición 8.2. *Sean: x^0 un punto cualquiera, $\{x^k\}$ la sucesión definida por el algoritmo anterior. Si $\Gamma = \{x : f(x) \leq f(x^0)\}$ es acotado y f, f', f'' son continuas en Γ, entonces*

$$\lim_{k\to\infty} f'(x^k) = 0.$$

8.3.2 Otra estrategia general

La estrategia presentadad a continuación, [NoWr99], es muy parecida a la anterior. Las principales diferencias son:

- Hay un dato $\Delta_{\max}$ correspondiente a un radio máximo.
- Si ρ_k es muy pequeño, también se disminuye el radio Δ, pero queda en función de $||d^k||$.
- Para aumentar el radio Δ_k, además de que ρ_k sea grande, se necesita también que la solución de (8.10) quede en la frontera.

REGIÓN DE CONFIANZA: SEGUNDA ESTRATEGIA
datos: $x^0,\ \Delta_0 > 0,\ 0 < \eta < \frac{1}{4},\ \Delta_{\max},\ \varepsilon_g,$ `MAXIT`
para $k = 0, 1, ...,$ `MAXIT`
  si $||f'(x^k)|| < \varepsilon_g$ **ent parar**
  $d^k =$ solución de (8.10)
  calcular ρ_k según (8.15)
  si $\rho_k < \frac{1}{4}$ **ent** $\Delta_{k+1} = \frac{1}{4}||d^k||$
  sino
    si $\rho_k > \frac{3}{4}$ y $||d^k|| = \Delta_k$ **ent** $\Delta_{k+1} = \min(2\Delta_k, \Delta_{\max})$
    sino $\Delta_{k+1} = \Delta_k$
  fin-sino
  si $\rho_k > \eta$ **ent** $x^{k+1} = x^k + d^k$
  sino $x^{k+1} = x^k$
fin-para k

8.3.3 El punto de Cauchy

Es simplemente una modificación del descenso más pendiente, la dirección es equivalente, pero el paso λ_k debe permitir que se minimice en $B[x^k, \Delta_k]$. Visto como una aproximación de la solución de (8.10), se consideran dos pasos:

- Minimizar la proximación lineal de f, en la bola $B[x^k, \Delta_k]$ para obtener la dirección $\bar{d}$.
- Minimizar la aproximación cuadrática de f en la bola $B[x^k, \Delta_k]$, pero en la dirección $\bar{d}$.

Más que considerarlo como una buena aproximación de la solución de (8.10), el punto de Cauchy sirve como punto de compararación para saber si una aproximación obtenida por otro método es buena o no lo es.

El problema de la aproximación lineal de f en $B[x^k, \Delta_k]$ es simplemente

$$\begin{aligned} \min\ \bar{q}_k(x^k + d) &= f(x^k) + f'(x^k)^{\mathrm{T}} d \\ ||d||_2 &\leq \Delta_k. \end{aligned} \tag{8.22}$$

Su solución es:

$$\bar{d} = -\frac{\Delta_k}{||f'(x^k)||} f'(x^k). \tag{8.23}$$

La dirección de Cauchy será simplemente $\tau\bar{d}$, solución de

$$\begin{aligned} \min\ q_k(x^k+\tau\bar{d}) = f(x^k) + f'(x^k)^{\mathrm{T}}(\tau\bar{d}) + \frac{1}{2}(\tau\bar{d})^{\mathrm{T}} f''(x^k)(\tau\bar{d}) \\ ||(\tau\bar{d})||_2 \leq \Delta_k. \end{aligned}$$

Este problema depende únicamente de τ,

$$\begin{aligned} \min\ f(x^k) + \tau f'(x^k)^{\mathrm{T}}\bar{d} + \frac{1}{2}\tau^2 \bar{d}^{\mathrm{T}} f''(x^k)\bar{d} \\ |\tau| \leq \frac{\Delta_k}{||\bar{d}||_2} = 1. \end{aligned}$$

Se puede reescribir así:

$$\begin{aligned} \min\ \frac{1}{2}a\tau^2 + b\tau + f(x^k) \\ |\tau| \leq 1, \end{aligned}$$

donde

$$a = \bar{d}^{\mathrm{T}} f''(x^k)\bar{d}, \tag{8.24}$$

$$b = f'(x^k)^{\mathrm{T}}\bar{d}\ <\ 0. \tag{8.25}$$

La solución es:

$$\tau^* = \begin{cases} -\dfrac{b}{a} & \text{si } a > 0 \quad \text{y} \quad -\dfrac{b}{a} \leq 1, \\ 1 & \text{en los demás casos.} \end{cases} \tag{8.26}$$

En resumen, la solución aproximada (algunas veces exacta) de (8.10), para obtener el punto de Cauchy, se logra mediante los siguientes pasos:

REGIÓN DE CONFIANZA: PUNTO DE CAUCHY
$d_N =$ solución de $f''(x^k)d = -f'(x^k)$
si $||d_N||_2 \leq \Delta_k$ y $f''(x^k)$ s.d.p. **ent** $d_C = d_N$
sino
    calcular $\bar{d}$ según (8.23)
    calcular τ^* según (8.26)
    $d_C = \tau^*\bar{d}$
fin-sino
$x_C^k = x_C = x^k + d_C$

Ejemplo 8.4. Calcular el punto de Cauchy para $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ en el punto $x^0 = (3, 4)$ con $\Delta_0 = 1$.

$$f(x^0) = 254, \quad f'(x^0) = \begin{bmatrix} 604 \\ -100 \end{bmatrix}, \quad f''(x^0) = \begin{bmatrix} 922 & -120 \\ -120 & 20 \end{bmatrix}.$$

La dirección de Newton está dada por la solución de $f''(x^k)d = -f'(x^k)$,

$$d_N = \begin{bmatrix} -0.0198 \\ 4.8812 \end{bmatrix}, \quad ||d_N|| = 4.8812 \not< \Delta_0 = 1.$$

Entonces es necesario calcular $\bar{d}$,

$$\bar{d} = \begin{bmatrix} -0.9866 \\ 0.1633 \end{bmatrix}.$$

Cálculo de d_C :

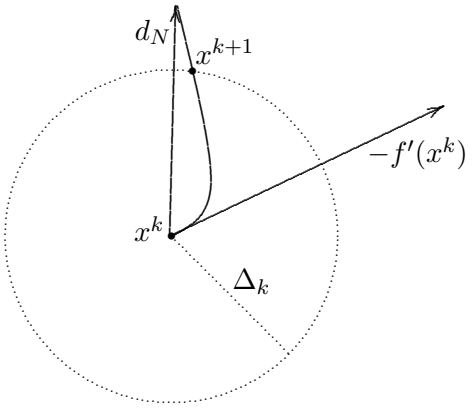
$$a = 936.6098, \quad b = -612.2222, \quad \tau^* = 0.653658, \quad d_C = \begin{bmatrix} -0.6449 \\ 0.1068 \end{bmatrix}.$$

$$x_C = \begin{bmatrix} 2.3551 \\ 4.1068 \end{bmatrix}, \quad f(x_C) = 22.5674\,. \quad \diamond$$

8.3.4 El método "dogleg"

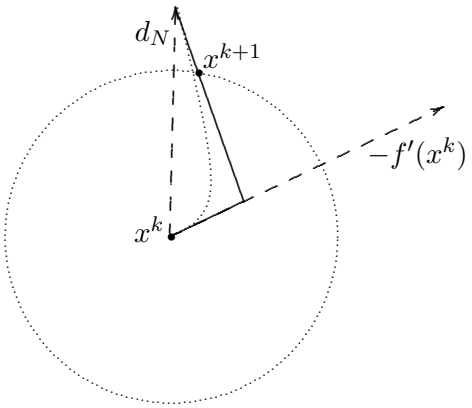
Este término viene de una palabra usada en el golf, cuya traducción literal sería pata de perro. Indica una curva aguda o un ángulo.

En la siguiente figura hay un ejemplo de la bola $B[x^k, \Delta_k]$, la dirección de Newton d_N, la dirección del descenso más pendiente (la opuesta al gradiente), lo que sería la trayectoria óptima de la solución de (8.10) para diferentes valores de Δ y el punto x^{k+1} dado por la solución exacta.



Inicialmente, cerca a x^k, la trayectoria casi coincide con la dirección $-f'(x^k)$. A medida que se aleja del centro, empieza a dar una curva y poco a poco se va acercando a la direccion de Newton. Para valores de Δ mayores que $||d_N||$, la solución es simplemente la dirección de Newton.

El método dogleg se puede usar cuando la matriz hessiana $f''(x^k)$ es semidefinida positiva. Si la dirección plena de Newton no sirve, se construye una trayectoria poligonal de dos tramos. El primero coincide con parte de la dirección del descenso más pendiente. El segundo tiende hacia la dirección de Newton.



Más precisamente, el primer tramo está dado por el mejor punto en la dirección de descenso:

$$d_D = -\frac{f'(x^k)^{\mathrm{T}} f'(x^k)}{f'(x^k)^{\mathrm{T}} f''(x^k) f'(x^k)} f'(x^k). \tag{8.27}$$

La trayectoria poligonal $\hat{d}(\tau)$ esta definida así:

$$\hat{d}(\tau) = \begin{cases} \tau d_D & \text{si} \quad 0 \leq \tau \leq 1, \\ d_D + (\tau - 1)(d_N - d_D) & \text{si} \quad 1 \leq \tau \leq 2. \end{cases} \tag{8.28}$$

Obviamente $\hat{d}(0) = 0, \;\; \hat{d}(1) = d_D, \;\; \hat{d}(2) = d_N$. Si $f''(x^k)$ es semidefinida positiva, esta trayectoria tiene dos propiedades importantes:

- $||\hat{d}(\tau)||$ es un función creciente en τ,
- $q_k(\hat{d}(\tau))$ es una función decreciente en τ.

Estas propiedades garantizan que para resolver el problema de minimizar $q_k(\hat{d}(\tau))$ con $||\hat{d}(\tau)|| \leq \Delta_k$, basta con encontrar el punto donde

$$||\hat{d}(\tau)|| = \Delta_k.$$

Si $||d_D|| \geq \Delta_k$, la solución es inmediata:

$$\tau^* = \frac{\Delta_k}{||d_D||}, \;\; d_P = \tau^* d_D. \tag{8.29}$$

En caso contrario, es necesario encontrar la solución de una ecuación cuadrática en τ:

$$||d_D + (\tau - 1)(d_N - d_D)||^2 = \Delta_k^2,$$

que se puede escribir $\alpha\tau^2 + \beta\tau + \gamma$. Su solución es:

$$\begin{aligned} &\alpha = ||d_N - d_D||^2, \;\; \beta = 2d_D^{\mathrm{T}}(d_N - d_D), \;\; \gamma = ||d_D||^2 - \Delta_k^2, \\ &\tau^* = 1 \; + \; \frac{-\beta + \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}, \\ &d_P = d_D + (\tau^* - 1)(d_N - d_D). \end{aligned} \tag{8.30}$$

Cuando $f''(x^k)$ no es semidefinida positiva, el método dogleg no se puede aplicar, se puede utilizar como salida alterna la dirección de Cauchy. En resumen, la solución aproximada (algunas veces exacta) de (8.10), por el método dogleg se logra mediante los siguientes pasos:

REGIÓN DE CONFIANZA: MÉTODO DOGLEG
si $f''(x^k)$ es s.d.p **ent**
  $d_N =$ solución de $f''(x^k)d = -f'(x^k)$
  si $||d_N||_2 \leq \Delta_k$ **ent** $d_P = d_N$
  sino
    calcular d_D según (8.29)
    si $||d_D|| \geq \Delta_k$ **ent** calcular d_P según (8.29)
    sino calcular d_P según (8.30)
  fin-sino
fin-ent
sino $d_P = d_C$
...
$x^{k+1} = x^k + d_P$

Ejemplo 8.5. Aplicar el método dogleg para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$ con $\Delta_0 = 1$.

$$f(x^0) = 254, \quad f'(x^0) = \begin{bmatrix} 604 \\ -100 \end{bmatrix}, \quad f''(x^0) = \begin{bmatrix} 922 & -120 \\ -120 & 20 \end{bmatrix}.$$

La matriz hessiana es semidefinida positiva. La dirección de Newton está dada por la solución de $f''(x^k)d = -f'(x^k)$,

$$d_N = \begin{bmatrix} -0.0198 \\ 4.8812 \end{bmatrix}, \quad ||d_N|| = 4.8812 \not< \Delta_0 = 1.$$

Para el primer tramo de la trayectoria,

$$d_D = \begin{bmatrix} -0.6449 \\ 0.1068 \end{bmatrix}, \quad ||d_D|| = 0.6537 < \Delta_0 = 1.$$

Entonces la solución está en el segundo tramo,

$$\alpha = 23.1858, \quad \beta = 0.2133, \quad \gamma = -0.5727,$$
$$\tau^* = 1.1526, \quad d_P = \begin{bmatrix} -0.5495 \\ 0.8355 \end{bmatrix}.$$

$$x + d_P = \begin{bmatrix} 2.4505 \\ 4.8355 \end{bmatrix}, \quad f(x + d_P) = 15.7833\,.$$

Si se calcula el punto de Cauchy

$$d_C = \begin{bmatrix} -0.6449 \\ 0.1068 \end{bmatrix}, \quad x + d_C = \begin{bmatrix} 2.3551 \\ 4.1068 \end{bmatrix}, \quad f(x + d_C) = 22.5674\,.$$

La reducción en f es mayor con d_P que con d_C, el cociente es aproximadamente 1.0293. Para pasar a la iteración siguiente,

$$\rho_0 = \frac{254 - 15.7833}{254 - 39.8242} = 1.1122\,.$$

entonces

$$x^1 = x^0 + d_P = \begin{bmatrix} 2.4505 \\ 4.8355 \end{bmatrix}, \quad \Delta_1 = 2.$$

$$f(x^1) = 15.7833, \quad f'(x^1) = \begin{bmatrix} 117.5450 \\ -23.3917 \end{bmatrix},$$
$$f''(x^1) = \begin{bmatrix} 529.1911 & -98.0212 \\ -98.0212 & 20.0000 \end{bmatrix}.$$

La dirección de Newton:

$$d_N = \begin{bmatrix} -0.0595 \\ 0.8781 \end{bmatrix}, \quad ||d_N|| = 0.8801 < 2.$$

Luego, esta dirección es la solución del subproblema.

$$\rho_1 = \frac{f(x^1) - f(x^1 + d^1)}{f(x^1) - q_1(x^1 + d^1)} = \frac{15.7833 - 1.9352}{15.7833 - 2.0178} = \frac{13.8481}{13.7655} = 1.0060,$$

entonces

$$x^2 = x^1 + d^1 = \begin{bmatrix} 2.3911 \\ 5.7136 \end{bmatrix}, \quad \Delta_2 = 2.$$

$$f(x^2) = 1.9352,\ f'(x^2) = \begin{bmatrix} 3.1204 \\ -0.0707 \end{bmatrix},\ f''(x^2) = \begin{bmatrix} 459.5155 & -95.6425 \\ -95.6425 & 20.0000 \end{bmatrix}.$$

La matriz hessiana es semidefinida positiva. La dirección de Newton es:

$$d_N = \begin{bmatrix} -1.2992 \\ -6.2093 \end{bmatrix}, \quad ||d_N|| = 6.3437 \not< \Delta_2 = 2.$$

Para el primer tramo de la trayectoria,

$$d_D = \begin{bmatrix} -0.0067 \\ 0.0002 \end{bmatrix}, \quad ||d_D|| = 0.0067 < \Delta_2 = 2.$$

Entonces la solución está en el segundo tramo,

$$\alpha = 40.2273, \quad \beta = 0.0155, \quad \gamma = -4.0000,$$
$$\tau^* = 1.3151, \quad d_P = \begin{bmatrix} -0.4140 \\ -1.9567 \end{bmatrix}.$$

$$x + d_P = \begin{bmatrix} 1.9770 \\ 3.7570 \end{bmatrix}, \quad f(x + d_P) = 1.1847\,.$$

Si se calcula el punto de Cauchy

$$d_C = \begin{bmatrix} -0.0067 \\ 0.0002 \end{bmatrix}, \quad x + d_C = \begin{bmatrix} 2.3843 \\ 5.7138 \end{bmatrix}, \quad f(x + d_C) = 1.9246\,.$$

La reducción en f es mayor con d_P (0.7505) que con d_C (0.0105), el cociente es aproximadamente 71.2341. Para pasar a la iteración siguiente,

$$\rho_2 = \frac{1.9352 - 1.1847}{1.9352 - 0.9706} = 0.7780\,.$$

entonces

$$x^3 = x^2 + d_P = \begin{bmatrix} 1.9770 \\ 3.7570 \end{bmatrix}, \quad \Delta_3 = 4.$$

Después de otras 8 iteraciones se obtiene $x = (1, 1)$. $\diamond$

Se puede mostrar que, bajo ciertas condiciones [Kel99], el método dog-leg tiene convergencia global (se tiene convergencia desde cualquier punto inicial) y además que la convergencia es superlineal.

Hay otras estrategias para hallar una solución aproximada de (8.10), por ejemplo, la minimización en un espacio bidimensional y el método de Steihaug. Este último está relacionado con el método del gradiente conjugado.

8.4 MÉTODO DEL GRADIENTE O DEL DESCENSO MÁS PENDIENTE

También se conoce con el nombre de método de Cauchy, método de la máxima pendiente, método del descenso más profundo y en inglés *steepest descent*. Es un método muy usado por su sencillez. Para ciertos tipos de funciones puede zigzaguear mucho y ser lento. Únicamente requiere primeras derivadas. Es un método de descenso.

Dado un punto no crítico x^k se escoge la dirección d^k como la mejor dirección local de descenso en x^k. Sean: λ un valor fijo, positivo y suficientemente pequeño, $d \neq 0$ una dirección tal que $||d||_2 = 1$,

$$f(x^k + \lambda d) = f(x^k) + \lambda f'(x^k)^{\mathrm{T}} d + \frac{\lambda^2}{2} d^{\mathrm{T}} f''(x^k) d^k + \cdots .$$

La expresión anterior depende únicamente de d y como λ es pequeño se puede aproximar por

$$\psi(d) = f(x^k) + \lambda f'(x^k)^{\mathrm{T}} d.$$

Para escoger d de manera óptima local, hay que resolver

$$\begin{aligned} \min \psi(d) &= f(x^k) + \lambda f'(x^k)^{\mathrm{T}} d, \\ ||d||_2^2 - 1 &= 0. \end{aligned}$$

La función ψ es afín, luego convexa y la función que define la igualdad es estrictamente convexa. Las condiciones necesarias de KKT dan:

$$\lambda f'(x^k) + 2v_1 d = 0.$$

Si se escogen

$$\begin{aligned} d &= -\frac{f'(x^k)}{||f'(x^k)||_2}, \\ v_1 &= \frac{\lambda ||f'(x^k)||_2}{2} > 0, \end{aligned}$$

se cumplen las condiciones necesarias y suficientes de KKT. Como las direcciones $-f'(x^k)/||f'(x^k)||_2$ y $-f'(x^k)$ son equivalentes se puede tomar la segunda como la dirección d^k, lo que garantiza que sea una dirección de descenso. El algoritmo se puede esquematizar así:

datos: $x^0,\ \varepsilon_g,\ \texttt{MAXIT},\ \ldots$
para $k = 0, ..., , \texttt{MAXIT}$
  si $||f'(x^k)|| \leq \varepsilon_g$ **ent parar**
  $d^k = -f'(x^k)$
  $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k),\ \lambda \geq 0$
  $x^{k+1} = x^k + \lambda_k^* d^k$
fin-para k

Ejemplo 8.6. Utilizar el método del descenso más pendiente para minimizar $f(x) = (x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$ partiendo de $x^0 = (3, 5)$,

k	x	$f(x)$	$f'(x)$	d	λ^*
0	3.0000	26.0000	8.0000	-8.0000	0.2500
	5.0000		12.0000	-12.0000	
1	1.0000	0.0000	0.0000		
	2.0000		0.0000		

No siempre este método funciona tan bien, ni siquiera para funciones cuadráticas. $\diamond$

Ejemplo 8.7. Utilizar el método del descenso más pendiente para minimizar $f(x) = 10x_1^2 + x_2^2$ partiendo de $x^0 = (2, 3)$.

k	x	$f(x)$	$f'(x)$	d	λ^*
0	2.0000	49.0000	40.0000	-40.0000	0.0510
	3.0000		6.0000	-6.0000	
1	-0.0404	7.2736	-0.8082	0.8082	0.4173
	2.6939		5.3879	-5.3879	
2	0.2969	1.0797	5.9377	-5.9377	0.0510
	0.4453		0.8906	-0.8906	
3	-0.0060	0.1603	-0.1200	0.1200	0.4173
	0.3999		0.7998	-0.7998	
4	0.0441	0.0238	0.8814	-0.8814	0.0510
	0.0661		0.1322	-0.1322	
7	-0.0001	0.0001	-0.0026	0.0026	0.4173
	0.0088		0.0176	-0.0176	$\diamond$

Aunque en los dos ejemplos se trata de funciones cuadráticas estrictamente convexas, la diferencia radica en la relación entre el valor propio más grande y el valor propio más pequeño de la matriz hessiana.

$$f''(x) = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \quad \begin{array}{l} \lambda_1 = 4 \\ \lambda_2 = 4 \end{array}, \quad f''(x) = \begin{bmatrix} 20 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{array}{l} \lambda_1 = 20 \\ \lambda_2 = 2. \end{array}$$

Para funciones cuadráticas se tienen los siguientes resultados: ([Lue89], [Gil81], [Per88]).

Proposición 8.3. *Sean:* $f(x) = \frac{1}{2}x^T Hx + c^T x$, *con* H *definida positiva,* x^* *el único minimizador global,* $\lambda_1 > 0$ *el valor propio de* H *más grande,* $\lambda_n > 0$ *el valor propio de* H *más pequeño,* $\{x^k\}$ *una sucesión determinada por el método del descenso más pendiente,* $\kappa = \kappa_2(H) = (\rho(H^T H))^{1/2} \geq 1$ *el condicionamiento o número de condición espectral de* H *(para matrices definidas positivas* $\kappa = \lambda_1/\lambda_n$*). Entonces*

$$\begin{aligned} f(x^{k+1}) - f(x^*) &\leq \frac{(\lambda_1 - \lambda_n)^2}{(\lambda_1 + \lambda_n)^2} \left(f(x^{k+1}) - f(x^*) \right), \\ f(x^{k+1}) - f(x^*) &\leq \frac{(\kappa - 1)^2}{(\kappa + 1)^2} \left(f(x^{k+1}) - f(x^*) \right), \\ x^k &\to x^*, \\ (x^{k+1} - x^k)^T (x^k - x^{k-1}) &= 0, \\ f'(x^k)^T f'(x^{k-1}) &= 0. \end{aligned}$$

Según los resultados anteriores, el método del descenso más pendiente tiene convergencia global. La sucesión $\{f(x^k)\}$ tiene convergencia lineal. La tasa de convergencia está acotada superiormente por $(\kappa - 1)^2/(\kappa + 1)^2 = (\lambda_1 - \lambda_n)^2/(\lambda_1 + \lambda_n)^2$. Es decir, entre más pequeño sea el condicionamiento (más cercano a 1), se garantiza una tasa de convergencia menor, o sea, una convergencia más rápida. Dicho de otra forma, la convergencia será mejor cuando λ_1 y λ_n sean semejantes. El caso "perfecto" se tiene cuando $\lambda_1 = \lambda_n$, o sea, $\kappa = 1$. Dos pasos consecutivos son ortogonales, o sea, la sucesión de puntos $\{x^k\}$ avanza formando ángulos de 90 grados.

El primer ejemplo representa el caso ideal, $\lambda_1 = \lambda_n$, $\kappa = 1$. En el segundo ejemplo $\kappa = 10$, la cota superior para la tasa de convergencia es 0.6694 y la tasa de convergencia real en este ejemplo fue 0.1484, que aunque relativamente pequeña, permite bastante zigzagueo.

Para funciones no necesariamente cuadráticas se tiene el siguiente resultado [Lue89]:

Proposición 8.4. *Sean: f con segundas derivadas parciales continuas, x^* un minimizador local de f, $f''(x^*)$ definida positiva, $\lambda_1 > 0$ su mayor valor propio, $\lambda_n > 0$ su menor valor propio. Si $\{x^k\}$ es una sucesión determinada por el método del descenso más pendiente que tiende a x^*, entonces la sucesión $\{f(x^k)\}$ tiene convergencia lineal con tasa de convergencia menor o igual a $(\lambda_1 - \lambda_n)^2/(\lambda_1 + \lambda_n)^2$.*

También hay resultados que garantizan la convergencia hacia un punto crítico si en lugar de calcular exactamente λ_k^*, se usa simplemente un λ_k que cumpla la regla de Armijo [Baz93].

Ejemplo 8.8. Utilizar el método del descenso más pendiente para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$, partiendo de $x^0 = (3, 4)$.

k	x	$f(x)$	$f'(x)$	d	λ^*
0	3.0000	254.0000	604.0000	-604.0000	0.0016
	4.0000		-100.0000	100.0000	
1	2.0335	1.0743	0.0419	-0.0419	0.1161
	4.1600		0.4979	-0.4979	
2	2.0286	1.0598	3.1241	-3.1241	0.0029
	4.1022		-0.2629	0.2629	
3	2.0195	1.0454	0.0389	-0.0389	0.1076
	4.1030		0.4952	-0.4952	
10	1.9796	0.9603	2.6522	-2.6522	0.0031
	3.9099		-0.1751	0.1751	
20	1.9286	0.8625	2.2930	-2.2930	0.0033
	3.7137		-0.1130	0.1130	
30	1.8851	0.7835	2.0918	-2.0918	0.0034
	3.5492		-0.0853	0.0853	
⋮					◇

8.5 MÉTODOS DE DIRECCIONES CONJUGADAS

Definición 8.1. Sea H una matriz simétrica. Un conjunto de vectores v^1, v^2, ..., v^m son **conjugados con respecto a** H o simplemente **H-conjugados** si

$$v^{i\,\mathrm{T}} H\, v^j = 0, \ \forall i \neq j.$$

Cuando la matriz H es simplemente la identidad se tiene el concepto usual de ortogonalidad. Si $H = 0$ cualquier conjunto de vectores es H-conjugado. Los métodos de direcciones conjugadas, para funciones cuadráticas $f(x) = \frac{1}{2}x^{\mathrm{T}}Hx + c^{\mathrm{T}}x$ estrictamente convexas (H definida positiva), construyen direcciones H-conjugadas y convergen en un número finito de pasos.

Proposición 8.5. *Sean: $f(x) = \frac{1}{2}x^T Hx + c^T x$ con H simétrica y definida positiva, d^1, d^2, ..., d^n direcciones H-conjugadas, x^1 un vector inicial cualquiera, x^2, x^3, ..., x^{n+1} obtenidos como minimizadores a lo largo de las direcciones d^1, d^2, ..., d^n, o sea, $x^{k+1} = x^k + \lambda_k^* d^k$ con $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k)$. Entonces para $k = 1, 2, ..., n$ se cumple:*

- $f'(x^{k+1})^T d^j = 0$ *para* $j = 1, 2, ..., k$,
- $f'(x^1)^T d^k = f'(x^k)^T d^k$,
- x^{k+1} *es el minimizador de* f *en* $\mathrm{Gen}(d^1, d^2, ..., d^k) + x^1$,
- $x^{n+1} = x^*$ *minimizador de* f *en* $\mathbb{R}^n$.

$\mathrm{Gen}(d^1, d^2, ..., d^k)$ es el subespacio vectorial generado por los vectores d^1, d^2, ..., d^k, es decir, el conjunto de combinaciones lineales de d^1, d^2, ..., d^k. La cuarta afirmación de la proposición es un caso particular de la tercera. Más aún, en algunos casos se puede obtener x^* antes de la iteración n. Lo interesante de los métodos de direcciones conjugadas es que, por lo menos teóricamente, a lo más en n iteraciones se obtiene el minimizador. Si por errores de redondeo, no se obtiene x^* en n iteraciones, entonces se utiliza x^{n+1} como nuevo x^1 para volver a empezar.

Los métodos de direcciones conjugadas más conocidos son: el método del gradiente conjugado, el método DFP (Davidon, Fletcher y Powell), el método BFGS (Broyden, Fletcher, Goldfarb y Shanno) y, en general, los métodos de Broyden.

Los métodos de direcciones conjugadas se pueden aplicar también a funciones que no sean cuadráticas, pero ya no se puede garantizar la convergencia finita. De todas formas son buenos métodos y en la medida en que la función f sea parecida a una función cuadrática, en esa misma medida la convergencia resulta mejor.

8.6 MÉTODO DEL GRADIENTE CONJUGADO: GC

Este método fue propuesto en 1952 por Hestenes y Stiefel para resolver sistemas de ecuaciones $Ax = b$ con matriz definida positiva. En 1964 Fletcher y Reeves lo adaptan para minimización de funciones cuadráticas y no cuadráticas (resolver $Ax = b$ es equivalente a minimizar $f(x) = \frac{1}{2}x^{\mathrm{T}}Ax - b^{\mathrm{T}}x$ si A es definida positiva). Aunque no es tan bueno como otros métodos de direcciones conjugadas, tiene dos grandes ventajas: su sencillez y su pequeña necesidad de memoria.

Para funciones cualesquiera, éste es el esquema del método GC:

datos: x^1, MAXIT , ε_g , ...
para $K = 1, ..., \texttt{MAXIT}$
  para $k = 1, ..., n$
    si $||f'(x^k)|| < \varepsilon_g$ **ent parar**
    si $k = 1$ **ent** $d^k = -f'(x^k)$
    sino
      $\alpha_k = ||f'(x^k)||_2^2 / ||f'(x^{k-1})||_2^2$
      $d^k = -f'(x^k) + \alpha_k d^{k-1}$
    fin-sino
    $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k), \lambda \geq 0$
    $x^{k+1} = x^k + \lambda_k^* d^k$
  fin-para k
  $x^1 = x^{n+1}$
fin-para K

Obviamente se desea que el algoritmo acabe porque el gradiente es casi nulo y no porque $K > \texttt{MAXIT}$. Una salida anormal es cuando $f(x^k + \lambda d^k)$ no parece tener minimizador, lo que indica que f no tendría minimizador global. Siempre en la primera iteración $d^1 = -f'(x^1)$, o sea, siempre se empieza con una iteración del descenso más pendiente. En cada iteración la dirección es la dirección del descenso más pendiente modificada por la dirección anterior para obtener, bajo ciertas condiciones, direcciones conjugadas.

Ejemplo 8.9. Utilizar el método del gradiente conjugado para minimizar $f(x_1, x_2) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$, partiendo de $x^1 = (3, 4)$.

K	k	x	$f(x)$	$f'(x)$	α	d	λ^*
1	1	3.0000	254.0000	604.0000		-604.0000	0.0016
		4.0000		-100.0000		100.0000	
	2	2.0336	1.0743	0.0798	6.5E-7	-0.0802	0.4355
		4.1600		0.4887		-0.4886	
2	1	1.9987	1.0201	5.8024		-5.8024	0.0030
		3.9472		-0.9519		0.9519	
	2	1.9815	0.9690	0.0780	0.0067	-0.1170	2.7545
		3.9501		0.4756		-0.4692	
3	1	1.6592	0.5253	7.6424		-7.6424	0.0042
		2.6575		-1.9058		1.9058	
	2	1.6271	0.3965	0.0893	0.0022	-0.1060	2.3540
		2.6655		0.3580		-0.3538	
⋮							
8	1	1.0000	0.0000	0.0000			
		1.0000		0.0000			◇

Proposición 8.6. *Para funciones cuadráticas estrictamente convexas, o sea, $f(x) = \frac{1}{2}x^T Hx + c^T x$, con H definida positiva, el método GC genera direcciones de descenso y conjugadas con respecto a H.*

Obviamente se tienen todas las propiedades ya vistas, de métodos de direcciones conjugadas, como por ejemplo convergencia en, a lo más, n iteraciones. Además, si se conoce explícitamente la matriz H, el cálculo del λ_k^* no requiere minimización en una variable sino simplemente aplicar una fórmula:

$$\begin{aligned}
\lambda_k^* &= -\frac{d^{k\mathrm{T}} f'(x^k)}{d^{k\mathrm{T}} H d^k} \\
&= -\frac{(-f'(x^k) + \alpha_k d^{k-1})^{\mathrm{T}} f'(x^k)}{d^{k\mathrm{T}} H d^k} \\
&= \frac{||f'(x^k)||_2^2 - \alpha_k d^{k-1\mathrm{T}} f'(x^k)}{d^{k\mathrm{T}} H d^k} \\
\lambda_k^* &= \frac{||f'(x^k)||_2^2}{d^{k\mathrm{T}} H d^k}. \qquad (8.31)
\end{aligned}$$

Si la función es cuadrática y estrictamente convexa, pero no se conoce explícitamente la matriz H, entonces de todas formas con una sola interpolación cuadrática se obtiene λ_k^*.

Ejemplo 8.10. Utilizar el método del gradiente conjugado para minimizar $f(x_1, x_2, x_3) = 10(x_1 + x_2 + x_3)^2 + 2x_2^2 + x_3^2$, partiendo de $x^1 = (3, 4, 5)$.

K	k	x	$f(x)$	$f'(x)$	α	d	λ^*
1	1	3.0000	1497.0000	240.0000		-240.0000	0.0161
		4.0000		256.0000		-256.0000	
		5.0000		250.0000		-250.0000	
	2	-0.8683	0.9795	-0.4773	1.8E-5	0.4730	0.4097
		-0.1262		-0.9820		0.9774	
		0.9706		1.4638		-1.4683	
	3	-0.6745	0.2963	-0.6238	0.1878	0.7126	0.9464
		0.2743		0.4734		-0.2898	
		0.3690		0.1141		-0.3899	
2	1	0.0000	0.0000	0.0000			
		0.0000		0.0000			
		0.0000		0.0000			

Se puede comprobar que las direcciones obtenidas: $d^1 = (-240, -256, -250)$, $d^2 = (0.4730, 0.9774, -1.4683)$, $d^3 = (0.7126, -0.2898, -0.3899)$, son conjugadas con respecto a H, es decir, ${d^1}^{\mathrm{T}} H d^2 = 0$, ${d^1}^{\mathrm{T}} H d^3 = 0$, ${d^2}^{\mathrm{T}} H d^3 = 0$, donde

$$H = f''(x) = \begin{bmatrix} 20 & 20 & 20 \\ 20 & 24 & 20 \\ 20 & 20 & 22 \end{bmatrix}.$$

8.7 MÉTODO DE DAVIDON, FLETCHER Y POWELL: DFP

Este método fue propuesto por Davidon en 1959 y desarrollado por Fletcher y Powell en 1963. También es un método de direcciones conjugadas para funciones cuadráticas. En el método de Newton $d^k = -\left(f''(x^k)\right)^{-1} f'(x^k)$, en el método DFP $d^k = -D^k f'(x^k)$. Por esta razón el método de DFP es uno de los métodos llamados **cuasi-Newton**. El método DFP también es conocido con el nombre de **métrica variable**. Bajo ciertas condiciones en el método DFP se puede llegar a $D^k = \left(f''(x^*)\right)^{-1}$. Si D^k es definida positiva y $f'(x^k) \neq 0$, entonces ${d^k}^{\mathrm{T}} f'(x^k) = -{f'(x^k)}^{\mathrm{T}} D^k f'(x^k) < 0$, luego d^k es una dirección de descenso.

De manera semejante al GC, una iteración completa está compuesta por n subiteraciones, se parte de x^1 y se llega a x^{n+1}. Si al final de las n subiteraciones no se tiene una buena aproximación de la solución, entonces el punto x^{n+1} obtenido se convierte en un nuevo x^1 para empezar otra iteración completa. Para funciones cualesquiera, este es el esquema del método DFP:

datos: $x^1,\ \varepsilon_g,\ \texttt{MAXIT}, ...$
para $K = 1, ..., \texttt{MAXIT}$
  para $k = 1, ..., n$
    si $||f'(x^k)|| < \varepsilon_g$ **ent** **parar**
    si $k = 1$ **ent** $D^1 = I$ o una matriz def. pos.
    sino
      $p = x^k - x^{k-1},\ \ q = f'(x^k) - f'(x^{k-1}),\ \ s = D^k q$
      $D^k = D^{k-1} + \dfrac{1}{p^{\mathrm{T}} q} p p^{\mathrm{T}} - \dfrac{1}{s^{\mathrm{T}} q} s s^{\mathrm{T}}$
    fin-sino
    $d^k = -D^k f'(x^k)$
    $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k), \lambda \geq 0$
    $x^{k+1} = x^k + \lambda_k^* d^k$
  fin-para k
  $x^1 = x^{n+1}$
fin-para K

Obviamente se desea que el algoritmo acabe porque el gradiente es casi nulo y no porque $K > \texttt{MAXIT}$. Una salida anormal es cuando $f(x^k + \lambda d^k)$ no parece tener minimizador, lo que indica que f no tendría minimizador global. Si $D^1 = I$, entonces $d^1 = -f'(x^1)$, o sea, se empieza con una iteración del descenso más pendiente. Si se calculara D^{n+1} sería una aproximación de la inversa de la matriz hessiana en puntos cercanos al minimizador. Para funciones cuadráticas $D^{n+1} = H^{-1}$.

La matriz D se modifica en cada iteración sumándole una matriz C, llamada matriz de corrección

$$\begin{aligned} D^k &= D^{k-1} + C^k, \\ C^k &= \frac{1}{p^{k\mathrm{T}} q^k} p^k p^{k\mathrm{T}} - \frac{1}{s^{k\mathrm{T}} q^k} s^k s^{k\mathrm{T}}. \end{aligned}$$

Esta matriz de corrección es igual a una suma de dos matrices simétricas de rango uno, por eso el método de DFP hace parte de los métodos con nombre **corrección de rango dos**.

Proposición 8.7. *En el método DFP para una función cualquiera:*

- *Las matrices D^k son definidas positivas.*
- *Las direcciones son de descenso.*
- *Si en cada iteración no se escoge un minimizador exacto, pero $p^T q > 0$, entonces las matrices D^k son definidas positivas.*
- *Si x^1 está suficientemente cerca de x^*, el método tiene convergencia superlineal.*
- *Si f es estrictamente convexa, la convergencia es global.*

La convergencia se llama **global** cuando no depende del punto inicial.

Ejemplo 8.11. Utilizar el método DFP para minimizar $f(x_1, x_2) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$, partiendo de $x^1 = (3, 4)$.

K	k	x	$f(x)$	$f'(x)$	$D_{\cdot 1}$	$D_{\cdot 2}$	d	λ^*
1	1	3.0000	254.0	604.00	1.0000	0.0000	-604.00	0.0016
		4.0000		-100.00	0.0000	1.0000	100.00	
	2	2.0335	1.074	0.0419	0.0285	0.1617	-0.0817	0.4574
		4.1600		0.4979	0.1617	0.9731	-0.4913	
2	1	1.9961	1.017	5.9255	1.0000	0.0000	-5.9255	0.0030
		3.9353		-0.9852	0.0000	1.0000	0.9852	
	2	1.9785	0.963	0.0789	0.0616	0.2346	-0.1162	2.7675
		3.9382		0.4746	0.2346	0.9414	-0.4653	
3	1	1.6570	0.522	7.6152	1.0000	0.0000	-7.6152	0.0042
		2.6505		-1.9014	0.0000	1.0000	1.9014	
	2	1.6250	0.394	0.0892	0.0866	0.2743	-0.1057	2.3577
		2.6585		0.3572	0.2743	0.9176	-0.3522	
⋮								
8	1	1.0000	0.000	0.0001	1.0000	0.0000	-0.0001	0.0098
		1.0000		-0.0000	0.0000	1.0000	0.0000	
	2	1.0000	0.000	0.0000				
		1.0000		0.0000				

Si se hubiera hecho el cálculo de las matrices D^{n+1}, se hubiera obtenido

$$\begin{bmatrix} 0.0238 & 0.1197 \\ 0.1197 & 0.6262 \end{bmatrix}, \quad \begin{bmatrix} 0.2029 & 0.7789 \\ 0.7789 & 3.0124 \end{bmatrix}, \quad \begin{bmatrix} 0.2467 & 0.7819 \\ 0.7819 & 2.5024 \end{bmatrix},$$

$$\dots \quad \begin{bmatrix} 0.5002 & 1.0008 \\ 1.0008 & 2.0526 \end{bmatrix},$$

que son aproximaciones de la inversa de la matriz hessiana:

$$f''(x^*) = \begin{bmatrix} 82 & -40 \\ -40 & 20 \end{bmatrix}, \quad \left(f''(x^*)\right)^{-1} = \begin{bmatrix} 0.50 & 1.00 \\ 1.00 & 2.05 \end{bmatrix}. \ \diamond$$

Proposición 8.8. *En el método DFP si $f(x) = \frac{1}{2}x^T Hx + c^T x$ con H definida positiva, entonces*

- *las direcciones son H-conjugadas,*
- *se obtiene x^* en a lo más n iteraciones,*
- *si $D^1 = I$ se obtienen direcciones equivalentes a las del GC y los mismos x^k.*

Ejemplo 8.12. Utilizar el método DFP para minimizar $f(x_1, x_2, x_3) = 10(x_1 + x_2 + x_3)^2 + 2x_2^2 + x_3^2$, a partir de $x^1 = (3, 4, 5)$.

Ver la tabla de resultados al final del capítulo.

Si se hubiera calculado la matriz D^{n+1} se hubiera obtenido

$$\begin{bmatrix} 0.8000 & -0.2500 & -0.5000 \\ -0.2500 & 0.2500 & 0.0000 \\ -0.5000 & 0.0000 & 0.5000 \end{bmatrix},$$

que es exactamente la inversa de la matriz hessiana.

$$f''(x^*) = \begin{bmatrix} 20 & 20 & 20 \\ 20 & 24 & 20 \\ 20 & 20 & 22 \end{bmatrix}.$$

Las 3 direcciones obtenidas son iguales o equivalentes a las obtenidas en el método GC para este mismo ejemplo. $\diamond$

Algunas de las ideas importantes de la presentación del método DFP hecha anteriormente son:

$$\begin{aligned} D^1 &= I, \\ d^k &= -D^k f'(x^k), \\ D^k &= D^{k-1} + C^k, \\ D^{n+1} &\approx (f''(x^*))^{-1}. \end{aligned}$$

Existe otra presentación del método DFP, obviamente equivalente a la anterior, pero donde se trabaja con matrices B que corresponden a las inversas de las matrices D:

$$\begin{aligned} B^1 &= I, \\ \text{resolver} \quad B^k d^k &= -f'(x^k), \\ B^k &= B^{k-1} + E^k, \\ B^{n+1} &\approx f''(x^*). \end{aligned}$$

Las fórmulas precisas que definen esta otra presentación del método DFP son:

$$\begin{aligned} B^1 &= I \\ \text{resolver} \quad B^k d^k &= -f'(x^k) \\ p &= x^k - x^{k-1} \\ q &= f'(x^k) - f'(x^{k-1}) \\ u &= B^k p \\ \gamma &= p^{\mathrm{T}} u \\ \rho &= p^{\mathrm{T}} q \\ w &= \frac{1}{\rho} q - \frac{1}{\gamma} u \\ B^k &= B^{k-1} - \frac{1}{\gamma} u u^{\mathrm{T}} + \frac{1}{\rho} q q^{\mathrm{T}} + \gamma w w^{\mathrm{T}}. \end{aligned}$$

Ejemplo 8.13. Utilizar el método DFP para minimizar $f(x_1, x_2, x_3) = 10(x_1 + x_2 + x_3)^2 + 2x_2^2 + x_3^2$, a partir de $x^1 = (3, 4, 5)$.

Ver la tabla de resultados al final del capítulo.

Si se hubiera calculado la matriz B^{n+1} se hubiera obtenido

$$\begin{bmatrix} 20 & 20 & 20 \\ 20 & 24 & 20 \\ 20 & 20 & 22 \end{bmatrix}. \quad \diamond$$

8.8 MÉTODO DE BROYDEN, FLETCHER, GOLDFARB Y SHANNO: BFGS

Es semejante al método DFP, construye direcciones conjugadas y tiene las propiedades de estos métodos. También es un método cuasi-Newton, con corrección de rango dos. El método DFP y el BFGS son casos particulares de la familia de métodos de Broyden. El método BFGS es bastante popular y, según muchos ensayos hechos, es numéricamente mejor que el DFP. Otra ventaja es que si en el BFGS no se hace minimización exacta en una dirección sino minimización imprecisa que cumpla el criterio de Goldstein y el de Wolfe-Powell se puede garantizar convergencia global. Para el DFP no hay resultado análogo.

La única diferencia con respecto al método DFP consiste en la manera como se hace la correción de D^{k-1} para obtener D^k.

$$\begin{aligned} C^k &= \frac{1}{p^{k\mathrm{T}} q^k} \left(1 + \frac{q^{k\mathrm{T}} s^k}{p^{k\mathrm{T}} q^k}\right) p^k p^{k\mathrm{T}} - \frac{1}{p^{k\mathrm{T}} q^k} \left(p^k s^{k\mathrm{T}} + s^k p^{k\mathrm{T}}\right), \\ D^k &= D^{k-1} + C^k. \end{aligned} \tag{8.32}$$

Ejemplo 8.14. Utilizar el método BFGS para minimizar $f(x_1, x_2) = (1 - x_1)^2 + 10(x_2 - x_1^2)^2$, a partir de $x^1 = (3, 4)$.

Ver tabla de resultados más adelante.

Si se hubiera calculado la matriz D^{n+1} se hubiera obtenido

$$\begin{bmatrix} 0.0265 & 0.1303 \\ 0.1303 & 0.6684 \end{bmatrix}, \quad \begin{bmatrix} 0.2350 & 0.8806 \\ 0.8806 & 3.3349 \end{bmatrix}, \quad \begin{bmatrix} 0.2793 & 0.8653 \\ 0.8653 & 2.7158 \end{bmatrix}$$

MÉTODO BFGS
datos: $x^1,\ \varepsilon_g,\ \texttt{MAXIT}, ...$
para $K = 1, ..., \texttt{MAXIT}$
  para $k = 1, ..., n$
    si $||f'(x^k)|| < \varepsilon_g$ **ent parar**
    si $k = 1$ **ent** $D^1 = I$ o una matriz def. pos.
    sino
      $p = x^k - x^{k-1},\ q = f'(x^k) - f'(x^{k-1}),\ s = D^k q,\ \gamma = p^{\mathrm{T}} q$
      $D^k = D^{k-1} + \frac{1}{\gamma}(1 + \frac{q^{\mathrm{T}} s}{\gamma}) p p^{\mathrm{T}} - \frac{1}{\gamma}(s p^{\mathrm{T}} + p s^{\mathrm{T}})$
    fin-sino
    $d^k = -D^k f'(x^k)$
    $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k), \lambda \geq 0$
    $x^{k+1} = x^k + \lambda_k^* d^k$
  fin-para k
  $x^1 = x^{n+1}$
fin-para K

$$... \qquad \begin{bmatrix} 0.5002 & 1.0008 \\ 1.0008 & 2.0526 \end{bmatrix}.$$

Los puntos x^k obtenidos son los mismos del método DFP, algunas matrices D^k cambian ligeramente. Como se verá en el ejemplo siguiente las matrices D^k no son siempre iguales en los dos métodos. $\diamond$

Ejemplo 8.15. Utilizar el método BFGS para minimizar la función $f(x_1, x_2, x_3) = 10(x_1 + x_2 + x_3)^2 + 2x_2^2 + x_3^2$, a partir de $x^1 = (3, 4, 5)$.

Ver la tabla de resultados al final del capítulo.

Si se hubiera calculado la matriz D^{n+1} se hubiera obtenido

$$\begin{bmatrix} 0.8000 & -0.2500 & -0.5000 \\ -0.2500 & 0.2500 & 0.0000 \\ -0.5000 & 0.0000 & 0.5000 \end{bmatrix},$$

que es exactamente la inversa de la matriz hessiana. $\diamond$

Para el método BFGS también existe una presentación con matrices B

$$B^1 = I$$

K	k	x	$f(x)$	$f'(x)$	$D_{\cdot 1}$	$D_{\cdot 2}$	d	λ^*
1	1	3.0000	254.0	604.00	1.0000	0.0000	-604.00	0.0016
		4.0000		-100.00	0.0000	1.0000	100.00	
	2	2.0335	1.074	0.0419	0.0285	0.1617	-0.0817	0.4574
		4.1600		0.4979	0.1617	0.9731	-0.4913	
2	1	1.9961	1.017	5.9255	1.0000	0.0000	-5.9255	0.0030
		3.9353		-0.9852	0.0000	1.0000	0.9852	
	2	1.9785	0.963	0.0789	0.0620	0.2361	-0.1169	2.7498
		3.9382		0.4746	0.2361	0.9474	-0.4683	
3	1	1.6570	0.522	7.6152	1.0000	0.0000	-7.6152	0.0042
		2.6505		-1.9014	0.0000	1.0000	1.9014	
	2	1.6250	0.394	0.0892	0.0868	0.2749	-0.1059	2.3525
		2.6585		0.3572	0.2749	0.9196	-0.3530	
⋮								
8	1	1.0000	0.000	0.0001	1.0000	0.0000	-0.0001	0.0098
		1.0000		-0.0000	0.0000	1.0000	0.0000	
	2	1.0000	0.000	0.0000				
		1.0000		0.0000				

Ejemplo 8.14

$$\begin{aligned} \text{resolver} \quad B^k d^k &= -f'(x^k) \\ p &= x^k - x^{k-1} \\ q &= f'(x^k) - f'(x^{k-1}) \\ B^k &= B^{k-1} + \frac{1}{q^{\mathrm{T}} p} q q^{\mathrm{T}} - \frac{1}{p^{\mathrm{T}} B^k p} B^k p p^{\mathrm{T}} B^k \end{aligned}$$

Es interesante observar la "dualidad" o "complementariedad" que existe entre las fórmulas de D y de B en los métodos DFP y BFGS:

$$D^{k+1} = D^k + \frac{1}{p^{\mathrm{T}} q} p p^{\mathrm{T}} - \frac{1}{q^{\mathrm{T}} D^k q} D^k q q^{\mathrm{T}} D^k, \tag{DFP}$$

$$B^{k+1} = B^k + \left(1 + \frac{p^{\mathrm{T}} B^k p}{q^{\mathrm{T}} p}\right) \frac{1}{q^{\mathrm{T}} p} q q^{\mathrm{T}} - \frac{1}{q^{\mathrm{T}} p} \left(q p^{\mathrm{T}} B^k + B^k p q^{\mathrm{T}}\right), \tag{DFP}$$

$$D^{k+1} = D^k + \left(1 + \frac{q^{\mathrm{T}} D^k q}{p^{\mathrm{T}} q}\right) \frac{1}{p^{\mathrm{T}} q} p p^{\mathrm{T}} - \frac{1}{p^{\mathrm{T}} q} \left(p q^{\mathrm{T}} D^k + D^k q p^{\mathrm{T}}\right), \tag{BFGS}$$

$$B^{k+1} = B^k + \frac{1}{q^{\mathrm{T}} p} q q^{\mathrm{T}} - \frac{1}{p^{\mathrm{T}} B^k p} B^k p p^{\mathrm{T}} B^k. \tag{BFGS}$$

8.9 MÉTODO CÍCLICO COORDENADO CONTINUO O EXACTO: CCC

Este es un método de descenso que no utiliza derivadas sino simplemente los valores de la función. La idea es muy sencilla, se trata de minimizar primero a lo largo de la primera dirección canónica $e^1 = [1 \;\; 0 \; ... \; 0]^{\mathrm{T}}$, enseguida a lo largo de $e^2 = [0 \;\; 1 \;\; 0 \; ... \; 0]^{\mathrm{T}}$ ... y, finalmente, a lo largo de la dirección $e^n = [0 \;\; 0 \; ... \; 0 \;\; 1]^{\mathrm{T}}$ y de nuevo volver a repetir el proceso hasta que haya convergencia o hasta que haya indicios de problemas.

datos: $x^0,\ \varepsilon_x,\ MAXIT,\ \ldots$
para para $k = 0, ..., \texttt{MAXIT}$
  $y^1 = x^k$
  para $j = 1, ..., n$
    $\lambda_j^* = \operatorname{argmin} f(y^j + \lambda e^j)$
    $y^{j+1} = y^j + \lambda_j^* e^j$
  fin-para j
  $x^{k+1} = y^{n+1}$
  si $||x^{k+1} - x^k|| < \varepsilon_x$ **ent parar**
fin-para k

Ejemplo 8.16. Utilizar el método cíclico coordenado continuo para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$.

k	x_1^k	x_2^k	$f(x^k)$	j	λ_j^*	y_1^{j+1}	y_2^{j+1}	$f(y^{j+1})$
0	3.0000	4.0000	254.00	1	-1.0062	1.9938	4.0000	0.994
				2	-0.0249	1.9938	3.9751	0.988
1	1.9938	3.9751	0.988	1	-0.0062	1.9875	3.9751	0.981
				2	-0.0248	1.9875	3.9502	0.975
2	1.9875	3.9502	0.975	1	-0.0062	1.9813	3.9502	0.969
				2	-0.0248	1.9813	3.9256	0.963
	...							
30	1.8135	3.2887	0.662	1	-0.0062	1.8073	3.2887	0.657
				2	-0.0223	1.8073	3.2663	0.652
	...							$\diamond$

Ejemplo 8.17. Utilizar el método de cíclico coordenado continuo para minimizar $f(x) = 10x_1^2 + x_2^2$ partiendo de $x^0 = (2, 3)$.

k	x_1^k	x_2^k	$f(x^k)$	j	λ_j^*	y_1^{j+1}	y_2^{j+1}	$f(y^{j+1})$
0	2.0000	3.0000	49.000	1	-2.0000	0.0000	3.0000	9.000
				2	-3.0000	0.0000	0.0000	0.000
1	0.0000	0.0000	0.000	1	0.0000	0.0000	0.0000	0.000
				2	0.0000	0.0000	0.0000	0.000

En el ejemplo 8.16 la convergencia es muy lenta. En el ejemplo actual la convergencia es muy rápida, las curvas de nivel de $f(x) = 10x_1^2 + x_2^2$ son elipses con ejes paralelos a los ejes coordenados. Si se aplica el método cíclico a $f(x) = 10(x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$, cuyas curvas de nivel son elipses, pero con ejes inclinados, la convergencia resulta lenta. $\diamond$

8.10 MÉTODO CÍCLICO COORDENADO DISCRETO O INEXACTO: CCD

Este es un método de descenso, que no utiliza derivadas sino simplemente los valores de la función. La idea es muy sencilla, se trata de encontrar un punto mejor, primero a lo largo de la primera dirección canónica $e^1 = [1 \;\; 0 \;... \; 0]^{\mathrm{T}}$, enseguida a lo largo de $e^2 = [0 \;\; 1 \;\; 0 \;... \; 0]^{\mathrm{T}}$ ... y, finalmente, a lo largo de la dirección $e^n = [0 \;\; 0 \;... \; 0 \;\; 1]^{\mathrm{T}}$. De nuevo se vuelve a repetir el proceso hasta que haya convergencia o hasta que haya indicios de problemas. Dicho de otra forma, para cada dirección canónica en el método cíclico coordenado continuo se encuentra λ_j^*, en el método discreto basta con encontrar un λ_j tal que

$$f(y^j + \lambda_j e^j) < f(y^j),$$

entonces

$$y^{j+1} = y^j + \lambda_j e^j.$$

Si no es posible encontrar el λ_j adecuado,

$$y^{j+1} = y^j.$$

La búsqueda de un punto mejor está limitada por el tamaño de λ, es decir, dado un ε (que puede ser un ε_j dependiente de j), se desea buscar un λ_j tal que $|\lambda_j| > \varepsilon$ y $f(y^j + \lambda_j e^j) < f(y^j)$.

Es muy probable que el método discreto requiera más iteraciones que el continuo, pero es posible que el número de evaluaciones de f sea menor, y así también el tiempo total sea menor.

datos: $x^0,\ \varepsilon_x,\ \varepsilon_1,\ ...,\ \varepsilon_n,\ MAXIT,\ \ldots$
para $k = 0, ..., \texttt{MAXIT}$
  $y^1 = x^k$
  para $j = 1, ..., n$
    buscar λ_j tal que $|\lambda_j| > \varepsilon_j$ y $f(y^j + \lambda_j e^j) < f(y^j)$
    si se encontró tal λ_j **ent** $y^{j+1} = y^j + \lambda_j e^j$
    sino $y^{j+1} = y^j$
  fin-para j
  $x^{k+1} = y^{n+1}$
  si $||x^{k+1} - x^k|| < \varepsilon_x$ **ent parar**
fin-para k

La búsqueda de λ_j puede hacerse por medio del algoritmo visto en la sección 7.9, en la parte correspondiente a encontrar un punto mejor en el método de los tres puntos, donde $x = y^j$, $d = e^j$, $\lambda_0 = 0$. El valor de η_0 no debe ser ni muy pequeño ni muy grande y puede estar relacionado con el último cambio en esa dirección, o sea, con $\lambda_j^k - \lambda_j^{k-1}$. Si el η_0 de la primera iteración es pequeño, en cada iteración se encuentra un punto mejor, pero muy cercano al anterior. Esto se puede evitar en parte, aumentando en cada iteración el tamaño del η_0, por ejemplo tomando $\eta_0 = 2(\lambda_j^k - \lambda_j^{k-1})$.

Ejemplo 8.18. Utilizar el método cíclico coordenado discreto para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$.

k	x_1^k	x_2^k	$f(x^k)$	j	λ_j	y_1^{j+1}	y_2^{j+1}	$f(y^{j+1})$
0	3.0000	4.0000	254.000	1	-1.0000	2.0000	4.0000	1.000
				2	0.0000	2.0000	4.0000	1.000
1	2.0000	4.0000	1.000	1	-0.0078	1.9922	4.0000	0.994
				2	-0.0312	1.9922	3.9688	0.984
2	1.9922	3.9688	0.984	1	-0.0078	1.9844	3.9688	0.979
				2	-0.0312	1.9844	3.9375	0.969
3	1.9844	3.9375	0.969	1	-0.0078	1.9766	3.9375	0.963
				2	-0.0312	1.9766	3.9062	0.954
	...							

8.11 MÉTODO DE HOOKE Y JEEVES CONTINUO O EXACTO: HJC

Este es un método de descenso, que no utiliza derivadas sino simplemente los valores de la función. Es una mejora del método cíclico coordenado continuo. Dado un x^0 se hace una iteración completa (n subiteraciones) del método cíclico coordenado para obtener x^1. A partir de x^1 se hace una minimización en la dirección $d^1 = x^1 - x^0$ y, en seguida, otra vez el método cíclico para obtener x^2. A partir de x^2 se hace una minimización en la dirección $d^2 = x^2 - x^1$ y de nuevo el método cíclico para obtener x^3, y así sucesivamente. La dirección $d^k = x^k - x^{k-1}$ representa el resumen de la iteración anterior y muestra hacia donde tiende a mejorar la función f.

datos: x^0, ε_x, `MAXIT`, ...
a partir de x^0 obtener x^1 (mét. CCC)
para $k = 1, ..., \texttt{MAXIT}$
  $d^k = x^k - x^{k-1}$
  si $||d^k|| < \varepsilon_x$ **ent parar**
  $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k)$
  $y^1 = x^k + \lambda_k^* d^k$
  a partir de y^1 obtener y^{n+1} (mét. CCC)
  $x^{k+1} = y^{n+1}$
fin-para k

Ejemplo 8.19. Utilizar el método de Hooke y Jeeves continuo para minimizar $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ partiendo de $x^0 = (3, 4)$.

Ver la tabla de resultados al final del capítulo.

También existe un método de Hooke y Jeeves discreto o inexacto. En cada iteración, en lugar de hallar λ_0^* minimizador de $f(x^k + \lambda d^k)$, se obtiene simplemente un punto $y^1 = x^k + \lambda_0 d^k$ mejor que x^k. Además, a partir de y^1 se utiliza el método cíclico coordenado discreto para obtener y^{n+1}.

EJERCICIOS

En los ejercicios 8.1 a 8.15 estudie el problema propuesto, averigüe cuáles métodos puede usar, use algunos de ellos, compare los resultados y la rapidez, verifique si el punto obtenido es punto crítico, minimizador, minimizador global. Use también puntos iniciales diferentes al propuesto.

8.1 Minimizar $f(x_1, x_2) = x_1^4 + x_2^2$, con $x^0 = (3, 4)$.

8.2 Minimizar $f(x_1, x_2) = (x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$, con $x^0 = (1, 2)$.

8.3 Minimizar $f(x_1, x_2) = 10x_1^2 + x_2^2$, con $x^0 = (2, 3)$.

8.4 Minimizar $f(x_1, x_2) = (x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$, con $x^0 = (3, 5)$.

8.5 Minimizar $f(x_1, x_2) = 10(x_1 + x_2 - 3)^2 + (x_1 - x_2 + 1)^2$, con $x^0 = (3, 5)$.

8.6 Minimizar $f(x_1, x_2, x_3) = 10(x_1 + x_2 + x_3)^2 + 2x_2^2 + x_3^2$, con $x^0 = (3, 4, 5)$.

8.7 Minimizar $f(x_1, x_2) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$, con $x^0 = (3, 4)$.

8.8 Minimizar $f(x_1, x_2) = (x_1 - x_2)^4 + (1 - x_1)^2$, con $x^0 = (3, 2)$.

8.9 Minimizar $f(x_1, x_2) = x_1^4 + 2x_1x_2 + 3x_2^2$, con $x^0 = (4, 5)$.

8.10 Minimizar $f(x_1, x_2) = x_1^3 + 2x_1x_2 + 3x_2^2$, con $x^0 = (4, 5)$.

8.11 Minimizar $f(x_1, x_2) = \operatorname{sen}(x_1x_2)$, con $x^0 = (0, 0)$.

8.12 Minimizar $f(x_1, x_2) = \operatorname{sen}(x_1x_2)$, con $x^0 = (1.5, -1.5)$.

8.13 Minimizar $f(x_1, x_2) = (x_1 - 1)^2 + e^{x_1} + x_2^2 - 2x_2 + 1$, con $x^0 = (2, 3)$.

8.14 Minimizar $f(x_1, x_2) = (x_1 + x_2)/(x_1^2 + x_2^2 + 1)$, con $x^0 = (2, -2)$.

8.15 Minimizar $f(x_1, x_2) = (x_1^2 - x_2)^2$, con $x^0 = (3, 4)$.

K	k	x	$f(x)$	$f'(x)$	$D_{\cdot 1}$	$D_{\cdot 2}$	$D_{\cdot 3}$	d	λ^*
1	1	3.0000	1497.00	240.0000	1.0000	0.0000	0.0000	-240.00	0.0161
		4.0000		256.0000	0.0000	1.0000	0.0000	-256.00	
		5.0000		250.0000	0.0000	0.0000	1.0000	-250.00	
	2	-0.8683	0.9795	-0.4743	0.6935	-0.3276	-0.3167	0.4730	0.4097
		-0.1262		-0.9820	-0.3276	0.6499	-0.3385	0.9773	
		0.9706		1.4638	-0.3167	-0.3385	0.6727	-1.4682	
	3	-0.6745	0.2963	-0.6238	0.7152	-0.2155	-0.4536	0.6000	1.1241
		0.2743		0.4734	-0.2155	0.2360	-0.0189	-0.2440	
		0.3690		0.1141	-0.4536	-0.0189	0.4746	-0.3282	
2	1	0.0000	0.0000	0.0000					
		0.0000		0.0000					
		0.0000		0.0000					

Ejemplo 8.12

K	k	x	$f(x)$	$f'(x)$	$B_{\cdot 1}$	$B_{\cdot 2}$	$B_{\cdot 3}$	d	λ^*
1	1	3.0000	1497.00	240.00	1.0000	0.0000	0.0000	-240.00	0.0161
		4.0000		256.00	0.0000	1.0000	0.0000	-256.00	
		5.0000		250.00	0.0000	0.0000	1.0000	-250.00	
	2	-0.8683	0.9795	-0.4743	20.0174	20.3233	19.6522	0.4730	0.4097
		-0.1262		-0.9820	20.3233	22.7188	21.0016	0.9773	
		0.9706		1.4638	19.6522	21.0016	21.3082	-1.4682	
	3	-0.6745	0.2963	-0.6238	20.0815	19.9381	19.9851	0.6000	1.1241
		0.2743		0.4734	19.9381	24.0469	20.0113	-0.2440	
		0.3690		0.1141	19.9851	20.0113	22.0027	-0.3282	
2	1	0.0000	0.0000	0.0000					
		0.0000		0.0000					
		0.0000		0.0000					

Ejemplo 8.13

K	k	x	$f(x)$	$f'(x)$	$D_{\cdot 1}$	$D_{\cdot 2}$	$D_{\cdot 3}$	d	λ^*
1	1	3.0000	1497.0000	240.0000	1.0000	0.0000	0.0000	-240.0000	0.0161
		4.0000		256.0000	0.0000	1.0000	0.0000	-256.0000	
		5.0000		250.0000	0.0000	0.0000	1.0000	-250.0000	
	2	-0.8683	0.9795	-0.4743	0.6935	-0.3276	-0.3167	0.4730	0.4097
		-0.1262		-0.9820	-0.3276	0.6500	-0.3385	0.9774	
		0.9706		1.4638	-0.3167	-0.3385	0.6727	-1.4683	
	3	-0.6745	0.2963	-0.6238	0.8435	-0.2677	-0.5238	0.7126	0.9464
		0.2743		0.4734	-0.2677	0.2572	-0.0097	-0.2898	
		0.3690		0.1141	-0.5238	-0.0097	0.5130	-0.3899	
2	1	0.0000	0.0000	0.0000					
		0.0000		0.0000					
		0.0000		0.0000					

Ejemplo 8.15

k	x_1^k	x_2^k	$f(x^k)$	d_1^k	d_2^k	j	λ_j^*	y_1^{j+1}	y_2^{j+1}	$f(y^{j+1})$
0	3.0000	4.0000	254.0000							
						1	-1.0062	1.9938	4.0000	0.9938
						2	-0.0249	1.9938	3.9751	0.9876
1	1.9938	3.9751	0.9876	-1.0062	-0.0249		0.0063	1.9874	3.9749	0.9813
						1	0.0000	1.9875	3.9749	0.9813
						2	-0.0248	1.9875	3.9501	0.9751
2	1.9875	3.9501	0.9751	-0.0063	-0.0250		50.8015	1.6685	2.6801	0.5545
						1	-0.0373	1.6312	2.6801	0.4021
						2	-0.0193	1.6312	2.6607	0.3984
3	1.6312	2.6607	0.3984	-0.3563	-1.2893		0.4138	1.4837	2.1272	0.2892
						1	-0.0306	1.4531	2.1272	0.2078
						2	-0.0156	1.4531	2.1116	0.2053
	⋮									
8	1.0043	1.0087	0.0000	-0.0361	-0.0738		0.1183	1.0001	1.0000	0.0000
						1	-0.0001	1.0000	1.0000	0.0000
						2	0.0000	1.0000	1.0000	0.0000

Ejemplo 8.19

Capítulo 9

MÉTODOS DE PENALIZACIÓN Y DE BARRERA

En estos métodos se trata de resolver problemas de minimización con restricciones mediante métodos de minimización para problemas sin restricciones. Para esto se construye una función F que tenga en cuenta tanto la función f como las restricciones. En el método de penalización se parte de puntos no admisibles y, poco a poco, se obtienen puntos cercanos a la frontera y si es del caso puntos admisibles interiores. En el método de barrera se inicia con un punto interior y si es necesario se construyen puntos interiores cercanos a la frontera, pero nunca en la frontera.

9.1 MÉTODO DE PENALIZACIÓN

Consideremos el PMDI, un problema de minimización con desigualdades e igualdades:

$$\begin{aligned} &\min f(x) \\ &\quad g_i(x) \leq 0,\ i = 1, ..., m \\ &\quad h_j(x) = 0,\ j = 1, ..., l. \end{aligned} \tag{PMDI}$$

Para construir la función F se necesita usar una función P que indique si un punto es admisible, más exactamente, $P(x) = 0$ si x es admisible, y $P(x) > 0$ si x no es admisible, y obviamente $P(x)$ debe aumentar en la

medida en que x incumple cada vez más las restricciones. Una manera de construir P es la siguiente:

$$P(x) = \sum_{i=1}^{m} \Phi_i(g_i(x)) + \sum_{j=1}^{l} \Psi_j(h_j(x)) \tag{9.1}$$

donde $\Phi_i, \Psi_j : \mathbb{R} \longrightarrow \mathbb{R}$ son continuas y de preferencia derivables una o dos veces. Las funciones Φ_i pueden ser iguales o diferentes, de igual forma las funciones Ψ_j pueden ser iguales o diferentes.
Las funciones Φ deben tener por lo menos las siguientes características:

- Φ es continua en $\mathbb{R}$,
- $\Phi(t) > 0$ si $t > 0$,
- $\Phi(t) = 0$ si $t \leq 0$,
- $\Phi(t)$ es estrictamente creciente si $t \geq 0$.

Ejemplos de funciones Φ son:

$$\begin{aligned} \Phi(t) &= \max\{0, t\}, \\ \Phi(t) &= (\max\{0, t\})^p \quad , \quad p \geq 1. \end{aligned}$$

Obviamente estas características se conservan si se multiplica por una constante positiva.

Las funciones Ψ deben tener por lo menos las siguientes características:

- Ψ es continua en $\mathbb{R}$,
- $\Psi(t) > 0$ si $t \neq 0$,
- $\Psi(t) = 0$ si $t = 0$,
- $\Psi(t)$ es estrictamente creciente si $t \geq 0$,
- $\Psi(t)$ es estrictamente decreciente si $t \leq 0$.

Ejemplos de funciones Ψ son:

$$\Psi(t) = |t|,$$

$$\begin{aligned} \Psi(t) &= t^2, \\ \Psi(t) &= |t|^p,\ p \geq 1. \end{aligned}$$

Obviamente estas características también se conservan si se multiplica por una constante positiva.

Es claro que la función P definida según 9.1, usando funciones Φ_i, Ψ_j adecuadas, puede "medir" la inadmisibilidad de un punto: si $P(x) = 0$, entonces x es admisible, y a mayor valor de $P(x)$ mayor inadmisibilidad.

La función F que va a ser minimizada, y que tiene en cuenta f y las restricciones es

$$F(x) = f(x) + \mu P(x),\ \mu > 0.$$

Si el valor de μ es grande, al minimizar F va a ser muy importante la inadmisibilidad. Si μ es cercano a cero al minimizar F va a preponderar f y es posible que el punto que se obtenga diste mucho de ser admisible. Para tratar de evitar estos inconvenientes, se empieza con un valor de μ pequeño y se obtiene un minimizador. Si éste es admisible o casi admisible se detiene el proceso. De lo contrario se aumenta el valor de μ y se vuelve a empezar. Más exactamente:

datos: $x^0,\ \varepsilon,\ \mu_0 > 0,\ \beta > 1,$ `MAXIT`
para $k = 0, ..., $`MAXIT`
    partiendo de x^k encontrar
        x^{k+1} minimizador de $F(x) = f(x) + \mu_k P(x)$
    si $\mu_k P(x^{k+1}) < \varepsilon$ **ent parar**
    sino $\mu_{k+1} = \beta \mu_k$
fin-para k

Ejemplo 9.1. Usar el método de penalización para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 - x_2^2 & \\ -x_1 + 2 &\leq 0 \\ -x_2 + 3 &\leq 0 \\ x_1 + x_2 - 10 &= 0, \end{aligned}$$

con $x^0 = (0, 0),\ \mu_0 = 0.1,\ \beta = 10$, $\Phi(t) = (\max\{0, t\})^2$, $\Psi(t) = t^2$.

$$\begin{aligned} F(x_1, x_2) &= x_1^2\text{-}x_2^2 + \mu\Big[(\max\{0, \text{-}x_1 + 2\})^2 + (\max\{0, \text{-}x_2 + 3\})^2 \\ &\quad +(x_1 + x_2\text{-}10)^2\Big] \end{aligned}$$

k	μ_k	x_1^{k+1}	x_2^{k+1}	$f(x^{k+1})$	$P(x^{k+1})$	$\mu_k P(x^{k+1})$
0	0.1	no hay	minimizador			
1	1	no hay	minimizador			
2	10	0.8989	10.1124	-101.4518	2.2351	22.3507
3	100	1.8992	8.1826	-63.3488	0.0169	1.6860
4	1000	1.9900	8.0180	-60.3287	0.0002	0.1644
5	10000	1.9990	8.0018	-60.0328	1.64E-6	0.0164
6	100000	1.9999	8.0002	-60.0033	1.64E-8	0.0016

$\diamond$

La principal ventaja del método de penalización es que fácilmente se puede adaptar un programa de computador de minimización sin restricciones para un PMDI. Además, se tiene convergencia global. Sin embargo, computacionalmente presenta dificultades y no es muy eficiente.

Proposición 9.1. *Sea $\{x^k\}$ una sucesión de vectores generada por el método de penalización. Entonces cualquier punto límite de la sucesión es solución del PMDI.*

9.2 MÉTODO DE BARRERA

El método de barrera se aplica a problemas sin igualdades:

$$\begin{aligned} \min\, & f(x) \\ & g_i(x) \;\leq\; 0,\ i = 1, ..., m \end{aligned}$$

o simplemente

$$\begin{aligned} \min\, & f(x) \\ & g(x) \;\leq\; 0, \end{aligned}$$

donde $g : \mathbb{R}^n \to \mathbb{R}^m$, $g(x) = (g_1(x), g_2(x), ..., g_m(x))$, las funciones f, g_i son continuas y, además, el conjunto admisible $\mathcal{A}$ tiene interior no vacío

$$\overset{\circ}{\mathcal{A}} = \{x : g(x) < 0\} \neq \emptyset.$$

El método de barrera trabaja en puntos de $\overset{\circ}{\mathcal{A}}$ y "castiga" a los puntos que se acercan a la frontera, es decir, a los puntos que se acercan a saturar

una desigualdad. Por esta razón hace parte de los métodos de punto interior. Actualmente, los métodos de punto interior son un tema importante de investigación. En particular, son muy útiles para problemas de programación lineal muy grandes.

Para trabajar con puntos interiores se define una función de barrera B que aumenta su valor cuando los puntos están cerca de la frontera

$$B(x) = \sum_{i=1}^{m} \chi_i(g_i(x)).$$

Las funciones χ deben tener por lo menos las siguientes características:

- χ está definida, es continua y estrictamente creciente en $]-\infty, 0[$
- $\chi(t) \longrightarrow +\infty$ cuando $t \longrightarrow 0^-$.

Ejemplos de funciones χ son:

$$\begin{aligned} \chi(t) &= -\frac{1}{t}, \\ \chi(t) &= -\log(-t). \end{aligned}$$

En el método de barrera se minimiza una función F que tiene en cuenta al mismo tiempo la función objetivo f y la función de barrera B:

$$F(x) = f(x) + \mu B(x).$$

El esquema del algoritmo es semejante al de penalización, con algunas diferencias importantes. Es necesario iniciar con un punto admisible, más exactamente $g(x^0) < 0$. El valor de μ disminuye en cada iteración.

datos: $x^0 \in \overset{\circ}{\mathcal{A}},\ \varepsilon,\ \mu_0 > 0,\ \beta \in]0,1[,$ `MAXIT`
para $k = 0, ..., $`MAXIT`
    partiendo de x^k encontrar
        x^{k+1} minimizador de $F(x) = f(x) + \mu_k B(x)$
    si $\mu_k B(x^{k+1}) < \varepsilon$ **ent parar**
    sino $\mu_{k+1} = \beta \mu_k$
fin-para k

Un detalle muy importante para tener en cuenta es que cuando x no es admisible, la función B puede tomar valores más pequeños que para puntos

admisibles, o peor aún, puede no estar definida. Entonces un programa de computador para implementar este método no debe permitir evaluar la función F en puntos no admisibles o de la frontera.

Ejemplo 9.2. Usar el método de barrera para el siguiente problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 - x_2^2 & \\ -x_1 + 2 & \leq 0 \\ -x_2 + 3 & \leq 0 \\ x_1 + x_2 - 10 & \leq 0, \end{aligned}$$

con $x^0 = (3.5, 3.5)$, $\mu_0 = 10$, $\beta = 0.1$, $\chi(t) = -1/t$.

$$F(x_1, x_2) = x_1^2 - x_2^2 + \mu \left[\frac{-1}{-x_1 + 2} + \frac{-1}{-x_2 + 3} + \frac{-1}{x_1 + x_2 - 10} \right].$$

k	μ_k	x_1^{k+1}	x_2^{k+1}	$f(x^{k+1})$	$B(x^{k+1})$	$\mu_k B(x^{k+1})$
0	10	2.7227	6.4229	-33.8402	2.8462	28.4623
1	1	2.2263	7.5162	-51.5378	8.5248	8.5248
2	0.1	2.0710	7.8492	-57.3211	26.8237	2.6824
3	0.01	2.0224	7.9525	-59.1528	84.7495	0.8475
4	0.001	2.0071	7.9850	-59.7320	267.8918	0.2679
5	0.0001	2.0022	7.9953	-59.9153	847.0252	0.0847
6	0.00001	2.0007	7.9985	-59.9731	2676.4114	0.0268
7	0.000001	2.0002	7.9995	-59.9915	8430.7449	0.0084

$\diamond$

EJERCICIOS

En los ejercicios 9.1 a 9.5 estudie el problema propuesto, averigüe cuáles métodos puede usar, use algunos de ellos. Obtenga un punto inicial.

9.1 Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$ sujeto a $x_1 + x_2 \geq 4$, $2x_1 + x_2 \geq 5$, $x \geq 0$.

9.2 Minimizar $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$ sujeto a $x_1^2 - x_2 + 2 \leq 0$, $x_1 \geq 0$, $-x_1 + 2 \leq 0$.

9.3 Minimizar $f(x_1, x_2) = (x_1 + 1)^2 + (x_2 + 1)^2$ sujeto a $x_1 + x_2 \leq 1$, $-x_1 - x_2 \leq -1$, $x \geq 0$.

9.4 Minimizar $f(x_1, x_2) = 3x_1 + 4x_2$ sujeto a $x_1 + 2x_2 \geq 2$, $x \geq 0$.

9.5 Minimizar $f(x_1, x_2) = x_1^2 + 2x_2^2 + x_1x_2 - 5x_1 - 6x_2$ sujeto a $x_1 + 2x_2 \leq 4$, $3x_1 + 2x_2 \leq 8$, $x \geq 0$.

Capítulo 10

MÉTODOS DE MINIMIZACIÓN CON RESTRICCIONES

Una clase de métodos para problemas de PNL es la de los métodos principales o "primales". En un método primal se trata de resolver el problema directamente en el conjunto admisible. Se empieza con un punto admisible, se construye una dirección admisible de descenso, se minimiza a lo largo de ella y se obtiene otro punto admisible mejor que el anterior. Si el proceso se detiene antes de llegar a la solución, de todas formas se tiene un punto factible, posiblemente no muy alejado del minimizador. Generalmente, cuando los métodos primales generan una sucesión convergente, su límite es un minimizador local. Con frecuencia se pueden aplicar a problemas sin una estructura especial, por ejemplo, sin que el conjunto admisible necesite ser convexo. Para poder empezar estos métodos hay que resolver otro problema: encontrar un punto admisible.

En los métodos duales las variables principales del problema son los coeficientes de KKT, o sea, los métodos duales no tratan de resolver directamente el problema inicial, sino el problema obtenido al plantear condiciones de KKT.

Otra clasificación de los métodos se refiere al tipo de restricciones, algunos se aplican a restricciones lineales, otros a restricciones no necesariamente lineales. Finalmente, algunos métodos de restricciones lineales se pueden adaptar para restricciones no lineales.

La estructura de los métodos de direcciones admisible es muy semejante a la de los métodos para problemas sin restricciones:

- tener un x^k admisible,
- construir una dirección d^k admisible y de descenso,
- hallar $\lambda_{\max} = \max\{\lambda : x^k + \lambda d^k \text{ es admisible }\}$,
- $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k), \lambda \in [0, \lambda_{\max}]$,
- $x^{k+1} = x^k + \lambda_k^* d^k$.

Obviamente, si $\lambda_{\max} = +\infty$ y además $f(x^k + \lambda d^k)$ no está acotado inferiormente cuando $\lambda > 0$, entonces, en el problema general, $f(x)$ tampoco está acotada inferiormente y no hay minimizador global.

10.1 MÉTODO DEL GRADIENTE REDUCIDO DE WOLFE

Antes de estudiar la deducción del método del gradiente reducido, consideremos un problema más sencillo en $\mathbb{R}^p$. Sea $\theta : \mathbb{R}^p \to \mathbb{R}$. El problema es:

$$\begin{aligned} \min\ & \theta(\xi) \\ & \xi \geq 0, \end{aligned}$$

equivalente a

$$\begin{aligned} \min\ & \theta(\xi) \\ & -\xi_i \leq 0,\ \ i = 1, ..., p. \end{aligned}$$

Sean: $\bar{\xi}$ un punto admisible, $I = \{i : \bar{\xi}_i = 0\}$. Al plantear condiciones de KKT se tiene:

$$\theta'(\bar{\xi}) + \sum_{i \in I} u_i(-e^i) = 0.$$

Tomando la anterior igualdad para cada una de las p componentes

$$\theta_i'(\bar{\xi}) = \frac{\partial \theta}{\partial \xi_i}(\bar{\xi}) \quad = \quad u_i \geq 0 \quad \text{para } i \in I$$

$$\theta_i'(\bar{\xi}) = \frac{\partial \theta}{\partial \xi_i}(\bar{\xi}) \quad = \quad 0 \qquad\qquad \text{para } i \notin I,$$

es decir, las condiciones de KKT para el punto $\bar{\xi}$ son:

$$\begin{aligned} \theta_i'(\bar{\xi}) &\geq 0 \qquad \text{si } \ \bar{\xi}_i = 0, \\ \theta_i'(\bar{\xi}) &= 0 \qquad \text{si } \ \bar{\xi}_i > 0. \end{aligned}$$

El método del descenso más pendiente se puede adaptar a este problema restringido no permitiendo que una componente de la dirección sea negativa cuando ξ_i sea nulo. Sea ξ^k un punto admisible; la dirección se construye así:

$$d_i^k = \begin{cases} 0 & \text{si } \ \theta_i'(\xi^k) > 0 \quad \text{y} \quad \xi_i^k = 0, \\ -\theta_i'(\xi^k) & \text{en los demás casos .} \end{cases} \tag{10.1}$$

Se puede verificar que si $d^k = 0$, entonces ξ^k es un punto de KKT (cumple condiciones necesarias de KKT). Si $d^k \neq 0$ la dirección es al mismo tiempo dirección admisible y de descenso.

El máximo valor positivo (puesto que d es dirección de descenso) que puede tomar λ se deduce sabiendo que $\xi_i^k + \lambda d_i^k \geq 0$ para todo i. La anterior desigualdad es importante cuando $d_i^k < 0$, caso en el cual λ puede aumentar hasta cuando $\xi_i^k + \lambda d_i^k = 0$, entonces λ no puede ser mayor de $-\xi_i^k / d_i^k$

$$\lambda_{\max} = \begin{cases} \infty & \text{si } d^k \geq 0, \\ \min\{-\dfrac{\xi_i^k}{d_i^k} : d_i^k < 0\} & \text{si } d^k \ngeq 0. \end{cases} \tag{10.2}$$

Cuando d^k se construye según las fórmulas anteriores y no es nula, se obtiene $\lambda_{\max} > 0$. Teniendo la dirección y el máximo valor de λ se construye el punto ξ^{k+1} de la manera usual:

$$\begin{aligned} \lambda_k^* &= \operatorname{argmin} \theta(\xi^k + \lambda d^k), \ \lambda \in [0, \lambda_{\max}], \\ \xi^{k+1} &= \xi^k + \lambda_k^* d^k. \end{aligned}$$

El método del gradiente reducido se aplica a un problema con restricciones lineales con la siguiente forma:

$$\min f(x) \tag{10.3}$$

$$\begin{aligned} Ax &= b \\ x &\geq 0, \end{aligned}$$

donde A es una matriz de tamaño $m \times n$ con $m \leq n$ y además rango$(A) = m$. Sea x un punto admisible. El método GR tiene varias semejanzas con el método simplex, hay m variables básicas, $n - m$ variables libres o independientes o no básicas. Sin perder generalidad, supongamos que las primeras m variables son las básicas y las variables libres son las $n-m$ que siguen (esto correspondería simplemente a un reordenamiento de las variables). Supongamos, además, que B submatriz de A, obtenida al tomar las primeras m columnas de A, es invertible. Sea L la matriz conformada con las $n - m$ columnas restantes. El sistema $Ax = b$ se puede expresar

$$Bx_B + Lx_L = b.$$

Para facilitar las fórmulas, sean:

$$\begin{aligned} x_B = \zeta &= [\zeta_1 \;\; \zeta_2 \;\; ... \;\; \zeta_m]^{\mathrm{T}} &&= [x_1 \;\; x_2 \;\; ... \;\; x_m]^{\mathrm{T}}, \\ x_L = \xi &= [\xi_1 \;\; \xi_2 \;\; ... \;\; \xi_{n-m}]^{\mathrm{T}} &&= [x_{m+1} \;\; x_{m+2} \;\; ... \;\; x_n]^{\mathrm{T}}. \end{aligned}$$

$$\begin{aligned} B\zeta + L\xi &= b \\ \zeta &= -B^{-1}L\xi + B^{-1}b \\ \zeta &= Q\xi + q. \end{aligned}$$

Supóngase, además, que $\zeta > 0$. El problema planteado se puede expresar

$$\begin{aligned} \min \theta(\xi) &= f(\zeta, \xi) \\ \xi &\geq 0. \end{aligned}$$

Sea ρ el gradiente de θ función de las variables ξ_1, ..., ξ_{n-m}.

$$\begin{aligned} \rho_j = \frac{\partial \theta}{\partial \xi_j} &= \frac{\partial f}{\partial \xi_j} + \sum_{i=1}^{m} \frac{\partial f}{\partial \zeta_i} \frac{\partial \zeta_i}{\partial \xi_j}, \quad j = 1, ..., n - m \\ &= \frac{\partial f}{\partial x_{m+j}} + \sum_{i=1}^{m} \frac{\partial f}{\partial x_i} q_{ij} \\ &= f'_{m+j} + \sum_{i=1}^{m} f'_i q_{ij} \end{aligned}$$

$$\begin{aligned} &= f'_{m+j} + [q_{1j} \ \ldots \ q_{mj}] \begin{bmatrix} f'_1 \\ \vdots \\ f'_m \end{bmatrix} \\ &= f'_{m+j} + (Q^{\mathrm{T}})_j . f'_B. \end{aligned}$$

Al considerar todas las $n - m$ componentes

$$\begin{aligned} \rho &= f'_L + Q^{\mathrm{T}} f'_B \\ \rho &= f'_L - L^{\mathrm{T}} B^{-1^{\mathrm{T}}} f'_B, \end{aligned} \tag{10.4}$$

donde f'_L es el vector columna formado por las componentes del gradiente $f'(x)$ correspondientes a las variables libres, y f'_B es el vector columna formado por las componentes del gradiente $f'(x)$ correspondientes a las variables básicas.

Entonces la dirección $d_L = \delta$ para las $n - m$ variables libres se define como en (10.1)

$$d_{m+i} = \delta_i = \begin{cases} 0 & \text{si } \rho_i > 0 \quad \text{y} \quad \xi_i = 0, \\ -\rho_i & \text{en los demás casos.} \end{cases} \tag{10.5}$$

El resto de coordenadas de d se obtienen teniendo en cuenta que para que d sea admisible se requiere que $Ad = 0$, entonces $Bd_B + Ld_L = 0$, luego

$$d_B = -B^{-1} L d_L.$$

Lo visto hasta ahora corresponde a la deducción simple del método del gradiente reducido. Usualmente la fórmula (10.5) se modifica para poder garantizar convergencia global [Lue89].

$$d_{m+i} = \delta_i = \begin{cases} -\rho_i & \text{si} \quad \rho_i \leq 0, \\ -\rho_i \xi_i & \text{si} \quad \rho_i > 0. \end{cases} \tag{10.6}$$

Dado un punto x^k admisible y una dirección d^k admisible ($Ad^k = 0$) el valor $\lambda_{\max}$ se calcula como en la fórmula (10.2)

$$\lambda_{\max} = \begin{cases} \infty & \text{si } d^k \geq 0, \\ \min\{-\dfrac{x_i^k}{d_i^k} : d_i^k < 0\} & \text{si } d^k \ngeq 0. \end{cases} \tag{10.7}$$

La fórmula (10.4) para el cálculo del gradiente ρ para las variables libres se puede reescribir utilizando r **gradiente reducido**, definido para las n variables, pero nulo para las variables básicas.

$$\begin{aligned} r &= f'(x^k) - A^{\mathrm{T}}(B^{-1})^{\mathrm{T}} f'_B(x^k), \\ r_B &= 0, \\ r_L &= f'_L(x^k) - L^{\mathrm{T}}(B^{-1})^{\mathrm{T}} f'_B(x^k). \end{aligned}$$

Entonces el cálculo de la dirección para las variables libres se convierte en

$$d_j = \begin{cases} -r_j & \text{si } r_j \leq 0 \quad \text{y} \quad x_j \text{ es libre,} \\ -x_j r_j & \text{si } r_j > 0 \quad \text{y} \quad x_j \text{ es libre.} \end{cases} \tag{10.8}$$

Las variables básicas se escogen como las m mayores componentes de x^k. En caso de empate se decide arbitrariamente.

$$\begin{aligned} \mathcal{I}_k &= \{\text{índices de las mayores } m \text{ componentes de } x^k\}, \\ B &= \begin{bmatrix} \cdots & A_{\cdot j} & \cdots \end{bmatrix} \quad j \in \mathcal{I}_k, \\ L &= \begin{bmatrix} \cdots & A_{\cdot j} & \cdots \end{bmatrix} \quad j \notin \mathcal{I}_k, \end{aligned}$$

es decir, B se obtiene tomando las columnas de A correspondientes a las variables básicas, o sea, las columnas cuyos índices están en $\mathcal{I}_k$. De manera semejante L se obtiene tomando las columnas de A correspondientes a las variables libres, o sea, las columnas cuyos índices no están en $\mathcal{I}_k$.

El esquema del algoritmo es:

datos: x^0 factible, ε, `MAXIT`
para $k = 0, ..., \texttt{MAXIT}$
  $\mathcal{I}_k = \{\text{índices de las } m \text{ más grandes componentes de } x^k\}$
  $B,\ L,\ ...,$ de acuerdo a $\mathcal{I}_k$
  $r_L = f'_L(x^k) - L^{\mathrm{T}}(B^{-1})^{\mathrm{T}} f'_B(x^k)$
  $d_L^k:$ según (10.8)
  $d_B^k = -B^{-1} L d_L^k$
  si $||d^k|| < \varepsilon$ **ent parar**
  $\lambda_{\max}:$ según (10.7)
  $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k),\ \lambda \in [0, \lambda_{\max}]$
  $x^{k+1} = x^k + \lambda_k^* d^k$
fin-para k

La utilización adecuada de este algoritmo hace que se cumplan las siguientes propiedades:

- $Ad^k = 0$,
- $f'(x^k)^{\mathrm{T}} d^k < 0$,
- $\lambda_{\max} > 0$,
- x^{k+1} es admisible,
- $f(x^{k+1}) < f(x^k)$.

La salida deseable se alcanza cuando $d^k \approx 0$, en este caso x^k es un punto de KKT. Las salidas no deseables son: B no es invertible, λ_k^* no existe ($f(x^k + \lambda d^k)$ no es acotado inferiormente), se llega hasta la iteración $k =$`MAXIT` sin obtener un punto de KKT.

Ejemplo 10.1. Utilizar el método GR para el problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 + x_2^2 & \\ x_1 + x_2 & \leq 2 \\ x_1 + 5x_2 & \leq 5 \\ x & \geq 0, \end{aligned}$$

partiendo de $x^0 = (3/2, 1/2)$.
Para poder resolver este problema por el método GR se necesita convertirlo a la forma 10.3 . Entonces introduciendo variables de holgura:

$$\begin{aligned} \min f(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 & \\ x_1 + x_2 + x_3 & = 2 \\ x_1 + 5x_2 + x_4 & = 5 \\ x & \geq 0. \end{aligned}$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad f'(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 0 \\ 0 \end{bmatrix}.$$

El punto admisible del problema original debe ser convertido en un punto de $\mathbb{R}^4$, admisible para el problema modificado

$$x^0 = \begin{bmatrix} 1.5 \\ 0.5 \\ 0 \\ 1 \end{bmatrix}, \quad f(x^0) = 2.5, \quad f'(x^0) = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_0 = \{1, 4\}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad d_B = \begin{bmatrix} d_1 \\ d_4 \end{bmatrix} = \begin{bmatrix} -5 \\ -5 \end{bmatrix}, \quad d = \begin{bmatrix} -5 \\ 2 \\ 3 \\ -5 \end{bmatrix},$$

$$\lambda_{\max} = 0.2, \quad \lambda_0^* = 0.2.$$

$$x^1 = \begin{bmatrix} 0.5 \\ 0.9 \\ 0.6 \\ 0 \end{bmatrix}, \quad f(x^1) = 1.06, \quad f'(x^1) = \begin{bmatrix} 1.0 \\ 1.8 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_1 = \{2, 3\}, \quad B = \begin{bmatrix} 1 & 1 \\ 5 & 0 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_1 \\ r_4 \end{bmatrix} = \begin{bmatrix} 0.64 \\ -0.36 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_1 \\ d_4 \end{bmatrix} = \begin{bmatrix} -0.32 \\ 0.36 \end{bmatrix}, \; d_B = \begin{bmatrix} d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} -0.008 \\ 0.328 \end{bmatrix}, \; d = \begin{bmatrix} -0.320 \\ -0.008 \\ 0.328 \\ 0.360 \end{bmatrix},$$

$$\lambda_{\max} = 1.5625, \quad \lambda_1^* = 1.5625.$$

$$x^2 = \begin{bmatrix} 0 \\ 0.8875 \\ 1.1125 \\ 0.5625 \end{bmatrix}, \quad f(x^2) = 0.787656, \quad f'(x^2) = \begin{bmatrix} 0 \\ 1.775 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_2 = \{2,3\}, \quad B = \begin{bmatrix} 1 & 1 \\ 5 & 0 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_1 \\ r_4 \end{bmatrix} = \begin{bmatrix} -0.355 \\ -0.355 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_1 \\ d_4 \end{bmatrix} = \begin{bmatrix} 0.355 \\ 0.355 \end{bmatrix}, \; d_B = \begin{bmatrix} d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} -0.142 \\ -0.213 \end{bmatrix}, \; d = \begin{bmatrix} 0.355 \\ -0.142 \\ -0.213 \\ 0.355 \end{bmatrix},$$

$$\lambda_{\max} = 5.223005, \quad \lambda_2^* = 0.862069.$$

$$x^3 = \begin{bmatrix} 0.306034 \\ 0.765086 \\ 0.928879 \\ 0.868534 \end{bmatrix}, \quad f(x^3) = 0.679013, \quad f'(x^3) = \begin{bmatrix} 0.612069 \\ 1.530172 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_3 = \{3,4\}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0.612069 \\ 1.530172 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} -0.187314 \\ -1.170714 \end{bmatrix}, \quad d_B = \begin{bmatrix} d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} 1.358028 \\ 6.040883 \end{bmatrix},$$
$$d = \begin{bmatrix} -0.187314 \\ -1.170714 \\ 1.358028 \\ 6.040883 \end{bmatrix},$$

$$\lambda_{\max} = 0.653521, \quad \lambda_3^* = 0.653521.$$

$$x^4 = \begin{bmatrix} 0.183621 \\ 0 \\ 1.816379 \\ 4.816379 \end{bmatrix}, \quad f(x^4) = 0.033717, \quad f'(x^4) = \begin{bmatrix} 0.367241 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_4 = \{3,4\}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0.367241 \\ 0 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} -0.067433 \\ 0 \end{bmatrix}, \quad d_B = \begin{bmatrix} d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} 0.067433 \\ 0.067433 \end{bmatrix},$$

$$d = \begin{bmatrix} -0.067433 \\ 0 \\ 0.067433 \\ 0.067433 \end{bmatrix},$$

$$\lambda_{\max} = 2.723005, \quad \lambda_4^* = 2.723005.$$

$$x^5 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 5 \end{bmatrix}, \quad f(x^5) = 0, \quad f'(x^5) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathcal{I}_5 = \{3,4\}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad r_L = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$d_L = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad d_B = \begin{bmatrix} d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Luego $x^5 = (0,0,2,5)$ es un punto de KKT. $\diamond$

En el método GR se puede buscar que siempre la matriz B sea la matriz identidad, lo cual simplifica algunas fórmulas. Para esto hay que hacer operaciones elementales sobre las filas de A. Antes de empezar la primera iteración hay que modificar A para que $B^0 = I$.

Si el punto inicial x^0 se obtuvo por programación lineal, seguramente la última tabla del simplex tiene esta propiedad. A partir de la primera iteración, cuando se construye un nuevo x^k se modifica la matriz A actual mediante pivoteos. Puede resultar que en la siguiente iteración el conjunto $\mathcal{I}$ no cambie, es decir, no se necesitaría hacer ningún pivoteo. Puede ser

necesario un pivoteo, o pueden ser necesarios varios pivoteos. Recordemos que en el método simplex, en cada iteración, hay exactamente un pivoteo, pues una y una sola variable deja de ser básica, y una variable libre se convierte en básica. Las fórmulas que se simplifican cuando $B = I$ son:

$$\begin{aligned} r_L &= f'_L(x^k) - L^{\mathrm{T}} f'_B(x^k), \\ d_B &= -L d_L. \end{aligned}$$

En palabras, estas fórmulas se pueden expresar así:

Si x_j es un variable libre, entonces r_j es igual a $f'_j(x^k)$ menos la suma de los productos, término a término, de la columna $A_{\cdot j}$ y el vector columna $f'_B(x^k)$.

Si x_j es la i-ésima variable básica, entonces d_j es igual a menos la suma de los productos, término a término, de la fila $L_{i\cdot}$ y el vector d_L.

Para problemas pequeños, resueltos a mano, se puede organizar una tabla un poco parecida a la del simplex, que facilite algunos cálculos. Supongamos, por facilidad en la presentación, que las variables básicas son las primeras m variables.

$$\begin{array}{c} x: \\ f'(x): \\ \\ x_B^k \\ \\ r: \\ d: \\ -x_j/d_j \end{array} \begin{bmatrix} & x_B^{\mathrm{T}} & x_L^{\mathrm{T}} & \\ & f'_B(x^k)^{\mathrm{T}} & f'_L(x^k)^{\mathrm{T}} & \\ f'_B & I & L & b \\ & & r_L & \\ & d_B & d_L & \\ & -\dfrac{x_i}{d_i} & -\dfrac{x_i}{d_i} & \end{bmatrix}$$

Para construir esta tabla, algunos valores simplemente se transcriben, otros provienen de cálculos sencillos dentro de la misma tabla. Para la construcción de la tabla el orden podría ser el siguiente:

1. Colocar en la primera fila los valores de x_j^k; colocar en la segunda fila los valores de $f'_j(x^k)$; colocar en las filas 3 hasta $m + 2$, la matriz A y el vector columna b.

2. Indicar en la primera columna de la izquierda cuáles son las variables básicas.

3. Colocar en la siguiente columna f'_B, es decir, los valores de las componentes del gradiente $f'(x^k)$ correspondientes a las variables básicas en el orden adecuado (el orden en que las columnas de B forman la identidad).

4. Calcular para las variables libres el valor r_j según la fórmula simplificada, o sea, restarle a $f'_j(x^k)$ (en la segunda fila) la suma de los productos de la columna $f'_B(x^k)$ y de la columna j de A.

5. Calcular para las variables libres el valor d_j según 10.8.

6. Calcular para las variables básicas el valor d_j según la fórmula simplificada, o sea, si x_j es la i-ésima variable básica, entonces d_j es menos la suma de los productos de la fila i-ésima de L (elementos de la fila i-ésima de A correspondientes a las variables libres) y del vector d_L recién calculado.

7. Efectuar los cocientes $-x_j^k/d_j^k$ cuando $d_j^k < 0$ y escoger el valor $\lambda_{\max}$ como el mínimo de estos cocientes.

Una vez calculada la tabla es necesario:

- Calcular el λ_k^*.
- Construir el punto x^{k+1}.
- Determinar las nuevas variables básicas.
- Efectuar los pivoteos necesarios sobre la matriz A y el vector columna b, para que de nuevo $B = I$.

Ejemplo 10.2. Utilizar el método GR para el problema:

$$\begin{aligned} \min f(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 & \\ x_1 + x_2 + x_3 &= 2 \\ x_1 + 5x_2 + x_4 &= 5 \\ x &\geq 0. \end{aligned}$$

con $x^0 = (1.5,\ 0.5,\ 0,\ 1)$.

$$\hat{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 2 \\ 1 & 5 & 0 & 1 & 5 \end{bmatrix}, \quad f'(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 0 \\ 0 \end{bmatrix}.$$

Como $\mathcal{I}_0 = \{1, 4\}$, entonces es necesario efectuar operaciones elementales sobre $\hat{A}$ para que $B = I$.

$$\hat{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 2 \\ 0 & 4 & -1 & 1 & 3 \end{bmatrix}.$$

Entonces se puede empezar a llenar la tabla:

$$\begin{array}{rc} x^0: & \left[\begin{array}{ccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ & 1 & 1 & 1 & 0 & 2 \\ & 0 & 4 & -1 & 1 & 3 \end{array}\right. \\ f'(x^0): & \\ & \\ & \\ r: & \\ d: & \\ -x_i^0/d_i^0 & \left.\begin{array}{ccccc} & & & & & \end{array}\right] \end{array}$$

Las variables básicas son x_1, x_4 y las componentes del gradiente para las variables básicas son 3 y 0 .

$$\begin{array}{rc} x^0: & \left[\begin{array}{ccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ 3 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 4 & -1 & 1 & 3 \end{array}\right. \\ f'(x^0): & \\ x_1 & \\ x_4 & \\ r: & \\ d: & \\ -x_i^0/d_i^0 & \left.\begin{array}{ccccc} & & & & & \end{array}\right] \end{array}$$

El valor r_2 es igual a 1 menos la suma de los productos de los elementos de las columnas $\begin{bmatrix} 3 & 0 \end{bmatrix}^{\mathrm{T}}$ y $\begin{bmatrix} 1 & 4 \end{bmatrix}^{\mathrm{T}}$, o sea, -2. El valor r_3 es igual a 0 menos la suma de los productos de los elementos de las columnas $\begin{bmatrix} 3 & 0 \end{bmatrix}^{\mathrm{T}}$ y $\begin{bmatrix} 1 & -1 \end{bmatrix}^{\mathrm{T}}$, o sea, -3.

$$\begin{array}{c} x^0: \\ f'(x^0): \\ x_1 \\ x_4 \\ r: \\ d: \\ -x_i^0/d_i^0 \end{array} \left[\begin{array}{cccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ 3 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 4 & -1 & 1 & 3 \\ & & -2 & -3 & & \\ & & & & & \\ & & & & & \end{array}\right]$$

Los valores d_j para las variables libres se calculan según 10.8

$$\begin{array}{c} x^0: \\ f'(x^0): \\ x_1 \\ x_4 \\ r: \\ d: \\ -x_i^0/d_i^0 \end{array} \left[\begin{array}{cccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ 3 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 4 & -1 & 1 & 3 \\ & & -2 & -3 & & \\ & & 2 & 3 & & \\ & & & & & \end{array}\right]$$

El valor de d_1, correspondiente a la primera variable básica, es igual a menos la suma de productos de los elementos de la primera fila de L, es decir, $\left[\begin{array}{cc} 1 & 1 \end{array}\right]$, y de los elementos de $d_L = \left[\begin{array}{cc} 2 & 3 \end{array}\right]^{\mathrm{T}}$, o sea, -5. El valor de d_4 correspondiente a la segunda variable básica es igual a menos la suma de productos de los elementos de la segunda fila de L, es decir, $\left[\begin{array}{cc} 4 & -1 \end{array}\right]$, y de los elementos de $d_L = \left[\begin{array}{cc} 2 & 3 \end{array}\right]^{\mathrm{T}}$, o sea, -5.

$$\begin{array}{c} x^0: \\ f'(x^0): \\ x_1 \\ x_4 \\ r: \\ d: \\ -x_i^0/d_i^0 \end{array} \left[\begin{array}{cccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ 3 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 4 & -1 & 1 & 3 \\ & & -2 & -3 & & \\ & -5 & 2 & 3 & -5 & \\ & & & & & \end{array}\right]$$

Para las componentes negativas de la dirección, se efectúa el cociente $-x_j^k/d_j^k$.

$$\begin{array}{r} x^0: \\ f'(x^0): \\ x_1 \\ x_4 \\ r: \\ d: \\ -x_i^0/d_i^0 \end{array} \left[\begin{array}{cccccc} & 1.5 & 0.5 & 0 & 1 & \\ & 3 & 1 & 0 & 0 & \\ 3 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 4 & -1 & 1 & 3 \\ & & -2 & -3 & & \\ & -5 & 2 & 3 & -5 & \\ & 0.3 & & & 0.2 & \end{array}\right]$$

Entonces $\lambda_{\max} = 0.2$ y al minimizar $f(x^0 + \lambda d^0)$ se obtiene $\lambda_0^* = 0.2$.

$$x^1 = (0.5, 0.9, 0.6, 0), \quad f(x^1) = 1.06, \quad \mathcal{I}_1 = \{2, 3\}.$$

Es necesario hacer dos pivoteos, sobre $a_{12} = 1$ y sobre $a_{23} = -1$.

$$\begin{array}{r} x^1: \\ f'(x^1): \\ x_2 \\ x_3 \\ r: \\ d: \\ -x_i^1/d_i^1 \end{array} \left[\begin{array}{cccccc} & 0.5 & 0.9 & 0.6 & 0 & \\ & 1 & 1.8 & 0 & 0 & \\ 1.8 & 0.2 & 1 & 0 & 0.2 & 1 \\ 0 & 0.8 & 0 & 1 & -0.2 & 1 \\ & 0.64 & & & -0.36 & \\ & -0.32 & -0.008 & 0.328 & 0.36 & \\ & 1.5625 & 112.5 & & & \end{array}\right]$$

Entonces $\lambda_{\max} = 1.5625$ y al minimizar $f(x^1 + \lambda d^1)$ se obtiene $\lambda_1^* = 1.5625$.

$$x^2 = (0, 0.8875, 1.1125, 0.5625), \quad f(x^2) = 0.7877, \quad \mathcal{I}_2 = \{2, 3\}.$$

Luego en este caso no es necesario pivotear.

$$\begin{array}{r} x^2: \\ f'(x^2): \\ x_2 \\ x_3 \\ r: \\ d: \\ -x_i^2/d_i^2 \end{array} \left[\begin{array}{cccccc} & 0 & 0.8875 & 1.1125 & 0.5625 & \\ & 0 & 1.7750 & 0 & 0 & \\ 1.775 & 0.2 & 1 & 0 & 0.2 & 1 \\ 0 & 0.8 & 0 & 1 & -0.2 & 1 \\ & -0.355 & & & -0.355 & \\ & 0.355 & -0.142 & -0.213 & 0.355 & \\ & & 6.25 & 5.223 & & \end{array}\right]$$

Entonces $\lambda_{\max} = 5.233$ y al minimizar $f(x^2 + \lambda d^2)$ se obtiene $\lambda_2^* = 0.8621$.

$$x^3 = (0.3060, 0.7651, 0.9289, 0.8685), \quad f(x^3) = 0.6790, \quad \mathcal{I}_3 = \{3, 4\}.$$

Luego en este caso es necesario pivotear sobre $a_{14} = 0.2$.

$$\begin{array}{c} x^3: \\ f'(x^3): \\ x_4 \\ x_3 \\ r: \\ d: \\ -x_i^3/d_i^3 \end{array} \left[\begin{array}{cccccc} & 0.3060 & 0.7651 & 0.9289 & 0.8685 & \\ & 0.6121 & 1.5302 & 0 & 0 & \\ 0 & 1 & 5 & 0 & 1 & 5 \\ 0 & 1 & 1 & 1 & 0 & 2 \\ & 0.6121 & 1.5302 & & & \\ & -0.1873 & -1.1707 & 1.3580 & 6.0409 & \\ & 1.6337 & 0.6535 & & & \end{array}\right]$$

Entonces $\lambda_{\max} = 0.6535$ y al minimizar $f(x^3 + \lambda d^3)$ se obtiene $\lambda_3^* = 0.6535$.

$$x^4 = (0.1836, 0, 1.8164, 4.8164), \;\; f(x^4) = 0.0337, \;\; \mathcal{I}_4 = \{3, 4\}.$$

Luego en este caso no es necesario pivotear.

$$\begin{array}{c} x^4: \\ f'(x^4): \\ x_4 \\ x_3 \\ r: \\ d: \\ -x_i^4/d_i^4 \end{array} \left[\begin{array}{cccccc} & 0.1836 & 0 & 1.8164 & 4.8164 & \\ & 0.3672 & 0 & 0 & 0 & \\ 0 & 1 & 5 & 0 & 1 & 5 \\ 0 & 1 & 1 & 1 & 0 & 2 \\ & 0.3672 & 0 & & & \\ & -0.0674 & 0 & 0.0674 & 0.0674 & \\ & 2.7230 & & & & \end{array}\right]$$

Entonces $\lambda_{\max} = 0.7230$ y al minimizar $f(x^4 + \lambda d^4)$ se obtiene $\lambda_4^* = 2.7230$.

$$x^5 = (0, 0, 2, 5), \;\; f(x^4) = 0, \;\; \mathcal{I}_4 = \{3, 4\}.$$

Luego en este caso no es necesario pivotear.

$$\begin{array}{c} x^5: \\ f'(x^5): \\ x_4 \\ x_3 \\ r: \\ d: \\ -x_i^5/d_i^5 \end{array} \left[\begin{array}{cccccc} & 0 & 0 & 2 & 5 & \\ & 0 & 0 & 0 & 0 & \\ 0 & 1 & 5 & 0 & 1 & 5 \\ 0 & 1 & 1 & 1 & 0 & 2 \\ & 0 & 0 & & & \\ & 0 & 0 & 0 & 0 & \\ & & & & & \end{array}\right]$$

Entonces $x^5 = (0, 0, 2, 5)$ es punto de KKT. $\diamond$

10.2 MÉTODO DEL GRADIENTE PROYECTADO DE ROSEN

Este es un método de direcciones admisibles, para problemas con restricciones lineales. Se aplica a un problema de la forma

$$\begin{aligned} \min & f(x) \\ Ax &\leq b \\ Ex &= e, \end{aligned}$$

donde A es un matriz $m \times n$, b es un vector columna $m \times 1$, E es una matriz $l \times n$, $l \leq n$, rango$(E) = l$ y e es un vector columna $l \times 1$. En un caso particular se puede tener que $m = 0$, es decir, no hay desigualdades, o también $l = 0$, o sea, no hay igualdades.

En cada iteración de este método, dado un x^k admisible, la dirección utilizada es la proyección de $-f'(x^k)$ sobre el espacio tangente de x^k.

Para proyectar un vector se utilizan matrices de proyección, caracterizadas de la siguiente forma. Sean: M una matriz $p \times n$, $0 \leq p \leq n$, rango$(M) = p$. Sea $\mathcal{T}$ el espacio nulo de M, o sea, el conjunto de vectores y tales que $My = 0$. Sea P la matriz $n \times n$ definida por

$$P = \begin{cases} I_n - M^{\mathrm{T}}(MM^{\mathrm{T}})^{-1}M & \text{si } \ p \geq 1, \\ I_n & \text{si } \ p = 0. \end{cases} \tag{10.9}$$

Proposición 10.1. *Bajo las anteriores condiciones, la matriz P se llama la* ***matriz de proyección ortogonal*** *sobre el espacio $\mathcal{T}$ y cumple las siguientes propiedades:*

- *P está bien definida, es decir, MM^T es invertible,*
- *$Px \in \mathcal{T}$ para todo $x \in \mathbb{R}^n$,*
- *Px es ortogonal a $x - Px$, o sea, $(x - Px)^T Px = 0$,*
- *P es simétrica,*
- *$PP = P$, luego $P(Px) = Px$ para todo $x \in \mathbb{R}^n$.*

Considerar el caso $p = 0$ indica que si M no tiene filas $\mathcal{T}$ es todo $\mathbb{R}^n$ y la matriz identidad "proyecta" cualquier vector sobre $\mathcal{T}$. Otro caso particular

se tiene cuando $p = n$ y se ve fácilmente que $\mathcal{T} = \{\mathbf{0}\} = \{(0, 0, ..., 0)\}$ y sin necesidad de efectuar operaciones arítméticas $P = 0$, o sea, la matriz nula (compuesta únicamente de ceros).

Dado x^k admisible, la proyección de $-f'(x^k)$ sobre el espacio tangente se efectúa de la siguiente forma. La matriz A se puede separar en dos submatrices: $A^<$ compuesta por las filas de A correspondientes a las desigualdades no saturadas, o sea, las filas tales que $A_i.x^k < b_i$ y otra submatriz $A^=$ compuesta por las filas de A correspondientes a las desigualdades saturadas, o sea, las filas tales que $A_i.x^k = b_i$. De manera semejante, el vector columna de términos independientes se puede separar en dos subvectores $b^<$ y $b^=$. Entonces, suponiendo una reordenación de las filas de A y de b, se puede escribir

$$A = \begin{bmatrix} A^< \\ A^= \end{bmatrix}, \quad b = \begin{bmatrix} b^< \\ b^= \end{bmatrix}.$$

Recuérdese que las filas de la matriz M, utilizada en la definición del espacio tangente en x^k, son los gradientes de las desigualdades activas y los gradientes de las igualdades. Entonces para el problema considerado en el método GP

$$\begin{aligned} M &= \begin{bmatrix} A^= \\ E \end{bmatrix}, k && (10.10) \\ P &: \text{definida según 10.9,} \\ d^k &= -Pf'(x^k). && (10.11) \end{aligned}$$

Proposición 10.2. *Si la dirección d^k se construye según las fórmulas anteriores y $d^k \neq 0$, entonces d^k es una dirección admisible y además ${d^k}^T f'(x^k) < 0$, luego d^k es también una dirección de descenso.*

Queda por estudiar el caso de $d^k = 0$. Cuando la dirección resultante es nula se puede calcular un vector con los coeficientes de KKT.

Proposición 10.3. *Los coeficientes de KKT están dados por:*

$$w = \begin{bmatrix} u \\ v \end{bmatrix} = -(MM^T)^{-1} M f'(x^k).$$

El vector columna w tiene tantas filas como tiene la matriz M. El vector columna u está formado por las componentes de w correspondientes a las desigualdades que hacen parte de M. A su vez, el vector columna v de

tamaño $l \times 1$ está formado por las componentes de w correspondientes a las igualdades que hacen parte de M. Al tener esto en cuenta, se deduce inmediatamente el siguiente resultado:

Proposición 10.4. *Sea $d^k = 0$. Si $\mathrm{nf}(u) = 0$ (número de filas de u), o si $\mathrm{nf}(u) > 0$ y $u \geq 0$, entonces x^k es un punto de KKT.*

Si $\mathrm{nf}(u) > 0$ y $u \ngeq 0$, entonces x^k no es un punto de KKT. En este caso se escoge un $u_i < 0$, generalmente el más negativo, y se suprime en la matriz M la fila correspondiente a este u_i negativo y se vuelve a calcular la matriz P, etc.

Si $d^k \neq 0$ se debe hallar $\lambda_{\max}$, para esto únicamente se consideran las desigualdades no activas ya que las desigualdades activas y las igualdades están consideradas en el cálculo de d^k. Obviamente, si no hay desigualdades inactivas, λ puede variar sin restricción. Si hay desigualdades inactivas, $x^k + \lambda d^k$ debe seguir cumpliendo estas desigualdades:

$$\begin{aligned} A^<(x^k + \lambda d^k) &\leq b^< \\ \lambda(A^< d^k) &\leq (b^< - A^< x^k) \\ \lambda \hat{d} &\leq \hat{b}, \end{aligned}$$

donde,

$$\begin{aligned} \hat{d} &= A^< d^k, \\ \hat{b} &= b^< - A^< x^k. \end{aligned} \tag{10.12}$$

Entonces

$$\lambda_{\max} = \begin{cases} +\infty & \text{si} \quad \mathrm{nf}(A^<) = 0, \\ +\infty & \text{si} \quad \mathrm{nf}(A^<) > 0 \quad \text{y} \quad \hat{d} \leq 0, \\ \min\{\dfrac{\hat{b}_i}{\hat{d}_i} : \hat{d}_i > 0\} & \text{si} \quad \mathrm{nf}(A^<) > 0 \quad \text{y} \quad \hat{d} \nleq 0. \end{cases} \tag{10.13}$$

La utilización adecuada del algoritmo GP hace que se cumplan las siguientes propiedades:

- $Md^k = 0$,
- $A^= d^k = 0$,

ALGORITMO DEL GRADIENTE PROYECTADO

datos: x^0 factible, ε, `MAXIT`
para $k = 0, ..., \texttt{MAXIT}$
  obtener $A^<,\ A^=,\ b^<$
  $M = \begin{bmatrix} A^= \\ E \end{bmatrix}$, $\text{fink} = 0$
  mientras $\text{fink} = 0$
    P : según 10.9 , $d^k = -Pf'(x^k)$
    si $||d^k|| > \varepsilon$ **ent**
      $\lambda_{\max}$: según 10.13
      $\lambda_k^* = \operatorname{argmin} f(x^k + \lambda d^k),\ \lambda \in [0, \lambda_{\max}]$
      $x^{k+1} = x^k + \lambda_k^* d^k$, $\text{fink} = 1$
    fin-ent
    sino
      si $\text{nf}(M) > 0$ **ent**
        $w = \begin{bmatrix} u \\ v \end{bmatrix} = -(MM^{\text{T}})^{-1} Mf'(x^k)$
        si $(\text{nf}(u) = 0)$ ó $(\text{nf}(u) > 0$ y $u \geq 0)$ **ent parar**
        sino
          buscar el u_i más pequeño
          quitar de M la fila correspondiente
        fin-sino
      fin-ent
      sino parar
    fin-sino
  fin-mientras
fin-para k

- $Ed^k = 0$,
- $f'(x^k)^{\text{T}} d^k < 0$,
- $\hat{b} > 0$,
- $\lambda_{\max} > 0$,
- x^{k+1} es admisible,
- $f(x^{k+1}) < f(x^k)$.

La salida deseable del algoritmo se alcanza cuando se obtiene un punto de KKT. Las salidas no deseables son:

- MM^{T} no es invertible,
- λ_k^* no existe: $f(x^k + \lambda d^k)$ no es acotado inferiormente,
- se llega hasta la iteración $k = \texttt{MAXIT}$ sin obtener un punto de KKT.

Ejemplo 10.3. Utilizar el método GP para el problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 + x_2^2 & \\ x_1 + x_2 & \leq 2 \\ x_1 + 5x_2 & \leq 5 \\ x & \geq 0, \end{aligned}$$

a partir de $x^0 = (3/2, 1/2)$.

Colocándolo en la forma adecuada

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 + x_2^2 & \\ x_1 + x_2 & \leq 2 \\ x_1 + 5x_2 & \leq 5 \\ -x_1 & \leq 0 \\ -x_2 & \leq 0, \end{aligned}$$

En este caso no hay igualdades, la matriz E y el vector e no tienen filas o simplemente no existen.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 5 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad f'(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix},$$

$$x^0 = \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix}, \quad f(x^0) = 2.5, \quad f'(x^0) = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

$$\mathcal{I}_0 = \{1\}, \quad A^< = \begin{bmatrix} 1 & 5 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad b^< = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \quad \widehat{b} = \begin{bmatrix} 1.0 \\ 1.5 \\ 0.5 \end{bmatrix},$$

$$M = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}, \quad d = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

$$\widehat{d} = \begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix}, \quad \lambda_{\max} = 0.25, \quad \lambda_0^* = 0.25.$$

$$x^1 = \begin{bmatrix} 1.25 \\ 0.75 \end{bmatrix}, \quad f(x^1) = 2.125, \quad f'(x^1) = \begin{bmatrix} 2.5 \\ 1.5 \end{bmatrix}, \quad \mathcal{I}_1 = \{1, 2\},$$

$$A^< = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad b^< = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \widehat{b} = \begin{bmatrix} 1.25 \\ 0.75 \end{bmatrix},$$

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$w = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -2.75 \\ 0.25 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 5 \end{bmatrix},$$

$$P = \begin{bmatrix} 0.961538 & -0.192308 \\ -0.192308 & 0.038462 \end{bmatrix}, \quad d = \begin{bmatrix} -2.115385 \\ 0.423077 \end{bmatrix},$$

$$\widehat{d} = \begin{bmatrix} 2.115385 \\ -0.423077 \end{bmatrix}, \quad \lambda_{\max} = 0.590909, \quad \lambda_1^* = 0.5.$$

$$x^2 = \begin{bmatrix} 0.192308 \\ 0.961538 \end{bmatrix}, \; f(x^2) = 0.961538, \; f'(x^2) = \begin{bmatrix} 0.384615 \\ 1.923077 \end{bmatrix}, \; \mathcal{I}_2 = \{2\},$$

$$A^< = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad b^< = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad \widehat{b} = \begin{bmatrix} 0.846154 \\ 0.192308 \\ 0.961538 \end{bmatrix},$$

$$M = \begin{bmatrix} 1 & 5 \end{bmatrix}, \quad P = \begin{bmatrix} 0.961538 & -0.192308 \\ -0.192308 & 0.038462 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$w = \begin{bmatrix} u_2 \end{bmatrix} = \begin{bmatrix} -0.384615 \end{bmatrix}, \quad M = \begin{bmatrix} \end{bmatrix},$$

$$P = \begin{bmatrix} rr1 & 0 \\ 0 & 1 \end{bmatrix}, \quad d = \begin{bmatrix} -0.384615 \\ -1.923077 \end{bmatrix}, \quad \widehat{d} = \begin{bmatrix} -2.307692 \\ 0.384615 \\ 1.923077 \end{bmatrix},$$

$$\lambda_{\max} = 0.5, \quad \lambda_2^* = 0.5.$$

$$x^3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad f(x^3) = 0, \quad f'(x^2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathcal{I}_3 = \{3, 4\},$$

$$A^< = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}, \quad b^< = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad \widehat{b} = \begin{bmatrix} 2 \\ 5 \end{bmatrix},$$

$$M = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$w = \begin{bmatrix} u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Luego el punto $x^3 = (0, 0)$ es un punto de KKT. $\diamond$

El método GP se puede modificar para usarlo en un problema general de PNL

$$\begin{array}{rcl} \min f(x) & & \\ g_i(x) & \leq & 0, \quad i = 1, ..., m \\ h_j(x) & = & 0, \quad j = 1, ..., l, \end{array}$$

mediante algunos cambios o adaptaciones, aunque presenta o puede presentar inconvenientes.

Las filas de la matriz M son los gradientes de las desigualdades activas y de las igualdades

$$M = \begin{bmatrix} g_i'(x^k)^{\mathrm{T}}, \ i \in \mathcal{I} \\ h_j'(x^k)^{\mathrm{T}}, \ \forall j \end{bmatrix}.$$

El cálculo de P, d^k y w se hace de la misma manera que en el problema con restricciones lineales. Como las restricciones no son lineales, es posible

que $x^k + \lambda d^k$ no sea admisible ni siquiera para valores muy pequeños de λ, en consecuencia no se puede calcular $\lambda_{\max}$, entonces

$$\begin{aligned} \lambda_k^* &= \operatorname{argmin} f(x^k + \lambda d^k), \quad \lambda \geq 0, \\ \tilde{x}^{k+1} &= x^k + \lambda_k^* d^k. \end{aligned}$$

Es muy posible que $\tilde{x}^{k+1}$ no sea admisible, entonces a partir de $\tilde{x}^{k+1}$ se obtiene un punto admisible x^{k+1}. Una manera de hacerlo consiste en minimizar una función de penalización $P(x)$, semejante a la utilizada en el capítulo 9, sin embargo, no siempre funciona bien pues este minimizador puede resultar peor que x^k.

La parte de regresar a la región admisible es la más delicada de la adaptación del método GP a restricciones no lineales.

EJERCICIOS

En los ejercicios 10.1 a 10.11 estudie el problema propuesto, averigüe cuáles métodos puede usar, use algunos de ellos. Si es necesario obtenga un problema equivalente al propuesto. Compare los resultados y la rapidez, verifique si el punto obtenido es punto de KKT, minimizador, minimizador global. Use también puntos iniciales diferentes al sugerido.

10.1 Minimizar $f(x_1, x_2) = x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_1 - 6x_2$ sujeto a $x_1 + x_2 \leq 2$, $-x_1 + 2x_2 \leq 2$, $x \geq 0$, con $x^0 = (0, 0)$.

10.2 Minimizar $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$ sujeto a $x_1^2 - x_2 + 2 \leq 0$, $x_1 \geq 0$, $-x_1 + 2 \leq 0$, con $x^0 = (0, 5/2)$.

10.3 Minimizar $f(x_1, x_2) = (x_1 + x_2 - 6)^2 + (2x_1 - x_2)^2$ sujeto a $x_1 + x_2 = 1$, $x \geq 0$, con $x^0 = (1, 0)$.

10.4 Minimizar $f(x_1, x_2) = 3x_1 + 4x_2$ sujeto a $x_1 + 2x_2 \geq 2$, $x \geq 0$. a) $x^0 = (2, 0)$; b) $x^0 = (0, 1)$.

10.5 Minimizar $f(x_1, x_2) = x_1 + x_2$ sujeto a $2x_1 + 3x_2 \geq 5$, $2x_1 + x_2 \geq 3$, $x \geq 0$, con $x^0 = (2.5, 0)$.

10.6 Minimizar $f(x_1, x_2) = x_1 + 2x_2$ sujeto a $-x_1 + x_2 = 4$, $x_2 \geq 3$, $x \geq 0$, con $x^0 = (0, 4)$.

10.7 Minimizar $f(x_1, x_2) = x_1^2 + x_2^2$ sujeto a $x_1 + x_2 \geq 4$, $2x_1 + x_2 \geq 5$, $x \geq 0$, con $x^0 = (4, 0)$.

10.8 Minimizar $f(x_1, x_2) = (x_1+1)^2+(x_2+1)^2$ sujeto a $x_1+x_2 = 1$, $x \geq 0$, con $x^0 = (1, 0)$.

10.9 Minimizar $f(x_1, x_2, x_3) = -x_1 - x_2 - x_3$ sujeto a $x_1^2 + x_2^2 + x_3^2 \leq 1$, $x_1 - x_2 + x_3 = 0$, $x \geq 0$, con $x^0 = (0, 0, 0)$.

10.10 Minimizar $f(x_1, x_2) = x_1^3 + x_2^2$ sujeto a $x_1 + x_2 \geq 4$, $2x_1 + x_2 \geq 5$, $x \geq 0$, con $x^0 = (4, 0)$.

10.11 Minimizar $f(x_1, x_2) = -x_1 - 1.4x_2$ sujeto a $x_1 + x_2 \leq 400$, $x_1 + 2x_2 \leq 580$, $x_1 \leq 300$, $x \geq 0$, con $x^0 = (300, 100)$.

Capítulo 11

MÉTODOS DE PUNTO INTERIOR

En este capítulo hay una presentación resumida de algunos métodos de punto interior, MPI, para programación lineal, para complementariedad lineal y para programación cuadrática convexa.

El método usual para resolver problemas de programación lineal, PL, es el método simplex. Este método empieza en una solución básica factible, es decir, un punto extremo o vértice del conjunto factible $\mathcal{F} = \{x : Ax = b, x \geq 0\}$, y en cada iteración va obteniendo otro vértice mejor que el anterior.

En el simplex, el número de "flops", operaciones de punto flotante, de cada iteración es aproximadamente $2mn$ o $2m^2$, dependiento de la variante utilizada. En la mayoría de los casos, [Dan63], el número de iteraciones no sobrepasa $3m$, más aún, con frecuencia el número de iteraciones no sobrepasa $3m/2$. Luego para la mayoría de los problemas el método simplex es polinomial. Un método se llama polinomial si el número de operaciones está acotado superiormente por un polinomio en función del tamaño del problema.

En 1972 Klee y Minty, [KlMn72], construyeron un ejemplo donde el número de iteraciones es 2^n. Así, en este caso artificial, el simplex es un método exponencial. En resumen, en la mayoría de los casos el simplex es polinomial, lo cual es bueno, pero en algunos casos puede ser exponencial, situación muy mala.

Quedaba entonces la inquietud de construir un algoritmo que fuera polinomial aún en el peor de los casos. En 1979 el matemático ruso L. G. Kachian o Kachiyan propone un nuevo método polinomial para problemas de PL basado en el método elipsoide pero desafortunadamente resultó menos

eficiente, en la práctica, que el simplex.

En 1984 Narendra Karmarkar, un matemático hindú, investigador de los laboratorios AT&T Bell, publicó el artículo *A New Polynomial-Time Algorithm for Linear Programming* [2], que marcó un hito en la Programación Lineal (PL). A partir del trabajo de Karmarkar se reactivó la investigación en PL, en especial en métodos de barrera y de punto interior (también para Programación No Lineal).

Actualmente hay métodos de punto interior para PL más eficientes que el método de Karmarkar, pero estos son consecuencia directa o indirecta del método de Karmarkar o de la redinamización de la investigación en MPI producida por su artículo.

11.0.1 Notación

En la literatura sobre MPI es usual utilizar la siguiente notación. Si x, s son vectores en $\mathbb{R}^n$,

$$x = (x_1, x_2, \ldots, x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

$$s = (s_1, s_2, \ldots, s_n) = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix},$$

se denota con letras mayúsculas las matrices diagonales

$$X = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & & 0 \\ \vdots & & \ddots & \\ 0 & & & x_n \end{bmatrix}, \quad S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & 0 \\ \vdots & & \ddots & \\ 0 & & & s_n \end{bmatrix}.$$

Además, e denota el vector de n unos:

$$e = (1, 1, \ldots, 1) = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

11.1 SOLUCIÓN DE UN SISTEMA FRECUENTE

En muchos métodos de punto interior, en cada iteración se presenta un sistema de ecuaciones semejante al siguiente:

$$\begin{bmatrix} 0_{n\times n} & A^{\mathrm{T}} & I_n \\ A & 0_{m\times m} & 0_{m\times n} \\ S & 0_{n\times m} & X \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f \\ g \\ h \end{bmatrix}$$

donde A es una matriz $m \times n$, $m \leq n$, $r(A) = m$, X, S son matrices diagonales positivas, $u, w, f, h \in \mathbb{R}^{n\times 1}$, $v, g \in \mathbb{R}^{m\times 1}$.

Obviamente este sistema, de tamaño $(2n + m) \times (2n + m)$, se podría resolver directamente por un método como el de Gauss o alguna de sus modificaciones. Sin embargo, es más eficiente utilizar su estructura por bloques. Este sistema se puede escribir

$$\begin{aligned} A^{\mathrm{T}}v + w &= f, && (11.1)\\ Au &= g, && (11.2)\\ Su + Xw &= h. && (11.3) \end{aligned}$$

De (11.1) y (11.3)

$$\begin{aligned} w &= f - A^{\mathrm{T}}v, && (11.4)\\ u &= S^{-1}(h - Xw). && (11.5) \end{aligned}$$

De (11.2) y (11.5)

$$\begin{aligned} Au - g = AS^{-1}(h - Xw) - g &= 0,\\ AS^{-1}(h - X(f - A^{\mathrm{T}}v)) - g &= 0,\\ AS^{-1}(h - Xf) + AS^{-1}XA^{\mathrm{T}}v - g &= 0,\\ AS^{-1}XA^{\mathrm{T}}\,v &= g + AS^{-1}(Xf - h) \quad (11.6) \end{aligned}$$

El sistema (11.6) se puede resolver fácilmente para obtener v. Como A es definida positiva, entonces se puede utilizar el método de Cholesky para resolverlo. Si al tratar de obtener la factorización de Cholesky, ésta no se puede conseguir, se puede concluir que $r(A) < m$ o que las matrices X, S no son positivas.

Utilizando (11.4) y (11.5) se obtiene el resto de la solución. En resumen:

$$\begin{aligned} \text{resolver} \quad AS^{-1}XA^{\mathrm{T}}v &= g + AS^{-1}(Xf - h), && (11.7)\\ w &= f - A^{\mathrm{T}}v, && (11.8)\\ u &= S^{-1}(h - Xw). && (11.9) \end{aligned}$$

Para el caso particular

$$\begin{bmatrix} 0_{n\times n} & A^{\mathrm{T}} & I_n \\ A & 0_{m\times m} & 0_{m\times n} \\ S & 0_{n\times m} & X \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ h \end{bmatrix}$$

los pasos a realizar son:

$$\begin{aligned} \text{resolver} \quad AS^{-1}XA^{\mathrm{T}}v &= -AS^{-1}h, && (11.10)\\ w &= -A^{\mathrm{T}}v, && (11.11)\\ u &= S^{-1}(h - Xw). && (11.12) \end{aligned}$$

11.2 MÉTODOS DE PUNTO INTERIOR PARA P.L.

El problema de PL que se va a resolver está en la forma estándar,

$$\begin{aligned} \min \quad & c^{\mathrm{T}}x \\ & Ax = b \\ & x \geq 0, \end{aligned} \qquad (11.13)$$

donde A es una matriz $m \times n$, $m \leq n$, rango$(A) = m$, c es un vector columna $n \times 1$, b es un vector columna $m \times 1$.

Los MPI trabajan con puntos interiores, es decir, puntos en $\mathcal{F}^\circ = \{x : Ax = b, x > 0\}$, interior relativo del conjunto factible $\mathcal{F} = \{x : Ax = b, x \geq 0\}$.

11.2.1 Condiciones de optimalidad

Al aplicar condiciones de KKT para el problema (11.13) reescrito de la siguiente manera,

$$\begin{aligned} \min\ & c^{\mathrm{T}}x \\ & -x \leq 0 \\ & Ax - b = 0, \end{aligned}$$

se obtiene

$$\begin{aligned} f'(x) + g'(x)u + h'(x)v = c - Iu + A^{\mathrm{T}}v = 0, \\ u \geq 0, \\ u^{\mathrm{T}}g(x) = u^{\mathrm{T}}(-x) = 0. \end{aligned}$$

Si se cambia u por s y v por $-y$, se obtiene

$$\begin{aligned} A^{\mathrm{T}}y + s - c = 0, \\ s \geq 0, \\ s^{\mathrm{T}}x = 0. \end{aligned}$$

Las dos primeras condiciones son simplemente las condiciones de factibilidad para el problema dual del problema (11.13).

Dado que $x \geq 0$ y $s \geq 0$, entonces $s^{\mathrm{T}}x$ es equivalente a $x_i s_i = 0\, \forall i$. Esto último se puede escribir

$$XSe = 0.$$

En resumen, las condiciones de factibilidad primal y las condiciones de KKT dan:

$$Ax - b = 0, \quad x \geq 0 \tag{11.14a}$$

$$A^{\mathrm{T}}y + s - c = 0, \quad s \geq 0 \tag{11.14b}$$

$$XSe = 0. \tag{11.14c}$$

La primera línea muestra simplemente las condiciones de factibilidad primal. La segunda da las condiciones de factibildad dual. La última igualdad recibe el nombre de condiciones de complementariedad ($x_i s_i = 0,\ \forall i$). Reordenando y relajando ligeramente las condiciones de no negatividad se obtiene

$$\begin{aligned} A^{\mathrm{T}}y + s - c &= 0, \\ Ax - b &= 0, \\ XSe &= 0, \\ (x, s) &> 0. \end{aligned} \tag{11.15}$$

La pareja (x, s) denota simplemente el vector de $\mathbb{R}^{2n}$

$$(x, s) = (x_1, x_2, ..., x_n, s_1, s_2, ..., s_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ s_1 \\ \vdots \\ s_n \end{bmatrix}$$

11.3 MÉTODO PRIMAL-DUAL AFÍN FACTIBLE

Definiendo una función $\Phi : \mathbb{R}^{2n+m} \to \mathbb{R}^{2n+m}$, las condiciones de optimalidad para PL se pueden escribir de manera simplificada:

$$\Phi(\xi) = \Phi(x, y, s) = 0, \quad (x, s) > 0.$$

Al anterior problema se le aplica el método de Newton con algunas modificaciones. El esquema del método de Newton para $\Phi(\xi) = 0$, es simplemente

$$\begin{aligned} \text{resolver} \quad \Phi'(\xi^k)\ \Delta\xi &= -\Phi(\xi^k), \\ \xi^{k+1} &= \xi^k + \Delta\xi, \end{aligned}$$

donde $\Phi'(\xi)$ es la matriz jacobiana de Φ. De manera explícita

$$\Phi'(x, y, s) = \begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix}.$$

En los métodos de punto interior todos los puntos (x^k, y^k, s^k), inclusive (x^0, y^0, s^0), deben cumplir

$$(x^k, s^k) > 0.$$

La aplicación directa del método de Newton $\xi^{k+1} = \xi^k + \Delta\xi$ haría, muy posiblemente, que no se cumpla la condición de positividad. Este inconveniente se obvia haciendo

$$\begin{aligned} \xi^{k+1} &= \xi^k + t_k \Delta\xi, \\ t_k &= \eta\, t_{\max}, \\ t_{\max} = t_{\max}((x^k, s^k), (\Delta x, \Delta s)) &= \max\{t : \xi^k + t\Delta\xi \geq 0\}, \\ 0 < \eta &< 1. \end{aligned}$$

Usualmente η varía en el intervalo [0.95, 0.999]. El valor máximo se puede calcular explícitamnte de la siguiente manera. Sean $u \in \mathbb{R}^p$, $u > 0$, $d \in \mathbb{R}^p$,

$$t_{\max} = t_{\max}(u, d) = \begin{cases} \infty & \text{si } d \geq 0, \\ \min\{\dfrac{u_i}{-d_i} : d_i < 0\} & \text{si } d \ngeq 0. \end{cases} \tag{11.16}$$

El valor $t_{\max}$ es el que hace exactamente que $\xi^k + t_{\max}\Delta\xi$ quede en la frontera. Al utilizar $t_k = \eta\, t_{\max}$ se obtiene un punto interior.

Las siguientes notaciones ayudan a hacer más compacta la escritura de las fórmulas y de los algoritmos.

Residuo primal:

$$r_p^k = Ax^k - b. \tag{11.17}$$

Residuo dual:

$$r_d^k = A^{\mathrm{T}} y^k + s^k - c. \tag{11.18}$$

Residuo de complementariedad:

$$r_c^k = X_k S_k e. \tag{11.19}$$

Aquí, X_k indica la matriz diagonal obtenida a partir de x^k.

Conjunto factible para las condiciones de factibilidad primal y dual:

$$\mathcal{F}_{PD} = \{(x, y, s) : Ax = b,\ A^{\mathrm{T}} y + s = c,\ (x, s) \geq 0\}.$$

Su interior relativo:

$$\mathcal{F}_{PD}^{\circ} = \{(x, y, s) : Ax = b,\ A^{\mathrm{T}}y + s = c,\ (x, s) > 0\}.$$

Medidas relativas de los residuos:

$$\begin{aligned} \rho_p &= \frac{\|r_p\|}{1 + \|b\|}, \\ \rho_d &= \frac{\|r_d\|}{1 + \|c\|}, \\ \rho_c &= \frac{\|r_c\|}{1 + \|x\|}. \end{aligned} \tag{11.20}$$

Una tripla (x, y, s) tal que $(x, s) \geq 0$ es factible si y solamente si $r_p = 0$ y $r_d = 0$, o lo que es lo mismo, si y solamente si $\rho_p = 0$ y $\rho_d = 0$.

Una tripla (x, y, s) factible es óptima si y solamente si $r_c = 0$. Una tripla en $\mathcal{F}_{PD}^{\circ}$ es aproximadamente óptima si $\rho_c \leq \varepsilon_c$ para un ε_c positivo y pequeño dado.

Sea (x^k, y^k, s^k) tal que $(x^k, s^k) > 0$. El sistema $\Phi'(\xi^k)\ \Delta\xi = -\Phi(\xi^k)$ expresado en x, y, s da:

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_d^k \\ -r_p^k \\ -r_c^k \end{bmatrix}. \tag{11.21}$$

Si $(x^k, y^k, s^k) \in \mathcal{F}_{PD}^{\circ}$, el sistema se simplifica a

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -r_c^k \end{bmatrix}. \tag{11.22}$$

Según lo visto en la primera sección de este capítulo, el sistema se resuelve aprovechando su estructura por bloques. Las ecuaciones (11.10), (11.11) y (11.12) se convierten en:

$$\text{resolver} \quad AS^{-1}XA^{\mathrm{T}}\,\Delta y = AS^{-1}r_c, \tag{11.23}$$

$$\Delta s = -A^{\mathrm{T}}\Delta y, \tag{11.24}$$

$$\Delta x = -S^{-1}(r_c + X\Delta s). \tag{11.25}$$

Como su nombre lo indica, el método primal dual afín factible parte de un punto factible $(x^0, y^0, s^0) \in \mathcal{F}_{PD}^{\circ}$. En cada iteración se calcula $(\Delta x, \Delta y, \Delta s)$, luego se calcula $t_{\max}$ y t_k. Finalmente se construye el nuevo punto.

MÉTODO PRIMAL-DUAL AFÍN FACTIBLE
datos: $(x^0, y^0, s^0) \in \mathcal{F}^\circ_{PD}$, ε_c, `MAXIT`, $\eta \in [0.95, 0.99]$
para $k = 0, ..., \texttt{MAXIT}$
    calcular ρ_c^k
    si $\rho_c^k \leq \varepsilon_c$ **ent parar**
    $(\Delta x, \Delta y, \Delta s)$ solución de (11.22)
    $t_{\max} = t_{\max}((x^k, s^k), (\Delta x, \Delta s))$
    $t_k = \min(1, \eta\, t_{\max})$
    $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + t_k(\Delta x, \Delta y, \Delta s)$
fin-para

Como los problemas primal y dual tienen puntos factibles, entonces no se puede presentar el caso de óptimo no acotado, luego no se presenta tampoco $t_{\max} = \infty$. Cerca de la solución algunas componentes de x son casi nulas, entonces la matriz $AS^{-1}XA^{\mathrm{T}}$ puede ser, numéricamente, casi singular o mal condicionada. Es necesario prever esta posible terminación no deseada del proceso, en el cálculo de $(\Delta x, \Delta y, \Delta s)$.

Dada una norma matricial $||\ \ ||$ en $\mathbb{R}^{n\times n}$ se define el número de condición o condicionamiento de una matriz A, cuadrada e invertible

$$\kappa(A) = ||A||\ ||A^{-1}||.$$

El condicionamiento siempre es mayor o igual a 1. Una matriz cuadrada es mal condicionada si su condicionamiento es grande. Es bien condicionada si el condicionamiento es cercano a 1.

Ejemplo 11.1. Aplicar el método primal-dual afín factible al siguiente problema

$$\begin{array}{rrrrrrr} \min & -x_1 & -1.4x_2 & & & & \\ & x_1 & +x_2 & +x_3 & & \leq & 4 \\ & x_1 & +2x_2 & & +x_4 & \leq & 5.8 \\ & x_1 & & & +x_5 & \leq & 3 \\ & & & & x & \geq & 0 \end{array}$$

con $x^0 = (1,\ 1,\ 2,\ 2.8,\ 2)$, $y^0 = (-0.3, -0.6, -0.3)$, $s^0 = (0.2,\ 0.1,\ 0.3,\ 0.6,\ 0.3)$, $\eta = 0.99$.

Después de verificar que $(x^0, y^0, s^0) \in \mathcal{F}^\circ_{PD}$, se calcula

$$r_c = (0.2,\ 0.1,\ 0.6,\ 1.68,\ 0.6),$$

$$\begin{aligned} ||r_c|| &= 1.8954, \\ \rho_c &= 0.3628. \end{aligned}$$

Como todavía no se tiene el óptimo, se resuelve el sistema (11.22). El primer paso consiste en resolver $AS^{-1}XA^{\mathrm{T}}\ \Delta y = AS^{-1}r_c$:

$$AS^{-1}XA^{\mathrm{T}} = \begin{bmatrix} 21.6667 & 25.0000 & 5.0000 \\ 25.0000 & 49.6667 & 5.0000 \\ 5.0000 & 5.0000 & 11.6667 \end{bmatrix}, \quad AS^{-1}r_c = \begin{bmatrix} 4.0000 \\ 5.8000 \\ 3.0000 \end{bmatrix}.$$

Entonces

$$\Delta y = (0.0637,\ 0.0644,\ 0.2023).$$

La dirección en s es simplemente $\Delta s = -A^{\mathrm{T}}\Delta y$:

$$\Delta s = (-0.3303,\ -0.1924,\ -0.0637,\ -0.0644,\ -0.2023).$$

Para la dirección en x, $\Delta x = -S^{-1}(r_c + X\Delta s)$,

$$r_c + X\Delta s = \begin{bmatrix} -0.1303 \\ -0.0924 \\ 0.4727 \\ 1.4998 \\ 0.1955 \end{bmatrix}, \quad \Delta x = \begin{bmatrix} 0.6515 \\ 0.9240 \\ -1.5756 \\ -2.4996 \\ -0.6515 \end{bmatrix}.$$

Cálculo del tamaño del paso:

$$\begin{aligned} t_{\max} &= 0.5197, \\ t_k &= 0.5145. \end{aligned}$$

El nuevo punto (x, y, s):

$$\begin{aligned} x^1 &= (1.3352,\ 1.4755,\ 1.1893,\ 1.5138,\ 1.6648), \\ y^1 &= (-0.2672,\ -0.5669,\ -0.1959), \end{aligned}$$

$$s^1 = (0.0300,\ 0.0010,\ 0.2672,\ 0.5669,\ 0.1959).$$

$$\begin{aligned} r_c &= (0.0401,\ 0.0015,\ 0.3178,\ 0.8582,\ 0.3262),\\ ||r_c|| &= 0.9723,\\ \rho_c &= 0.2298. \end{aligned}$$

Las direcciones:

$$\begin{aligned} \Delta x &= (1.1043,\ 0.1885,\ -1.2928,\ -1.4813,\ -1.1043),\\ \Delta y &= (-0.0233,\ 0.0122,\ 0.0660),\\ \Delta s &= (-0.0549,\ -0.0011,\ 0.0233,\ -0.0122,\ -0.0660). \end{aligned}$$

Cálculo del tamaño del paso:

$$\begin{aligned} t_{\max} &= 0.5473,\\ t_k &= 0.5419. \end{aligned}$$

El nuevo punto (x, y, s):

$$\begin{aligned} x^2 &= (1.9336,\ 1.5776,\ 0.4888,\ 0.7112,\ 1.0664),\\ y^2 &= (-0.2798,\ -0.5603,\ -0.1602),\\ s^2 &= (0.0003,\ 0.0004,\ 0.2798,\ 0.5603,\ 0.1602). \end{aligned}$$

k	ρ_c	x_1	x_2	x_3	x_4	x_5
2	0.1181	2.4410	1.5541	0.0049	0.2508	0.5590
3	0.0373	2.3579	1.6421	0.0000	0.1580	0.6421
4	0.0187	2.2437	1.7563	0.0000	0.0437	0.7563
5	0.0044	2.2028	1.7972	0.0000	0.0028	0.7972
6	0.0003	2.2000	1.8000	0.0000	0.0000	0.8000
7	0.0000					

11.4 TRAYECTORIA CENTRAL

En la práctica los métodos afines no son muy eficentes. Generalmente en las primeras iteraciones se obtienen puntos muy cercanos a la frontera pero alejados de la solución. En estos puntos el avance hacia el punto óptimo es lento.

Los métodos más eficientes buscan que los puntos no se alejen de la trayectoria central. Ésta se define a continuación.

Sean $\tau > 0$ y $(x(\tau), y(\tau), s(\tau))$ solución del siguiente sistema:

$$A^{\mathrm{T}}y + s - c = 0, \tag{11.26a}$$
$$Ax - b = 0, \tag{11.26b}$$
$$XSe - \tau e = 0, \tag{11.26c}$$
$$(x, s) > 0. \tag{11.26d}$$

Se puede demostrar que $(x(\tau), y(\tau), s(\tau))$ está definido de manera única para $\tau > 0$ si y solamente si $\mathcal{F}^{\circ}_{PD} \neq \emptyset$. Bajo esta condición se define la **trayectoria central** como el conjunto de todas estas triplas para $\tau > 0$,

$$\mathcal{C} = \{(x(\tau), y(\tau), s(\tau)) : \tau > 0\}.$$

Es claro que

$$\mathcal{C} \subset \mathcal{F}^{\circ}_{PD},$$
$$\lim_{\tau \to 0^+} = x^*.$$

En la definición de $(x(\tau), y(\tau), s(\tau))$ únicamente en la tercera igualdad hay un cambio con respecto a las condiciones de optimalidad (11.15) para el problema (11.13).

Ejemplo 11.2. Considere el problema

$$\begin{aligned} \min \ & -x_1 - 1.4x_2 \\ & x_1 + x_2 \leq 4 \\ & x_1 + 2x_2 \leq 5.8 \\ & x_1 \leq 3 \\ & x \geq 0. \end{aligned}$$

Al introducir variables de holgura se tiene exactamente el problema en la forma estándar del ejemplo 11.1. A continuación hay dos ejemplos de puntos de la trayectoria central.

$$
\begin{aligned}
&\tau = 1,\\
&x = (1.0913,\ 1.5758,\ 1.3329,\ 1.5572,\ 1.9087),\\
&y = (-0.7502,\ -0.6422,\ -0.5239),\\
&s = (0.9163,\ 0.6346,\ 0.7502,\ 0.6422,\ 0.5239).
\end{aligned}
$$

$$
\begin{aligned}
&\tau = 0.01,\\
&x = (2.1896,\ 1.7931,\ 0.0173,\ 0.0242,\ 0.8104),\\
&y = (-0.5789,\ -0.4134,\ -0.0123),\\
&s = (0.0046,\ 0.0056,\ 0.5789,\ 0.4134,\ 0.0123).
\end{aligned}
$$

Si se toman todos los puntos de la trayectoria central del problema (en la forma estándar) y se consideran únicamente las dos primeras componentes de x, o sea x_1, x_2, es decir, las variables del problema original (sin variables de holgura) se tiene la Figura 11.1.

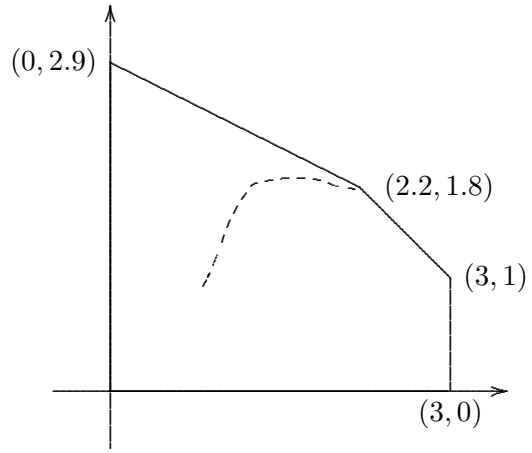


Figura 11.1

La mayoría de los métodos eficientes buscan obtener puntos que no se alejen de $\mathcal{C}$. Para esto se definen vecindades de $\mathcal{C}$ donde deben permanecer

los puntos. Hay varias estrategias, por ejemplo, las de pasos cortos y las de pasos largos. A su vez, hay varias subestrategias.

Es usual utilizar la siguiente medida de dualidad

$$\mu = \frac{x^{\mathrm{T}} s}{n}. \tag{11.27}$$

Esta medida de dualidad está relacionada con el salto o brecha dual (dual gap), ya que si $(x, y, s) \in \mathcal{F}_{PD}$, entonces se comprueba fácilmente que

$$x^{\mathrm{T}} s = c^{\mathrm{T}} x - b^{\mathrm{T}} y.$$

Por teoría de dualidad para PL se sabe que si x es factible para el problema primal y y es factible para el dual, entonces x, y son óptimos si $c^{\mathrm{T}} x - b^{\mathrm{T}} y = 0$. Esto se puede reescribir así: $(x, y, s) \in \mathcal{F}^{\circ}_{PD}$ es aproximadamente la solución si $\mu \leq \varepsilon$.

Un ejemplo típico de un teorema de complejidad para un algoritmo de punto interior es el presentado en [NoWr99], que simplificado dice: Sean $\varepsilon > 0$, $(x^0, y^0, s^0) \in \mathcal{F}^{\circ}_{PD}$ en una vecindad adecuada de $\mathcal{C}$. Entonces existe K, con $K = O(n \log(1/\varepsilon))$, tal que

$$\mu_k \leq \varepsilon, \quad \text{para todo } k \geq K.$$

11.5 MÉTODO PREDICTOR-CORRECTOR DE MEHROTRA

La mayoría de los programas eficicientes para problemas de PL de tipo general están basados en el algoritmo de Mehrotra. En este algoritmo no es necesario que $(x^0, y^0, s^0) \in \mathcal{F}^{\circ}_{PD}$. Simplemente se necesita que $x^0 > 0$ y que $s^0 > 0$. Obviamente si además se tiene factibilidad, pues mucho mejor.

En cada iteración del método de Mehrotra se calculan dos triplas de direcciones, una afín o predictora $(\Delta x^a, \Delta y^a, \Delta s^a)$ y una de corrección y centrado $(\Delta x^c, \Delta y^c, \Delta s^c)$. Con ellas se construye la dirección de cada iteración $(\Delta x, \Delta y, \Delta s)$.

Para no recargar la notación, en la iteración k, no se utilizará el superíndice k, o sea en lugar de utilizar (x^k, y^k, s^k), se utilizará simplemente (x, y, s).

El proceso acaba cuando se tiene un punto factible, $(x, y, s) \in \mathcal{F}^{\circ}_{PD}$, tal que $XSe \approx 0$. El método garantiza que siempre, en cualquier iteración, $(x, s) > 0$. Para decidir sobre la optimalidad se utilizan los residuos primal, dual y de complementariedad. Entonces el proceso termina cuando,

dados ε_p, ε_d, ε_c valores positivos pequeños, se cumplen las tres condiciones siguientes:

$$\begin{array}{rcl} \rho_p & \leq & \varepsilon_p, \\ \rho_d & \leq & \varepsilon_d, \\ \rho_c & \leq & \varepsilon_c. \end{array} \tag{11.28}$$

La tercera condición se puede remplazar por

$$\mu \leq \varepsilon_\mu.$$

Las direcciones afines se calculan mediante la solución de

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^a \\ \Delta y^a \\ \Delta s^a \end{bmatrix} = \begin{bmatrix} -r_d \\ -r_p \\ -r_c \end{bmatrix}. \tag{11.29}$$

Con estas direcciones se calculan los pasos afines primal y dual:

$$\begin{array}{l} t_p^a = \min\{1,\ t_{\max}(x, \Delta x^a)\}, \\ t_d^a = \min\{1,\ t_{\max}(s, \Delta s^a)\}. \end{array} \tag{11.30}$$

El calculo de μ^a (μ afín) se hace con lo que serían los "puntos afines". Con este valor se obtiene σ:

$$\mu^a = \frac{(x + t_p^a \Delta x^a)^{\mathrm{T}}(s + t_d^a \Delta s^a)}{n}, \tag{11.31}$$

$$\sigma = \left(\frac{\mu^a}{\mu}\right)^3. \tag{11.32}$$

Esto permite calcular lo que sería el residuo de complementariedad para la corrección y centrado:

$$r_c^c = -\sigma\mu\, e + \Delta X^a\, \Delta S^a\, e. \tag{11.33}$$

El cálculo de las direcciones de corrección y centrado se hace resolviendo el siguiente sistema. Como la matriz de coeficientes es la misma, no se requiere volver a hacer la factorización de Cholesky de $AS^{-1}XA^{\mathrm{T}}$, que es la parte más demorada de cada iteración.

$$\begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^c \\ \Delta y^c \\ \Delta s^c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -r_c^c \end{bmatrix}. \tag{11.34}$$

La dirección definitiva es la suma de las dos:

$$(\Delta x, \Delta y, \Delta s) = (\Delta x^a, \Delta y^a, \Delta s^a) + (\Delta x^c, \Delta y^c, \Delta s^c). \tag{11.35}$$

Los pasos definitivos primal y dual están dados por

$$\begin{aligned} t_p &= \min\{1,\ \eta\, t_{\max}(x, \Delta x)\}, \\ t_d &= \min\{1,\ \eta\, t_{\max}(s, \Delta s)\}, \end{aligned} \tag{11.36}$$

donde $\eta \in [0.9,\ 0.995]$ sirve para garantizar que $(x^{k+1}, s^{k+1}) > 0$.

Finalmente se calcula el siguiente punto:

$$\begin{aligned} x^{k+1} &= x + t_p \Delta x, \\ (y^{k+1},\ s^{k+1}) &= (y, s) + t_d(\Delta y, \Delta s). \end{aligned} \tag{11.37}$$

ALGORITMO DE MEHROTRA PARA P.L.
datos: $c,\ A,\ b,\ (x^0, y^0, s^0),\ (x^0, s^0) > 0,$
$\varepsilon_p,\ \varepsilon_d,\ \varepsilon_c,$ `MAXIT`
$(x, y, s) = (x^0, y^0, s^0)$
para $k = 0, .., $`MAXIT`
calcular $r_p,\ r_d,\ r_c$
calcular $\rho_p,\ \rho_d,\ \rho_c$
si $\rho_p \le \varepsilon_p$ y $\rho_d \le \varepsilon_d$ y $\rho_c \le \varepsilon_c$ **ent parar**
calcular $(\Delta x^a, \Delta y^a, \Delta s^a)$ resolviendo (11.29)
calcular $t_p^a,\ t_d^a$
calcular $\mu_a,\ \sigma$
calcular $(\Delta x^c, \Delta y^c, \Delta s^c)$ resolviendo (11.34)
$(\Delta x, \Delta y, \Delta s) = (\Delta x^a, \Delta y^a, \Delta s^a) + (\Delta x^c, \Delta y^c, \Delta s^c)$
calcular $t_p,\ t_d$
construir $(x^{k+1}, y^{k+1},\ s^{k+1})$
fin-para k

Ejemplo 11.3. Aplicar el método de Mehrotra al siguiente problema

$$\begin{array}{rrrrrrr} \min & -x_1 & -1.4x_2 & & & & \\ & x_1 & +x_2 & +x_3 & & \leq & 4 \\ & x_1 & +2x_2 & & +x_4 & \leq & 5.8 \\ & x_1 & & & +x_5 & \leq & 3 \\ & & & & x & \geq & 0 \end{array}$$

con $x^0 = (1.2, 1.2, 1.2, 1.2, 1.2)$, $y^0 = (-1, -1, -1)$, $s^0 = (0.5, 0.5, 0.5, 0.5, 0.5)$, $\eta = 0.99$.

El cálculo de los residuos da:

$$\begin{aligned} r_p &= \begin{bmatrix} -0.4000 & -1.0000 & -0.6000 \end{bmatrix}^{\mathrm{T}}, \\ r_d &= \begin{bmatrix} -1.5000 & -1.1000 & -0.5000 & -0.5000 & -0.5000 \end{bmatrix}^{\mathrm{T}}, \\ r_c &= \begin{bmatrix} 0.6000 & 0.6000 & 0.6000 & 0.6000 & 0.6000 \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

Además,

$$\rho_p = 0.1424, \quad \rho_d = 0.7542, \quad \rho_c 0.3643, \quad \mu = 0.6000.$$

Entonces el proceso continúa con el cálculo de las direcciones afines:

$$\begin{aligned} \Delta x^a &= \begin{bmatrix} 0.6080 & 0.6560 & -0.8640 & -0.9200 & -0.0080 \end{bmatrix}^{\mathrm{T}}, \\ \Delta y^a &= \begin{bmatrix} 0.6400 & 0.6167 & 0.9967 \end{bmatrix}^{\mathrm{T}}, \\ \Delta s^a &= \begin{bmatrix} -0.7533 & -0.7733 & -0.1400 & -0.1167 & -0.4967 \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

Longitudes de paso para las direcciones afines:

$$t_p^a = 1.0000, \quad t_d^a = 0.6466$$

$$\begin{aligned} x + t_p^a \Delta x^a &= \begin{bmatrix} 1.8080 & 1.8560 & 0.3360 & 0.2800 & 1.1920 \end{bmatrix}^{\mathrm{T}}, \\ s + t_d^a \Delta s^a &= \begin{bmatrix} 0.0129 & 0.0000 & 0.4095 & 0.4246 & 0.1789 \end{bmatrix}^{\mathrm{T}}, \end{aligned}$$

$$\mu^a = 0.0986, \quad \sigma = 0.0044.$$

El residuo de complementariedad para corrección y centrado:

$$r_c^c = \begin{bmatrix} -0.4607 & -0.5100 & 0.1183 & 0.1047 & 0.0013 \end{bmatrix}^{\mathrm{T}}.$$

Las direcciones de centrado:

$$\begin{aligned}\Delta x^c &= \begin{bmatrix}0.2129 & 0.1727 & -0.3857 & -0.5584 & -0.2129\end{bmatrix}^{\mathrm{T}},\\ \Delta y^c &= \begin{bmatrix}-0.0621 & -0.1454 & -0.0876\end{bmatrix}^{\mathrm{T}},\\ \Delta s^c &= \begin{bmatrix}0.2952 & 0.3530 & 0.0621 & 0.1454 & 0.0876\end{bmatrix}^{\mathrm{T}}.\end{aligned}$$

Las direcciones definitivas:

$$\begin{aligned}\Delta x &= \begin{bmatrix}0.8209 & 0.8287 & -1.2497 & -1.4784 & -0.2209\end{bmatrix}^{\mathrm{T}},\\ \Delta y &= \begin{bmatrix}0.5779 & 0.4712 & 0.9090\end{bmatrix}^{\mathrm{T}},\\ \Delta s &= \begin{bmatrix}-0.4581 & -0.4203 & -0.0779 & 0.0288 & -0.4090\end{bmatrix}^{\mathrm{T}}.\end{aligned}$$

Las longitudes de paso finales:

$$\begin{aligned}t_{\max}(x, \Delta x) &= 0.8117,\\ t_{\max}(s, \Delta s) &= 1.0914,\\ t_p &= 0.8036,\\ t_d &= 1.0000\end{aligned}$$

Los nuevos puntos:

$$\begin{aligned}x^1 &= \begin{bmatrix}1.8597 & 1.8659 & 0.1958 & 0.0120 & 1.0225\end{bmatrix}^{\mathrm{T}},\\ y^1 &= \begin{bmatrix}-0.4221 & -0.5288 & -0.0910\end{bmatrix}^{\mathrm{T}},\\ s^1 &= \begin{bmatrix}0.0419 & 0.0797 & 0.4221 & 0.5288 & 0.0910\end{bmatrix}^{\mathrm{T}}.\end{aligned}$$

k	ρ_p	ρ_d	μ	x_1	x_2	x_3	x_4	x_5
1	0.0280	0.0000	0.0817	2.2039	1.7847	0.0020	0.0031	0.7820
2	0.0034	0.0000	0.0035	2.2000	1.7998	0.0000	0.0000	0.7998
3	0.0000	0.0000	0.0000	2.2000	1.8000	0.0000	0.0000	0.8000
4	0.0000	0.0000	0.0000					

11.6 COMPLEMENTARIEDAD LINEAL

Sean M una matriz cuadrada de orden n y $q \in \mathbb{R}^n$. El problema de complementariedad lineal consiste en encontrar x y s en $\mathbb{R}^n$ tales que

$$Mx + q = s, \tag{11.38a}$$
$$(x, s) \geq 0, \tag{11.38b}$$
$$x^{\mathrm{T}} s = 0. \tag{11.38c}$$

Dado que $(x, s) \geq 0$, entonces $x^{\mathrm{T}} s = 0$ es equivalente a $x_i s_i = 0 \; \forall i$. Si se exige que M sea definida positiva se tiene un problema de complementariedad lineal monótona. En este libro se supondrá que se cumple esta última condición.

De manera semejante a PL, las condiciones (11.38) se puede reescribir como $\Phi(\xi) = 0$, $\xi \geq 0$.

$$\begin{bmatrix} Mx - s + q \\ XSe \end{bmatrix} = 0, \quad (x, s) \geq 0. \tag{11.39}$$

La mayoría de los conceptos y métodos de punto interior para PL se pueden adaptar para CL, por ejemplo, el conjunto factible y su interior relativo

$$\mathcal{F}_{CL} = \mathcal{F} = \{(x, s) : Mx + q = s, \; (x, s) \geq 0\},$$
$$\mathcal{F}^{\circ}_{CL} = \mathcal{F}^{\circ} = \{(x, s) : Mx + q = s, \; (x, s) > 0\}.$$

La trayectoria central es el conjunto de puntos que satisfacen

$$Mx - s + q = 0,$$
$$XSe - \tau e = 0,$$
$$(x, s) \geq 0,$$

para $\tau > 0$.

El método de Mehrotra para PL se puede adaptar para CL, algunas pasos y variables son exactamente los mismos. Otros son simplemente adaptaciones.

El residuo del sistema lineal, el residuo de complementariedad, los valores ρ, el valor μ,

$$\begin{aligned} r_L &= Mx - s + q, \\ r_c &= XSe, \\ \rho_L &= \frac{||r_L||}{1 + ||q||}, \\ \rho_c &= \frac{||r_c||}{1 + ||x||}, \\ \mu &= \frac{x^{\mathrm{T}} s}{n}\,. \end{aligned} \tag{11.40}$$

Las direcciones afines:

$$\begin{bmatrix} M & -I \\ S & X \end{bmatrix} \begin{bmatrix} \Delta x^a \\ \Delta s^a \end{bmatrix} = \begin{bmatrix} -r_L \\ -r_c \end{bmatrix}. \tag{11.41}$$

Los tamaños de paso afines:

$$\begin{aligned} t_x^a &= \min\{1,\ t_{\max}(x, \Delta x^a)\}, \\ t_s^a &= \min\{1,\ t_{\max}(s, \Delta s^a)\}. \end{aligned} \tag{11.42}$$

Los valores μ^a (μ afín) y σ:

$$\begin{aligned} \mu^a &= \frac{(x + t_x^a \Delta x^a)^{\mathrm{T}} (s + t_s^a \Delta s^a)}{n}, \\ \sigma &= \left(\frac{\mu^a}{\mu}\right)^3. \end{aligned} \tag{11.43}$$

El residuo de complementariedad para la corrección y centrado:

$$r_c^c = -\sigma\mu\, e + \Delta X^a\, \Delta S^a\, e.$$

Las direcciones de centrado:

$$\begin{bmatrix} M & -I \\ S & X \end{bmatrix} \begin{bmatrix} \Delta x^c \\ \Delta s^c \end{bmatrix} = \begin{bmatrix} 0 \\ -r_c^c \end{bmatrix}. \tag{11.44}$$

La dirección definitiva es la suma de las dos:

$$(\Delta x, \Delta s) = (\Delta x^a, \Delta s^a) + (\Delta x^c, \Delta s^c). \tag{11.45}$$

Las longitudes definitivas:

$$\begin{aligned} t_x &= \min\{1,\ \eta\, t_{\max}(x, \Delta x)\}, \\ t_s &= \min\{1,\ \eta\, t_{\max}(s, \Delta s)\}. \end{aligned} \tag{11.46}$$

Los puntos para la iteracion siguiente:

$$\begin{aligned} x^{k+1} &= x^k + t_x \Delta x, \\ s^{k+1} &= s^k + t_s \Delta s. \end{aligned} \tag{11.47}$$

ALGORITMO DE MEHROTRA PARA C.L.
datos: $M,\ q,\ (x^0, s^0) > 0,\ \varepsilon_L,\ \varepsilon_c,$ `MAXIT`
$(x, s) = (x^0, s^0)$
para $k = 0, .., $`MAXIT`
    calcular $r_L,\ r_c$
    calcular $\rho_L,\ \rho_c$
    si $\rho_L \leq \varepsilon_L$ y $\rho_c \leq \varepsilon_c$ **ent parar**
    calcular $(\Delta x^a, \Delta s^a)$
    calcular $t_x^a,\ t_s^a$
    calcular $\mu_a,\ \sigma$
    calcular $(\Delta x^c, \Delta s^c)$
    $(\Delta x, \Delta s) = (\Delta x^a, \Delta s^a) + (\Delta x^c, \Delta s^c)$
    calcular $t_x,\ t_s$
    construir $(x^{k+1},\ s^{k+1})$
fin-para k

11.7 PROGRAMACIÓN CUADRÁTICA CONVEXA

Los problemas de programación cuadrática se pueden expresar de varias maneras, una de ellas toma las restricciones lineales en la forma estándar:

$$\begin{aligned} \min\ \ f(x) = \frac{1}{2} x^{\mathrm{T}} H x + c^{\mathrm{T}} x \\ Ax = b \\ x \geq 0, \end{aligned} \tag{11.48}$$

donde H es una matriz simétrica $n \times n$, c es un vector columna $n \times 1$, A es una matriz $m \times n$, $m \leq n$, rango$(A) = m$, b es un vector columna $m \times 1$. En el caso convexo H es semidefinida positiva. La convexidad estricta de f se tiene si y solamente si H es definida positiva. En este caso, si el conjunto factible no es vacío, entonces la solución existe y es única. En esta sección se supondrá que H es semidefinida positiva.

Al aplicar condiciones de optimalidad (incluyendo factibilidad) se tiene:

$$\begin{aligned} Ax - b = 0, \quad x \geq 0, \\ Hx + c - Iu + A^{\mathrm{T}} v = 0, \quad u \geq 0, \\ u^{\mathrm{T}} x = 0. \end{aligned}$$

Si se cambia u por s y v por $-y$ y se reescribe, se obtiene

$$\begin{aligned} -Hx + A^{\mathrm{T}} y + s - c = 0, \\ Ax - b = 0, \\ XSe = 0, \\ (x, s) \geq 0. \end{aligned} \tag{11.49}$$

Estas condiciones son muy parecidas a las condiciones para PL. Entonces, de manera natural, se pueden adaptar los métodos de punto interior. Por ejemplo para el método de Mehrotra las fórmulas son:

$$\begin{aligned} \mu &= \frac{x^{\mathrm{T}} s}{n}, \\ r_p &= Ax - b, \\ r_d &= -Hx + A^{\mathrm{T}} y + s - c, \\ r_c &= XSe, \\ \rho_p &= \frac{\|r_p\|}{1 + \|b\|}, \\ \rho_d &= \frac{\|r_d\|}{1 + \|c\|}, \\ \rho_c &= \frac{\|r_c\|}{1 + \|x\|}. \end{aligned} \tag{11.50}$$

Las direcciones afines:

$$\begin{bmatrix} -H & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^a \\ \Delta y^a \\ \Delta s^a \end{bmatrix} = \begin{bmatrix} -r_d \\ -r_p \\ -r_c \end{bmatrix}. \tag{11.51}$$

Los pasos afines primal y dual:

$$t_p^a = \min\{1,\ t_{\max}(x, \Delta x^a)\},$$
$$t_d^a = \min\{1,\ t_{\max}(s, \Delta s^a)\}.$$

Calculo de μ^a (μ afín) y σ:

$$\mu^a = \frac{(x + t_p^a \Delta x^a)^{\mathrm{T}}(s + t_d^a \Delta s^a)}{n},$$
$$\sigma = \left(\frac{\mu^a}{\mu}\right)^3.$$

Residuo de complementariedad para la corrección y centrado:

$$r_c^c = -\sigma\mu\, e + \Delta X^a\, \Delta S^a\, e.$$

Las direcciones de corrección y centrado:

$$\begin{bmatrix} -H & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^c \\ \Delta y^c \\ \Delta s^c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -r_c^c \end{bmatrix}. \tag{11.52}$$

La dirección definitiva:

$$(\Delta x, \Delta y, \Delta s) = (\Delta x^a, \Delta y^a, \Delta s^a) + (\Delta x^c, \Delta y^c, \Delta s^c).$$

Los pasos definitivos primal y dual:

$$t_p = \min\{1,\ \eta\, t_{\max}(x, \Delta x)\},$$
$$t_d = \min\{1,\ \eta\, t_{\max}(s, \Delta s)\},$$

donde $\eta \in [0.9,\ 0.995]$. Finalmente el siguiente punto:

$$x^{k+1} = x + t_p \Delta x,$$
$$(y^{k+1},\ s^{k+1}) = (y, s) + t_d(\Delta y, \Delta s).$$

ALGORITMO DE MEHROTRA PARA P.C.C.
datos: $H,\ c,\ A,\ b,\ (x^0, y^0, s^0),\ (x^0, s^0) > 0,$
$\varepsilon_p,\ \varepsilon_d,\ \varepsilon_c,$ `MAXIT`
$(x, y, s) = (x^0, y^0, s^0)$
para $k = 0, .., $`MAXIT`
calcular $r_p,\ r_d,\ r_c$ según (11.50)
calcular $\rho_p,\ \rho_d,\ \rho_c$
si $\rho_p \leq \varepsilon_p$ y $\rho_d \leq \varepsilon_d$ y $\rho_c \leq \varepsilon_c$ **ent parar**
calcular $(\Delta x^a, \Delta y^a, \Delta s^a)$ resolviendo (11.51)
calcular $t_p^a,\ t_d^a$
calcular $\mu_a,\ \sigma$
calcular $(\Delta x^c, \Delta y^c, \Delta s^c)$ resolviendo (11.52)
$(\Delta x, \Delta y, \Delta s) = (\Delta x^a, \Delta y^a, \Delta s^a) + (\Delta x^c, \Delta y^c, \Delta s^c)$
calcular $t_p,\ t_d$
construir $(x^{k+1}, y^{k+1},\ s^{k+1})$
fin-para k

Ejemplo 11.4. Utilizar el método de punto interior descrito anteriormente para el problema:

$$\begin{aligned} \min f(x_1, x_2) = x_1^2 + x_2^2 & \\ x_1 + x_2 &\leq 2 \\ x_1 + 5x_2 &\leq 5 \\ x &\geq 0. \end{aligned}$$

Para poder resolver este problema se necesita convertir las restricciones en la forma estándar. Entonces introduciendo variables de holgura:

$$\begin{aligned} \min f(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 & \\ x_1 + x_2 + x_3 &= 2 \\ x_1 + 5x_2 + x_4 &= 5 \\ x &\geq 0. \end{aligned}$$

$$H = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ c = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{bmatrix},\ b = \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

Utilicemos $x^0 = (1,1,1,1)$, $y^0 = (0,0)$, $s^0 = (2,2,2,2)$.

$$r_p = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad r_d = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 2 \end{bmatrix}, \quad r_c = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$$\rho_p = 0.3502, \quad \rho_d = 2.8284, \quad \rho_c = 1.3333, \quad \mu = 2.0000$$

$$\Delta x^a = \begin{bmatrix} -0.5263 \\ -0.3158 \\ -0.1579 \\ 0.1053 \end{bmatrix}, \quad \Delta y^a = \begin{bmatrix} -0.3158 \\ 0.2105 \end{bmatrix}, \quad \Delta s^a = \begin{bmatrix} -0.9474 \\ -1.3684 \\ -1.6842 \\ -2.2105 \end{bmatrix}$$

$$t_p^a = 1, \ t_d^a = 0.9048, \ x + t_p^a \Delta x^a = \begin{bmatrix} 0.4737 \\ 0.6842 \\ 0.8421 \\ 1.1053 \end{bmatrix}, \ s + t_d^a \Delta s^a = \begin{bmatrix} 1.1429 \\ 0.7619 \\ 0.4762 \\ 0.0000 \end{bmatrix}$$

$$\mu^{\text{af}} = 0.3659, \quad \sigma = 0.0061, \quad r_c^c = \begin{bmatrix} 0.4864 \\ 0.4199 \\ 0.2537 \\ -0.2449 \end{bmatrix}$$

$$\Delta x^c = \begin{bmatrix} -0.0331 \\ -0.0177 \\ 0.0508 \\ 0.1218 \end{bmatrix}, \quad \Delta y^c = \begin{bmatrix} 0.3553 \\ -0.0013 \end{bmatrix}, \quad \Delta s^c = \begin{bmatrix} -0.4202 \\ -0.3844 \\ -0.3553 \\ 0.0013 \end{bmatrix}$$

$$\Delta x = \begin{bmatrix} -0.5594 \\ -0.3335 \\ -0.1071 \\ 0.2271 \end{bmatrix}, \quad \Delta y = \begin{bmatrix} 0.0395 \\ 0.2092 \end{bmatrix}, \quad \Delta s = \begin{bmatrix} -1.3676 \\ -1.7528 \\ -2.0395 \\ -2.2092 \end{bmatrix}$$

$$t_{\max}^p = 1.7876, \quad t_{\max}^d = 0.9053, \quad t_p = 1.0000, \quad t_d = 0.8962$$

$$x^1 = \begin{bmatrix} 0.4406 \\ 0.6665 \\ 0.8929 \\ 1.2271 \end{bmatrix}, \quad y^1 = \begin{bmatrix} 0.0354 \\ 0.1875 \end{bmatrix}, \quad s^1 = \begin{bmatrix} 0.7743 \\ 0.4291 \\ 0.1721 \\ 0.0200 \end{bmatrix}$$

$$r_p = \begin{bmatrix} -0.0000 \\ 0.0000 \end{bmatrix}, \quad r_d = \begin{bmatrix} 0.1161 \\ 0.0692 \\ 0.2075 \\ 0.2075 \end{bmatrix}, \quad r_c = \begin{bmatrix} 0.3412 \\ 0.2860 \\ 0.1537 \\ 0.0245 \end{bmatrix}$$

$$\rho_p = 0.0000, \quad \rho_d = 0.3231, \quad \rho_c = 0.1737, \quad \mu = 0.2013$$

$$\Delta x^a = \begin{bmatrix} -0.1949 \\ -0.3944 \\ 0.5893 \\ 2.1670 \end{bmatrix}, \quad \Delta y^a = \begin{bmatrix} 0.0781 \\ -0.1522 \end{bmatrix}, \quad \Delta s^a = \begin{bmatrix} -0.4318 \\ -0.1751 \\ -0.2857 \\ -0.0553 \end{bmatrix}$$

$$t_p^a = 1, \ t_d^a = 0.3615, \ x + t_p^a \Delta x^a = \begin{bmatrix} 0.2457 \\ 0.2720 \\ 1.4822 \\ 3.3941 \end{bmatrix}, \ s + t_d^a \Delta s^a = \begin{bmatrix} 0.6182 \\ 0.3657 \\ 0.0688 \\ 0.0000 \end{bmatrix}$$

$$\mu^{\text{af}} = 0.0883, \quad \sigma = 0.0845, \quad r_c^c = \begin{bmatrix} 0.0671 \\ 0.0521 \\ -0.1854 \\ -0.1369 \end{bmatrix}$$

$$\Delta x^c = \begin{bmatrix} -0.1015 \\ -0.2514 \\ 0.3529 \\ 1.3587 \end{bmatrix}, \quad \Delta y^c = \begin{bmatrix} -0.1396 \\ -0.0894 \end{bmatrix}, \quad \Delta s^c = \begin{bmatrix} 0.0260 \\ 0.0837 \\ 0.1396 \\ 0.0894 \end{bmatrix}$$

$$\Delta x = \begin{bmatrix} -0.2964 \\ -0.6458 \\ 0.9422 \\ 3.5256 \end{bmatrix}, \quad \Delta y = \begin{bmatrix} -0.0614 \\ -0.2416 \end{bmatrix}, \quad \Delta s = \begin{bmatrix} -0.4058 \\ -0.0914 \\ -0.1461 \\ 0.0341 \end{bmatrix}$$

$$t^p_{\max} = 1.0319, \quad t^d_{\max} = 1.1779, \quad t_p = 1.0000, \quad t_d = 1.0000$$

$$x^2 = \begin{bmatrix} 0.1442 \\ 0.0206 \\ 1.8352 \\ 4.7527 \end{bmatrix}, \quad y^2 = \begin{bmatrix} -0.0260 \\ -0.0541 \end{bmatrix}, \quad s^2 = \begin{bmatrix} 0.3685 \\ 0.3377 \\ 0.0260 \\ 0.0541 \end{bmatrix}$$

k	ρ_p	ρ_d	μ	x_1	x_2	x_3	x_4
2	0.0000	0.0000	0.0912	0.0611	0.0169	1.9219	4.8543
3	0.0000	0.0034	0.0032	0.0234	0.0070	1.9697	4.9419
4	0.0000	0.0008	0.0003	0.0089	0.0028	1.9883	4.9772
5	0.0000	0.0000	0.0000	0.0034	0.0011	1.9955	4.9911
6	0.0000	0.0000	0.0000	0.0013	0.0004	1.9983	4.9966
7	0.0000	0.0000	0.0000	0.0005	0.0002	1.9994	4.9987
8	0.0000	0.0000	0.0000	0.0002	0.0001	1.9998	4.9995
9	0.0000	0.0000	0.0000	0.0001	0.0000	1.9999	4.9998
10	0.0000	0.0000	0.0000				

EJERCICIOS

11.1 Considere el siguiente problema de PL:

$$\begin{aligned} \min \quad z = 3x_1 + 4x_2 & \\ x_1 + 2x_2 &\geq 4 \\ 5x_1 + 2x_2 &\geq 12 \\ x &\geq 0. \end{aligned}$$

Llévelo a la forma estándar. A partir de $x_1 = 3$, $x_2 = 3$, $y_1 = 0.2$, $y_2 = 0.2$ construya x^0, y^0, s^0 y utilice el método primal-dual afín factible.

11.2 Aplique el método de Mehrotra al siguiente problema:

$$\begin{aligned} \min \quad z = 10x_1 + 3x_2 & \\ x_1 + 2x_2 - x_3 &= 4 \\ 5x_1 + 2x_2 - x_4 &= 12 \\ x &\geq 0\,. \end{aligned}$$

11.3 Aplique el método de Mehrotra al siguiente problema

$$\begin{aligned} \min \quad z = 10x_1 + 4x_2 & \\ x_1 + 2x_2 - x_3 &= 4 \\ 5x_1 + 2x_2 - x_4 &= 12 \\ x &\geq 0\,, \end{aligned}$$

partiendo de $x^0 = (1,1,1,1)$, $y^0 = (0,0)$, $s^0 = (1,1,1,1)$. Compare con la solución obtenida con el simplex.

11.4 Escoja un problema de PL tal que el valor de la función objetivo no esté acotado inferiormente (su dual no tiene solución). Observe lo que pasa al aplicar el método de Mehrotra.

11.5 Escoja un problema de PL que no tenga solución y que su dual tampoco tenga solución. Observe lo que pasa al aplicar el método de Mehrotra.

11.6 Resuelva el siguiente problema:

$$\begin{aligned} \min \; f(x_1,x_2) = (x_1-7)^2 \;&+\; (x_2-7)^2 \\ x_1 + x_2 &\leq 10 \\ x_1 &\geq 1 \\ x_2 &\leq 8 \\ x &\geq 0. \end{aligned}$$

11.7 Resuelva el siguiente problema:

$$\begin{aligned} \min \; f(x_1,x_2) = (x_1-7)^2 \;&+\; (x_2-7)^2 \\ x_1 + x_2 &\leq 10 \\ x_1 &\geq 1 \\ x_2 &\leq 4 \\ x &\geq 0. \end{aligned}$$

Capítulo 12

PROGRAMACIÓN DINÁMICA

Este capítulo, pretende presentar la principal idea de la programación dinámica: *toda subpolítica de una política óptima también debe ser óptima.* El anterior principio se conoce como el principio de optimalidad de Richard Bellman.

El nombre de programación dinámica se debe a que inicialmente el método se aplicó a la optimización de algunos sistemas dinámicos, es decir, sistemas que evolucionan con el tiempo. Sin embargo, el tiempo no es indispensable, se requiere simplemente que los sistemas se puedan expresar por etapas o por fases.

Dicho de otra forma, la idea básica de la programación dinámica consiste en convertir un problema de n variables en una sucesión de problemas más simples, por ejemplo, de una variable, y para más sencillez, una variable discreta.

Desde un punto de vista recurrente[1]: un problema complejo se resuelve mediante el planteamiento de problemas más sencillos pero análogos al problema general. Estos problemas más sencillos se resuelven mediante el planteamiento de problemas aún más sencillos pero que siguen guardando la misma estructura. Este proceso recurrente se aplica hasta encontrar problemas de solución inmediata. Una vez resueltos estos problemas supersencillos se pasa a la solución de los problemas un poquito mas complejos y así sucesivamente hasta calcular la solución del problema general.

Aunque el principio es muy sencillo y aplicable a muchos problemas,

[1]Con cierta frecuencia, en lugar de recurrente, se utiliza el término "recursivo", anglicismo usado para la traduccion de "recursive".

se aplica de manera específica a cada problema. Es decir, no existe un algoritmo (o un programa de computador) único que se pueda aplicar a todos los problemas.

En lugar de presentar fórmulas, definiciones o conceptos generales pero abstractos, se presentan ejemplos típicos con sus soluciones. Todos los ejemplos presentados son determinísticos, es decir, se supone que todos los datos del problema son conocidos de manera precisa.

12.1 EL PROBLEMA DE LA RUTA MÁS CORTA

12.1.1 Enunciado del problema

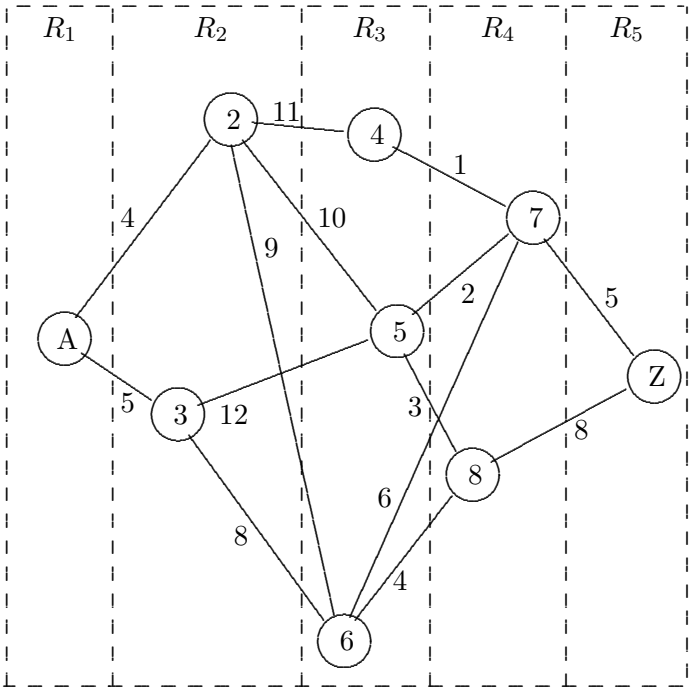
El Ministerio de Obras desea construir una autopista entre la ciudad de Girardot, que denotaremos simplemente por A y la ciudad de Barranquilla denotada por$Z$. Globalmente, la autopista sigue la dirección del río Magdalena. El valle del río y su zona de influencia directa se dividió en una sucesión de n regiones adyacentes que, por facilidad, llamaremos R_1, R_2, ... R_n. La ciudad A está en R_1 y Z está en R_n. La autopista pasa por todas las n regiones, pero está previsto que pase solamente por una ciudad de cada región. Sin embargo en algunas regiones hay varias ciudades importantes y cada una de ellas podría ser la ciudad de la región por donde pasa la autopista. Con este esquema, el Ministerio está estudiando muchos tramos de autopista, cada tramo va de una ciudad en una región, a otra ciudad en la región siguiente. Sin embargo, de cada ciudad en una región no hay necesariamente tramos a todas las ciudades de la siguiente región. Pero por otro lado, para cada ciudad de la regiones intermedias, de la 2 a la $n-1$, hay por lo menos un tramo proveniente de una ciudad de la región anterior y por lo menos un tramo que va hasta una ciudad de la siguiente región.

Para cada uno de estos tramos posibles el Ministerio ha calculado un costo total que tiene en cuenta, entre otros aspectos, la distancia, las dificultades específicas de la construcción, el sobrecosto del transporte desde otras ciudades de cada región hasta la ciudad por donde pasa la autopista.

El objetivo del Ministerio es encontrar una sucesión de tramos concatenados, que van desde A hasta Z, con costo mínimo.

Las condiciones del problema se pueden formalizar de la siguiente manera:

- $N = \{A, ..., Z\}$ conjunto de ciudades o nodos. N es simplemente un conjunto finito cualquiera (no necesariamente un subconjunto del abecedario), en el cual están A y Z.



- R_1, R_2, ..., R_n forman una partición de N, es decir,

$$\begin{aligned} R_1 \cup R_2 \cup ... \cup R_n &= N, \\ R_i &\neq \emptyset \ , \quad i = 1, ..., n, \\ R_i \cap R_j &= \emptyset \ \text{ si } \quad i \neq j. \end{aligned}$$

- $R_1 = \{A\}$.
- $R_n = \{Z\}$.
- $\Gamma \subseteq N \times N$ conjunto de tramos posibles ("flechas" del grafo[2]).
- Si $(i, j) \in \Gamma$ entonces existe $1 \leq k \leq n-1$ tal que $i \in R_k$ y $j \in R_{k+1}$.
- Para $1 \leq k \leq n-1$, si $i \in R_k$ entonces existe $j \in R_{k+1}$ tal que $(i, j) \in \Gamma$.

[2]Un grafo G es una pareja $G = (N, \Gamma)$, donde N es el conjunto finito de vértices y $\Gamma \subseteq N \times N$ es el conjunto de flechas

- Para $2 \leq k \leq n$, si $i \in R_k$ entonces existe $j \in R_{k-1}$ tal que $(j, i) \in \Gamma$.
- Si $(i, j) \in \Gamma$ entonces se conoce $c(i, j) = c_{ij} > 0$.

En teoría de grafos se habla de predecesor, sucesor, conjunto de predecesores, conjunto de sucesores. La utilización de estos conceptos, con sus respectivas notaciones, facilita y hace más compacto el planteamiento del problema. Si la flecha (i, j) está en el grafo se dice que i es un **predecesor** de j, y que j es un **sucesor** de i. Se denota por $\Gamma^-(i)$ el conjunto de predecesores de i y por $\Gamma^+(i)$ el conjunto de sus sucesores. Entonces:

- $\Gamma^-(i) \neq \emptyset,\ \forall i \neq A.$
- $\Gamma^+(i) \neq \emptyset,\ \forall i \neq Z.$
- $i \in R_k \Rightarrow \Gamma^-(i) \subseteq R_{k-1},\ k = 2, ..., n.$
- $i \in R_k \Rightarrow \Gamma^+(i) \subseteq R_{k+1},\ k = 1, ..., n-1.$

12.1.2 Planteamiento del problema de optimización

El problema se puede presentar de la siguiente manera: encontrar i_1, i_2, ..., i_n para minimizar una función con ciertas restricciones:

$$\begin{aligned} \min c(i_1, i_2) + c(i_2, i_3) + ... + c(i_{n-1}, i_n) = \sum_{k=1}^{n-1} c(i_k, i_{k+1}) \\ (i_k, i_{k+1}) \in \Gamma,\ k = 1, ..., n-1 \\ i_k \in R_k,\ k = 1, ..., n. \end{aligned}$$

En realidad el problema depende únicamente de $n-2$ variables: i_2, i_3, ..., i_{n-1} ya que $i_1 = A$, $i_n = Z$. Además la última condición se puede quitar puesto que ya está implícita en las propiedades de Γ.

A cada una de las ciudades de R_2 llega por lo menos un tramo desde una ciudad de R_1, pero como en R_1 solo hay una ciudad entonces: a cada una de las ciudades de R_2 llega por lo menos una ruta desde A. A su vez, a cada una de las ciudades de R_3 llega por lo menos un tramo desde una ciudad de R_2 y como a cada ciudad de R_2 llega una ruta desde A entonces: a cada una de las ciudades de R_3 llega por lo menos una ruta desde A. Repitiendo este proceso se puede deducir que a cada una de las ciudades llega por lo menos una ruta desde A. En particular existe por lo menos una ruta desde A hasta Z.

Una manera de resolver este problema es utilizando la **fuerza bruta**: hacer una lista de todas las rutas posibles desde A hasta Z, para cada ruta evaluar el costo (sumando los costos de cada tramo), y buscar la ruta (o una de las rutas) de menor costo.

De ahora en adelante, cuando se hable del mejor (camino, procedimiento, política, ...) se entenderá que es el mejor, en el sentido estricto cuando hay uno solo, o uno de los mejores cuando hay varios.

12.1.3 Solución por programación dinámica

La forma recurrente de resolver este problema es muy sencilla. Para conocer la mejor ruta de A hasta Z basta con conocer la mejor ruta a cada una de las ciudades de R_{n-1}. Al costo de cada ruta óptima hasta una de ciudad de R_{n-1} se le agrega el costo del tramo entre esa ciudad de R_{n-1} y Z en R_n y finalmente se escoge la menor suma. Definir algunas funciones permite escribir el razonamiento anterior de manera más corta y precisa.

Sean :

$$\begin{aligned} C_n^*(Z) &: \text{costo mínimo de todas las rutas desde } A \text{ hasta } Z. \\ C_{n-1}^*(i) &: \text{costo mínimo de todas las rutas desde } A \text{ hasta la} \\ &\quad \text{ciudad } i \in R_{n-1}. \end{aligned}$$

Entonces la solución recurrente dice:

$$C_n^*(Z) = \min_{i \in R_{n-1}} \{C_{n-1}(i) + c(i, Z)\}.$$

Obviamente esto presupone que se conocen todos los valores $C_{n-1}^*(i)$, y entonces la pregunta inmediata es: *¿Cómo se calculan los valores $C_{n-1}^*(i)$?* La respuesta es de nuevo recurrente: *Utilizando los costos de las rutas mínimas desde A hasta las ciudades de R_{n-2}.*
Sea :

$$\begin{aligned} C_{n-2}^*(i) &: \text{costo mínimo de todas las rutas desde } A \text{ hasta la} \\ &\quad \text{ciudad } i \in R_{n-2}. \end{aligned}$$

Entonces la solución recurrente dice:

$$C_{n-1}^*(j) = \min\{C_{n-2}(i) + c(i, j) : i \in \Gamma^-(j)\}, \quad j \in R_{n-1}.$$

Este proceso se repite hasta poder utilizar valores "inmediatos" o de muy fácil obtención. Para este problema de la autopista, puede ser

$$C_2^*(j) = c(A, j), \quad j \in R_2.$$

donde $C_2^*(j)$indica costo mínimo de todas las rutas (hay una sola) desde A hasta la ciudad $j \in R_2$.

La deducción de la solución recurrente se hizo hacia atrás (desde n hasta 2) pero el cálculo se hace hacia adelante (de 2 hasta n). En resumen,

- definir una función que permita la recurrencia,
- definir el objetivo final,
- definir las condiciones iniciales (fáciles de evaluar),
- definir una relación o fórmula recurrente,
- calcular los valores iniciales,
- hacer cálculos recurrentes hasta encontrar la solución

Para este problema :

$$\begin{aligned} C_k^*(i) &= \text{costo mínimo de todas las rutas desde } A \text{ hasta la} \\ &\qquad \text{ciudad } i \in R_k, \;\; k = 2, ..., n. \\ C_n^*(Z) &= \; ? \\ C_2^*(j) &= c(A, j), \quad j \in R_2 \\ C_{k+1}^*(j) &= \min\{C_k^*(i) + c(i, j) : i \in \Gamma^-(j)\}, \;\; 2 \leq k \leq n-1, \;\; j \in R_{k+1} \end{aligned}$$

Un planteamiento ligeramente diferente podría ser:

$$\begin{aligned} C_k^*(i) &= \text{costo mínimo de todas las rutas desde } A \text{ hasta la} \\ &\qquad \text{ciudad } i \in R_k, \;\; k = 2, ..., n. \\ C_n^*(Z) &= \; ? \\ C_1^*(A) &= 0 \\ C_{k+1}^*(j) &= \min\{C_k^*(i) + c(i, j) : i \in \Gamma^-(j)\}, \;\; 1 \leq k \leq n-1, \;\; j \in R_{k+1} \end{aligned}$$

La diferencia está simplemente en el "sitio inicial".

12.1.4 Resultados numéricos

Los valores iniciales se obtienen inmediatamente:

$$\begin{aligned} C_2^*(2) &= c_{A2} = 4, \\ C_2^*(3) &= c_{A3} = 5. \end{aligned}$$

El proceso recurrente empieza realmente a partir de las ciudades de R_3:

$$\begin{aligned} C_3^*(4) &= \min_{i \in \Gamma^-(4)} \{C_2^*(i) + c_{i4}\}, \\ &= \min\{C_2^*(2) + c_{24}\}, \\ &= \min\{4 + 11\}, \\ C_3^*(4) &= 15. \end{aligned}$$

$$\begin{aligned} C_3^*(5) &= \min_{i \in \Gamma^-(5)} \{C_2^*(i) + c_{i5}\}, \\ &= \min\{C_2^*(2) + c_{25}, C_2^*(3) + c_{35}\}, \\ &= \min\{4 + 10, 5 + 12\}, \\ C_3^*(5) &= 14. \end{aligned}$$

$$\begin{aligned} C_3^*(6) &= \min_{i \in \Gamma^-(6)} \{C_2^*(i) + c_{i6}\}, \\ &= \min\{C_2^*(2) + c_{26}, C_2^*(3) + c_{36}\}, \\ &= \min\{4 + 9, 5 + 8\}, \\ C_3^*(6) &= 13. \end{aligned}$$

En resumen para la región R_3:

$$\begin{aligned} C_3^*(4) &= 15, \\ C_3^*(5) &= 14, \\ C_3^*(6) &= 13. \end{aligned}$$

Para la región R_4:

$$\begin{aligned} C_4^*(7) &= \min_{i\in\Gamma^-(7)}\{C_3^*(i)+c_{i7}\},\\ &= \min\{C_3^*(4)+c_{47}, C_3^*(5)+c_{57}, C_3^*(6)+c_{67}\},\\ &= \min\{15+1, 14+2, 13+6\},\\ C_4^*(7) &= 16. \end{aligned}$$

$$\begin{aligned} C_4^*(8) &= \min_{i\in\Gamma^-(8)}\{C_3^*(i)+c_{i8},\}\\ &= \min\{C_3^*(5)+c_{58}, C_3^*(6)+c_{68}\},\\ &= \min\{14+3, 13+4\},\\ C_4^*(8) &= 17. \end{aligned}$$

En resumen para la región R_4:

$$\begin{aligned} C_4^*(7) &= 16,\\ C_4^*(8) &= 17. \end{aligned}$$

Finalmente para Z en R_5:

$$\begin{aligned} C_5^*(Z) &= \min_{i\in\Gamma^-(Z)}\{C_4^*(i)+c_{iZ}\},\\ &= \min\{C_4^*(7)+c_{7Z}, C_4^*(8)+c_{8Z}\},\\ &= \min\{16+5, 17+8\},\\ C_5^*(Z) &= 21. \end{aligned}$$

Ya se obtuvo el costo mínimo, pero es necesario conocer también las ciudades por donde debe pasar una autopista de costo mínimo. Con la información que se tiene no se puede reconstruir la mejor ruta. Entonces es necesario volver a resolver el problema, pero esta vez es necesario, para cada ciudad intermedia j, no solo conocer $C_k^*(j)$, sino también saber desde que ciudad i^* de la región R_{k-1} se obtiene este costo mínimo.

Obviamente

$$C_2^*(2) = 4,\ \ i^* = A,$$

$$C_2^*(3) = 5, \;\; i^* = A.$$

$C_3^*(4)$ se obtiene viniendo de la única ciudad predecesora: 2.

$$C_3^*(4) = 15, \;\; i^* = 2.$$

$$\begin{aligned} C_3^*(5) &= \min\{C_2^*(2) + c_{25}, C_2^*(3) + c_{35}\}, \\ &= \min\{\overline{4+10}, 5+12\}, \\ C_3^*(5) &= 14, \;\; i^* = 2. \end{aligned}$$

Para $C_3^*(7)$ hay empate ya que se obtiene el costo mínimo viniendo de 2 o viniendo de 3, sin embargo basta con tener información sobre una de las mejores ciudades precedentes, por ejemplo, la primera encontrada.

$$\begin{aligned} C_3^*(6) &= \min\{C_2^*(2) + c_{26}, C_2^*(3) + c_{36}\}, \\ &= \min\{\overline{4+9}, \overline{5+8}\}, \\ C_3^*(6) &= 13, \;\; i^* = 2. \end{aligned}$$

Toda la información necesaria para poder calcular el costo mínimo desde A hasta Z y poder reconstruir una ruta mínima es:

j	$C_2^*(j)$	i^*	j	$C_3^*(j)$	i^*	j	$C_4^*(j)$	i^*	j	$C_5^*(j)$	i^*
2	4	A	4	15	2	7	16	4	Z	21	7
3	5	A	5	14	2	8	17	5			
			6	13	2						

Luego según la tabla, el costo mínimo es 21. Además a Z se llega proveniente de 7, a 7 se llega proveniente de 4, a 4 se llega proveniente de 2, y finalmente a 2 se llega proveniente de A. Entonces la autopista de costo mínimo (o una de las autopistas de costo mínimo) es: $(A, 2, 4, 7, Z)$

12.1.5 Solución hacia atrás

La solución presentada en los dos numerales anteriores se conoce como la solución **hacia adelante**. También se tiene una solución análoga hacia atrás. Se busca el costo mínimo desde cada ciudad hasta Z empezando con el costo desde las ciudades en la región R_{n-1}.

$C_k^*(i) =$ costo mínimo de todas las rutas desde la ciudad $i \in R_k$ hasta Z, $k = n-1, n-2, ..., 1$.

Objetivo final:

$$C_1^*(A) = \ ?$$

Condiciones iniciales:

$$C_{n-1}^*(j) = c(j, Z), \ \ j \in R_{n-1}.$$

Relación recurrente:

$$C_{k-1}^*(i) = \min_{j \in \Gamma^+(i)} \{c(i,j) + C_k^*(j)\}, \quad n-1 \geq k \geq 2, \ \ i \in R_{k-1}.$$

Obviamente

$$\begin{aligned} C_4^*(7) &= 5, \ \ j^* = Z, \\ C_4^*(8) &= 8, \ \ j^* = Z. \end{aligned}$$

$$\begin{aligned} C_3^*(4) &= \min_{j \in \Gamma^+(4)} \{c(4,j) + C_4^*(j)\}, \\ &= \min\{c(4,7) + C_4^*(7)\}, \\ &= \min\{\overline{1+5}\}, \\ C_3^*(4) &= 6, \ \ j^* = 7. \end{aligned}$$

$$\begin{aligned} C_3^*(5) &= \min_{j \in \Gamma^+(5)} \{c(5,j) + C_4^*(j)\}, \\ &= \min\{c(5,7) + C_4^*(7), c(5,8) + C_4^*(8)\}, \\ &= \min\{\overline{2+5}, 3+8\}, \\ C_3^*(5) &= 7, \ \ j^* = 7. \end{aligned}$$

y así sucesivamente.

i	$C_4^*(i)$	j^*	i	$C_3^*(i)$	j^*	i	$C_2^*(i)$	j^*	i	$C_1^*(i)$	j^*
7	5	Z	4	6	7	2	17	4	A	21	2
8	8	Z	5	7	7	3	19	5			
			6	11	7						

Luego según la tabla, el costo mínimo es 21. Además se llega desde A pasando por 2, se llega desde 2 pasando por 4, se llega desde 4 pasando por 7, y finalmente se llega desde 7 pasando por Z. Entonces la autopista de costo mínimo (o una de las autopistas de costo mínimo) es: $(A, 2, 4, 7, Z)$

12.2 EL PROBLEMA DE ASIGNACIÓN DE MÉDICOS

12.2.1 Enunciado del problema

(Adaptación de un ejemplo de Hillier y Lieberman). La OMS (Organización Mundial de la Salud) tiene un equipo de m médicos especialistas en salud pública y los desea repartir en n paises P_1, P_2, ..., P_n para que desarrollen campañas educativas tendientes a disminuir la mortalidad infantil. De acuerdo con las condiciones específicas de cada país, de la tasa de natalidad, de la mortalidad antes de los dos años, del número de habitantes, la OMS posee evaluaciones bastante precisas de los valores $b(i,j) \equiv b_{ij}$, $i = 0, ..., m$, $j = 1, ..., n$ que indican el beneficio de asignar i médicos al país P_j. Este beneficio b_{ij} indica (en cientos de mil) la disminución en el número de niños muertos antes de los dos años de vida, durante los próximos 5 años. Así por ejemplo, $b_{23} = 6$ indica que si se asignan 2 médicos al país P_3 se espera que en los próximos 5 años haya una disminución de 600000 en el número de niños muertos antes de los dos años de vida.

Los beneficios no son directamente proporcionales al número de médicos, es decir, si $b_{23} = 6$ no se cumple necesariamente que $b_{43} = 12$.

Se supone además que no es obligatorio asignar médicos en cada uno de los paises y también se supone que es posible asignar todos los médicos a un solo país.

También se supone que al aumentar el número de médicos asignados a un país el beneficio no disminuye. Pensando en un problema más general se podría pensar que cuando hay demasiadas personas asignadas a una labor, se obstruye el adecuado funcionamiento y el resultado global podría disminuir. Sin embargo se puede suponer que b_{ij} indica el mayor beneficio obtenido en el país P_j al asignar a los más i médicos.

La OMS desea saber cuantos médicos debe asignar a cada país para maximizar el beneficio total, o sea, para maximizar la disminución total en la mortalidad infantil en los n paises.

12.2.2 Planteamiento del problema de optimización

Se necesita conocer el número de médicos que se asigna a cada país para maximizar el beneficio total, sin sobrepasar el número de médicos disponibles. Si x_j indica el número de médicos que se asignan al pais P_j, entonces el problema de optimización es:

$$\begin{aligned} \max \sum_{j=1}^{n} b(x_j, j) \\ \sum_{j=1}^{n} x_j \leq m, \\ 0 \leq x_j \leq m, \;\; j = 1, ..., n, \\ x_j \in \mathbb{Z}, \;\; j = 1, ..., n. \end{aligned}$$

Este problema se puede resolver por la fuerza bruta construyendo todas las combinaciones, verificando si cada combinación es factible ($\sum_{j=1}^{n} x_j \leq m$) y escogiendo la mejor entre las factibles. De esta forma cada variable puede tomar $m+1$ valores: 0, 1, 2, ..., m, o sea, es necesario estudiar $(m+1)^n$ combinaciones. Es claro que para valores grandes de m y n el número de combinaciones puede ser inmanejable.

12.2.3 Solución recurrente

Para asignar optimamente m médicos a los n paises hay que asignar una parte de los médicos a los paises P_1, P_2, ..., P_{n-1} y el resto al país P_n. Supongamos que sabemos asignar óptimamente 0, 1, 2, 3, ..., m médicos a los paises P_1, P_2, ..., P_{n-1}. Entonces para asignar óptimamente m médicos a los n paises hay que considerar $m + 1$ posibilidades:

- 0 médicos a los paises P_1,..., P_{n-1} y m médicos a P_n
- 1 médico a los paises P_1,..., P_{n-1} y $m - 1$ médicos a P_n
- 2 médicos a los paises P_1,..., P_{n-1} y $m - 2$ médicos a P_n

  ...

- m médicos a los paises P_1,..., P_{n-1} y 0 médicos a P_n

Al escoger la mejor combinación se tiene la solución del problema. Obviamente para conocer las soluciones óptimas en los primeros $n-1$ paises se requiere conocer las soluciones óptimas en los primeros $n-2$ paises y así sucesivamente. O sea, primero se resuelve el problema de asignar óptimamente médicos al primer país, con estos resultados se puede obtener la asignación óptima de médicos a los 2 primeros paises, y así sucesivamente hasta obtener la solución global.

$$\begin{aligned} B_k^*(i) = &\text{ beneficio máximo obtenido al asignar } i \text{ médicos} \\ &\text{a los paises } P_1,\ P_2,\ ...,\ P_k,\quad k=1,...,n,\ \ i=0,...,m. \end{aligned}$$

Objetivo final:

$$B_n^*(m) = \ ?$$

Condiciones iniciales:

$$B_1^*(i) = b_{i1},\ \ i=0,...,m.$$

Relación recurrente:

$$B_{k+1}^*(i) = \max_{0\le j\le i}\{B_k^*(i-j) + b_{j,k+1}\},\quad k=1,...,n-1,\ \ i=0,...,m.$$

Esta relación recurrente dice que la mejor manera de asignar i médicos a los paises P_1, P_2, ..., P_{k+1} es estudiando todas las posibilidades consistentes en asignar j médicos al país P_{k+1} y el resto, $i-j$, a los paises P_1, P_2, ..., P_k.

12.2.4 Resultados numéricos

Consideremos los siguientes datos: $m=5,\ \ n=4$ y los siguientes beneficios:

i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
0	0	0	0	0
1	2	2	1	4
2	4	3	4	5
3	6	4	7	6
4	8	8	9	7
5	10	12	11	8

Condiciones iniciales:

$$\begin{aligned}
B_1^*(0) &= 0, \;\; j^* = 0,\\
B_1^*(1) &= 2, \;\; j^* = 1,\\
B_1^*(2) &= 4, \;\; j^* = 2,\\
B_1^*(3) &= 6, \;\; j^* = 3,\\
B_1^*(4) &= 8, \;\; j^* = 4,\\
B_1^*(5) &= 10, \;\; j^* = 5.
\end{aligned}$$

$$\begin{aligned}
B_2^*(0) &= \max_{0\leq j\leq 0}\{B_1^*(0-j)+b_{j2}\},\\
&= \max\{B_1^*(0)+b_{02}\},\\
&= \max\{0+0\},\\
B_2^*(0) &= 0, \;\; j^* = 0.
\end{aligned}$$

$$\begin{aligned}
B_2^*(1) &= \max_{0\leq j\leq 1}\{B_1^*(1-j)+b_{j2}\},\\
&= \max\{B_1^*(1)+b_{02}, B_1^*(0)+b_{12}\},\\
&= \max\{\overline{2+0}, 0+2\},\\
B_2^*(1) &= 2, \;\; j^* = 0.
\end{aligned}$$

$$\begin{aligned}
B_2^*(2) &= \max_{0\leq j\leq 2}\{B_1^*(2-j)+b_{j2}\},\\
&= \max\{B_1^*(2)+b_{02}, B_1^*(1)+b_{12}, B_1^*(0)+b_{22}\},\\
&= \max\{\overline{4+0}, 2+2, 0+3\},\\
B_2^*(2) &= 4, \;\; j^* = 0.
\end{aligned}$$

Para los dos primeros paises se tiene:

$$\begin{aligned}
B_2^*(0) &= 0, \;\; j^* = 0,\\
B_2^*(1) &= 2, \;\; j^* = 0,\\
B_2^*(2) &= 4, \;\; j^* = 0,
\end{aligned}$$

$$B_2^*(3) = 6, \;\; j^* = 0,$$
$$B_2^*(4) = 8, \;\; j^* = 0,$$
$$B_2^*(5) = 12, \;\; j^* = 5.$$

Un ejemplo de un cálculo para los tres primeros paises es:

$$\begin{aligned} B_3^*(4) &= \max_{0 \le j \le 4} \{B_2^*(4-j) + b_{j3}\}, \\ &= \max\{B_2^*(4) + b_{03}, B_2^*(3) + b_{13}, B_2^*(2) + b_{23}, \\ &\qquad B_2^*(1) + b_{33}, B_2^*(0) + b_{43}\}, \\ &= \max\{8+0, 6+1, 4+4, \overline{2+7}, 0+9\}, \\ B_3^*(4) &= 9, \;\; j^* = 3. \end{aligned}$$

En la siguiente tabla están los resultados importantes:

i	$B_1^*(i)$	j^*	$B_2^*(i)$	j^*	$B_3^*(i)$	j^*	$B_4^*(i)$	j^*
0	0	0	0	0	0	0		
1	2	1	2	0	2	0		
2	4	2	4	0	4	0		
3	6	3	6	0	7	3		
4	8	4	8	0	9	3		
5	10	5	12	5	12	0	13	1

Se observa que no es necesario calcular $B_4^*(0)$, $B_4^*(1)$, ..., $B_4^*(4)$, ya que no se utilizan para calcular $B_4^*(5)$. El beneficio máximo es 13. Este beneficio máximo se obtiene asignando 1 médico al cuarto país ($j^* = 1$). Entonces quedan 4 médicos para los primeros 3 paises. El j^* correspondiente a $B_3^*(4)$ es 3, esto indica que hay que asignar 3 médicos al tercer país. Entonces queda 1 médico para los primeros 2 paises. El j^* correspondiente a $B_2^*(1)$ es 0, esto indica que hay que asignar 0 médicos al segundo país. Entonces queda 1 médico para el primer país.

$$x_1^* = 1,$$
$$x_2^* = 0,$$
$$x_3^* = 3,$$
$$x_4^* = 1,$$
$$B_4^*(5) = 13.$$

12.2.5 Problema de asignación de médicos con cotas superiores

El planteamiento de este problema es una generalización del anterior, la única diferencia es que para cada país P_j hay una cota superior v_j para el número de médicos que se pueden asignar allí. Obviamente se debe cumplir que $v_j \leq m$. Para garantizar lo anterior, basta con considerar una nueva cota $v'_j = \min\{v_j, m\}$. Entonces los datos para este problema de médicos con cota superior son:

- m número de médicos
- n número de paises
- v_1, v_2, ..., v_n cotas superiores para el número de médicos en cada país
- para cada país P_j los valores de los beneficios $b_{ij} \equiv b(i,j)$: b_{0j}, b_{1j}, ..., $b_{v_j j}$

El planteamiento del problema de optimización es el siguiente: encontrar x_1, x_2, ..., x_n para maximizar el beneficio total con ciertas restricciones:

$$\begin{aligned} \max \sum_{j=1}^{n} b(x_j, j) & \\ \sum_{j=1}^{n} x_j \leq m & \\ 0 \leq x_j \leq v_j, \;\; j = 1, ..., n & \\ x_j \in \mathbb{Z}, \;\; j = 1, ..., n & \end{aligned}$$

Antes de plantear la solución recurrente es necesario considerar lo siguiente: el número de médicos que se asignan al conjunto de paises P_1, P_2, ..., P_k, no puede ser superior a m y tampoco puede ser superior a la suma de sus cotas superiores. Para esto se introducen unos nuevos valores

$$V_k = \min\{m, \sum_{j=1}^{k} v_j\}, \;\; k = 1, ..., n.$$

$B_k^*(i)$ = beneficio máximo obtenido al asignar i médicos

$$\text{a los paises } P_1,\ P_2,\ ...,\ P_k,\quad k=1,...,n,\ \ i=0,...,V_k.$$

Objetivo final:

$$B_n^*(m) = \ ?$$

Condiciones iniciales:

$$B_1^*(i) = b_{i1},\ \ i=0,...,V_1.$$

Relación recurrente:

$$B_{k+1}^*(i) = \max_{\substack{0\le i-j\le V_k \\ i-j\le i \\ 0\le j\le v_{k+1} \\ j\le i}} \{B_k^*(i-j) + b_{j,k+1}\},\quad k=1,...,n-1,\ \ i=0,...,V_{k+1}.$$

Los límites para la variación de $i-j$ y de j han sido presentados de la manera más natural y más segura posible, pero obviamente hay algunas redundancias y admiten simplificaciones.

$$\max\{i-V_k,0\} \le j \le \min\{i,v_k\}$$

Consideremos los siguientes datos: $m=5$, $n=4$, $v_j : 3,4,2,3$ y los siguientes beneficios:

i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
0	0	0	0	0
1	2	2	1	4
2	4	3	4	5
3	6	4	7	6
4		8		

Entonces

$$V_1=3,\ \ V_2=5,\ \ V_3=5,\ \ V_4=5.$$

Condiciones iniciales:

$$B_1^*(0)=0,\ \ j^*=0,$$

$$B_1^*(1) = 2, \;\; j^* = 1,$$
$$B_1^*(2) = 4, \;\; j^* = 2,$$
$$B_1^*(3) = 6, \;\; j^* = 3.$$

$\vdots$

$$\begin{aligned} B_2^*(2) &= \max_{0 \le j \le 2} \{B_1^*(2-j) + b_{j2}\} \\ &= \max\{B_1^*(2) + b_{02}, B_1^*(1) + b_{12}, B_1^*(0) + b_{22}\} \\ &= \max\{\overline{4+0}, 2+2, 0+3\} \\ B_2^*(2) &= 4, \;\; j^* = 0. \end{aligned}$$

$\vdots$

$$\begin{aligned} B_2^*(5) &= \max_{2 \le j \le 4} \{B_1^*(5-j) + b_{j2}\} \\ &= \max\{B_1^*(3) + b_{22}, B_1^*(2) + b_{32}, B_1^*(1) + b_{42}\} \\ &= \max\{6+3, 4+4, \overline{2+8}\} \\ B_2^*(5) &= 10, \;\; j^* = 4. \end{aligned}$$

$\vdots$

i	$B_1^*(i)$	j^*	$B_2^*(i)$	j^*	$B_3^*(i)$	j^*	$B_4^*(i)$	j^*
0	0	0	0	0	0	0		
1	2	1	2	0	2	0		
2	4	2	4	0	4	0		
3	6	3	6	0	6	0		
4			8	1	8	0		
5			10	4	10	0	12	1

El beneficio máximo es 12. Este beneficio máximo se obtiene asignando 1 médico al cuarto país ($j^* = 1$). Entonces quedan 4 médicos para los primeros 3 paises. El j^* correspondiente a $B_3^*(4)$ es 0, esto indica que hay que asignar 0 médicos al tercer país. Entonces quedan 4 médicos para los primeros 2 paises. El j^* correspondiente a $B_2^*(4)$ es 1, esto indica que hay que asignar 1 médico al segundo país. Entonces quedan 3 médicos para el primer país.

$$x_1^* = 3$$
$$x_2^* = 1$$
$$x_3^* = 0$$
$$x_4^* = 1$$
$$B_4^*(5) = 12$$

12.2.6 Problema de asignación de médicos con cotas inferiores y superiores

El planteamiento de este problema es una generalización del anterior, ahora para cada país P_j hay una cota inferior u_j y una cota superior v_j para el número de médicos que se pueden asignar allí. Obviamente se debe cumplir que $0 \leq u_j < v_j \leq m$.
Los datos para este problema son:

- m número de médicos,
- n número de paises,
- u_1, ..., u_n cotas inferiores para el número de médicos,
- v_1, ..., v_n cotas superiores para el número de médicos,
- para cada país P_j los valores de los beneficios $b_{ij} \equiv b(i,j)$: $b_{u_j j}$, $b_{u_j+1,j}$, ..., $b_{v_j j}$.

Para que no haya información redundante, los datos deben cumplir la siguiente propiedad: la cota superior para el número de médicos en el país P_j debe permitir asignar el número mínimo de médicos en los otros paises, es decir:

$$v_j \leq m - \sum_{\substack{i=1 \\ i \neq j}}^{n} u_i.$$

En caso contrario

$$v'_j = \min\{v_j, m - \sum_{\substack{i=1 \\ i \neq j}}^{n} u_i\}.$$

El problema tiene solución si y solamente si el número total de médicos disponibles alcanza para cumplir con las cotas inferiores:

$$\sum_{j=1}^{n} u_j \leq m.$$

El planteamiento del problema de optimización es el siguiente: encontrar x_1, x_2, ..., x_n para maximizar el beneficio total con ciertas restricciones:

$$\begin{aligned} \max \sum_{j=1}^{n} b(x_j, j) & \\ \sum_{j=1}^{n} x_j \leq m, & \\ u_j \leq x_j \leq v_j, \;\; j = 1, ..., n, & \\ x_j \in \mathbb{Z}, \;\; j = 1, ..., n. & \end{aligned}$$

Antes de plantear la solución recurrente es necesario considerar lo siguiente: el número de médicos que se asignan al conjunto de paises P_1, P_2, ..., P_k, no puede ser inferior a la suma de sus cotas inferiores. Por otro lado, no puede ser superior a m, tampoco puede ser superior a la suma de sus cotas superiores y debe permitir satisfacer las cotas mínimas para el resto de paises, es decir, para los paises P_{k+1}, P_{k+2}, ..., P_n. Para esto se introducen unos nuevos valores

$$\begin{aligned} U_k &= \sum_{j=1}^{k} u_j \qquad\qquad\qquad\qquad\qquad\qquad , \;\; k = 1, ..., n \\ V_k &= \min\{m, \;\; \sum_{j=1}^{k} v_j, \;\; m - \sum_{j=k+1}^{n} u_j\}, \;\; k = 1, ..., n \\ V_k &= \min\{\sum_{j=1}^{k} v_j, \;\; m - \sum_{j=k+1}^{n} u_j\} \end{aligned}$$

Aquí se sobreentiende que el valor de una sumatoria es cero, cuando el límite inferior es más grande que el límite superior, por ejemplo,

$$\sum_{i=4}^{2} a_i = 0.$$

$B_k^*(i)$ = beneficio máximo obtenido al asignar i médicos
a los paises $P_1,\ P_2,\ ...,\ P_k,\ \ k = 1, ..., n,\ \ i = U_k, ..., V_k$.

Objetivo final:

$$B_n^*(m) = \ ?$$

Condiciones iniciales:

$$B_1^*(i) = b_{i1},\ \ i = U_1, ..., V_1.$$

Relación recurrente:

$$B_{k+1}^*(i) = \max_{\substack{U_k \le i-j \le V_k \\ i-j \le i \\ u_{k+1} \le j \le v_{k+1} \\ j \le i}} \{B_k^*(i-j) + b_{j,k+1}\}, \quad k = 1, ..., n-1,\ \ i = U_{k+1}, ..., V_{k+1}.$$

En resumen, la variación de j en la relación recurrente está dada por:

$$\max\{i - V_k, u_{k+1}\} \le j \le \min\{i - U_k, v_{k+1}\}.$$

Consideremos los siguientes datos: $m = 5,\ \ n = 4,\ \ u_j : 0, 0, 1, 2$, $v_j : 3, 4, 2, 3$ y los siguientes beneficios:

i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
0	0	0		
1	2	2	1	
2	4	3	4	5
3	6	4		6
4		8		

Si se considera el segundo país, para los demás paises en total se necesitan por lo menos 0+1+2=3 médicos, luego en el segundo país no se pueden asignar más de $5 - 3 = 2$ médicos, entonces los valores $b_{32} = 4$ y $b_{42} = 8$ nunca se van a utilizar. En realidad los datos con los cuales se va a trabajar son:

$$u_j : 0, 0, 1, 2, \quad v_j : 2, 2, 2, 3.$$

i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
0	0	0		
1	2	2	1	
2	4	3	4	5
3				6

$$
\begin{aligned}
&U_1 = 0, \;\; V_1 = \min\{2, 5-3\} = 2,\\
&U_2 = 0, \;\; V_2 = \min\{4, 5-3\} = 2,\\
&U_3 = 1, \;\; V_3 = \min\{6, 5-2\} = 3,\\
&U_4 = 3, \;\; V_4 = \min\{9, 5-0\} = 5.
\end{aligned}
$$

Condiciones iniciales:

$$
\begin{aligned}
&B_1^*(0) = 0, \;\; j^* = 0,\\
&B_1^*(1) = 2, \;\; j^* = 1,\\
&B_1^*(2) = 4, \;\; j^* = 2.
\end{aligned}
$$

$$
\begin{aligned}
B_2^*(1) &= \max_{0\le j\le 1}\{B_1^*(1-j) + b_{j2}\}\\
&= \max\{B_1^*(1) + b_{02}, B_1^*(0) + b_{12}\}\\
&= \max\{\overline{2+0}, 0+2\}\\
B_2^*(1) &= 2, \;\; j^* = 0.
\end{aligned}
$$

$\vdots$

$$
\begin{aligned}
B_3^*(3) &= \max_{1\le j\le 2}\{B_2^*(3-j) + b_{j3}\}\\
&= \max\{B_2^*(2) + b_{13}, B_2^*(1) + b_{23}\}\\
&= \max\{4+1, \overline{2+4}\}\\
B_3^*(3) &= 6, \;\; j^* = 2.
\end{aligned}
$$

$\vdots$

i	$B_1^*(i)$	j^*	$B_2^*(i)$	j^*	$B_3^*(i)$	j^*	$B_4^*(i)$	j^*
0	0	0	0	0				
1	2	1	2	0	1	1		
2	4	2	4	0	4	2		
3					6	2		
4								
5							11	2

El beneficio máximo es 11. Este beneficio máximo se obtiene asignando 2 médicos al cuarto país ($j^* = 2$). Entonces quedan 3 médicos para los primeros 3 paises. El j^* correspondiente a $B_3^*(3)$ es 2, esto indica que hay que asignar 2 médicos al tercer país. Entonces queda 1 médico para los primeros 2 paises. El j^* correspondiente a $B_2^*(1)$ es 0, esto indica que hay que asignar 0 médicos al segundo país. Entonces queda 1 médico para el primer país.

$$\begin{aligned} x_1^* &= 1, \\ x_2^* &= 0, \\ x_3^* &= 2, \\ x_4^* &= 2, \\ B_4^*(5) &= 11. \end{aligned}$$

El problema anterior también se puede resolver como un problema con cotas superiores (sin cotas inferiores) considerando unicamente los datos por encima de lo exigido por las cotas mínimimas, es decir, de los m médicos disponibles en realidad hay únicamente

$$m' = m - \sum_{i=1}^{n} u_i$$

médicos para distribuir, pues de todas formas hay que asignar $\sum_{i=1}^{n} u_i$ para satisfacer las cotas mínimas.

La cota máxima para el número de médicos adicionales en el país P_j es naturalmente

$$v_j' = v_j - u_j, \;\; j = 1, ..., n.$$

De manera análoga se puede pensar en un beneficio b_{ij}' correspondiente al beneficio adicional al asignar i médicos, por encima de de la cota mínima u_j en el país P_j

$$b'_{ij} = b(i + u_j, j) - b(u_j, j), \quad 1 \leq j \leq n, \;\; 0 \leq i \leq v'_j.$$

Al aplicar estos cambios a los datos del problema se tiene:

$$m' = 5 - 3 = 2, \quad n = 4, \quad v'_j : 2, 2, 1, 1.$$

i	b'_{i1}	b'_{i2}	b'_{i3}	b'_{i4}
0	0	0	0	0
1	2	2	3	1
2	4	3		

La solución de este problema modificado es:

$$\begin{aligned} x_1^{*'} &= 1, \\ x_2^{*'} &= 0, \\ x_3^{*'} &= 1, \\ x_4^{*'} &= 0, \\ B_4^{*'}(2) &= 5. \end{aligned}$$

Entonces

$$\begin{aligned} x_1^* &= 1 + u_1 = 1 + 0 = 1, \\ x_2^* &= 0 + u_2 = 0 + 0 = 0, \\ x_3^* &= 1 + u_3 = 1 + 1 = 2, \\ x_4^* &= 0 + u_4 = 0 + 2 = 2, \\ B_4^*(5) &= 5 + \sum_{i=1}^{n} b(u_i, i) = 5 + 6 = 11. \end{aligned}$$

Como observación final sobre **estos** datos numéricos, se puede ver que para cada uno de los tres problemas hay varias soluciones, en particular para este último problema hay otra solución:

$$\begin{aligned} x_1^* &= 0, \\ x_2^* &= 1, \end{aligned}$$

$$x_3^* = 2,$$
$$x_4^* = 2,$$
$$B_4^*(5) = 11.$$

12.3 EL PROBLEMA DEL MORRAL (KNAPSACK)

12.3.1 Enunciado del problema

Un montañista está planeando una excursión muy especial. Evaluando la capacidad de su morral, la dificultad de la excursión, algunos implementos indispensables y sus fuerzas, cree que tiene en su morral una capacidad de $C \in \mathbb{Z}$ kilos (u otra unidad de peso, o de manera más precisa, de masa) disponibles para alimentos. De acuerdo con su experiencia, sus necesidades y sus gustos ha escogido n tipos de alimentos A_1, A_2, ..., A_n, todos más o menos equilibrados. Estos alimentos vienen en paquetes indivisibles (por ejemplo en lata) y $p_i \in \mathbb{Z}$ indica el peso de cada paquete del alimento A_i. Teniendo en cuenta la composición de cada alimento, las calorías, las vitaminas, los minerales, el sabor, el contenido de agua, etc., el montañista asignó a cada paquete del alimento A_i un beneficio global b_i.

El montañista desea saber la cantidad de paquetes de cada alimento que debe llevar en su morral, de tal manera que maximice su beneficio, sin sobrepasar la capacidad destinada para alimentos.

En este problema se supone que no es obligación llevar paquetes de cada uno de los alimentos. También se supone que no hay cotas inferiores ni superiores para el número de paquetes de cada alimento.

Tal vez ningún montañista ha tratado de resolver este problema para organizar su morral, seguramente ni siquiera ha tratado de plantearlo. Lo que si es cierto es que hay muchos problemas, de gran tamaño y de mucha importancia, que tienen una estructura análoga. Hay libros y muchos artículos sobre este problema.

12.3.2 Planteamiento del problema de optimización

Si x_j indica el número de paquetes del alimento A_j que el montañista debe llevar en su morral, entonces se debe maximizar el beneficio, bajo ciertas restricciones:

$$\max \sum_{j=1}^{n} b_j x_j$$
$$\sum_{j=1}^{n} p_j x_j \leq C.$$
$$x_j \in \mathbb{Z}, \;\; j = 1, ..., n.$$

Este problema se puede resolver por la fuerza bruta construyendo todas las combinaciones, haciendo variar x_j entre 0 y $\lfloor C/p_j \rfloor$, verificando si cada combinación es factible ($\sum_{j=1}^{n} p_j x_j \leq C$) y escogiendo la mejor entre las factibles. El significado de $\lfloor\, t \,\rfloor$ es simplemente la parte entera inferior, o parte entera usual, es decir, el mayor entero menor o igual a t. Es claro que para valores grandes de n el número de combinaciones puede ser inmanejable.

La función objetivo (la función que hay que maximizar) es lineal, la restricción también es lineal, las variables deben ser enteras y se puede suponer que los coeficientes b_j y p_j también son enteros. Entonces este problema también se puede resolver por métodos de programación entera (programación lineal con variables enteras).

12.3.3 Solución recurrente

Para conocer optimamente el número de paquetes de cada uno de los n alimentos, se divide el morral con capacidad $i = C$ en dos "submorrales", uno con capacidad j utilizado únicamente para el alimento A_n y otro submorral con capacidad $i - j$ utilizado para los alimentos A_1, A_2,, ..., A_{n-1}. Si se conoce la solución optima de un morral con capacidad $c = 0, ..., C$ para los alimentos A_1, A_2,, ..., A_{n-1} entonces basta con estudiar todas las posibilidades de variación de j y escoger la mejor para obtener la respuesta global. Las posibles combinaciones son:

- C kilos para A_1,...,A_{n-1}, 0 kilos para A_n
- $C - 1$ kilos para A_1,...,A_{n-1}, 1 kilo para A_n

  $\vdots$

- 0 kilos para A_1,...,A_{n-1}, C kilos para A_n

El razonamiento anterior se puede aplicar para los primeros $n-1$ alimentos con un morral de i kilos de capacidad, y así sucesivamente. Precisando más:

$$B_k^*(i) = \text{beneficio máximo utilizando los primeros } k \text{ alimentos en un morral con capacidad de } i \text{ kilos, } 1 \leq k \leq n,\ 0 \leq i \leq C.$$

Objetivo final:

$$B_n^*(C) = \ ?$$

Condiciones iniciales:

$$B_1^*(i) = b_1 \left\lfloor \frac{i}{p_1} \right\rfloor,\ \ 0 \leq i \leq C$$

Relación recurrente:

$$B_{k+1}^*(i) = \max_{0 \leq j \leq i} \left\{ B_k^*(i-j) + b_{k+1} \left\lfloor \frac{j}{p_{k+1}} \right\rfloor \right\},\ \ 1 \leq k \leq n-1,\ \ 0 \leq i \leq C.$$

Esta relación recurrente dice que la mejor manera de conocer el número de paquetes de los alimentos A_1, A_2, ..., A_{k+1} en un morral con una capacidad de i kilos es estudiando todas las posibilidades consistentes en dejar j kilos para el alimento A_{k+1} y el resto, $i-j$ kilos, para los alimentos A_1, A_2, ..., A_k.

12.3.4 Resultados numéricos

Consideremos los siguientes datos: $C = 11,\ \ n = 4,\quad p_i : 4,\ 3,\ 2,\ 5,\quad b_i : 14, 10, 6, 17.8.$

$$\begin{aligned}
B_1^*(0) &= 0,\ \ j^* = 0,\\
B_1^*(1) &= 0,\ \ j^* = 1,\\
B_1^*(2) &= 0,\ \ j^* = 2,\\
B_1^*(3) &= 0,\ \ j^* = 3,\\
B_1^*(4) &= 14,\ \ j^* = 4,\\
B_1^*(5) &= 14,\ \ j^* = 5,\\
B_1^*(6) &= 14,\ \ j^* = 6,
\end{aligned}$$

i	$B_1^*(i)$	j^*	$B_2^*(i)$	j^*	$B_3^*(i)$	j^*	$B_4^*(i)$	j^*
0	0	0	0	0	0	0		
1	0	1	0	0	0	0		
2	0	2	0	0	6	2		
3	0	3	10	3	10	0		
4	14	4	14	0	14	0		
5	14	5	14	0	16	2		
6	14	6	20	6	20	0		
7	14	7	24	3	24	0		
8	28	8	28	0	28	0		
9	28	9	30	9	30	0		
10	28	10	34	6	34	0		
11	28	11	38	3	38	0	38	0

$$
\begin{aligned}
B_1^*(7) &= 14, \;\; j^* = 7,\\
B_1^*(8) &= 28, \;\; j^* = 8,\\
B_1^*(9) &= 28, \;\; j^* = 9,\\
B_1^*(10) &= 28, \;\; j^* = 10,\\
B_1^*(11) &= 28, \;\; j^* = 11.
\end{aligned}
$$

$\vdots$

$$
\begin{aligned}
B_2^*(3) &= \max_{1\le j\le 3}\left\{B_1^*(3-j) + b_2\left\lfloor\frac{j}{p_2}\right\rfloor\right\},\\
&= \max_{1\le j\le 3}\left\{B_1^*(3-j) + 10\left\lfloor\frac{j}{3}\right\rfloor\right\},\\
&= \max\{0+0, 0+0, 0+0, \overline{0+10}\},\\
B_2^*(3) &= 10, \quad j^* = 3.
\end{aligned}
$$

En la tabla está el resumen de los resultados. El beneficio máximo es 38. Este beneficio máximo se obtiene asignando 0 kilos al cuarto alimento ($j^* = 0$). Entonces quedan 11 kilos para los primeros 3 alimentos. El j^* correspondiente a $B_3^*(11)$ es 0, esto indica que hay que asignar 0 kilos al tercer alimento. Entonces quedan 11 kilos para los primeros 2 alimentos. El j^* correspondiente a $B_2^*(11)$ es 3, esto indica que hay que asignar 3 kilos al segundo alimento, o sea, 1 paquete del segundo alimento. Entonces quedan 8 kilos para el primer alimento, o sea, 2 paquetes.

$$
\begin{aligned}
x_1^* &= 2, \\
x_2^* &= 1, \\
x_3^* &= 0, \\
x_4^* &= 0, \\
B_4^*(11) &= 38.
\end{aligned}
$$

Uno podría pensar que una manera simple de resolver este problema es buscar el alimento con mayor beneficio por kilo, y asignar la mayor cantidad posible de este alimento. Para los datos anteriores los beneficios por kilo son: 3.5, 3.33, 3 y 3.56 . Entonces se deberían llevar dos paquetes del cuarto alimento para un beneficio de 35.6 . Es claro que esta no es la mejor solución.

El problema del morral se puede convertir en uno análogo al problema de los médicos, introduciendo b_{ij} el beneficio obtenido con i kilos dedicados al alimento A_j

$$b_{ij} = b_j \left\lfloor \frac{i}{p_j} \right\rfloor$$

Para los datos anteriores se tendría la tabla

i	b_{i1}	b_{i2}	b_{i3}	b_{i4}
0	0	0	0	0
1	0	0	0	0
2	0	0	6	0
3	0	10	6	0
4	14	10	12	0
5	14	10	12	17.8
6	14	20	18	17.8
7	14	20	18	17.8
8	28	20	24	17.8
9	28	30	24	17.8
10	28	30	30	35.6
11	28	30	30	35.6

y la solución óptima

$$x_1^{*'} = 8,$$

$$x_2^{*'} = 3,$$
$$x_3^{*'} = 0,$$
$$x_4^{*'} = 0,$$
$$B_4^*(11) = 38.$$

Es conveniente recordar que x_i' indica el número de kilos dedicados al producto i, luego $x_1 = 2,\ \ x_2 = 1,\ \ x_3 = 0,\ \ x_4 = 0,\ \ B_4^*(11) = 38$.

Este paso por el problema de los médicos, presenta un inconveniente cuando C tiene un valor grande: es necesario construir una tabla muy grande de datos (los beneficios), cuando en realidad el conjunto de datos es pequeño: n, C, los p_i, los b_i.

El problema del morral también se puede generalizar al caso con cotas inferiores y superiores para el número de paquetes de cada alimento.

12.4 PROBLEMA DE UN SISTEMA ELÉCTRICO

12.4.1 Enunciado del problema

Un sistema eléctrico está compuesto por n partes. Para que el sistema funcione se requiere que cada parte funcione. En cada parte hay que colocar por lo menos una unidad, pero se pueden colocar varias unidades para aumentar la probabilidad de que esa parte funcione. La probabilidad de que todo el sistema funcione es igual al producto de las probabilidades de que cada parte funcione. Los datos del problema son:

- n : número de partes
- v_i : número máximo de unidades que se pueden colocar en la parte i, $1 \leq i \leq n$
- $p_{ij} \equiv p(i,j)$: probabilidad de que la parte i funcione si se colocan j unidades, $1 \leq i \leq n$, $1 \leq j \leq v_i$
- $c_{ij} \equiv c(i,j)$: costo de colocar j unidades en la parte i del sistema, $1 \leq i \leq n$, $1 \leq j \leq v_i$
- $C \in \mathbb{Z}$: cantidad disponible para la construcción del sistema eléctrico

Se desea conocer el número de unidades que hay que colocar en cada parte de manera que se maximice la probabilidad de que todo el sistema funcione si se dispone de un capital de C pesos para la fabricación.

Se supone, lo cual es totalmente acorde con la realidad, que p_{ij}, c_{ij} son crecientes con respecto a j, es decir, si el número de unidades aumenta entonces ni el costo ni la probabilidad pueden disminuir, o sea, $p_{ij} \leq p_{i,j+1}$, $c_{ij} \leq c_{i,j+1}$ para $1 \leq j \leq v_i - 1$. También se supone que $1 \leq v_i$. Además el dinero disponible C debe alcanzar para colocar en cada parte el número máximo de unidades posible, o sea, $c(i, v_i) \leq C$ Si esto no es así, se puede modificar el valor v_i de la siguiente manera:

$$v_i' = j_i \quad \text{donde} \quad c(i, j_i) = \max_{1 \leq j \leq v_i} \{c_{ij} \mid c_{ij} \leq C\}.$$

Además se debería cumplir

$$c(i, v_i) \leq C - \sum_{\substack{j=1 \\ j \neq i}}^{n} c(j, 1), \ \ \forall i.$$

El problema tiene solución si y solamente si

$$\sum_{i=1}^{n} c_{i1} \leq C$$

12.4.2 Planteamiento del problema de optimización

Sea x_i el número de unidades colocadas en la parte i

$$\begin{aligned} \max f(x_1, x_2, ..., x_n) &= \prod_{i=1}^{n} p(i, x_i) \\ \sum_{i=1}^{n} c(i, x_i) &\leq C, \\ 1 \leq x_i &\leq v_i, \quad i = 1, ..., n, \\ x_i &\in \mathbb{Z}, \quad i = 1, ..., n. \end{aligned}$$

Este problema se puede resolver con fuerza bruta, construyendo todas las posibilidades realizables y escogiendo la que maximice la probabilidad de que todo el sistema funcione.

12.4.3 Solución recurrente

La idea recurrente es la misma de los problemas anteriores, el problema se puede resolver dinámicamente, considerando inicialmente las mejores políticas para la primera parte, luego para la primera y segunda partes, en seguida para la primera, segunda y tercera partes y así sucesivamente hasta considerar todo el sistema eléctrico.

La cantidad de dinero que se gasta en las primeras k partes tiene que ser suficiente para colocar por lo menos una unidad en cada una de las primeras k partes, además debe permitir que con el resto se pueda colocar por lo menos una unidad en las restantes $n-k$ partes y no debe ser superior a la cantidad necesaria para colocar el número máximo de unidades en cada una de las primeras k partes. Entonces los límites de variación serán:

$$C_k = \sum_{i=1}^{k} c_{i1} \qquad 1 \leq k \leq n$$

$$D_k = \min \left\{ \sum_{i=1}^{k} c(i, v_i), C - \sum_{i=k+1}^{n} c_{i1} \right\}$$

$P_k^*(t) =$ probabilidad máxima de que el subsistema formado por las primeras k partes funcione, si se gastan t pesos en su construcción, donde $1 \leq k \leq n,\ C_k \leq t \leq D_k$.

Objetivo final:

$$P_n^*(C) = \ ?$$

Condiciones iniciales:

$$P_1^*(t) = \max_{1 \leq j \leq v_1} \{p_{1j} \mid c_{1j} \leq t\},\ C_1 \leq t \leq D_1$$

Relación recurrente: Si se dispone de t pesos para el subsistema formado por las partes 1, 2, ..., k, $k+1$, entonces se puede disponer de s pesos para las primeras k partes y el resto, $t-s$ pesos, para la parte $k+1$. Haciendo variar s se escoge la mejor posibilidad. Si $C_{k+1} \leq t \leq D_{k+1}$ entonces la relación recurrente es:

$$P_{k+1}^*(t) = \max\{P_k^*(t-s)(\max_{1 \leq j \leq v_{k+1}} \{p_{k+1,j} \mid c_{k+1,j} \leq s\})\}.$$

$0 \le s \le t$

$0 \le t-s \le t$

$C_k \le t-s \le D_k$

$c(k+1,1) \le s \le c(k+1, v_{k+1})$

En esta fórmula resultan 4 cotas inferiores y 4 cotas superiores para s. Algunas resultan redundantes. Sin embargo, ante la duda, es preferible escribir dos veces la misma cosa, que olvidar alguna restricción importante. Entonces s varía entre la mayor cota inferior y la menor cota superior.

El planteamiento puede resultar más claro si se define π_{it} como la máxima probabilidad de que la parte i funcione si se dispone de t pesos para su construcción,

$$\pi(i,t) = \pi_{it} := \max_{1 \le j \le v_i} \{p_{ij} \mid c_{ij} \le t\}, \;\; 1 \le i \le n, \;\; c(i,1) \le t \le c(i, v_i).$$

Basados en esta función π, la condiciones iniciales y la relación de recurrencia son:

$$\begin{aligned} P_1^*(t) &= \pi_{1t}, \quad C_1 \le t \le D_1, \\ P_{k+1}^*(t) &= \max\{P_k^*(t-s)\pi(k+1,s)\}, \;\; C_{k+1} \le t \le D_{k+1}. \end{aligned}$$

$t-D_k \le s \le t-C_k$

$c(k+1,1) \le s \le c(k+1, v_{k+1})$

12.4.4 Resultados numéricos

Consideremos los siguientes datos: $C = 10$, $n = 4$, v_i : 3, 3, 2, 3

$j \to$	1	2	3
p_{1j}	0.4	0.6	0.8
p_{2j}	0.5	0.6	0.7
p_{3j}	0.7	0.8	
p_{4j}	0.4	0.6	0.8

$j \to$	1	2	3
c_{1j}	2	3	5
c_{2j}	2	4	5
c_{3j}	2	5	
c_{4j}	1	3	6

entonces:

$$\begin{aligned} C_1 &= 2, \quad D_1 = 5, \\ C_2 &= 4, \quad D_2 = 7, \\ C_3 &= 6, \quad D_3 = 9, \\ C_4 &= 7, \quad D_4 = 10. \end{aligned}$$

Condiciones iniciales: con 2 pesos se puede colocar una unidad en la primera parte y la probabilidad de que la primera parte funcione es 0.4, con 3 pesos se pueden colocar dos unidades en la primera parte y la probabilidad de que esta parte funcione es 0.6, con 4 pesos se pueden colocar únicamente dos unidades en la primera parte y la probabilidad sigue siendo 0.6, con 5 pesos se pueden colocar tres unidades en la primera parte y la probabilidad de que funcione la primera parte pasa a ser 0.6.

$$\begin{aligned} P_1^*(2) &= \pi_{12} = 0.4, \quad s^* = 2, \\ P_1^*(3) &= \pi_{13} = 0.6, \quad s^* = 3, \\ P_1^*(4) &= \pi_{14} = 0.6, \quad s^* = 4, \\ P_1^*(5) &= \pi_{15} = 0.8, \quad s^* = 5. \end{aligned}$$

$\vdots$

Para el cálculo de $P_2^*(6)$ la variación de s está dada por $6 - D_1 \leq s \leq 6 - C_1$, $c_{21} \leq s \leq c(2, v_2)$, o sea, $6 - 5 \leq s \leq 6 - 2,\ \ 2 \leq s \leq 5$,

$$\begin{aligned} P_2^*(6) &= \max_{2 \leq s \leq 4} \{P_1^*(6-s)\pi_{2s}\} \\ &= \max\{P_1^*(4)\pi_{22},\ P_1^*(3)\pi_{23},\ P_1^*(2)\pi_{24}\} \\ &= \max\{0.6 \times 0.5,\ 0.6 \times 0.5,\ 0.4 \times 0.6\} \\ P_2^*(6) &= 0.3, \quad s^* = 2. \end{aligned}$$

$\vdots$

La tabla con el resumen de los reultados es la siguiente:

i	$P_1^*(i)$	s^*	$P_2^*(i)$	s^*	$P_3^*(i)$	s^*	$P_4^*(i)$	s^*
2	.4	2						
3	.6	3						
4	.6	4	.2	2				
5	.8	5	.3	2				
6			.3	2	.14	2		
7			.4	2	.21	2		
8					.21	2		
9					.28	2		
10							.126	3

Con 10 pesos disponibles, la probabilidad máxima de que el sistema funcione es 0.126, asignando 3 pesos para la cuarta parte, o sea, con 2 unidades en la cuarta parte. Quedan 7 pesos y el s^* correspondiente a $P_3^*(7)$ es 2, luego con 2 pesos se coloca una unidad en la tercera parte. Quedan 5 pesos y el s^* correspondiente a $P_2^*(5)$ es 2, luego con 2 pesos se coloca una unidad en la segunda parte. Quedan 3 pesos y el s^* correspondiente a $P_1^*(5)$ es 3, luego con 3 pesos se colocan dos unidades en la primera parte.

$$
\begin{aligned}
x_1^* &= 2, \\
x_2^* &= 1, \\
x_3^* &= 1, \\
x_4^* &= 2, \\
P_4^*(10) &= 0.126\,.
\end{aligned}
$$

12.5 PROBLEMA DE MANTENIMIENTO Y CAMBIO DE EQUIPO POR UNO NUEVO

12.5.1 Enunciado del problema

Hoy 31 de diciembre del año 0, el señor Tuta y su socio el señor Rodríguez, propietarios de un bus de e_0 años de edad, desean planificar su política de mantenimiento y compra de su equipo de trabajo, es decir de su bus, durante n años. La decisión de seguir con el mismo equipo o comprar uno nuevo se toma una vez al año, cada primero de enero. El señor Tuta tiene mucha experiencia y puede evaluar de manera bastante precisa los siguientes valores.

- u : vida útil del equipo, en años,
- p_i : precio de un equipo nuevo al empezar el año i, $1 \leq i \leq n$,
- $m_{ij} = m(i,j)$: precio del mantenimiento durante el año i, desde el 2 de enero hasta el 31 de diciembre, de un equipo que al empezar ese año (el 2 de enero), tiene j años de edad, $1 \leq i \leq n$, $0 \leq j \leq u-1$,
- $v_{ij} = v(i,j)$: precio de venta, el 1 de enero del año i, de un equipo que en esta fecha tiene j años de edad, $1 \leq i \leq n+1$, $1 \leq j \leq u$.

La decisión de comprar un equipo nuevo o guardar el que se tiene, se toma y se lleva a cabo el 1 de enero de cada año. Al final de los n años la compañía vende el equipo que tenga.

12.5.2 Solución recurrente

Sea E_k el conjunto de valores correspondientes a la edad que puede tener el equipo al finalizar el año k. Si $e_0 < u$ entonces la edad que puede tener el equipo al finalizar el primer año es 1 (si se compra uno nuevo) o $e_0 + 1$ (si se continua con el mismo equipo durante el primer año), entonces $E_0 = \{1, e_0 + 1\}$. Si $e_0 = u$ entonces la edad del equipo al finalizar el primer año es necesariamente 1, luego $E_0 = \{1\}$. De manera análoga, dependiendo de e_0 las edades al finalizar el segundo año pueden ser: $E_2 = \{1, 2, e_0 + 2\}$ o $E_2 = \{1, 2\}$. En general

$$\begin{aligned} E_0 &= \{e_0\}, \\ E_{k+1} &= \{\,1\,\} \cup \{j+1 : j \in E_k,\ j < u\}. \end{aligned}$$

Sea

$C_k^*(e) =$ costo mínimo de la política de compra y mantenimiento del equipo desde el 31 de diciembre del año 0 hasta el 31 de diciembre del año k, de tal manera que el 31 de diciembre del año k el equipo tiene e años de edad, $1 \leq k \leq n$, $e \in E_k$.

Si el 31 de diciembre del año n el equipo tiene e años entonces hay que vender el 1 de enero del año $n+1$ a un precio $v(n+1, e)$. Entonces se debe escoger la mejor opción entre las posibles. O sea, el objetivo final es conocer el valor:

$$\min_{e\in E_n}\{C_n^*(e) - v_{n+1,e}\} = ?$$

Condiciones iniciales: Si al empezar el primer año se compra un bus nuevo, bien sea porque $e_0 = u$ o bien sea porque se tomó esa decisión, se recupera el dinero de la venta, se compra un bus nuevo y durante el primer año se gasta en el mantenimiento de un bus con 0 años, así al final del primer año el bus tiene 1 año. Si no se compra bus entonces el único gasto es el manteniemiento de un bus de e_0 años, y al final del primer año el bus tiene $e_0 + 1$ años.

$$C_1^*(e) = \begin{cases} m(1, e_0) & \text{si } e = e_0 + 1, \\ -v(1, e_0) + p_1 + m_{1,0} & \text{si } e = 1. \end{cases}$$

Relación de recurrencia: Si e indica la edad del bus al final del año $k+1$, entonces $e = 1$ o e toma otros valores en E_{k+1}. Si $e > 1$ entonces al costo de la política óptima de los primeros k años se le aumenta el costo del mantenimiento de un bus con $e - 1$ años. Si $e = 1$ entonces al final del año k el bus podía tener d años, luego para cada edad d se toma el valor $C_k^*(d)$, se le resta lo de la venta, se le suma el valor de la compra y se le agrega el costo del mantemiento, y finalmente se escoge el menor valor.

$$C_{k+1}^*(e) = \begin{cases} \min\limits_{d\in E_k}\{C_k^*(d) - v_{k+1,d}\} + p_{k+1} + m_{k+1,0} & \text{si } e = 1, \\ C_k^*(e-1) + m_{k+1,e-1} & \text{si } e > 1. \end{cases}$$
$$k = 1, ..., n-1, \;\; e \in E_{k+1}.$$

12.5.3 Resultados numéricos

Consideremos los siguientes datos: $n = 4$, $u = 4$, $e_0 = 2$, p_i : 10, 12, 14, 15.

i	m_{i0}	m_{i1}	m_{i2}	m_{i3}
1	1	3	4	6
2	1	3	4	5
3	1	2	4	6
4	2	3	4	5

i	v_{i1}	v_{i2}	v_{i3}	v_{i4}
1	6	4	3	2
2	6	4	3	1
3	7	5	3	2
4	7	4	3	2
5	6	4	2	0

Entonces:

$$\begin{aligned} E_1 &= \{1, 3\}, \\ E_2 &= \{1, 2, 4\}, \\ E_3 &= \{1, 2, 3\}, \\ E_4 &= \{1, 2, 3, 4\}. \end{aligned}$$

Condiciones iniciales:

$$\begin{aligned} C_1^*(1) &= -4 + 10 + 1 = 7, \\ C_1^*(3) &= 4. \end{aligned}$$

$$\begin{aligned} C_2^*(1) &= \min_{d \in E_1} \{C_1^*(d) - v_{2d}\} + p_2 + m_{20} \\ &= \min\{C_1^*(1) - v_{21}, C_1^*(3) - v_{23}\} + p_2 + m_{20} \\ &= \min\{\overline{7 - 6}, 4 - 3\} + 12 + 1 \\ C_2^*(1) &= 14, \;\; d^* = 1. \end{aligned}$$

$$\begin{aligned} C_2^*(4) &= C_1^*(3) + m_{23} \\ &= 4 + 5 = 9, \;\; d^* = 3. \end{aligned}$$

$\vdots$

La tabla con el resumen de los reultados es la siguiente:

e	$C_1^*(e)$	$s*$	$C_2^*(e)$	s^*	$C_3^*(e)$	s^*	$C_4^*(e)$	s^*
1	7	2	14	1	20	2	28	3
2			10	1	16	1	23	1
3	4	2			14	2	20	2
4			9	3			19	3

Ahora es necesario encontrar $\min_e\{C_4^*(e) - v_{5e}\}$, es decir, el mínimo de $28-6$, $23-4$, $20-2$, $19-0$. Este valor mínimo es 18 y se obtiene para $e = 3$. Si e_k indica la edad del bus al finalizar el año k, entonces $e_4 = 3$, $e_3 = 2$, $e_2 = 1$. Al mirar en la tabla, para $C_2^*(1)$ se tiene que $d^* = 1$, o sea, $e_1 = 1$.

Si x_k indica la decisión tomada al empezar el año k, con la convención

$x_k = 0$: se compra un bus nuevo,

$x_k = 1$: se mantiene el bus que se tiene,

entonces se puede decir que $e_k = 1 \Rightarrow x_k = 0$ y que $e_k > 1 \Rightarrow x_k = 1$.

$$x_1^* = 0,$$
$$x_2^* = 0,$$
$$x_3^* = 1,$$
$$x_4^* = 1.$$

12.6 PROBLEMA DE PRODUCCION Y ALMACENAMIENTO

12.6.1 Enunciado del problema

Considere una compañía que fabrica un bien no perecedero. Esta compañía estimó de manera bastante precisa las demandas d_1, d_2, ..., d_n de los n períodos siguientes. La producción en el período i, denotada por p_i, puede ser utilizada, en parte para satisfacer la demanda d_i, o en parte puede ser almacenada para satisfacer demandas posteriores. Para facilitar la comprensión del problema, supóngase que la demanda de cada período se satisface en los últimos dias del período. Sea x_i el inventario al final del período $i-1$ despues de satisfacer la demanda d_{i-1}, es decir, el inventario al empezar el período i. El costo $c_i(x_i, p_i)$ de almacenar x_i unidades y producir p_i unidades durante el período i se supone conocido. También se conoce x_1 el inventario inicial y x_{n+1} el inventario deseado al final de los n períodos.

Se desea planear la producción y el almacenamiento de cada período, de manera que permitan cumplir con las demandas previstas y se minimice el costo total de almacenamiento y producción.

Para facilitar el planteamiento se puede suponer que el inventario deseado al final de los n periodos se puede incluir en la demanda del último periodo d_n, o sea, hacer $d_n \leftarrow d_n + x_{n+1}$, $x_{n+1} \leftarrow 0$.

12.6.2 Planteamiento del problema de optimización

Las variables de este problema son: x_2, x_3, ..., x_n, p_1, p_2, ..., p_n. Recordemos que x_1, x_{n+1} son datos del problema. Las variables están relacionadas por las siguientes igualdades

$$\begin{aligned} x_1 + p_1 - d_1 &= x_2 \\ x_2 + p_2 - d_2 = x_1 + p_1 - d_1 + p_2 - d_2 &= x_3 \\ x_3 + p_3 - d_3 = x_1 + p_1 - d_1 + p_2 - d_2 + p_3 - d_3 &= x_4 \end{aligned}$$

En general,

$$x_i + p_i - d_i = \sum_{j=1}^{i} p_j - \sum_{j=1}^{i} d_j + x_1 = x_{i+1}, \;\; i = 1, ..., n. \tag{12.1}$$

Es claro que para $i = n$,

$$x_n + p_n - d_n = \sum_{j=1}^{n} p_j - \sum_{j=1}^{n} d_j + x_1 = 0. \tag{12.2}$$

Sea

$$D_i = \sum_{j=1}^{i} d_j - x_1, \quad i = 1, ..., n, \tag{12.3}$$

es decir, la demanda neta acumulada de los primeros i períodos, descontando el inventario inicial. En particular, toda la producción esta dada por $\sum_{j=1}^{n} p_j = D_n$. Entonces las igualdades que relacionan las variables son:

$$\sum_{j=1}^{i-1} p_j - D_{i-1} = x_i, \;\; i = 2, ..., n, \tag{12.4}$$

$$\sum_{j=1}^{n} p_j - D_n = 0. \tag{12.5}$$

El problema de optimización es entonces:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n} c_i(x_i, p_i) \\
& \sum_{j=1}^{i-1} p_j - D_{i-1} = x_i, \quad i = 2, \ldots, n \\
& \sum_{j=1}^{n} p_j - D_n = 0, \\
& x_2, \ldots, x_n, p_1, \ldots, p_n \geq 0, \\
& x_2, \ldots, x_n, p_1, \ldots, p_n \in \mathbb{Z}.
\end{aligned}$$

El problema se puede plantear únicamente con las variables p_1, ..., p_{n-1}. A partir de (12.2) y (12.4) se tiene

$$\begin{aligned}
-x_n + d_n &= p_n, \\
-\sum_{j=1}^{n-1} p_j + D_{n-1} + d_n &= p_n, \\
D_n - \sum_{j=1}^{n-1} p_j &= p_n.
\end{aligned}$$

La función objetivo se puede agrupar en tres partes:

$$\min \ c_1(x_1, p_1) + \sum_{i=2}^{n-1} c_i(x_i, p_i) + c_n(x_n, p_n)$$

Entonces el problema, expresado únicamente con las variables p_1, ..., p_{n-1}, es :

$$\begin{aligned}
\min \ c_1(x_1, p_1) & + \sum_{i=2}^{n-1} c_i\big(\sum_{j=1}^{i-1} p_j - D_{i-1}, p_i\big) \\
& + c_n\big(\sum_{j=1}^{n-1} p_j - D_{n-1}, D_n - \sum_{j=1}^{n-1} p_j\big) \\
\sum_{j=1}^{i-1} p_j - D_{i-1} \geq 0, & \quad i = 2, ..., n,
\end{aligned}$$

$$D_n - \sum_{j=1}^{n-1} p_j \geq 0,$$
$$p_i \geq 0,\ p_i \in \mathbb{Z},\quad i = 1, ..., n-1.$$

Para resolver este problema por la fuerza bruta, estudiando todas las posibilidades, basta con considerar que el mayor valor de p_i se tiene cuando esta producción satisface por sí sola toda la demanda de los períodos i, $i+1$, ..., n. En ese caso no habría producción en los períodos $i+1$, ..., n. Dicho de otra forma, la producción en el período i no debe ser mayor que la demanda acumulada de los períodos i, ..., n.

Sea

$$E_i = \sum_{j=i}^{n} d_j,\quad i = 1, ..., n,$$

entonces los siguientes son límites para la variación de las variables p_i

$$0 \leq p_i \leq E_i,\quad i = 1, ..., n-1$$

Obviamente los valores p_1, ..., p_{n-1} deben cumplir las otras restricciones. Igualmente es claro que para las variables x_i también se tiene la misma restricción, o sea, ni el inventario al empezar el período i, ni la producción durante el período i pueden ser superiores a E_i

De la definición de D_i y E_i se deduce inmediatamente que

$$D_i + E_{i+1} = D_n,\quad i = 1, ..., n-1$$

12.6.3 Solución recurrente

Sea

$$C_k^*(q) = \text{costo mínimo de producir en total } q \text{ unidades durante los primeros } k \text{ períodos tal manera que se satisfagan las demandas de estos períodos, } 1 \leq k \leq n.$$

La producción acumulada de los primeros k períodos debe ser suficiente para satisfacer la demanda total de estos k períodos (descontando x_1) y no debe sobrepasar la demanda total de los n períodos, entonces en la definición de $C_k^*(q)$ la variación de q está dada por

$$D_k \leq q \leq D_n.$$

Objetivo final:

$$C_n^*(D_n) = ?$$

Condiciones iniciales:

$$C_1^*(q) = c_1(x_1, q), \;\; D_1 \leq q \leq D_n.$$

Si en los primeros $k+1$ períodos la producción es de q unidades y la producción en el período $k+1$ es de y unidades, entonces la producción en los primeros k períodos es de $q-y$ unidades y el inventario durante el período $k+1$ es $q-y-D_k$.

Relación recurrente:

$$C_{k+1}^*(q) = \min_{\substack{0 \leq q-y \leq q \\ D_k \leq q-y \leq D_n \\ 0 \leq q-y-D_k \leq E_{k+1} \\ 0 \leq y \leq E_{k+1}}} \{C_k^*(q-y) + c_{k+1}(q-y-D_k, y)\},$$

definida para $k = 1, \ldots, n-1, \;\; D_{k+1} \leq q \leq D_n$.

De las anteriores cotas inferiores y superiores para y y para $q-y$ se deduce fácilmente que $0 \leq y \leq q - D_k$. En resumen

$$C_{k+1}^*(q) = \min_{0 \leq y \leq q-D_k} \{C_k^*(q-y) + c_{k+1}(q-y-D_k, y)\}.$$

12.6.4 Resultados numéricos

Consideremos los siguientes datos: $n = 4, \;\; d_i = 2, 3, 2, 5, \;\; c_i(x_i, p_i) = hx_i + \gamma_i p_i, \;\; h = 2, \;\; \gamma_i = 1, 4, 3, 6, \;\; x_1 = 1, \;\; x_5 = 0.$

Entonces

$$D_1 = 1, \;\; D_2 = 4, \;\; D_3 = 6, \;\; D_4 = 11.$$

$$\begin{aligned} C_1^*(q) &= c_1(x_1, q) \\ &= 2x_1 + 1q \\ &= 2 + q, \quad 1 \leq q \leq 11. \end{aligned}$$

$$\begin{aligned} C_2^*(q) &= \min_{0 \leq y \leq q-1} \{C_1^*(q-y) + c_2(q-y-, y)\} \\ &= \min_{0 \leq y \leq q-1} \{2 + (q-y) + 2(q-y-1) + 4y\} \\ &= \min_{0 \leq y \leq q-1} \{3q + y\}, \quad y^* = 0 \\ &= 3q, \quad 4 \leq q \leq 11. \end{aligned}$$

$$\begin{aligned} C_3^*(q) &= \min_{0 \leq y \leq q-4} \{C_2^*(q-y) + c_3(q-y-4, y)\} \\ &= \min_{0 \leq y \leq q-4} \{3(q-y) + 2(q-y-4) + 3y\} \\ &= \min_{0 \leq y \leq q-4} \{5q - 2y - 8\}, \quad y^* = q - 4 \\ &= 3q, \quad 6 \leq q \leq 11. \end{aligned}$$

$$\begin{aligned} C_4^*(11) &= \min_{0 \leq y \leq 11-6} \{C_3^*(11-y) + c_4(11-y-6, y)\} \\ &= \min_{0 \leq y \leq 5} \{3(11-y) + 2(5-y) + 6y\} \\ &= \min_{0 \leq y \leq 5} \{43 + y\}, \quad y^* = 0 \\ &= 43. \end{aligned}$$

La producción óptima en el cuarto período es 0, luego la producción en los tres primeros períodos es 11. La producción óptima en el tercer período es $q - 4 = 11 - 4 = 7$, luego la producción en los dos primeros períodos es 4. La producción óptima en el segundo período es 0, entonces la producción óptima para el primer período es 4

$$p_4^* = 0,$$

$$
\begin{aligned}
p_3^* &= 7, \\
p_2^* &= 0, \\
p_1^* &= 4, \\
C_4^*(11) &= 43.
\end{aligned}
$$

EJERCICIOS

12.1 Una corporación tiene n plantas de producción. De cada una de ellas recibe propuestas para una posible expansión de las instalaciones. La corporación tiene un presupuesto de D mil millones de pesos para asignarlo a las n plantas. Cada planta P_i envía sus propuestas, indicando n_i el número de propuestas, $c_i(j)$ el costo de la propuesta p_{ij} y $g_i(j)$ la ganancia adicional total acumulada al cabo de m años. Obviamente en cada planta se lleva a cabo una sola propuesta. Una propuesta válida para una planta consiste en no invertir en expansión, siendo su costo y ganancia nulos. Más aún, se podría pensar que en este caso la ganancia podría ser negativa.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

Resuelva el problema para los siguientes valores numéricos: $n = 3$, $D = 5$,

	Planta 1		Planta 2		Planta 3	
	$c_1(j)$	$g_1(j)$	$c_2(j)$	$g_2(j)$	$c_3(j)$	$g_3(j)$
Prop. 1	0	0	0	0	0	0
Prop. 2	2	8	1	5	1	3
Prop. 3	3	9	2	6		
Prop. 4	4	12				

12.2 El proceso de manufactura de un bien perecedero es tal que el costo de cambiar el nivel de producción de un mes al siguiente es $\$a$ veces el cuadrado de la diferencia de los niveles de producción. Cualquier cantidad del producto que no se haya vendido al final del mes se desperdicia con un costo de $\$b$ por unidad. Si se conoce el pronóstico de ventas d_1 , d_2 , ..., d_n para los próximos n meses y se sabe que en el último mes (el mes pasado) la producción fue de x_0 unidades.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

Resuelva el problema para los siguientes valores numéricos: $n = 4$, $a = 1$, $b = 2$, $d_i = 42$, 44, 39,36, $x_0 = 40$.

12.3 Considere el problema del morral con cotas inferiores y superiores con los siguientes datos: C = capacidad en kilos del morral (entero posotivo); n número de alimentos; p_1, p_2, ..., p_n, donde p_i es un entero positivo que indica el peso, en kilos, de un paquete del alimento i; b_1, b_2, ..., b_n, donde b_i indica el beneficio de un paquete del alimento i; u_1, u_2, ..., u_n, donde u_i es un entero no negativo que indica el número mínimo de paquetes del alimento i que el montañista debe llevar; v_1, v_2, ..., v_n, donde v_i es un entero que indica el número máximo de paquetes del alimento i que el montañista debe llevar. Los datos cumplen con la condición $u_i < v_i \ \forall i$.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

12.4 A partir del lunes próximo usted tiene n exámenes finales y $m \geq n$ días para prepararlos. Teniendo en cuenta el tamaño y la dificultad del tema usted a evaluado c_{ij} la posible nota o calificación que usted obtendría si dedica i días para estudiar la materia j. Usted ha estudiado regularmente durante todo el semestre y en todas sus materias tiene buenas notas de tal forma que aún obteniendo malas notas en los exámenes finales usted aprobará todas las materias. En consecuencia su único interés es obtener el mejor promedio posible. De todas maneras piensa dedicar por lo menos un día para cada materia.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

Resuelva el problema para los siguientes valores numéricos: $n = 4$ materias, $m = 7$ días,

c_{0j}	10	10	05	05
c_{1j}	20	15	35	10
c_{2j}	23	25	30	20
c_{3j}	25	30	40	35
c_{4j}	40	35	42	40
c_{5j}	45	40	45	45
c_{6j}	48	48	45	48
c_{7j}	50	48	48	50

12.5 Una empresa agrícola tiene actualmente x_0 empleados y conoce de

manera bastante precisa las necesidades de mano de obra para las siguientes n semanas de cosecha, es decir, conoce los valores e_i, $i = 1, ..., n$, donde e_i es el número de empleados necesarios durante la semana i. También conoce: $c_i(j)$ el precio de contratar j empleados nuevos al empezar la semana i, $i = 1, ..., n$; $d_i(j)$ el costo de despedir j empleados al finalizar la semana i, $i = 0, ..., n$; y $m_i(j)$ el costo de mantener sin trabajo (pero con sueldo) j empleados durante la semana i, $i = 1, ..., n$. Después de las n semanas de cosecha, la empresa únicamente necesita e_{n+1}, un número pequeño de empleados que se quedan trabajando por varias semanas. La empresa desea saber cuantos empleados debe tener durante cada una de las n semanas.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

Resuelva el problema para los siguientes valores numéricos: $n = 4$, $x_0 = 10$, $e_i = 30, 45, 40, 25$, $e_{n+1} = 5$, $c_i(j) = 10 + 3j$, $d_i(j) = 6j$, $m_i(j) = 8j$.

12.6 Un zoocriadero de chigüiros tiene actualmente x_0 animales y tiene capacidad para una gran cantidad de ellos. En un año el número de chiguiros se multiplica por $a > 1$. Al principio de cada año (del año i) el gerente toma la decisión de vender algunos chigüiros al precio unitario p_i, $i = 1, ..., n + 1$. Después de n años se venden todos los chigüiros. El gerente desea saber cuantos chigüiros debe vender al comienzo de cada uno de los n años.

Plantee el problema de optimización. Resuelva el problema por PD: defina una función que permita la recurrencia, dé las condiciones iniciales, la relación de recurrencia.

Resuelva el problema para los siguientes valores numéricos: $n = 4$, $x_0 = 5$, $a = 3$, $p_i = 50, 10, 60, 25, 45$.

12.7 Resuelva por PD el siguiente problema de optimización:

$$\min \quad f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 4)^2 + (x_1 - x_2)^2$$

Sugerencia. Fije una variable y halle la solución (en función de la variable fija). Haga variar la variable que estaba fija.

12.8 Resuelva por PD el siguiente problema de PL:

$$\begin{aligned} \max \quad z = x_1 + 1.4x_2 & \\ x_1 + x_2 &\leq 40 \\ x_1 + 2x_2 &\leq 58 \\ x &\geq 0. \end{aligned}$$

BIBLIOGRAFÍA

[ArGa85] Armitano O., Edelman J., García U., *Programación no lineal*, Limusa, México, 1985.

[Aro86] Arora Jasbir S., *Introduction to Optimum Design*, McGraw-Hill, New York, 1989.

[Atk78] Atkinson Kendall E., *An Introduction to Numerical Analysis*, Wiley, New York, 1978.

[Avr76] Avriel Mordecai, *Nonlinear Programming, Analysis and Methods*, Prentice-Hall, Englewood Cliffs, 1976.

[Bea96] Beasley John E. ed., *Advances in Linear and Integer Programming*, Oxford U. Press, Oxford, 1996.

[Ber95] Bertsekas Dimitri, *Nonlinear Programming*, Athena Scientific, Belmont, 1995.

[BGLS97] Bonnans J. Frédéric, Gilbert Jean Charles, Lemarechal Claude, Sagastizábal Claudia, *Optimisation Numérique, Aspects théoriques et pratiques*, Springer, Berlin, 1997.

[BSS93] Bazaraa M. S., Sherali H. D., Shetty C. M., *Nonlinear Programming, Theory and Algorithms*, 2nd ed., Wiley, New York, 1993.

[Bjo96] Björck Ake, *Numerical Methods for Least Square Problems*, SIAM, Philadelphia, 1996.

[Cia82] Ciarlet P. G., *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, Paris, 1982.

[Dan93] Dantzig G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, 1963

[den94] den Hertog D., *Interior Point Approach to Linear, Quadratic, and Convex Programming, Algorithms and Complexity*, Kluwer, Dordrecht, 1994.

[DeSc83] Dennis J. E. Jr., Schnabel R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*,
Prentice Hall, Englewood Cliffs, 1983.

[Des76] Desbazeille Gérard, *Exercices et problèmes de recherche opérationnelle*, Dunod, Paris, 1976.

[FiMc68] Fiacco Anthony V., McCormick Garth P., *Nonlinear Programming, Sequential Unscontrained Minimization Techniques*, Wiley, New York, 1968, reimpreso por SIAM, Philadelphia, 1990.

[Flm69] Fleming Wendell H., *Funciones de diversas variables*, Centro Regional de Ayuda Técnica AID, México, 1969.

[Fle87] Fletcher R., *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester , 1987.

[GMWl81] Gill Ph., Murray W., Wright M., *Practical Optimization*, Academic Press, London, 1981.

[GoVa96] Golub Gene H., Van Loan Charles F., *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.

[Gri00] Griewank Andreas, *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*, SIAM, Philadelphia, 2000.

[HaHo91] Hammerlin Gunther, Hoffmann Karl-Heinz, *Numerical Mathematics*, Springer-Verlag, New York, 1991.

[HePe83] Helary Jean Michel, Pedrono René, *Recherche opérationnelle, Exercices corrigés*, Hermann, Paris, 1983.

[HiLi74] Hillier Frederick S., Lieberman Gerald J., *Operations Research*, 2nd ed., Holden Day, San Francisco, 1974.

[Kel99] Kelley C.T., *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[KlMi72] Klee V., Minty G. J.,, *How good is the simplex algorithm*, en SHISHA O., *Inequalities III*, Academic Press, New York, 1972.

[LaHa74] Lawson Charles L., Hanson Richard J., *Solving Least Square Problems*, Prentice Hall, Englewood Cliffs, 1974, reimpreso por SIAM, Philadelphia, 1996.

[Lue89] Luenberger David E., *Programación lineal y no lineal*, Addison Wesley Iberoamericana, Wilmington, 1989.

[Man69] Mangasarian Olvi L., *Nonlinear Programming*, Tata McGraw Hill, Bombay, 1969.

[Mar87] Márquez Javier, *Fundamentos de teoría de optimización*, Limusa, México, 1987.

[Min83] Minoux Michel, *Programmation Mathématique, Théorie et algorithmes*, tome 1,2, Dunod, Paris, 1983.

[MoWr93] Moré Jorge J., Wright Stephen J., *Optimization Software Guide*, SIAM, Philadelphia, 1993.

[NaSo96] Nash Stephen G., Sofer Ariela, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.

[NoWr99] Nocedal Jorge, Wright Stephen J., *Numerical Optimization*, Springer, New York, 1999.

[PSU88] Peressini A. L., Sullivan F. E., Uhl J. J. Jr., *The Mathematics of Nonlinear Programming*, Springer-Verlag, New York, 1988.

[Pra81] Prawda Juan, *Métodos y Modelos de Investigación de Operaciones*, Vol. 1, *Modelos Determinísticos*, Limusa, México, 1981.

[Pre86] Press William H. et al., *Numerical Recipes, The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.

[Roc70] Rockafellar R. Tyrrell, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[RTV97] Roos Kees, Terlaky Tamás, Vial Jean-Philippe, *Theory ans Algorithms for Linear Optimization, An Interior Point Approach*, Wiley, Chichester, 1997.

[Ros85] Roseaux (groupe), *Exercices et problèmes résolus de recherche opérationnelle*, Tome 3, *Programmation linéaire et extensions, problèmes classiques*, Masson, Paris, 1985.

[Sim75] Simmons Donald M., *Nonlinear Programming for Operations Research*, Prentice-Hall, Englewood Cliffs, 1975.

[Spe94] Spedicato Emilio, ed., *Algorithms for Continuous Optimization, The State of the Art*, Kluwer, Dordrecht, 1994.

[StBu93] Stoer J., Bulirsch R., *Introduction to Numerical Analysis*, 2nd ed., Springer-Verlag, New York, 1993.

[Ter96] Terlaky Tamás, ed., *Interior Point Methods of Mathematical Programming*, Kluwer, Dordrecht, 1996.

[Tah82] Taha Handy A., *Operations Research*, 3rd ed., Macmillan, New York, 1982.

[Van96] Vanderbei Robert J., *Linear Programming, Foundations and Extensions*, Kluwer, Boston, 1996.

[Var78] Varela Jaime E., *Introducción a la investigación de operaciones*, Fondo Editorial Colombiano, Bogotá, 1978.

[Wri97] Wright Stephen J., *Primal-Dual Interior Point Methods* SIAM, Philadelphia, 1997.

[Ye97] Ye Yinyu, *Interior Point Algorithms, Theory and Analysis*, Wiley, New York, 1997.