

## Capítulo 2

# Optimización no lineal

### 2.1 Introducción

En este tema se desarrollan las condiciones necesarias y suficientes para la solución de un problema de optimización no lineal de tipo general. Tal y como se introdujo un problema de programación no lineal se puede expresar como:

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \\ & h_i(\mathbf{x}) = 0 \quad \forall i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad \forall j = 1, \dots, p \end{array} \quad (2.1)$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$  son funciones reales de varias variables y  $A \subseteq \mathbb{R}^n$  es un conjunto abierto. Habrá entonces  $m$  restricciones de igualdad y  $p$  restricciones de desigualdad.

Si  $\Omega$  es el conjunto factible del problema anterior. Resolver el problema de optimización general es encontrar los óptimos (máximos y/o mínimos) de la función  $f(\mathbf{x})$  no sobre todo el conjunto  $A$  donde está definida sino sobre el conjunto  $\Omega$  de los puntos que cumplen las restricciones.

Si  $m = p = 0$ , entonces no hay restricciones y el problema es de optimización no lineal sin restricciones.

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ & \mathbf{x} \in A \end{array} \quad (2.2)$$

En otro caso ( $n \neq 0$  ó  $p \neq 0$ ), el problema se dice con restricciones y a sus extremos locales o globales, se les califica como *extremos condicionados*, para distinguirlos de los extremos de los problemas sin restricciones.

Por último si  $m \neq 0$  y  $p = 0$ , es decir, solamente hay restricciones de igualdad en el problema, entonces el problema con restricciones es un *problema de Lagrange*.

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \\ & h_i(\mathbf{x}) = 0 \quad \forall i = 1, \dots, m \end{array} \quad (2.3)$$

El comportamiento de las restricciones de igualdad y las de desigualdad en un problema de optimización no lineal es ligeramente distinto.

**Definición 2.1 (Restricciones activas)** Dado el problema de optimización con restricciones descrito en 2.1; diremos que una restricción de desigualdad  $g_j(\mathbf{x}) \leq 0$  es activa o saturada en el punto factible  $\mathbf{x}^* \in \Omega \iff g_j(\mathbf{x}^*) = 0$ ; en caso contrario la restricción es inactiva o no saturada en  $\mathbf{x}^*$ .

Las restricciones activas se comportan como las restricciones de igualdad, que por su propia naturaleza son activas en cualquier punto factible.

**Definición 2.2** Dado el problema de optimización con restricciones descrito en 2.1 y sea  $\mathbf{x}^* \in \Omega$ , un punto factible. Se define el conjunto de actividad asociado a  $\mathbf{x}^*$ ,  $J(\mathbf{x}^*)$  como

$$J(\mathbf{x}^*) = \{j \in \{1, \dots, p\} \mid g_j(\mathbf{x}^*) = 0\}$$

es decir,  $J(\mathbf{x}^*)$  es el conjunto de los índices de las restricciones activas en  $\mathbf{x}^*$ .

Si  $\mathbf{x}^*$  es una solución óptima del problema 2.1, las restricciones que sean no activas en él serán irrelevantes puesto que no se alcanza la limitación impuesta por la restricción. Sería posible eliminar las restricciones no saturadas de la formulación del problema, siempre que fueran conocidas previamente pero esto, en general, no es posible.

**Definición 2.3 (Punto regular)** Dado el problema de optimización con restricciones 2.1. Sea  $\mathbf{x}^* \in \Omega$ , un punto factible. Se dice que  $\mathbf{x}^*$  es regular para las restricciones si la familia de vectores

$$\left\{ \{\nabla h_j(\mathbf{x}^*)\}_{j=1,\dots,m}, \{\nabla g_i(\mathbf{x}^*)\}_{i \in J(\mathbf{x}^*)} \right\}$$

es linealmente independiente.

1. Caso sin restricciones ( $m = p = 0$ ): En un problema sin restricciones, todos los puntos son regulares.
2. Caso con restricciones de desigualdad ( $m = 0, p \neq 0$ ): El punto  $\mathbf{x}^*$  también será regular si no hay ninguna restricción activa en él, es decir,  $\mathbf{x}^*$  es regular si  $J(\mathbf{x}^*) = \emptyset$ .

**Definición 2.4 (Espacio tangente)** Sea  $\mathbf{x}^*$  un punto factible para el problema de optimización 2.1. Definimos el espacio tangente en  $\mathbf{x}^*$  al conjunto de vectores definido como

$$M(\mathbf{x}^*) = \{ \mathbf{d} \in \mathbb{R}^n \mid \nabla^T h_i(\mathbf{x}) \mathbf{d} = 0; \quad \nabla^T g_j(\mathbf{x}^*) \mathbf{d} = 0, \quad i = 1, \dots, m; j \in J(\mathbf{x}^*) \}$$

o utilizando derivadas

$$M(\mathbf{x}^*) = \left\{ \mathbf{d} \in \mathbb{R}^n \mid \sum_{k=1}^n \frac{\partial h_i}{\partial x_k}(\mathbf{x}^*) d_k = 0, \quad i = 1, \dots, m; \sum_{k=1}^n \frac{\partial g_j}{\partial x_k}(\mathbf{x}^*) d_k = 0, \quad j \in J(\mathbf{x}^*) \right\}$$

donde de nuevo  $J(\mathbf{x}^*)$  el conjunto de índices de las restricciones de desigualdad que son activas en  $\mathbf{x}^*$ .

**Caso Particular:** Si  $m = p = 0$ , es decir, si el problema no tiene restricciones, entonces el espacio tangente coincide con el espacio  $\mathbb{R}^n$ .

## 2.2 Condiciones de Karush-Kuhn-Tucker

Para los problemas generales de optimización no lineal es posible obtener condiciones necesarias que deban cumplir sus posibles soluciones óptimas, estas condiciones son las llamadas condiciones de Karush-Kuhn-Tucker (CKKT).

Aunque la teoría es indentica en ambos casos, vamos a distinguir entre las condiciones de Karush-Kuhn-Tucker para un problema con objetivo de minimizar y aquellas que se deben cumplir cuando el objetivo es de maximizar.

### 2.2.1 Minimización

**Definición 2.5** Dado el problema

$$\begin{aligned} &\text{Minimizar} && f(\mathbf{x}) \\ &\text{Sujeto a} && \\ &&& h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ &&& g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{aligned} \tag{2.4}$$

con  $f, h_i, g_j : A \rightarrow \mathbb{R}$  funciones de clase  $\mathcal{C}^1(A)$  y  $A \subseteq \mathbb{R}^n$  un conjunto abierto. Diremos que  $\mathbf{x}^* \in A$  es un punto de Karush-Kuhn-Tucker o que cumple las condiciones de Karush-Kuhn-Tucker de Mínimo (CKKTMín) para el problema si y sólo si  $\exists \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$ , de forma que se cumplen las siguientes condiciones:

1. Condición estacionaria

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\mathbf{x}^*) = 0$$

## 2. Condición de factibilidad

$$\begin{aligned} h_i(\mathbf{x}^*) &= 0, & i &= 1, \dots, m \\ g_j(\mathbf{x}^*) &\leq 0 & j &= 1, \dots, p \end{aligned}$$

## 3. Condición de positividad

$$\mu_j \geq 0, \quad i = 1, \dots, m$$

## 4. Condición de holgura

$$\mu_j g_j(\mathbf{x}^*) = 0 \quad j = 1, \dots, p$$

Los valores de  $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$  son los llamados *multiplicadores de Lagrange* del problema 2.4 en el punto  $\mathbf{x}^*$ .

Los puntos  $\mathbf{x}^* \in A \cap \Omega$  que cumplen la *condición estacionaria* se dice que son *puntos críticos o estacionarios*, que serán *condicionados* cuando en el problema hay restricciones de algún tipo.

## 2.2.2 Maximización

Se obtienen condiciones equivalentes para un problema de maximización de la forma

$$\begin{aligned} &\text{Maximizar } f(x) \\ &\text{Sujeto a} \\ &\quad h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ &\quad g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{aligned}$$

con la idea de que maximizar una función es equivalente a minimizar su opuesta. Es posible comprobar, aplicando la definición de punto de Karush-Kuhn-Tucker al problema

$$\begin{aligned} &\text{Minimizar } g(\mathbf{x}) = -f(x) \\ &\text{Sujeto a} \\ &\quad h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ &\quad g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{aligned}$$

que la única condición que cambia es la condición de positividad. Para máximo, esta condición se transforma es una *condición de negatividad*. Las condiciones y definiciones correspondientes son las siguientes:

**Definición 2.6** A partir del problema de maximización

$$\begin{aligned} &\text{Maximizar } f(\mathbf{x}) \\ &\text{Sujeto a} \\ &\quad h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ &\quad g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{aligned} \tag{2.5}$$

con  $f, h_i, g_j : A \rightarrow \mathbb{R}$ , funciones de clase  $\mathcal{C}^1(A)$  y  $A \subseteq \mathbb{R}^n$  un conjunto abierto. Diremos que  $\mathbf{x}^* \in A$  es un punto de Karush-Kuhn-Tucker para este problema, o que cumple las condiciones de Karush-Kuhn-Tucker de Máximo (CKKTMax) para el problema si y sólo si  $\exists \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$ , de forma que se cumplen las siguientes condiciones:

## 1. Condición estacionaria

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\mathbf{x}^*) = 0$$

## 2. Condición de factibilidad

$$\begin{aligned} h_i(\mathbf{x}^*) &= 0, & i &= 1, \dots, m \\ g_j(\mathbf{x}^*) &\leq 0 & j &= 1, \dots, p \end{aligned}$$

3. Condición de negatividad

$$\mu_j \leq 0 \quad j = 1, \dots, p$$

4. Condición de holgura

$$\mu_j g_j(\mathbf{x}^*) = 0 \quad j = 1, \dots, p$$

### 2.2.3 Casos Particulares

Cuando en el problema no hay restricciones o solamente hay restricciones de igualdad, las condiciones KKT tienen formas particulares que indicamos a continuación.

#### Problemas sin restricciones

Si el problema no tiene restricciones de ningún tipo, entonces  $m = p = 0$ . El planteamiento del problema sería como en 2.2 y los multiplicadores no serán necesarios, ni tampoco las condiciones relacionadas con ellos. La única condición utilizada es la estacionaria

$$\nabla f(\mathbf{x}^*) = 0$$

que es la misma para ambos objetivos de maximizar y minimizar.

Cabe distinguir además el caso  $n = 1$ , es decir, en el caso de la optimización de una función real de variable real, en el que la condición estacionaria nos conduce a la conocida expresión

$$x^* \text{ es un extremo local} \Rightarrow f'(x) = 0$$

#### Problemas de Lagrange

Si el problema sólo tiene restricciones de igualdad, es decir  $p = 0$  y  $m \neq 0$ , el problema considerado es el problema clásico de Lagrange representado en 2.3 y las condiciones de Karush-Kuhn-Tucker se obtienen eliminando aquellos términos que dependen de las restricciones de desigualdad

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) = 0$$

$$h_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m$$

que es el resultado que proporciona el teorema clásico de los *multiplicadores de Lagrange*, siendo estos los valores de los  $\lambda_i$ . Se observa como en el caso anterior que las condiciones de KKT para ambos objetivos de maximizar y minimizar, coinciden.

### 2.2.4 Ejemplos

**Ejemplo 2.7** Encuentra las soluciones factibles que cumplen las condiciones de Karush-Kuhn-Tucker para el problema

$$\text{Optimizar } x^2 + y^2$$

$$\text{Sujeto a } x + y = 6$$

$$x^2 + y^2 \leq 26$$

$$x - 1 \geq 0$$

**Solución:** Las funciones del problema, tanto la función objetivo como las restricciones, son funciones polinomiales y por tanto tienen derivadas parciales continuas respecto de todas sus variables, y por tanto se da la primera condición para establecer las condiciones de Karush-Kuhn-Tucker.

El primer paso antes de resolver un problema de optimización es presentar el problema en la forma estándar que se ha utilizado para la presentación de los resultados durante el tema. Es decir, los términos independientes de las restricciones deben ser cero y utilizamos solamente

Calculamos en primer lugar los vectores gradiente de todas y cada una de las funciones implicadas en el problema.

$$f(x, y) = x^2 - y \quad \Rightarrow \quad \nabla f(x, y) = \begin{pmatrix} 2x \\ -1 \end{pmatrix}$$

$$h_1(x, y) = x + y - 6 \quad \Rightarrow \quad \nabla h_1(x, y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$g_1(x, y) = x^2 + y^2 - 26 \quad \Rightarrow \quad \nabla g_1(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

$$g_2(x, y) = 1 - x \quad \Rightarrow \quad \nabla g_2(x, y) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

Donde se le ha cambiado el sentido a la segunda restricción de desigualdad para ponerlo en la forma usual utilizada en este capítulo

$$x - 1 \geq 0 \Rightarrow 1 - x \leq 0$$

1. *Condición estacionaria*

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\mathbf{x}^*) \Leftrightarrow \begin{pmatrix} 2x^* \\ -1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu_1 \begin{pmatrix} 2x^* \\ 2y^* \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = 0$$

que en ecuaciones

$$2x^* + \lambda + 2x^* \mu_1 - \mu_2 = 0$$

$$-1 + \lambda + 2y^* \mu_1 = 0$$

2. *Condición de factibilidad*

$$x^* + y^* = 6$$

$$x^* - 1 \geq 0 \Rightarrow 1 - x^* \leq 0$$

$$(x^*)^2 + (y^*)^2 \leq 26$$

3. *Condición de positividad o negatividad.*

Los multiplicadores asociados a desigualdades en los puntos de KKT deben tener el mismo signo.

$$\mu_j \geq 0 \quad j = 1, \dots, p \quad (CKKTMin)$$

ó

$$\mu_j \leq 0 \quad j = 1, \dots, p \quad (CKKTMax)$$

4. *Condiciones de holgura.*

$$\mu_j g_j(\mathbf{x}^*) \Leftrightarrow \begin{cases} \mu_1 ((x^*)^2 + (y^*)^2 - 26) = 0 \\ \mu_2 (1 - x^*) = 0 \end{cases}$$

Para encontrar los puntos de Karush-Kuhn-Tucker, en primer lugar hay que resolver el conjunto de ecuaciones que se producen al aplicar las condiciones de KKT.

$$2x + \lambda + \mu_1 2x - \mu_2 = 0$$

$$-1 + \lambda + \mu_1 2y = 0$$

$$x + y = 6$$

$$\mu_1 (x^2 + y^2 - 26) = 0$$

$$\mu_2 (1 - x) = 0$$

Los puntos de KKT son soluciones factibles, por tanto hay que comprobar la factibilidad de las soluciones obtenidas al resolver el sistema anterior y por último se aplican las condiciones de negatividad o positividad.

La forma usual de resolver el sistema es observando las ecuaciones que se producen al aplicar las condiciones de holgura, ya que estas son un producto de dos factores que se anula; luego alguno de los dos factores debe ser cero. Este análisis produce dos opciones por cada ecuación, con un total de cuatro casos:

$$\left\{ \begin{array}{l} \mu_1 = 0 \Rightarrow \left\{ \begin{array}{ll} \mu_2 = 0 & \text{Caso I} \\ (1 - x) = 0 & \text{Caso II} \end{array} \right. \\ x^2 + y^2 - 26 = 0 \Rightarrow \left\{ \begin{array}{ll} \mu_2 = 0 & \text{Caso III} \\ (1 - x) = 0 & \text{Caso IV} \end{array} \right. \end{array} \right.$$

A continuación se desarrolla cada caso individualmente.

1. *Caso I:*  $\mu_1 = \mu_2 = 0$ . El sistema para estos valores queda

$$2x + \lambda = 0$$

$$-1 + \lambda = 0$$

$$x + y = 6$$

con solución

$$\begin{aligned} \lambda &= 1 \\ x &= -\frac{\lambda}{2} = -\frac{1}{2} \\ y &= 6 - x = 6 + \frac{1}{2} = \frac{13}{2} \end{aligned}$$

y los puntos que resuelven este caso son

$$P_1 = \left( -\frac{1}{2}, \frac{13}{2} \right) \quad \lambda = 1 \quad \mu_1 = \mu_2 = 0$$

A continuación se utilizan las condiciones de factibilidad para comprobar que los puntos obtenidos pertenecen al conjunto factible. Al sustituir en las restricciones de desigualdad

$$x^2 + y^2 = \left( \frac{1}{2} \right)^2 + \left( \frac{13}{2} \right)^2 = \frac{1}{4} + \frac{169}{4} \not\leq 26 \Rightarrow \text{No se cumple}$$

$$1 - x = 1 + \frac{1}{2} = \frac{3}{2} \not\leq 0 \Rightarrow \text{No se cumple}$$

luego este punto no está en el conjunto factible y por tanto no es de KKT.

2. *Caso II:*  $\mu_1 = 0$ ,  $x = 1$ . Con estos datos el sistema queda

$$2 + \lambda - \mu_2 = 0$$

$$-1 + \lambda = 0$$

$$1 + y = 6$$

cuya solución es

$$y = 5$$

$$\lambda = 1$$

$$\mu_2 = 2 + \lambda = 2 + 1 = 3$$

y el punto que se obtiene es

$$P_2 = (1, 5) \quad \lambda = 1 \quad \mu_1 = 0 \quad \mu_2 = 3$$

Comprobamos a continuación si es un punto factible

$$x^2 + y^2 - 26 = 1 + 25 - 26 \leq 0 \quad \text{Sí cumple la primera restricción}$$

$$1 - x = 1 - 1 = 0 \leq 0 \quad \text{Sí cumple la segunda restricción}$$

como además se cumple la condición de positividad,  $P_2$  es un punto que cumple las condiciones de KKT de mínimo (*CKKTMín*).

3. *Caso III:*  $x^2 + y^2 = 26$ ,  $\mu_2 = 0$ . Con estos valores el sistema es

$$2x + \lambda + \mu_1 2x = 0$$

$$-1 + \lambda + \mu_1 2y = 0$$

$$x + y = 6$$

$$x^2 + y^2 - 26 = 0$$

cuya solución se obtiene fácilmente despejando en primer lugar una de las variables de la tercera ecuación y sustituyendo en la cuarta para obtener

$$y = 6 - x$$

$$x^2 + (6 - x)^2 - 26 = 0$$

desarrollando

$$x^2 + (6 - x)^2 - 26 = 0 \Leftrightarrow 2x^2 - 12x + 10 = 0$$

con soluciones

$$x_1 = 5 \quad \text{y} \quad x_2 = 1$$

A continuación se obtiene el valor de  $y$  a partir de la variable  $x$  para obtener dos puntos

$$P_3 = (5, 1)$$

y

$$P_4 = (1, 5)$$

Comprobamos a continuación su factibilidad

$$\begin{aligned} P_3 &= (5, 1) \Rightarrow \begin{cases} 5^2 + 1^2 = 26 \leq 26 \\ 5 - 1 = 4 \geq 0 \end{cases} \\ P_4 &= (1, 5) \Rightarrow \begin{cases} 1^2 + 5^2 = 26 \leq 26 \\ 1 - 1 = 0 \geq 0 \end{cases} \end{aligned}$$

Y por último, es necesario obtener el valor de los multiplicadores para determinar si cumplen las condiciones de KKT de mínimo o de máximo. Utilizando las dos primeras ecuaciones, que forman un sistema lineal en  $\lambda$  y  $\mu_1$ , obtenemos

$$\begin{aligned} \mu_1 &= \frac{1+2x}{2(y-x)} = \begin{cases} \mu_1(P_3) = -\frac{11}{8} \\ \mu_1(P_4) = \frac{3}{8} \end{cases} \\ \lambda &= \frac{x(1+2y)}{x-y} = \begin{cases} \lambda(P_3) = \frac{15}{4} \\ \lambda(P_4) = -\frac{11}{4} \end{cases} \end{aligned}$$

En resumen, los puntos con sus respectivos multiplicadores son:

$$P_3 = (5, 1) \quad \lambda = \frac{15}{4} \quad \mu_1 = -\frac{11}{8} \leq 0 \quad \mu_2 = 0$$

y

$$P_4 = (1, 5) \quad \lambda = -\frac{11}{4} \quad \mu_1 = \frac{3}{8} \geq 0 \quad \mu_2 = 0 \geq 0$$

de donde se obtiene que  $P_4$  es un punto de KKT para el problema de minimización ( $\mu_j \geq 0$ ), mientras que para  $P_3$  cumple las condiciones de KKT para máximo (*CKKTM*ax).

4. *Caso IV*:  $x^2 + y^2 = 6$ ,  $x = 1$ . En este último caso queda el siguiente sistema:

$$2 + \lambda + 2\mu_1 - \mu_2 = 0$$

$$-1 + \lambda + \mu_1 2y = 0$$

$$1 + y = 6$$

$$1 + y^2 - 26 = 0$$

De la tercera y cuarta ecuación tenemos el punto

$$P_5 = (1, 5)$$

que es uno de los puntos encontrados en el apartado anterior y ya se ha discutido. Sin embargo, el cálculo de los multiplicadores se obtiene a partir de las dos primeras ecuaciones, en las que al sustituir por los valores de  $x$  e  $y$  correspondientes obtenemos el sistema

$$2 + \lambda + 2\mu_1 - \mu_2 = 0$$

$$-1 + \lambda + \mu_1 10 = 0$$



que es lineal con más incógnitas que ecuaciones, por tanto, será indeterminado y la solución es

$$\lambda = \lambda; \quad \mu = \left( \frac{1-\lambda}{10}, \frac{11+4\lambda}{5} \right)$$

Notar que si

$$\lambda = 1 \Rightarrow \mu = (0, 3)$$

$$\lambda = -\frac{11}{4} \Rightarrow \mu = \left( \frac{3}{8}, 0 \right)$$

que corresponden a los multiplicadores de los puntos  $P_2$  y  $P_4$ . No obstante se trata en los tres casos del mismo punto. El hecho de que existan diversos multiplicadores para el mismo punto es debido a que éste es un caso singular.

**Ejemplo 2.8** *Calcula los puntos de Karush-Kuhn-Tucker del siguiente problema*

$$\begin{array}{ll} \text{Minimizar} & f(x, y) = x^2 + y^2 \\ \text{Sujeto a} & h_1(x, y) = 2x + y - 2 = 0 \end{array}$$

**Solución:** En este caso  $m = 1$  y  $p = 0$ , es decir hay solamente una restricción de igualdad y el problema es de Lagrange. Las condiciones de Karush-Kuhn-Tucker son:

$$\nabla f(x) + \lambda_1 \nabla h_1(x) = 0 \Rightarrow \begin{pmatrix} 2x \\ 2y \end{pmatrix} + \lambda_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0$$

$$h_1(x) = 0 \Rightarrow 2x + y - 2 = 0$$

y un punto será de KKT si es solución del sistema

$$2x + 2\lambda_1 = 0 \Rightarrow \lambda_1 = -x$$

$$2y + \lambda_1 = 0 \Rightarrow \lambda_1 = -2y$$

$$2x + y - 2 = 0$$

Resolviendo el sistema obtenemos los valores de  $(x, y)$  y de  $\lambda_1$

$$-2\lambda_1 - \frac{\lambda_1}{2} - 2 = 0 \Leftrightarrow -\frac{5\lambda_1}{2} = 2 \Leftrightarrow \lambda_1 = -\frac{4}{5}$$

$$x = \frac{4}{5} \quad y = \frac{2}{5}$$

### 2.2.5 Hipótesis de cualificación de las restricciones

Las condiciones de Karush-Kuhn-Tucker que se han presentado en la sección anterior son necesarias para la mayoría de los problemas de optimización no lineal, es decir, si  $\mathbf{x}^*$  es un óptimo local para un problema no lineal, entonces tendrá que cumplirlas. No obstante, en ocasiones, para garantizar que un óptimo local debe cumplir estas condiciones, las restricciones en dicho punto tienen que cumplir determinadas propiedades denominadas *Hipótesis de Cualificación de las Restricciones (H.C.R.)*. El estudio de estas hipótesis no cae dentro del ámbito de este curso y solamente indicaremos algunas de ellas.

**Definición 2.9 (Sin restricciones):** *En un problema de optimización no lineal sin restricciones, todos los puntos cumplen la primera hipótesis de cualificación.*

**Definición 2.10 (Condición de Karlin):** En un problema de optimización no lineal donde solamente hay restricciones de tipo lineal, todos los puntos factibles cumplen la hipótesis de cualificación de Karlin.

**Definición 2.11 (Condición de Slater):** En un problema de optimización no lineal en el que el conjunto factible,  $\Omega$ , es un conjunto convexo con interior no vacío, todos los puntos factibles cumplen la hipótesis de cualificación de Slater.

**Definición 2.12 (Condición de Fiacco-McKormik).** En un problema de optimización no lineal, todos los puntos factibles que sean regulares cumplen la hipótesis de cualificación de Fiacco-McKormik.  $\quad$

### 2.2.6 Condiciones necesarias de primer orden

Estamos en condiciones de establecer las condiciones necesarias de primer orden que deben cumplir los extremos locales de un problema de optimización.

**Teorema 2.13** Dado el problema general de optimización

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{array} \quad (2.6)$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$  son funciones de clase  $\mathcal{C}^1(A)$  en  $A \subseteq \mathbb{R}^n$  un conjunto abierto. Sea  $\Omega$  el conjunto factible del problema anterior y  $\mathbf{x}^* \in \Omega$  un punto donde las restricciones del problema cumplen alguna de las hipótesis de cualificación y en el que la función  $f(\mathbf{x})$  alcanza un mínimo [máximo respectivamente] relativo  $\implies \mathbf{x}^*$  cumple las condiciones de Karush-Kuhn-Tucker de Minimización [Maximización respectivamente].

Como se ha comentado anteriormente, las condiciones de Karush-Kuhn-Tucker son necesarias para la mayoría de los problemas de optimización, sin embargo, es posible encontrar problemas en los que la solución óptima no cumple estas condiciones.

**Ejemplo 2.14** Consideremos el problema

$$\begin{array}{ll} \text{Maximizar} & -x^2 + y \\ \text{Sujeto a} & y^3 = 0 \end{array}$$

El conjunto factible del problema es

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid (x, 0)\}$$

y el problema es equivalente a:

$$\begin{array}{ll} \text{Maximizar} & -x^2 \\ \text{Sujeto a} & x \in \mathbb{R} \end{array}$$

cuya solución óptima podemos obtener fácilmente.

$$x = 0$$

La solución óptima del problema en  $\mathbb{R}^2$  es

$$\mathbf{x}^* = (0, 0)$$

Vamos a comprobar si  $\mathbf{x}^*$  es un punto de Karush-Kuhn-Tucker. Como el problema solamente tiene restricciones de igualdad, las condiciones de KKT indican que debería existir un multiplicador  $\lambda$  de forma que se cumplan las siguientes ecuaciones:

$$\begin{aligned} \nabla f(\mathbf{x}) + \lambda \nabla h(\mathbf{x}) &= 0 \implies \begin{pmatrix} -2x \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 3y^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ h(\mathbf{x}) &= 0 \end{aligned}$$

es decir

$$-2x = 0$$

$$1 + 3\lambda y^2 = 0$$

$$y^3 = 0$$

pero este sistema no tiene solución, ya que de la última ecuación  $y = 0$  y al sustituir en la segunda obtenemos una contradicción.

El punto  $\mathbf{x}^* = (0, 0)$  es un máximo local (de hecho es global) que no cumple las condiciones de Karush-Kuhn-Tucker. Comprobamos que no se cumplen ninguna de las hipótesis de cualificación indicadas anteriormente:

1. Condición sin restricciones: Está claro que esta hipótesis no se cumple por la presencia de la restricción:  $y^3 = 0$ .
2. Condición de Karlin: Esta condición no se cumple puesto que la única restricción del problema,  $h(x, y) = y^3$  es no lineal.
3. Condición de Slater: El conjunto factible es

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid (x, 0)\}$$

es decir eje  $X$ , su interior será vacío, puesto que cualquier bola de centro un punto de  $\Omega$  tiene puntos fuera de  $\Omega$  y por tanto tampoco se cumple esta hipótesis.

4. Condición de Fiacco-McKormik: En este caso tenemos que comprobar si el punto es o no regular para las restricciones del problema, es decir, habrá que comprobar si el conjunto de vectores formado por los gradientes de las restricciones activas en  $\mathbf{x}^*$  es linealmente independiente. Como solamente tenemos una restricción que es además activa en  $\mathbf{x}^*$

$$h(x, y) = y^3 \Rightarrow h(\mathbf{x}^*) = h(0, 0) = 0$$

luego la familia de vectores estará formada por el vector

$$\{\nabla h(\mathbf{x}^*)\}$$

como  $\nabla h(x, y) = \begin{pmatrix} 0 \\ 3y^2 \end{pmatrix}$ , al evaluar en  $\mathbf{x}^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  obtenemos

$$\{\nabla h(\mathbf{x}^*)\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

que es linealmente dependiente y por tanto  $\mathbf{x}^* = (0, 0)$  es no regular para las restricciones.

Si en el problema se cambia la restricción  $y^3 = 0$  por

$$y = 0$$

que es equivalente, la solución será la misma, pero ahora sí se cumple la condición de Karlin, puesto que todas las restricciones del problema son lineales. Ahora es posible comprobar que el punto  $\mathbf{x}^* = (0, 0)$ , cumple las condiciones de KKT, puesto que el sistema es

$$-2x = 0$$

$$1 + \lambda = 0$$

$$y = 0$$

de donde

$$\lambda = -1$$

**Ejemplo 2.15** Sea el programa

$$\begin{array}{ll} \text{Maximizar} & f(x, y, z) = y \\ \text{Sujeto a} & h_1(x, y, z) = (x-1)^2 + y^2 - 1 = 0 \\ & h_2(x, y, z) = (x+1)^2 + y^2 - 1 = 0 \end{array}$$

El conjunto factible para este problema es

$$\Omega = \{(x, y, z) \in \mathbb{R}^3 \mid (0, 0, z)\}$$

y la solución óptima del problema es cualquier punto del conjunto  $\Omega$ , puesto que  $f(x, y, z)$  es constante en él.

Al plantear la condición estacionaria de Karush-Kuhn-Tucker para este problema

$$\nabla f(\mathbf{x}) + \lambda_1 \nabla h_1(\mathbf{x}) + \lambda_2 \nabla h_2(\mathbf{x}) = 0$$

obtenemos

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2(x-1) \\ 2y \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2(x+1) \\ 2y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Los puntos críticos deben cumplir entonces el siguiente sistema de ecuaciones:

$$2\lambda_1(x-1) + 2\lambda_2(x+1) = 0$$

$$1 + 2\lambda_1 y + 2\lambda_2 y = 0$$

$$0 = 0$$

sin embargo, ningún punto de la forma  $(0, 0, z)$  es solución del sistema anterior, puesto que al sustituir estos puntos en la segunda ecuación obtenemos una contradicción, y ninguno de los extremos locales del problema cumple las condiciones de KKT.

Podemos comprobar de nuevo como en el caso anterior que ninguno de ellos cumple alguna de las hipótesis de cualificación. Es un problema con restricciones no lineales donde  $\overset{\circ}{\Omega} = \emptyset$ , es decir, no se cumplen ninguna de las tres primeras condiciones dadas. Para la condición de regularidad (Fiacco-McKormik) observamos que el conjunto de vectores que son gradiente de las restricciones activas (en este caso son todas puesto que es un problema con sólo igualdades) está dado por

$$\{\nabla h_1(x, y, z), \nabla h_2(x, y, z)\} = \left\{ \begin{pmatrix} 2(x-1) \\ 2y \\ 0 \end{pmatrix}, \begin{pmatrix} 2(x+1) \\ 2y \\ 0 \end{pmatrix} \right\}$$

y al evaluarlo en los puntos óptimos  $(0, 0, z)$  obtenemos

$$\left\{ \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \right\}$$

que son vectores linealmente dependientes y por tanto ningún punto de la forma  $(0, 0, z)$  es regular.

También puede suceder que un punto donde las restricciones no cumplan ninguna de las hipótesis de cualificación, sea un extremo local de la función objetivo donde se cumplan las condiciones de Karush-Kuhn-Tucker. Si consideramos, por ejemplo, las restricciones del ejemplo anterior y cambiamos la función objetivo por  $f(x, y, z) = x$ , las condiciones de KKT serían

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2(x-1) \\ 2y \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2(x+1) \\ 2y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

o equivalentemente serán soluciones del sistema

$$1 + 2\lambda_1(x - 1) + 2\lambda_2(x + 1) = 0$$

$$2\lambda_1 y + 2\lambda_2 y = 0$$

$$0 = 0$$

y teniendo en cuenta que el conjunto factible está formado por los puntos  $(0, 0, z)$

$$1 - 2\lambda_1 + 2\lambda_2 = 0$$

$$0 = 0$$

$$0 = 0$$

que tiene por solución

$$\lambda_1 - \lambda_2 = \frac{1}{2}$$

Tomando cualquier solución de esta ecuación, por ejemplo  $\lambda = (\lambda_1, \lambda_2) = (1, 1/2)$ , vemos que todos los puntos extremos cumplen las condiciones de Lagrange, sin embargo, como se ha comprobado en ninguno de ellos las restricciones cumplen alguna de las hipótesis de cualificación.

Todos estos ejemplos son atípicos y en general sucederá que los extremos locales del problema tendrán que cumplir las condiciones de Karush-Kuhn-Tucker.

**Observación 2.16** *Es posible comprobar que en el punto  $(5, 1)$  del ejemplo 2.7 es un punto donde no se cumple ninguna de las hipótesis de cualificación.*

Por último hay que indicar que estas condiciones son necesarias, en el sentido de que bajo las hipótesis del teorema, los extremos de un problema de optimización deben ser puntos de Karush-Kuhn-Tucker. Sin embargo, las condiciones no son suficientes, ya que podemos encontrar puntos que aún cumpliendo las condiciones de Karush-Kuhn-Tucker, no son extremos, por ejemplo la función  $f(x) = x^3$ , tiene como único punto estacionario  $x = 0$ , que no es extremo puesto que la función es siempre creciente.

**Definición 2.17** *Los puntos que cumplen la condición estacionaria pero que no son extremos de la función se denominan puntos de silla (que son condicionados si hay presencia de restricciones en el problema). Para funciones reales de una variable a estos puntos se les conoce mejor por puntos de inflexión.*

### 2.2.7 Más Ejemplos

Emplearemos el teorema de las condiciones de primer orden para resolver en esta sección algunos problemas de optimización con restricciones.

**Ejemplo 2.18** *Resolver el problema*

$$\begin{array}{ll} \text{Minimizar} & xy + yz + zx \\ \text{Sujeto a} & x + y + z = 3 \end{array}$$

**Solución:** Como solamente tiene una restricción de igualdad se trata de un problema de Lagrange; además dicha restricción es lineal, luego la condición de Karlin indica que se cumplen las hipótesis de cualificación de las restricciones en todos los puntos del conjunto factible, luego cualquier mínimo local debe cumplir las condiciones de Karush-Kuhn-Tucker. En este caso, estas condiciones se reducen a la condición estacionaria y a la condición de factibilidad:

$$\nabla f(x, y, z) + \lambda \nabla h(x, y, z) = \begin{pmatrix} y + z \\ z + x \\ x + y \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$

$$h(x, y, z) = x + y + z - 3 = 0$$

Ambas condiciones se transforman en el siguiente sistema:

$$y + z + \lambda = 0$$

$$x + z + \lambda = 0$$

$$y + x + \lambda = 0$$

$$x + y + z - 3 = 0$$

La única solución de dicho sistema es

$$x = y = z = 1 \quad \lambda_1 = -2$$

y por tanto este es el único punto que cumple las condiciones de KKT.

Notar que aunque  $\lambda_1 = -2 < 0$  y el objetivo es minimizar, el punto sigue siendo de KKT puesto que es un multiplicador asociado a una restricción de igualdad.

**Ejemplo 2.19** Plantear y resolver el problema de construir una caja de cartón de volumen máximo y área fija  $A$ .

**Solución:** Si  $x, y, z$  son las dimensiones de la caja, el problema no lineal se puede expresar como

$$\begin{array}{ll} \text{Maximizar} & xyz \\ \text{Sujeto a} & xy + yz + zx = \frac{A}{2} \end{array}$$

siendo  $A > 0$  el área fija de la caja.

La restricción no cumple ni la condición de Karlin, ni la de Slater. Comprobamos que se cumple la condición de regularidad en todos los puntos. Para ello calculamos el gradiente de la restricción:

$$\nabla h(x, y, z) = \begin{pmatrix} y + z \\ x + z \\ x + y \end{pmatrix}$$

Este vector será linealmente dependiente sólo cuando sea el vector nulo, es decir si

$$\begin{aligned} y + z &= 0 \\ x + z &= 0 \\ x + y &= 0 \end{aligned}$$

cuya única solución es

$$x = y = z = 0$$

Sin embargo, sustituyendo en la restricción, podemos comprobar fácilmente que este punto es infactible; de donde se deduce que todos los puntos de  $\Omega$  (puntos factibles) son regulares. Este resultado implica que cualquier extremo local del problema debe cumplir las condiciones de Karush-Kuhn-Tucker:

1. Condición estacionaria

$$yz + \lambda(y + z) = 0$$

$$xz + \lambda(x + z) = 0$$

$$xy + \lambda(x + y) = 0$$

2. Condición de factibilidad:

$$xy + yz + zx - \frac{A}{2} = 0$$

El sistema formado por las cuatro ecuaciones anteriores tiene como única solución a:

$$x = y = z = \sqrt{\frac{A}{6}} \quad \lambda = -\sqrt{\frac{A}{24}}$$

**Ejemplo 2.20 (Entropía)** *Caracteriza las distribuciones de probabilidad que aparecen naturalmente como distribuciones de entropía máxima.*

**Solución:** Sea una densidad de probabilidad discreta correspondiente a un valor medido tomando uno de los  $n$  valores  $x_1, \dots, x_m$ . La probabilidad asociada a  $x_j$  es  $p_j$  con  $p_j \geq 0$  y  $\sum_{j=1}^m p_j = 1$ . Se define la entropía de la densidad como

$$\varepsilon = - \sum_{j=1}^m p_j \ln p_j$$

mientras que el valor medio es

$$\sum_{j=1}^m x_j p_j$$

Si sabemos cual es la media ( $\bar{x}$ ), el argumento de entropía máxima requiere que la densidad debe tomarse como aquella que resuelve el problema

$$\text{Maximizar} \quad - \sum_{i=1}^m p_j \ln p_j$$

$$\text{Sujeto a} \quad \sum_{i=1}^m p_j = 1$$

$$\sum_{j=1}^m x_j p_j = \bar{x}$$

$$p_j \geq 0$$

Como todas las restricciones son lineales se cumple la condición de Karlin, luego cualquier solución óptima debe cumplir las condiciones de Karush-Kuhn-Tucker. Es decir, deben existir los multiplicadores de KKT:  $\lambda_1, \lambda_2, \mu_1, \dots, \mu_n$  cumpliendo las condiciones.

1. *Condición estacionaria*

$$\nabla f(\mathbf{p}) + \lambda_1 \nabla h_1(\mathbf{p}) + \lambda_2 \nabla h_2(\mathbf{p}) + \sum_{j=1}^n \mu_j \nabla g_j(\mathbf{p}) = 0$$

donde en este caso

$$f(\mathbf{p}) = - \sum_{i=1}^n p_i \ln p_i \Rightarrow \nabla f(\mathbf{p}) = (-\ln p_1 - 1, \dots, -\ln p_n - 1)$$

$$h_1(\mathbf{p}) = \sum_{i=1}^n p_i - 1 \Rightarrow \nabla h_1(\mathbf{p}) = (1, \dots, 1)$$

$$h_2(\mathbf{p}) = \sum_{i=1}^n x_i p_i - m \Rightarrow \nabla h_2(\mathbf{p}) = (x_1, \dots, x_n)$$

$$g_j(\mathbf{p}) = -p_j \Rightarrow \nabla g_j(\mathbf{p}) = \left( 0, \dots, \overbrace{-1}^j, \dots, 0_n \right) \quad j = 1, \dots, n$$

y las ecuaciones que tienen que cumplir los extremos son

$$-\ln p_i - 1 + \lambda_1 + \lambda_2 x_i - \mu_i = 0 \quad i = 1, \dots, n \quad (2.7)$$

## 2. Condición de factibilidad

$$\begin{aligned}\sum_{i=1}^n p_i &= 1 \\ \sum_{i=1}^n x_i p_i &= \bar{x} \\ p_i &\geq 0\end{aligned}$$

3. *Condición de Negatividad:* Puesto que estamos buscando el máximo de la función, los multiplicadores asociados a las restricciones de desigualdad deben ser negativos, es decir, debe ocurrir

$$\mu_j \leq 0$$

## 4. Condición de holgura:

$$\mu_j g_j(\mathbf{p}) = 0 \Leftrightarrow \mu_j (-p_j) = 0$$

De esta condición de holgura se deduce

$$\mu_j = 0$$

ya que en caso contrario  $p_j = 0$  y en ese punto la función  $\ln(x)$  no está definida. Por tanto todos los multiplicadores asociados a desigualdades deben ser nulos. Y sustituyendo en la ecuación

$$-\ln p_j - 1 + \lambda_1 + \lambda_2 x_j - \mu_j = 0 \quad j = 1, \dots, n$$

obtenemos

$$-\ln p_j - 1 + \lambda_1 + \lambda_2 x_j = 0 \quad j = 1, \dots, n$$

Despejando  $p_j$  en el conjunto de ecuaciones 2.7

$$-\ln p_j = -1 + \lambda_1 + \lambda_2 x_j \Rightarrow p_j = e^{\lambda_2 x_j + (\lambda_1 - 1)}$$

Observamos que  $p_j > 0$ . Los valores de  $\lambda_1$  y  $\lambda_2$  deben seleccionarse de forma que se cumplan las 2 restricciones

$$\sum p_j = 1$$

$$\sum p_j x_i = m$$

## 2.2.8 Condiciones necesarias de segundo orden

Si las funciones son suficientemente derivables, entonces, podemos utilizar información relativa a las segundas derivadas para obtener el siguiente resultado:

**Teorema 2.21 (Condiciones necesarias de 2º orden)** Dado el problema general de optimización

$$\begin{aligned}&\text{Minimizar [Maximizar]} \quad f(\mathbf{x}) \\&\text{Sujeto a} \quad \begin{aligned} h_i(\mathbf{x}) &= 0 \quad i = 1, \dots, m \\ g_j(\mathbf{x}) &\leq 0 \quad j = 1, \dots, p \end{aligned}\end{aligned}$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$  son funciones de clase  $\mathcal{C}^2(A)$  en el conjunto abierto  $A \subseteq \mathbb{R}^n$ . Sea  $\Omega$  su conjunto factible y  $\mathbf{x}^* \in \Omega$  un punto donde las restricciones del problema cumplen alguna de las hipótesis de cualificación de las restricciones (h.c.r.) y en el que la función  $f(\mathbf{x})$  alcanza un mínimo [máximo] relativo condicionado  $\Rightarrow \mathbf{x}^*$  es un punto que cumple las condiciones de Karush-Kuhn-Tucker para el problema de minimizar [maximizar] es decir  $\exists \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$  de forma que se cumplen las siguientes condiciones:



1. Condición estacionaria

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\mathbf{x}^*) = 0$$

2. Condición de factibilidad

$$\begin{aligned} h_i(\mathbf{x}^*) &= 0, & i &= 1, \dots, m \\ g_j(\mathbf{x}^*) &\leq 0 & j &= 1, \dots, p \end{aligned}$$

3. Condición de positividad

$$\mu_j \geq 0 \quad [\mu_j \leq 0 \text{ para máximo}] \quad j = 1, \dots, p$$

4. Condición de holgura

$$\mu_j g_j(\mathbf{x}^*) = 0 \quad j = 1, \dots, p$$

Además se cumple la siguiente condición:

5. Condición del Hessiano: La matriz  $HL(\mathbf{x}^*, \boldsymbol{\lambda})$  definida como

$$HL(\mathbf{x}^*, \boldsymbol{\lambda}) = Hf(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i Hh_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j Hg_j(\mathbf{x}^*) \quad (2.8)$$

es semidefinida positiva [semidefinida negativa respectivamente] sobre el espacio tangente  $M(\mathbf{x}^*)$  en  $\mathbf{x}^*$ .

#### Casos Particulares:

1. Sin restricciones y una variable ( $m = p = 0$ ,  $n = 1$ ): En el caso de problemas con una sola variable, el Hessiano de la función  $f(x)$  es su segunda derivada y la condición 2.8 se convierte en

$$f''(x^*) \geq 0$$

2. Sin restricciones y varias variables ( $m = p = 0$ ): En este caso  $M(\mathbf{x}^*) = \mathbb{R}^n$  y la condición es

$$Hf(\mathbf{x}^*) \text{ es semidefinida positiva [negativa respectivamente]}$$

### 2.2.9 Condiciones suficientes

Para dar las condiciones de suficiencia de extremo relativo condicionado de un problema no lineal con restricciones de igualdad y desigualdad, es necesario exigir que la forma cuadrática correspondiente sea definida positiva (o definida negativa respectivamente para máximo) en un espacio mayor que el espacio tangente. Se da en primer la definición de restricción de desigualdad degenerada.

**Definición 2.22** Dado el problema general con restricciones

$$\begin{aligned} &\text{Minimizar [Maximizar]} \quad f(\mathbf{x}) \\ &\text{Sujeto a} \\ &\quad h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ &\quad g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{aligned}$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$ , son funciones de clase  $C^1(A)$  en  $A \subseteq \mathbb{R}^n$  abierto. Sea  $\Omega$  el conjunto factible del problema anterior y  $\mathbf{x}^* \in \Omega$  un punto de Karush-Kuhn-Tucker para el problema. Diremos que una restricción de desigualdad  $g_j(\mathbf{x})$  es degenerada en  $\mathbf{x}^* \iff g_j(\mathbf{x}^*) = 0$  y  $\mu_j = 0$ .

**Definición 2.23** Definimos el conjunto de índices de restricciones no degeneradas en un punto como

$$\tilde{J}(\mathbf{x}^*) = \{j \in \{1, \dots, p\} \mid g_j(\mathbf{x}^*) = 0 \text{ y } \mu_j \neq 0\}$$

**Definición 2.24** Definimos el espacio tangente ampliado como

$$\widetilde{M}(\mathbf{x}^*) = \left\{ \mathbf{d} \in \mathbb{R}^n \mid \nabla^T h_i(\mathbf{x}^*) \mathbf{d} = 0; \quad \nabla^T g_j(\mathbf{x}^*) \mathbf{d} = 0, \quad i = 1, \dots, m; j \in \widetilde{J}(\mathbf{x}^*) \right\}$$

En el caso de un problema sin restricciones

$$\widetilde{M}(\mathbf{x}^*) = M(\mathbf{x}^*) = \mathbb{R}^n$$

**Teorema 2.25 (Condiciones suficientes)** Dado el problema general de optimización

$$\begin{array}{ll} \text{Minimizar [Maximizar]} & f(\mathbf{x}) \\ \text{Sujeto a} & \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{array}$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$  son funciones de clase  $\mathcal{C}^2(A)$  en  $A \subseteq \mathbb{R}^n$  un conjunto abierto. Sea  $\Omega$  el conjunto factible y  $\mathbf{x}^* \in \Omega$  un punto donde las restricciones del problema cumplen alguna de las hipótesis de cualificación. Si  $\mathbf{x}^*$  es un punto de Karush-Kuhn-Tucker para el problema de minimizar [maximizar], es decir  $\exists \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$  de forma que se cumplen las siguientes condiciones:

1. Condición estacionaria

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla g_j(\mathbf{x}^*) = 0$$

2. Condición de factibilidad

$$\begin{array}{ll} h_i(\mathbf{x}^*) & = 0, \quad i = 1, \dots, m \\ g_j(\mathbf{x}^*) & \leq 0 \quad j = 1, \dots, p \end{array}$$

3. Condición de positividad o negatividad

$$\mu_j \geq 0 \quad [\mu_j \leq 0] \quad j = 1, \dots, p$$

4. Condición de holgura

$$\mu_j g_j(\mathbf{x}^*) = 0 \quad j = 1, \dots, p$$

y además se cumple la condición:

5. Condición del Hessiano: la matriz  $HL(\mathbf{x}^*)$  definida como

$$HL(\mathbf{x}^*) = Hf(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i Hh_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j Hg_j(\mathbf{x}^*)$$

es definida positiva [definida negativa respectivamente] sobre el espacio tangente ampliado  $\widetilde{M}(\mathbf{x}^*)$ . Entonces en  $\mathbf{x}^*$  hay un mínimo [máximo] relativo condicionado estricto de  $f$  sobre  $\Omega$ . Si la matriz  $HL(\mathbf{x}^*)$  es indefinida sobre  $\widetilde{M}(\mathbf{x}^*)$  entonces en  $\mathbf{x}^*$  hay un punto de silla condicionado.

### Casos Particulares:

1. Sin restricciones y una variable ( $m = p = 0, n = 1$ ): En el caso de problemas con una sola variable, la condición del Hessiano se convierte en

$$f''(x^*) > 0$$

2. Sin restricciones y varias variables ( $m = p = 0$ ): En este caso  $M(\mathbf{x}^*) = \mathbb{R}^n$  y la condición es:

Si la matriz  $Hf(\mathbf{x}^*)$  es definida positiva [negativa]  $\Rightarrow \mathbf{x}^*$  es un mínimo [máximo respectivamente] local estricto

**Ejemplo 2.26** Resolver el problema

$$\begin{array}{ll} \text{Minimizar} & xy + yz + zx \\ \text{Sujeto a} & x + y + z = 3 \end{array}$$

**Solución:** Se comprobó anteriormente que el único punto crítico obtenido era:

$$x = y = z = 1 \quad \lambda_1 = -2$$

Si ahora tratamos de emplear las condiciones suficientes descritas en la proposición anterior, tendremos:

$$\mathbf{H}f(x, y, z) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

que no es ni definida positiva, ni definida negativa si consideramos todos los vectores de  $\mathbb{R}^3$ , sin embargo si restringimos la matriz a los puntos del espacio tangente

$$\begin{aligned} M(\mathbf{x}^*) &= \{ \mathbf{d} \in \mathbb{R}^3 \mid \nabla h^T(\mathbf{x}^*) \cdot \mathbf{d} = 0 \} = \{ \mathbf{d} \in \mathbb{R}^3 \mid (1, 1, 1)|_{(1,1,1)} \cdot \mathbf{d} = 0 \} = \\ &= \left\{ \mathbf{d} \in \mathbb{R}^3 \mid (1, 1, 1) \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0 \right\} = \{ \mathbf{d} \in \mathbb{R}^3 \mid d_1 + d_2 + d_3 = 0 \} \end{aligned}$$

la forma cuadrática asociada será

$$\begin{aligned} \varphi|_{Hf(\mathbf{x}^*)}(\mathbf{d}) &= (d_1, d_2, d_3) \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = (d_1, d_2, -d_1 - d_2) \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{pmatrix} d_1 \\ d_2 \\ -d_1 - d_2 \end{pmatrix} = \\ &= (d_1, d_2, -d_1 - d_2) \begin{pmatrix} -d_1 \\ -d_2 \\ d_1 + d_2 \end{pmatrix} = -d_1^2 - d_2^2 - (d_1 + d_2)^2 = -(d_1^2 + d_2^2 + (d_1 + d_2)^2) \geq 0 \end{aligned}$$

y solamente será 0, cuando

$$d_1 = d_2 = (d_1 + d_2) = 0$$

y por tanto

$$\mathbf{d} = (0, 0, 0)$$

Por tanto la forma cuadrática  $\varphi|_{Hf(\mathbf{x}^*)}(\mathbf{d})$  asociada a la matriz  $\mathbf{H}f(\mathbf{x}^*)$  es definida negativa sobre el espacio tangente  $M(\mathbf{x}^*)$  y por la proposición anterior el punto  $\mathbf{x}^*$  será un máximo local estricto.

En el caso de los problemas convexos las condiciones necesarias de primer orden son también suficientes.

**Proposición 2.27 (Programas convexos)** Dado el problema

$$\begin{array}{ll} \text{Minimizar [Maximizar]} & f(\mathbf{x}) \\ \text{Sujeto a} & \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{array}$$

con  $f : A \rightarrow \mathbb{R}$ ,  $h_i : A \rightarrow \mathbb{R}^m$ ,  $g_j : A \rightarrow \mathbb{R}^p$ , funciones de clase  $\mathcal{C}^1(A)$ ,  $A \subseteq \mathbb{R}^n$  un conjunto abierto y  $\Omega$  su conjunto factible. Si  $\Omega$  es convexo y  $f(\mathbf{x})$  es convexa [cóncava respectivamente] sobre  $\Omega$ , entonces, si existe  $\mathbf{x}^* \in \Omega$  y multiplicadores  $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p \in \mathbb{R}$  tales que el punto  $\mathbf{x}^*$ , junto con estos multiplicadores cumple las condiciones necesarias de primer orden para ser un mínimo [máximo respectivamente] local, entonces  $\mathbf{x}^*$  es un mínimo [máximo] global del problema.

Aunque es posible extender las condiciones necesarias y suficientes a órdenes superiores, en la práctica la aplicación de estas condiciones requiere de un excesivo esfuerzo y solamente tienen una utilidad práctica en el caso de funciones reales de variable real, es decir, cuando  $n = 1$  y  $m = p = 0$ .

**Teorema 2.28 (Condición suficiente de óptimo local)** *Supongamos que, para  $x^* \in I$ , la función  $f : I \rightarrow \mathbb{R}$  es suficientemente derivable y verifica*

$$\begin{aligned} f^{(k)}(x^*) &= 0 & k = 1, \dots, n-1 \\ f^{(n)}(x^*) &\neq 0 \end{aligned}$$

Donde  $f^{(k)}(x)$  es la derivada  $k$ -ésima de la función

1. Si  $n$  es impar  $\implies x^*$  es un punto de inflexión.
2. Si  $n$  es par  $\implies x^*$  es un óptimo local. Además
  - (a) Si  $f^{(n)}(x^*) > 0 \implies x^*$  es un mínimo local estricto.
  - (b) Si  $f^{(n)}(x^*) < 0 \implies x^*$  es un máximo local estricto.

## 2.3 Interpretación de los multiplicadores de KKT

En esta sección trataremos de explicar de forma no rigurosa el significado de los multiplicadores que aparecen en las condiciones de Karush-Kuhn-Tucker para un problema con restricciones y sus aplicaciones en el análisis de la sensibilidad de los problemas no lineales.

Planteemos en primer lugar un problema no lineal con restricciones de igualdad y de desigualdad de la siguiente forma

$$\begin{aligned} &\text{Minimizar [Maximizar]} && f(\mathbf{x}) \\ &\text{Sujeto a} && \\ & && h_i(\mathbf{x}) = b_i \quad i = 1, \dots, m \\ & && g_j(\mathbf{x}) \leq c_j \quad j = 1, \dots, p \end{aligned} \tag{2.9}$$

donde  $f, h_i, g_j : A \rightarrow \mathbb{R}$ , son funciones de clase  $\mathcal{C}^2(A)$  en  $A \subseteq \mathbb{R}^n$  abierto; y  $b_1, \dots, b_m, c_1, \dots, c_p \in \mathbb{R}$ .

Está claro que el conjunto factible del problema 2.9, dependerá de los vectores  $\mathbf{b} = (b_1, \dots, b_m)$  y  $\mathbf{c} = (c_1, \dots, c_p)$ , es decir

$$\Omega \equiv \Omega(\mathbf{b}, \mathbf{c})$$

y también es obvio que los puntos óptimos del problema, si existen, dependerán de estos valores

$$\mathbf{x}^* = \mathbf{x}^*(\mathbf{b}, \mathbf{c})$$

Supongamos que para ciertos valores de estos parámetros,  $(\mathbf{b}^*, \mathbf{c}^*)$ , el problema general con restricciones 2.9 posee un óptimo en el punto  $\mathbf{x}^*$ , con multiplicadores de Karush-Kuhn-Tucker  $(\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p)$  asociados. Podemos definir una función

$$F : U_{(\mathbf{b}^*, \mathbf{c}^*)} \longrightarrow \mathbb{R}, \text{ con } U_{(\mathbf{b}^*, \mathbf{c}^*)} \text{ un entorno de } (\mathbf{b}^*, \mathbf{c}^*) \in \mathbb{R}^{m+p}$$

con  $U_{(\mathbf{b}^*, \mathbf{c}^*)}$ , un entorno de  $(\mathbf{b}^*, \mathbf{c}^*) \in \mathbb{R}^{m+p}$ , de forma que

$$F(\bar{\mathbf{b}}, \bar{\mathbf{c}}) = f(\mathbf{x}^*(\bar{\mathbf{b}}, \bar{\mathbf{c}})) \quad \forall \bar{\mathbf{b}}, \bar{\mathbf{c}} \in U_{(\mathbf{b}^*, \mathbf{c}^*)}$$

siendo  $\mathbf{x}^*(\bar{\mathbf{b}}, \bar{\mathbf{c}})$  el óptimo del programa para cuando se utilizan en el problema 2.9, los términos independientes  $(\bar{\mathbf{b}}, \bar{\mathbf{c}}) \in U_{(\mathbf{b}^*, \mathbf{c}^*)}$ .

El siguiente teorema da una relación entre las variaciones del término independiente y las variaciones que experimenta el valor óptimo de la función objetivo.

**Teorema 2.29** *Dado el programa de optimización con restricciones dado en la ecuación 2.9. Si para ciertos valores de los parámetros  $\mathbf{b}$  y  $\mathbf{c}$ ,  $(\mathbf{b}^*, \mathbf{c}^*) = (b_1^*, \dots, b_m^*, c_1^*, \dots, c_p^*)$ , el punto  $\mathbf{x}^*$  es un punto de Karush-Kuhn-Tucker y junto con los multiplicadores asociados,  $\lambda_1, \dots, \lambda_m$  y  $\mu_1, \dots, \mu_p$ ; cumple las condiciones de suficiencia*

para que la función  $f(\mathbf{x})$  posea en ese punto un extremo relativo sobre el conjunto  $\Omega(\mathbf{b}^*, \mathbf{c}^*)$  y si no hay restricciones de desigualdad activas degeneradas, entonces

$$-\lambda_i = \frac{\partial F(\mathbf{b}^*, \mathbf{c}^*)}{\partial b_i} = \frac{\partial f(\mathbf{x}_{\mathbf{b}^*, \mathbf{c}^*}^*)}{\partial b_i} \quad i = 1, \dots, m$$

$$-\mu_j = \frac{\partial F(\mathbf{b}^*, \mathbf{c}^*)}{\partial c_j} = \frac{\partial f(\mathbf{x}_{\mathbf{b}^*, \mathbf{c}^*}^*)}{\partial c_j} \quad j = 1, \dots, p$$

Los multiplicadores  $\lambda_i$  y  $\mu_j$ , asociados a la  $i$ -ésima restricción de igualdad y a la  $j$ -ésima restricción de desigualdad respectivamente, nos mide la tasa de variación del valor de la función objetivo  $f(x, y)$ , en el punto óptimo respecto a la variación de su correspondiente término independiente ( $b_i, c_j$ ).

Notar finalmente que utilizando diferencias finitas obtenemos

$$\Delta F(\mathbf{b}^*, \mathbf{c}^*) = - \sum_{i=1}^m \lambda_i \Delta b_i - \sum_{j=1}^p \mu_j \Delta c_j = - \sum_{i=1}^m \lambda_i (\bar{b}_i - b_i^*) - \sum_{j=1}^p \mu_j (\bar{c}_j - c_j^*)$$

La ecuación anterior nos proporciona un valor aproximado del incremento que se producirá en el valor del objetivo óptimo al variar el término independiente de las restricciones de  $(\mathbf{b}^*, \mathbf{c}^*)$  a  $(\bar{\mathbf{b}}, \bar{\mathbf{c}})$ .

## 2.4 Métodos numéricos

Desarrollamos en esta sección la cuestión dinámica, es decir, cómo obtener a partir de una aproximación inicial, el verdadero valor de la solución óptima o al menos una aproximación mejor.

Teniendo en cuenta la equivalencia entre los problemas de maximización y minimización y puesto que con estos métodos sólo será posible encontrar a lo sumo una solución cada vez, en lo sucesivo consideraremos solamente el objetivo de la minimización.

Prácticamente todos los algoritmos descritos son iterativos y en términos generales consisten en tres pasos básicos:

1. Existe un conjunto de valores iniciales para las variables de decisión,  $\mathbf{x}^0 \in \mathbb{R}^n$ . Esta es la primera aproximación a la solución óptima buscada. También se le denomina *punto de partida*.
2. Un método o *algoritmo* para alcanzar la aproximación  $k$ -ésima,  $\mathbf{x}^{k+1} \in \mathbb{R}^n$ , a partir de los valores de la aproximación  $k$ -ésima  $\mathbf{x}^k$ , para  $k = 0, \dots, N$ .
3. Un criterio de parada; es decir, una regla para decidir cuando  $N$  es suficientemente grande o cuando  $\mathbf{x}^N$  está suficientemente cerca de la solución exacta.

En primer lugar se presentan los métodos de búsqueda de óptimos de funciones de una sola variable real. Distinguiendo para ellos varios métodos en función de las propiedades de continuidad y derivabilidad de la función que estemos utilizando. Estos métodos constituyen una parte importante de la implementación de algunos de los métodos empleados para la optimización de funciones de varias variables, que serán el objetivo de la segunda parte de esta sección.

### 2.4.1 Optimización en una variable

En esta sección proponemos un conjunto de métodos de búsqueda univariante destinados a la localización de puntos óptimos en un intervalo dado. Dependiendo del tipo de información sobre la función utilizada se distinguen en general 3 tipos de métodos:

- 1.- Métodos de reducción de intervalo.
- 2.- Métodos de aproximación polinomial.
- 3.- Métodos con derivación.

Todos los métodos de optimización para funciones de una variable que se describen a continuación, son métodos locales, en el sentido de que los puntos encontrados son mínimos locales. Esta restricción implica por una parte que el punto encontrado puede no ser la solución óptima que se pretende encontrar. Por otra parte la solución encontrada dependerá de la primera aproximación, por ejemplo, puede que la función tenga varios mínimos locales y que el método nos proporcione uno u otro en función del punto de partida.

### Métodos de reducción de intervalo

Este tipo de métodos se basa en la búsqueda de soluciones óptimas mediante sucesivas reducciones del intervalo de estudio y en la eliminación de subintervalos. Para poder desarrollar los métodos supondremos algunas limitaciones en cuanto a la función  $f(x)$ , que se pretende minimizar.

Dado el problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ & x \in I \end{array}$$

donde  $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ ,  $I = [a, b] \subseteq \mathbb{R}$  y  $f \in \mathcal{C}(I)$ . Supondremos que  $f(x)$  tiene las siguientes propiedades

$$\left. \begin{array}{l} \exists x^* \in I \text{ tal que } x \text{ es un mínimo local de } f(x) \text{ en } I \\ \nexists I_1 \subseteq I, I_1 = [c, d] \text{ tal que } f(x^*)|_{I_1} = \text{cte.} \end{array} \right\} \quad (2.10)$$

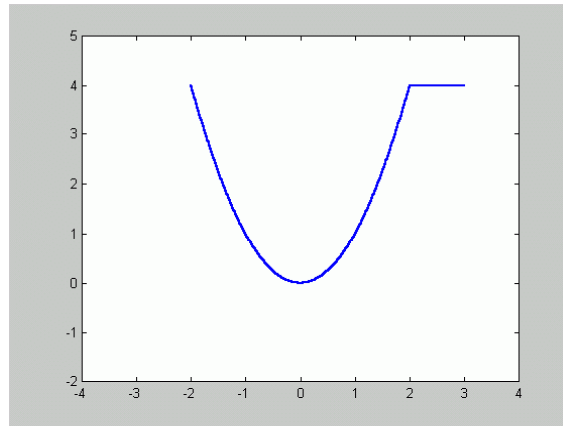
La primera propiedad implica que la función  $f(x)$  tiene un único mínimo en el intervalo compacto  $[a, b]$  y que además el máximo se alcanza en los extremos. La segunda propiedad establece que la función no es constante en ningún intervalo de longitud finita que esté incluido en  $I$ . La aplicación conjunta de estas dos propiedades implica que  $f(x)$  es siempre decreciente a la izquierda de  $x^*$ , mientras que es siempre creciente a la derecha de  $x^*$ , es decir, si  $x^*$  es el único punto de mínimo de  $f(x)$  en el intervalo  $[a, b]$ , para cualquier par de puntos  $x_1$  y  $x_2$  en  $[a, b]$  sucede

$$x^* < x_1 < x_2 \implies f(x^*) \leq f(x_1) \leq f(x_2)$$

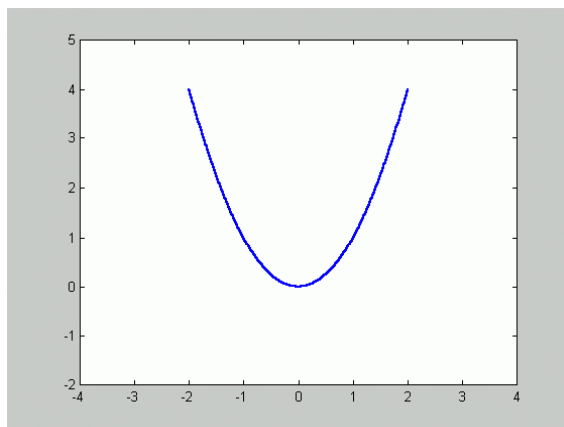
y

$$x^* > x_2 > x_1 \implies f(x^*) \leq f(x_2) \leq f(x_1)$$

La figura 2.1 representa una función que cumple la primera propiedad, pero no la segunda (este tipo de funciones se llaman *unimodales*). Mientras que la figura 2.2 representa una función con ambas propiedades (llamadas funciones estrictamente unimodales).



**Figura 2.1: Función unimodal**



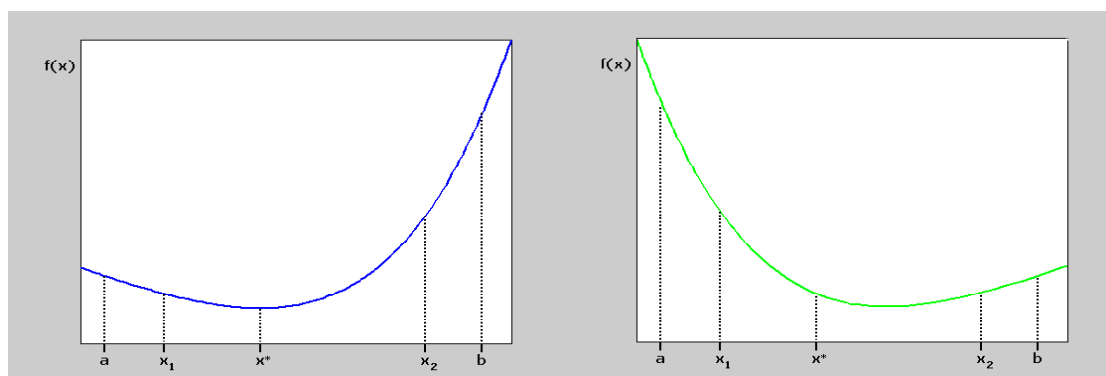
**Figura 2.2:** Función estrictamente unimodal

Una función con las propiedades descritas tiene la llamada *propiedad de eliminación* y que se describe a continuación:

**Propiedad de eliminación:** Supongamos que  $f : I = [a, b] \rightarrow \mathbb{R}$  cumple las propiedades descritas en 2.10, entonces se cumple

$$\forall x_1, x_2 \in I : a < x_1 < x_2 < b \implies \begin{cases} f(x_1) < f(x_2) \implies x^* \in [a, x_2] \\ f(x_1) > f(x_2) \implies x^* \in [x_1, b] \end{cases}$$

Podemos observar en la gráfica 2.3 ambas opciones.



**Figura 2.3:** Propiedad de eliminación (2.4.1).

**Observación 2.30** Cuando  $f(x_1) = f(x_2)$ , podríamos eliminar ambos extremos,  $[a, x_1]$  y  $[x_2, b]$ , y el mínimo se encontrará en el intervalo  $(x_1, x_2)$ .

La utilidad de esta propiedad radica en el hecho de que si  $f(x)$  es una función que la cumple, entonces solamente es necesario comparar  $f(x)$  en dos puntos diferentes para predecir en cuál de los subintervalos definidos por esos puntos **no** se va a encontrar el óptimo.

Utilizando la propiedad de eliminación podemos organizar una búsqueda en la que el óptimo se puede encontrar, eliminando recursivamente regiones del intervalo inicial. Cuando el subintervalo “*superviviente*” tenga una amplitud suficientemente pequeña, la búsqueda termina. Como se ha dicho anteriormente, la gran ventaja de estos métodos de búsqueda es que solamente requieren evaluaciones de la función y no necesitamos ninguna

hipótesis adicional acerca de la derivabilidad de la misma y pueden aplicarse a funciones tanto continuas como discontinuas, e incluso a funciones discretas. Las operaciones que se realizan se basan en la simple comparación de los valores que toma la función en dos puntos.

Generalmente en estos métodos de búsqueda se distinguen dos fases:

1. **Fase de acotación:** Un proceso de búsqueda inicial para acotar o localizar el óptimo en un intervalo inicial.
2. **Fase de reducción de intervalo:** Es una serie de reducciones o refinamientos más precisos, para reducir el intervalo de búsqueda inicial a la amplitud deseada.

Veamos cada fase por separado.

**Fase de acotación** En esta fase y a partir de un punto inicial cualquiera,  $x_0$ , se acota o encierra el óptimo en un intervalo finito haciendo uso de la propiedad de eliminación. Normalmente, esta acotación se realiza mediante un patrón de búsqueda heurística, aunque se pueden utilizar otros métodos más efectivos. Un ejemplo de patrón de búsqueda es el llamado *método de Swann*, en el que el  $(k + 1)$ -ésimo punto de prueba se obtiene mediante la ecuación recursiva:

$$x_{k+1} = x_k + 2^k \Delta \text{ para } k = 0, 1, 2, \dots$$

siendo  $x_0$  el punto de partida, elegido arbitrariamente, y  $\Delta \neq 0$ , es el tamaño de paso elegido.

Para determinar el signo de  $\Delta$ , se hace uso de la propiedad de eliminación y se comparan los valores que toma la función en los puntos  $x_0$ ,  $x_0 + |\Delta|$  y  $x_0 - |\Delta|$ . Podemos distinguir los siguientes casos:

1. Si

$$f(x_0 - |\Delta|) \geq f(x_0) \geq f(x_0 + |\Delta|)$$

entonces, por la hipótesis de unimodalidad, el mínimo debe estar a la derecha de  $x_0$  y  $\Delta > 0$ .

2. Si

$$f(x_0 - |\Delta|) \leq f(x_0) \leq f(x_0 + |\Delta|)$$

entonces, el mínimo debe estar a la izquierda de  $x_0$  y  $\Delta < 0$ .

3. Si

$$f(x_0 - |\Delta|) \geq f(x_0) \leq f(x_0 + |\Delta|)$$

el mínimo estará en el intervalo  $[x_0 - |\Delta|, x_0 + |\Delta|]$ , y puesto que ya tenemos la solución dentro de un intervalo, podremos terminar la fase de acotación.

4. Si

$$f(x_0 - |\Delta|) \leq f(x_0) \geq f(x_0 + |\Delta|)$$

entonces la función  $f(x)$  puede tener un máximo local y hay que buscar el mínimo fuera del intervalo  $[x_0 - |\Delta|, x_0 + |\Delta|]$ . Como podemos buscar a la izquierda de  $x_0 - |\Delta|$  o a la derecha de  $x_0 + |\Delta|$ , repetimos el proceso dos veces tomando como puntos de partida un valor  $x'_0 < x_0 - |\Delta|$ , y otro valor  $x''_0 > x_0 + |\Delta|$ .

El algoritmo del método se describe a continuación:

#### *Método de Swann*

**Inicialización.** Se eligen  $x_0$  y  $\Delta > 0$

**Paso 1.-** Se calculan  $f(x_0)$ ,  $f(x_0 + |\Delta|)$  y  $f(x_0 - |\Delta|)$

**Paso 2.-** Si

$$f(x_0 - |\Delta|) \geq f(x_0) \geq f(x_0 + |\Delta|) \implies \Delta = |\Delta| > 0$$

y se continúa en el paso 6 de la **Fase de reducción**.



**Paso 3.-** Si

$$f(x_0 - |\Delta|) \leq f(x_0) \leq f(x_0 + |\Delta|) \implies \Delta = -|\Delta| < 0$$

y se continúa en el paso 6 de la **Fase de reducción**.

**Paso 4.-** Si

$$f(x_0 - |\Delta|) \geq f(x_0) \leq f(x_0 + |\Delta|) \implies x^* \in [x_0 - |\Delta|, x_0 + |\Delta|]$$

y el algoritmo termina.

**Paso 5.-** Si

$$f(x_0 - |\Delta|) \leq f(x_0) \geq f(x_0 + |\Delta|) \Rightarrow \text{No hay mínimo}$$

el algoritmo termina y se determina que la función  $f(x)$  no cumple las propiedades exigidas.

**Fase de Reducción.** Se define  $k = 0$

**Paso 6.-** Se construye la sucesión de Swann

$$\begin{aligned} x_{k+1} &= x_k + 2^k \Delta \\ x_{k+2} &= x_{k+1} + 2^{k+1} \Delta \end{aligned}$$

**Paso 7.-** Se evalúan  $f(x_{k+1})$  y  $f(x_{k+2})$

**Paso 8.-** Si

$$f(x_{k+1}) > f(x_{k+2}) \Rightarrow k = k + 1$$

y se vuelve al **Paso 6**.

**Paso 9.-** Si

$$f(x_{k+1}) < f(x_{k+2}) \Rightarrow x^* \in [x_k, x_{k+2}]$$

y el algoritmo termina.

A continuación se aplica el método de Swann a un ejemplo.

**Ejemplo 2.31** Consideremos el problema de minimizar  $f(x) = (100 - x)^2$  utilizando  $x_0 = 30$  y un tamaño de paso  $|\Delta| = 5$ .

**Solución:** El signo de  $\Delta$  se determina al comparar

$$\begin{aligned} f(x_0) &= f(30) = 4900 \\ f(x_0 + |\Delta|) &= f(35) = 4225 \\ f(x_0 - |\Delta|) &= f(25) = 5625 \end{aligned}$$

Puesto que

$$f(x_0 - |\Delta|) \geq f(x_0) \geq f(x_0 + |\Delta|)$$

$\Delta > 0$ , y el punto mínimo  $x^*$  debe ser mayor que 30. Así  $x_1 = x_0 + \Delta = 35$ .

El punto  $x_2$  se obtiene

$$\begin{aligned} x_2 &= x_1 + 2\Delta = 45 \\ f(45) &= 3025 < f(x_1) \end{aligned}$$

por tanto,  $x^* > 35$

$$x_3 = x_2 + 2^2 \Delta = 65; \quad f(65) = 1225 < f(x_2)$$

por tanto,  $x^* > 45$

$$x_4 = x_3 + 2^3 \Delta = 105; \quad f(105) = 25 < f(x_3)$$

por tanto,  $x^* > 65$

$$x_5 = x_4 + 2^4 \Delta = 185; \quad f(185) = 7225 > f(x_4)$$

por tanto,  $x^* < 185$ . Por tanto, en 6 evaluaciones de  $f(x)$ , el mínimo  $x^*$ , se ha acotado dentro del intervalo

$$65 \leq x^* \leq 185$$

Hay que resaltar que la efectividad de esta acotación depende directamente del tamaño del paso  $\Delta$ . Si  $\Delta$  es grande, se produce una acotación inicial pobre, es decir, un intervalo inicial muy grande. Por otra parte, si  $\Delta$  es pequeño, es posible que se necesiten muchas evaluaciones de la función antes de encontrar el intervalo buscado.

**Fase de reducción del intervalo** Una vez que se ha determinado el intervalo que contiene al mínimo de la función, se necesita un esquema más eficiente de reducción del intervalo, con el fin de obtener una estimación mejor del punto óptimo de forma más rápida.

La cantidad de subintervalo eliminado en cada paso depende de la localización de los puntos  $x_1$  y  $x_2$  dentro del intervalo de búsqueda. Puesto la ubicación del óptimo (mínimo) es desconocida, sería razonable esperar que la colocación de los puntos  $x_1$  y  $x_2$  sea tal, que el intervalo que quede después del proceso se haya reducido en la misma proporción, independientemente del subintervalo eliminado. Sin embargo, por eficiencia, esa cantidad debería ser tan grande como fuera posible.

Se describen a continuación dos métodos de reducción de intervalo, el primero por su sencillez y el segundo por su eficacia.

**Método de bipartición** En este método se elimina exactamente la mitad del intervalo en cada paso. Se utilizan para ello 3 puntos equidistantes. Dentro del intervalo  $I = [a, b]$ , localizamos aquellos dos puntos  $x_1$  y  $x_2$  que están situados a una distancia  $L/4$  de cada extremo, siendo  $L = (b - a)$  la longitud del intervalo  $I$ , es decir

$$a < x_1 < x_2 < b \text{ con } |x_1 - a| = |b - x_2| = \frac{L}{4}$$

Junto con estos dos puntos hallamos el punto medio del intervalo  $x_m = \frac{a+b}{2}$ , y por tanto

$$|x_m - x_1| = |x_2 - x_m| = \frac{L}{4}$$

A continuación utilizamos la propiedad de eliminación comparando los valores que la función  $f(x)$  toma en los puntos anteriores.

Si  $f(x_1) < f(x_m)$ , podemos garantizar que  $x^* \notin [x_m, b]$ , y por tanto podemos utilizar como nuevo intervalo de búsqueda a  $[a, x_m]$ , siendo en este caso  $x_1$  el punto medio de dicho intervalo.

Si por el contrario  $f(x_1) \geq f(x_m)$ , podemos comparar ahora qué ocurre con  $f(x_2)$  y  $f(x_m)$ . Si  $f(x_2) < f(x_m)$ , la propiedad de eliminación permite eliminar el intervalo  $[a, x_m]$  y dejar como nuevo a  $[x_m, b]$  como nuevo intervalo de búsqueda, siendo en este caso el punto  $x_2$  el nuevo punto medio. En caso contrario, estaríamos en el caso  $f(x_1) \geq f(x_m) \leq f(x_2)$  y usando de nuevo la propiedad de eliminación es posible eliminar los subintervalos  $[a, x_1]$  y  $[x_2, b]$ , es decir  $x^* \in [x_1, x_2]$  y podemos continuar ahora el proceso con este nuevo subintervalo, siendo en este caso el punto medio el mismo que el del intervalo inicial. En cualquiera de los casos el intervalo que contenga a  $x^*$  tendrá una longitud igual a la mitad de la longitud del intervalo de partida.

El algoritmo de este método se puede desarrollar como sigue:

#### *Algoritmo del método de la bipartición*

**Inicialización.** Se elige  $\varepsilon > 0$ , tamaño final del intervalo. Dar  $a$  y  $b$  el intervalo inicial de búsqueda. Hacemos  $k = 0$ ,  $a_0 = a$  y  $b_0 = b$

**Paso 1.-** Calculamos

$$x_m = \frac{(a_k + b_k)}{2}$$

$$L_k = b_k - a_k$$

$$f(x_m)$$

**Paso 2.-** Calculamos

$$x_1 = a_k + \frac{L_k}{4}$$

$$x_2 = b_k - \frac{L_k}{4}$$

Notar que los puntos  $x_1, x_m$  y  $x_2$  son equidistantes

**Paso 3.-** Se comparan los valores de  $f(x_1)$  y  $f(x_m)$

- (a) Si  $f(x_1) < f(x_m)$ , eliminamos el intervalo  $(x_m, b_k)$  tomando  $b_{k+1} = x_m$  y  $a_{k+1} = a_k$ . El punto medio de la nueva búsqueda es  $x_1$ . Hacemos, por tanto,  $x_m = x_1$ . Ir al **Paso 5**.
- (b) Si  $f(x_1) \geq f(x_m)$ , entonces, ir al **Paso 4**.

**Paso 4.-** Se comparan los valores de  $f(x_2)$  y  $f(x_m)$

- (a) Si  $f(x_2) < f(x_m)$ , eliminamos el intervalo  $(a_k, x_m)$  tomando  $a_{k+1} = x_m$ ,  $b_{k+1} = b_k$ . El punto medio de la nueva búsqueda será ahora  $x_2$ . Hacemos, por tanto,  $x_m = x_2$ . Ir al **Paso 5**.
- (b) Si  $f(x_2) \geq f(x_m)$ , eliminamos el intervalo  $(a, x_1)$  y el intervalo  $(x_2, b)$ . Ponemos  $a_{k+1} = x_1$ , y  $b_{k+1} = x_2$ . En este caso el punto medio del nuevo intervalo sigue siendo  $x_m$ . Ir al **Paso 5**.

**Paso 5.-** Calculamos  $L_{k+1} = b_{k+1} - a_{k+1}$ . Si  $|L_{k+1}| \leq \varepsilon$  terminar, la solución aproximada es  $x_m \simeq x^*$ , con un error  $|x^* - x_m| \leq \frac{\varepsilon}{2}$ . En otro caso, hacer  $k = k + 1$ , ir al **Paso 2**.

**Observación 2.32** Del método de bipartición se puede deducir:

1. En cada paso del algoritmo se elimina exactamente la mitad del intervalo.
2. El punto medio de los sucesivos intervalos es siempre igual a alguno de los puntos de prueba  $x_1, x_2$  o  $x_m$ . Por tanto, se necesitan como mucho dos evaluaciones de la función en cada paso.
3. Después de  $n$  **iteraciones**, el intervalo final,  $I_n = [a_n, b_n]$ , tendrá una amplitud  $L_n = (b_n - a_n) = \left(\frac{1}{2}\right)^n L_0$ , siendo  $L_0 = (b_0 - a_0)$  la longitud del intervalo inicial. En cada una de estas iteraciones se han utilizado dos evaluaciones de la función.
4. Si en el intervalo final,  $I_n = [a_n, b_n]$ , se elige como solución su punto medio  $x^* = (a_n + b_n)/2$ , el error cometido estará acotado por  $L_n/2$ , siendo  $L_n = b_n - a_n$ , la longitud de este intervalo.

Aplicamos a continuación, el algoritmo de bipartición a un ejemplo.

**Ejemplo 2.33** Utiliza tres iteraciones del método de bipartición para resolver el siguiente problema

$$\begin{aligned} \text{Minimizar } f(x) &= (100 - x)^2 \\ 60 &\leq x \leq 150 \end{aligned}$$

**Solución:** En este caso  $a_0 = 60$ ,  $b_0 = 150$ , y  $L_0 = 150 - 60 = 90$

**1ª Iteración.-**

$$\begin{aligned} x_m &= \frac{60 + 150}{2} = 105 \\ x_1 &= a_0 + \frac{L_0}{4} = 60 + \frac{90}{4} = 82.5 \quad x_2 = b_0 - \frac{L_0}{4} = 150 - \frac{90}{4} = 127.5 \\ f(82.5) &= 306.25 > f(105) = 25 \quad f(127.5) = 756.25 > f(105) \end{aligned}$$

Por tanto eliminamos los intervalos  $(60, 82.5)$  y  $(127.5, 150)$ . La longitud del intervalo de búsqueda  $I_1 = [82.5, 127.5]$  se reduce de 90 a 45.

**2ª Iteración.-**

$$\begin{aligned} a_1 &= 82.5 \quad b_1 = 127.5 \quad x_m = 105 \\ L_1 &= 127.5 - 82.5 = 45 \\ x_1 &= 82.5 + \frac{45}{4} = 93.75 \quad x_2 = 127.5 - \frac{45}{4} = 116.25 \\ f(93.75) &= 39.06 > f(105) = 25 \quad f(116.25) = 264.06 > f(105) \end{aligned}$$

Por tanto el intervalo de incertidumbre es ahora  $I_2 = [93.75, 116.25]$

**3ª Iteración.-**

$$a_2 = 93.75 \quad b_2 = 116.25 \quad x_m = 105$$

$$L_2 = 116.25 - 93.75 = 22.5$$

$$x_1 = 99.375 \quad x_2 = 110.625$$

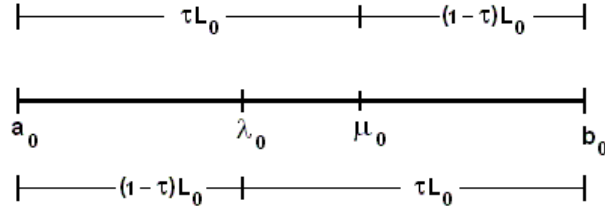
$$f(99.375) = 0.39 < f(105) = 25$$

Por tanto, eliminamos el intervalo  $(105, 116.25)$ . El nuevo intervalo es ahora  $I_3 = [93.75, 105]$  y su punto medio es 99.375. Así, en 3 iteraciones (6 evaluaciones de la función), la longitud del intervalo inicial de longitud 90 ha sido reducido exactamente a  $L(I_3) = (90) \left(\frac{1}{2}\right)^3 = 11.25$

**Método de la sección áurea** En la discusión anterior sobre el método de reducción de intervalo y de la estrategia de búsqueda, se observa lo siguiente

1. Si se disponen de dos intentos, es mejor localizarlos equidistantes al centro del intervalo.
2. Por sencillez del método, la estrategia sugiere que los puntos donde se ha de evaluar la función estén localizados simétricamente en el intervalo de forma que el subintervalo eliminado sea de la misma longitud independientemente del resultado de la evaluación.
3. Mejorar el esquema de búsqueda pasa por utilizar un sólo punto nuevo en cada iteración.

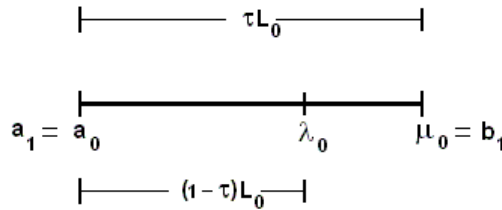
Con estos objetivos, consideramos la localización simétrica de dos puntos sobre un intervalo  $I = [a_0, b_0]$ , como en la figura 2.4.



**Figura 2.4: Búsqueda de la sección áurea**

Partimos de un intervalo  $[a_0, b_0]$  de longitud  $L_0 = b_0 - a_0$  y localizamos dos puntos, cada uno a una fracción  $\tau$  de cada extremo. Con esta simetría, independientemente de qué valor de la función sea el más pequeño, la longitud del intervalo que permanece es siempre de  $\tau L_0$ .

Supongamos que utilizando la propiedad de eliminación, descartamos el subintervalo de la derecha. Se puede observar a partir de la figura 2.5

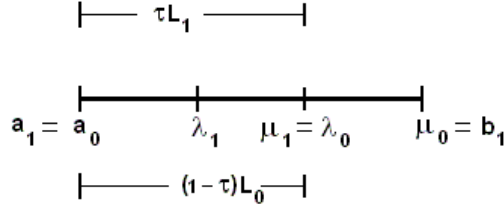


**Figura 2.5: Intervalos de la sección áurea**

que el punto que queda, de los dos anteriores, está situado a una distancia  $(1 - \tau) L_0$  de uno de los extremos.

Para mantener la simetría del patrón de búsqueda, la distancia  $(1 - \tau) L_0$  debe corresponder a una fracción  $\tau$  de la longitud  $L_1$  de este intervalo, que recordemos es de longitud  $\tau L_0$ . En este caso y con esta elección para  $\tau$ , el siguiente punto debe localizarse a una fracción  $\tau$  de la longitud del intervalo desde el extremo de la parte derecha.

Por tanto, con la elección de  $\tau$  que satisfaga  $(1 - \tau) L_0 = \tau L_1$ , el patrón de búsqueda de la figura permanece en el intervalo reducido de la figura 2.6



**Figura 2.6: Simetría de la sección áurea**

Teniendo en cuenta el valor de  $L_1$  en términos de  $L_0$ , obtenemos

$$(1 - \tau) L_0 = \tau L_1 = \tau (\tau L_0) = \tau^2 L_0$$

y simplificando

$$(1 - \tau) = \tau^2$$

Luego  $\tau$  es solución de la ecuación cuadrática

$$\tau^2 + \tau - 1 = 0$$

Esta ecuación se resuelve de forma directa para obtener

$$\tau = \frac{-1 \pm \sqrt{5}}{2}$$

y tomamos como valor de  $\tau$

$$\tau = \frac{-1 + \sqrt{5}}{2} \simeq 0.61803\dots$$

puesto que al ser  $\tau$  una fracción de longitud, debe ocurrir  $\tau > 0$ .

El esquema de búsqueda para la localización de los puntos iniciales siguiendo esta proporción es la llamada *búsqueda de la sección áurea*<sup>1</sup>. Después de las primeras 2 evaluaciones, cada evaluación posterior eliminará  $(1 - \tau)$  del intervalo restante. Por tanto, el intervalo que queda después de  $n$  iteraciones, asumiendo que el intervalo inicial es de longitud  $L_0$ , será de longitud  $\tau^n L_0$ .

Para desarrollar este método, partimos de un intervalo  $[a_0, b_0]$  y situamos 2 puntos a una distancia  $\tau$  de los extremos del intervalo

$$\lambda_0 = a_0 + (1 - \tau)(b_0 - a_0)$$

$$\mu_0 = a_0 + \tau(b_0 - a_0)$$

Utilizamos ahora la propiedad de eliminación y comparamos los valores que la función  $f(x)$  toma en  $\lambda_0$  y  $\mu_0$

<sup>1</sup>Los griegos llamaban a  $\tau^{-1} = \phi$ , razón áurea.

Si  $f(\lambda_0) < f(\mu_0)$  podemos garantizar que  $x^* \notin [\mu_0, b_0]$  y por tanto podemos utilizar como nuevo intervalo de búsqueda a  $[a_1, b_1] = [a_0, \mu_0]$ , en este caso y teniendo en cuenta las propiedades de  $\tau$ , de los dos puntos que hay que situar en el nuevo intervalo,  $\lambda_1$  y  $\mu_1$ , uno ya es conocido y solamente habrá que calcular el otro

$$\lambda_1 = a_1 + (1 - \tau)(b_1 - a_1)$$

$$\mu_1 = a_1 + \tau(b_1 - a_1)$$

Podemos comprobar que  $\mu_1 = \lambda_0$

$$\begin{aligned}\mu_1 &= a_1 + \tau(b_1 - a_1) = a_0 + \tau(\mu_0 - a_0) = a_0 + \tau(\{a_0 + \tau(b_0 - a_0)\} - a_0) = \\ &= a_0 + \tau^2(b_0 - a_0) = a_0 + (1 - \tau)(b_0 - a_0) = \lambda_0\end{aligned}$$

Si por el contrario  $f(\lambda_0) \geq f(\mu_0)$ , la propiedad de eliminación permite eliminar ahora el intervalo  $[a_0, \lambda_0]$  y queda como nuevo intervalo  $[a_1, b_1] = [\lambda_0, b_0]$ . Como en el caso anterior no es necesario calcular los valores de los 2 nuevos puntos dentro de este intervalo, puesto que uno viene de la fase anterior. En este caso:

$$\lambda_1 = a_1 + (1 - \tau)(b_1 - a_1) = \mu_0$$

$$\mu_1 = a_1 + \tau(b_1 - a_1)$$

El proceso se repite con el nuevo intervalo.

**Observación 2.34** Del método de la sección áurea se deduce:

1. En cada paso del algoritmo se elimina exactamente la cantidad  $(1 - \tau)$  del intervalo.
2. Solamente utilizamos un nuevo punto y por tanto se necesita una sola evaluación de la función en cada iteración.
3. Después de  $n$  **iteraciones**, el intervalo final,  $I_n = [a_n, b_n]$ , tendrá una amplitud  $L_n(b_n - a_n) = \tau^n L_0$ , siendo  $L_0 = b_0 - a_0$ , la longitud del intervalo inicial; y como se ha comentado antes, en cada iteración se evalúa la función una sola vez.
4. Si en el intervalo final, se elige como solución su punto medio  $x^* = (a_n + b_n)/2$ , el error cometido estará acotado por  $L_n/2$ , siendo  $L_n = b_n - a_n$ , la longitud de este intervalo.

#### Algoritmo sección áurea

Se incluye a continuación el algoritmo para la búsqueda del mínimo de una función mediante la sección o búsqueda áurea. Utilizaremos valores aproximados para los valores de los parámetros:

$$\begin{aligned}\tau &= 0.618 \\ 1 - \tau &= 0.382\end{aligned}$$

- **Inicio.** Se elige,  $\epsilon > 0$ , la longitud final del intervalo. Si  $[a_0, b_0]$  es el intervalo inicial de búsqueda, se definen  $\lambda_0 = a_0 + (1 - \tau)(b_0 - a_0)$  y  $\mu_0 = a_0 + \tau(b_0 - a_0)$ . Se evalúan  $f(\lambda_0)$  y  $f(\mu_0)$  y se inicializa el valor del contador de iteraciones,  $k = 0$ .

- **Paso Principal.**

**Paso 1.-** Si  $b_k - a_k < \epsilon$ , parar; la solución óptima está en el intervalo  $[a_k, b_k]$ . En caso contrario ir al **Paso 2** si  $f(\lambda_k) > f(\mu_k)$ , o al **Paso 3** si  $f(\lambda_k) \leq f(\mu_k)$ .

**Paso 2.-** Hacer  $a_{k+1} = \lambda_k$  y  $b_{k+1} = b_k$ ; además  $\lambda_{k+1} = \mu_k$  y  $\mu_{k+1} = a_{k+1} + \tau(b_{k+1} - a_{k+1})$ . Evaluar  $f(\mu_{k+1})$  e ir al **Paso 4**.

**Paso 3.-** Hacer  $a_{k+1} = a_k$  y  $b_{k+1} = \mu_k$ ; además  $\mu_{k+1} = \lambda_k$  y  $\lambda_{k+1} = a_{k+1} + (1 - \tau)(b_{k+1} - a_{k+1})$ . Evaluar  $f(\lambda_{k+1})$  e ir al **Paso 4**.

**Paso 4.-** Reemplazar  $k$  por  $k + 1$  e ir al **Paso 1**.

Aplicaremos a continuación el método de la sección áurea a un ejemplo.

**Ejemplo 2.35** Dado el siguiente problema

$$\begin{array}{ll} \text{Minimizar} & f(x) = x^2 + 2x \\ \text{Sujeto a} & -3 \leq x \leq 5 \end{array}$$

Utiliza el método de la sección áurea hasta conseguir un intervalo de longitud 0.2.

**Solución:** El intervalo inicial tiene longitud  $L_0 = 8$ . Vamos a reducir el intervalo hasta llegar a uno cuya longitud sea como mucho 0.2.

Las primeras dos observaciones se realizan sobre:

$$\lambda_0 = -3 + 0.382(8) = 0.056 \quad \mu_0 = -3 + 0.618(8) = 1.944$$

Notar que  $f(\lambda_1) < f(\mu_1)$  y por tanto, el nuevo intervalo de incertidumbre es  $[-3.000, 1.944]$

Iteración $k$	$a_k$	$b_k$	$\lambda_k$	$\mu_k$	$f(\lambda_k)$	$f(\mu_k)$
0	-3.000	5.000	0.056	1.944	0.115	7.667
1	-3.000	1.944	-1.112	0.056	-0.987	0.115
2	-3.000	0.056	-1.832	-1.112	-0.308	-0.987
3	-1.832	0.056	-1.112	-0.664	-0.987	-0.887
4	-1.832	-0.664	-1.384	-1.112	-0.853	-0.987
5	-1.384	-0.664	-1.112	-0.936	-0.987	-0.996
6	-1.112	-0.664	-0.936	-0.840	-0.996	-0.974
7	-1.112	-0.840	-1.016	-0.936	-1.000	-0.996
8	-1.112	-0.936				

Tabla 2.1: Método de la sección áurea

El proceso se repite en cada nuevo intervalo. En la tabla 2.1 se presenta un resumen de los datos obtenidos en cada iteración. Después de 8 iteraciones, que implican 9 evaluaciones de la función, el intervalo resultante es  $[-1.112, -0.936]$ , de longitud  $L_8 = 0.176$ . Tomamos como solución aproximada el punto medio del intervalo  $-1.024$ , con un error máximo  $|\epsilon_x| \leq \frac{0.176}{2} = 0.088$ .

### Métodos de aproximación polinomial

La idea básica que motiva esta técnica es que si la función  $f(x)$  es suficientemente suave (derivable) entonces puede aproximarse mediante un polinomio; dicho polinomio entonces, se puede utilizar para predecir la localización del óptimo de  $f(x)$ . El teorema de aproximación de Weierstrass garantiza que si una función es continua en un intervalo, entonces dicha función puede aproximarse suficientemente mediante polinomios de orden adecuado, y entonces el óptimo de la función se puede predecir razonablemente bien mediante el óptimo del polinomio.

**Métodos de estimación cuadrática** La interpolación polinomial más simple es la aproximación cuadrática. Esta basada en la observación de que una función continua que alcanza un mínimo en el interior de un intervalo debe ser al menos cuadrática. Si fuera lineal, el valor óptimo se obtendría en uno u otro de los extremos del intervalo. Así, un esquema de estimación cuadrática asume que dentro de un intervalo acotado la función puede aproximarse mediante una función cuadrática.

Dados tres puntos consecutivos  $x_0 < x_1 < x_2$  y sus correspondientes imágenes,  $f_0 = f(x_0)$ ,  $f_1 = f(x_1)$  y  $f_2 = f(x_2)$ ; buscamos 3 constantes  $a_0, a_1$  y  $a_2$  tales que la función cuadrática

$$q(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1)$$

coincida con  $f(x)$  en esos puntos. Es decir

$$f_j = q(x_j) \quad j = 0, 1, 2$$

De las relaciones anteriores obtenemos

$$\begin{aligned} f_0 &= a_0 & \implies a_0 &= f_0 \\ f_1 &= a_0 + a_1(x_1 - x_0) = f_0 + a_1(x_1 - x_0) & \implies a_1 &= \frac{f_1 - f_0}{x_1 - x_0} \\ f_2 &= a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) & \implies a_2 &= \frac{1}{x_2 - x_1} \left( \frac{f_2 - f_0}{x_2 - x_0} - \frac{f_1 - f_0}{x_1 - x_0} \right) \end{aligned} \quad (2.11)$$

Con tres puntos y los valores de la función en dichos puntos, hemos construido una función cuadrática (polinomio de 2º grado) que pasa por esos puntos y es próximo a la función  $f(x)$ . Si esta aproximación es suficientemente “buena”, entonces podemos utilizarla para obtener una aproximación del óptimo de  $f(x)$ , utilizando para ello el óptimo de  $q(x)$ .

Si  $\bar{x}$  es el óptimo de  $q(x)$

$$\bar{x} \text{ óptimo de } q(x) \in C^1(\mathbb{R}) \implies q'(\bar{x}) = 0$$

Derivando en la expresión de  $q(x)$  y sustituyendo por  $\bar{x}$ , se obtiene la ecuación

$$q'(\bar{x}) = a_1 + a_2(\bar{x} - x_1) + a_2(\bar{x} - x_0) = 0$$

que tiene por solución

$$\bar{x} = \frac{x_1 + x_0}{2} - \frac{a_1}{2a_2} \quad (2.12)$$

Podemos tomar ahora

$$x^* \simeq \bar{x}$$

siendo  $x^*$  el óptimo de  $f(x)$  buscado.

Es posible utilizar este esquema junto con la información obtenida en esta primera aproximación para obtener una aproximación mejor. La única precaución que hay que tomar es la elección de los tres valores  $x_0$ ,  $x_1$  y  $x_2$ , para que la función  $q(x)$  que tomamos como aproximación tenga un mínimo y no un máximo. Una forma de conseguir este hecho es, por ejemplo, elegir los puntos  $x_0 < x_1 < x_2$ , de forma que se cumplan las relaciones

$$f(x_1) < f(x_0) \text{ y } f(x_1) < f(x_2)$$

En problemas acotados, es decir, para los problemas del tipo

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &&& x \in [a, b] \end{aligned}$$

podemos tomar como primera terna de valores

$$\begin{aligned} x_0 &= a \\ x_1 &= x_m = \frac{a+b}{2} \\ x_2 &= b \end{aligned}$$

Veamos su aplicación a un ejemplo:

**Ejemplo 2.36** Consideremos la estimación del mínimo de la función

$$f(x) = 2x^2 + \frac{16}{x}$$

en el intervalo  $1 \leq x \leq 5$ . Tomaremos  $x_0 = 1$ ,  $x_2 = 5$  y elegimos  $x_1$  el punto medio  $x_1 = 3$ . Evaluando la función en esos puntos obtenemos

$$f_0 = 18 \quad f_1 = 23.33 \quad f_2 = 53.2$$



Para obtener el valor  $\bar{x}$ , calculamos  $a_1$  y  $a_2$

$$a_1 = \frac{23.33 - 18}{3 - 1} = \frac{8}{3}$$

$$a_2 = \frac{1}{5 - 3} \left\{ \frac{53.2 - 18}{5 - 1} - \frac{8}{3} \right\} = \frac{46}{15}$$

y ahora calculamos  $\bar{x}$

$$\bar{x} = \frac{3 + 1}{2} - \frac{8/3}{2(46/15)} = 1.565$$

y si evaluamos  $f(x)$  en ese punto

$$f(\bar{x}) = 15.1221$$

Podemos obtener una nueva aproximación, eligiendo 3 de los cuatros puntos actuales y realizar un nuevo ajuste.

Las relaciones entre los puntos y los valores de la función son

$$x_0 < \bar{x} < x_1 < x_2$$

$$f(x_0) > f(\bar{x}) < f(x_1) < f(x_2)$$

De donde se deduce que el próximo ajuste podría hacerse con los puntos  $x'_0 = x_0$ ,  $x'_1 = \bar{x}$  y  $x'_2 = x_1$ . El mínimo exacto es  $x^* = 1.5874$ .

### Métodos con derivación

Es de suponer que si además de continuidad en las funciones que queremos optimizar, podemos utilizar también la propiedad de derivabilidad de las mismas, podremos incrementar la eficiencia de los algoritmos de búsqueda. Nos referiremos en esta sección a métodos de búsqueda de óptimos de funciones derivables. Describiremos en primer lugar el esquema clásico para la búsqueda de raíces de ecuaciones no lineales definido en primer lugar por Newton y revisado posteriormente por Raphson.

**Método de Newton-Raphson** Sea  $f : \mathbb{R} \rightarrow \mathbb{R}$ , con  $f \in \mathcal{C}^2(\mathbb{R})$ . Si  $f(x)$  tiene un mínimo en el punto  $x^*$ , en primer lugar y utilizando las condiciones necesarias de primer orden,  $x^*$  será un punto estacionario de  $f(x) \Rightarrow$

$$f'(x^*) = 0$$

Por otra parte, como  $f(x)$  es dos veces derivable, dado un punto cualquiera  $x_0$ , que podemos considerar como la primera aproximación o aproximación inicial al punto  $x^*$ , podemos construir el polinomio de Taylor de orden 2 centrado en  $x_0$  para  $f(x)$

$$T_2 f(x_0; x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2$$

de forma que

$$f(x) \simeq T_2 f(x_0; x) \quad \forall x \in [x_0 - \epsilon, x_0 + \epsilon]$$

para un cierto  $\epsilon > 0$ ; y por tanto

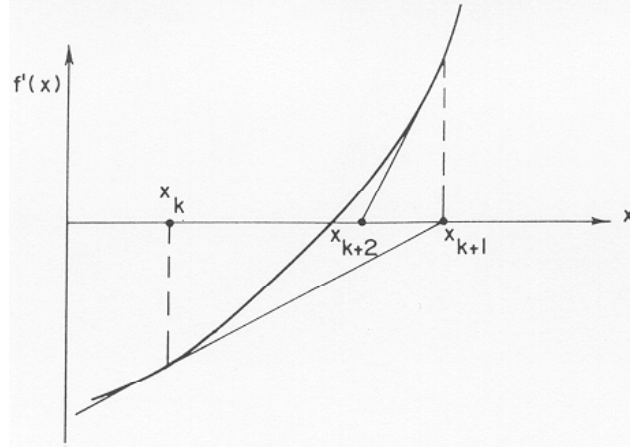
$$f'(x) \simeq T'_2 f(x_0; x) \quad \forall x \in [x_0 - \epsilon, x_0 + \epsilon]$$

Si  $x_0$  está próximo al punto  $x^*$  y  $T_2 f(x_0; x)$  tiene un punto estacionario  $x_1$ , podemos suponer que dicho punto es una aproximación mejor que  $x_0$  para  $x^*$ , el punto estacionario de  $f(x)$ .

$$T'_2 f(x_0; x) = f'(x_0) + f''(x_0)(x - x_0) \Rightarrow T'_2 f(x_0; x_1) = f'(x_0) + f''(x_0)(x_1 - x_0) = 0$$

de donde, siempre que  $f''(x_0) \neq 0$ , obtenemos

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}$$



**Figura 2.7: Método de Newton-Raphson (Convergencia)**

El proceso se repite tomando ahora como punto de partida el punto  $x_1$ .

Para el caso general, dada una función  $f : \mathbb{R} \rightarrow \mathbb{R}$ , si en la aproximación  $k$ -ésima,  $x_k$ , es posible calcular  $f(x_k)$ ,  $f'(x_k)$  y  $f''(x_k)$  ( $f \in \mathcal{C}^2(x_k)$ ), entonces se puede obtener la siguiente aproximación  $x_{k+1}$ , como el punto estacionario del polinomio de Taylor de orden 2 de  $f(x)$  en el punto  $x_k$

$$T_2 f(x_k, x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f''(x_k)}{2!}(x - x_k)^2$$

Si  $x_{k+1}$  es un punto estacionario de  $T_2 f(x_k; x)$

$$0 = T_2' f(x_k; x_{k+1}) = f'(x_k) + f''(x_k)(x_{k+1} - x_k)$$

y despejando  $x_{k+1}$  resulta

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

el proceso se repite hasta que el valor de la primera derivada sea suficientemente pequeño

$$|f'(x_k)| \leq \epsilon$$

con  $\epsilon \geq 0$ , el error máximo admitido.

La figura 2.7 ilustra los pasos generales del método de Newton, notar que  $x_{k+1}$  es el corte con el eje  $x$ , de la recta tangente a  $f'(x)$  en el punto  $x_k$ .

Desafortunadamente el método depende de la elección del punto de partida y de la naturaleza de la función y es bastante posible que este método no converja hacia el verdadero punto estacionario de  $f(x)$ . La figura 2.8 ilustra esta dificultad. Si comenzamos en un punto a la derecha de  $x_0$ , las aproximaciones sucesivas se alejarán del punto estacionario  $z$ .

Iteración $k$	$x_k$	$f'(x_k)$	$f''(x_k)$	$x_{k+1}$
1	0.400000	1.152000	3.840000	0.100000
2	0.100000	0.108000	2.040000	0.047059
3	0.047059	0.025324	1.049692	0.022934
4	0.022934	0.006167	0.531481	0.011331
5	0.011331	0.001523	0.267322	0.005634
6	0.005634	0.000379	0.134073	0.002807

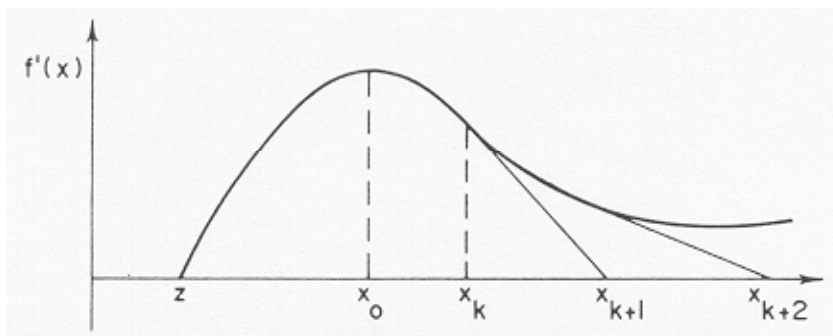
Tabla 2.2: Método de Newton-Raphson partiendo de  $\lambda_1 = 0.4$ 

Figura 2.8: Método de Newton-Raphson (Divergencia)

En general si se utiliza un punto inicial que se encuentre demasiado lejos del punto estacionario, la sucesión creada mediante el método de Newton, en general, no convergerá hacia el punto estacionario. No obstante, si el punto inicial está suficientemente cerca del punto estacionario, el método sí será convergente y además con convergencia muy rápida (convergencia cuadrática).

**Ejemplo 2.37** Considerar la función  $f(x)$  definida como

$$f(x) = \begin{cases} 4x^3 - 3x^4 & x \geq 0 \\ 4x^3 + 3x^4 & x < 0 \end{cases}$$

**Solución:** En primer lugar comprobamos que  $f(x)$  es dos veces derivable. El único punto conflictivo es  $x = 0$ , pero en ese punto

$$f'(x) = \begin{cases} 12x^2 - 12x^3 & x \geq 0 \\ 12x^2 + 12x^3 & x < 0 \end{cases} \Rightarrow \lim_{x \rightarrow 0^-} f'(0) = \lim_{x \rightarrow 0^+} f'(0)$$

$$f''(x) = \begin{cases} 24x - 36x^2 & x \geq 0 \\ 24x + 36x^2 & x < 0 \end{cases} \Rightarrow \lim_{x \rightarrow 0^-} f''(0) = \lim_{x \rightarrow 0^+} f''(0)$$

y la función  $f(x) \in \mathcal{C}^2(\mathbb{R})$ .

Por comodidad y puesto que  $f'(x)$  y  $f''(x)$  son conocidas en todos los puntos, podemos dar la expresión analítica de la sucesión de Newton

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \Rightarrow \begin{cases} x_{k+1} = \frac{12x_k - 24x_k^2}{24 - 36x_k} & \text{si } x_k \geq 0 \\ x_{k+1} = \frac{12x_k + 24x_k^2}{24 + 36x_k} & \text{si } x_k < 0 \end{cases}$$

Utilizamos la expresión anterior, utilizando como partida dos puntos distintos. En el primer caso,  $x_0 = 0.40$ ; y como se muestra en la tabla 2.2, se obtiene el punto 0.002807 después de seis iteraciones; la sucesión  $\{x_k\}$  parece convergente hacia  $x^* = 0$ .

Iteración $k$	$x_k$	$f'(x_k)$	$f''(x_k)$	$x_{k+1}$
1	0.600	1.728	1.440	-0.6000
2	-0.6000	1.728	-1.440	0.6000
3	0.600	1.728	1.440	-0.6000
4	-0.600	1.728	-1.440	0.6000

Tabla 2.3: Método de Newton-Raphson partiendo de  $\lambda_1 = 0.6$ 

El segundo punto de partida utilizado será  $x_0 = 0.60$  que como se puede apreciar en la tabla 2.3, produce una sucesión oscilante entre los valores 0.60 y  $-0.60$ , y por tanto no convergente.

A pesar de no haber una diferencia demasiado grande entre 0.4 y 0.6, los resultados son completamente distintos.

**Ejemplo 2.38 (Método de Newton-Raphson)** *Considera el problema*

$$\text{Minimizar } f(x) = 2x^2 + \frac{16}{x}.$$

Emplea el método de Newton-Raphson con  $\epsilon = 0.01$  y  $x_0 = 1$ .

**Solución:** Calculamos en primer lugar las expresiones de  $f'(x)$  y  $f''(x)$  que vamos a necesitar en cada iteración.

$$f'(x) = 4x - \frac{16}{x^2} \quad f''(x) = 4 + \frac{32}{x^3}$$

**Paso 1.**

$$|f'(x_0)| = 12 > 0.01$$

$$x_0 = 1 \quad f'(x_0) = -12 \quad f''(x_0) = 36 \quad \Rightarrow x_1 = 1 - \frac{-12}{36} = 1.3333$$

**Paso 2.**

$$|f'(x_1)| = 3.6673 > 0.01$$

$$x_1 = 1.3333 \quad f'(x_1) = -3.6673 \quad f''(x_1) = 17.5010 \quad \Rightarrow x_2 = 1.3333 - \frac{-3.6673}{17.5010} = 1.5428$$

**Paso 3.**

$$|f'(x_2)| = 0.5508 > 0.01$$

$$x_2 = 1.5428 \quad f'(x_2) = -0.5508 \quad f''(x_2) = 12.7141 \quad \Rightarrow x_3 = 1.5428 - \frac{-0.5508}{12.7141} = 1.5861$$

**Paso 4.**

$$|f'(x_3)| = 0.0156 > 0.01$$

$$x_3 = 1.5861 \quad f'(x_3) = -0.0156 \quad f''(x_3) = 12.0197 \quad \Rightarrow x_4 = 1.5861 - \frac{-0.0156}{12.0197} = 1.5874$$

**Paso 5.**

$$|f'(x_4)| = -1.2624e - 005 < 0.01$$

La solución obtenida es

$$x^* = x_4 = 1.5874$$

### Método de bisección

Si la función  $f(x) \in \mathcal{C}^1(I)$ , cumple la propiedad 2.10 en  $I$ , entonces el punto de óptimo se alcanzará donde  $f'(x) = 0$ . Si disponemos de ambas funciones,  $f(x)$  y su derivada  $f'(x)$ , entonces podemos utilizar un método de reducción del intervalo utilizando únicamente un punto, en lugar de dos, para identificar el punto donde  $f'(x) = 0$ . Por ejemplo, si en el punto  $z$ , ocurre  $f'(z) < 0$ , ese mínimo no podrá alcanzarse a la izquierda de  $z$ . En otras palabras, podemos prescindir del intervalo  $x \leq z$ . Si ocurre  $f'(z) > 0$ , entonces el mínimo no se va a encontrar a la derecha de  $z$ , es decir, eliminaremos el intervalo  $x \geq z$ . Utilizando esta filosofía podemos construir el método de bisección (o método de búsqueda de Bolzano).

Sirve para funciones  $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ , con derivada continua, y tales que

$$f'(a) \cdot f'(b) < 0$$

Se basa en el cálculo de soluciones de

$$g(x) = f'(x) = 0$$

aplicando el teorema de Bolzano.

Se considera el punto medio del intervalo  $x_m = (a + b)/2$ . Puede ocurrir alguno de los siguientes casos:

**Caso 1.** Si  $f'(x_m) = 0$ , entonces  $\lambda$  es el punto estacionario buscado.

**Caso 2.** Si  $f'(x_m) > 0$ , entonces podemos eliminar el intervalo  $[x_m, b]$  y se repite el proceso tomando como nuevo intervalo  $[a, x_m]$ .

**Caso 3.** Si  $f'(x_m) < 0$ , entonces podemos eliminar el intervalo  $[a, x_m]$  y se repite el proceso tomando como nuevo intervalo  $[x_m, b]$

Como en cada paso se reduce a la mitad la longitud del intervalo de incertidumbre, el algoritmo es lineal de razón 0.5.

Al tratarse de un método de reducción de intervalo, en el que en cada paso reducimos la longitud del intervalo a la mitad, podemos calcular el número de pasos,  $n$ , necesarios para alcanzar una determinada precisión,  $\varepsilon > 0$ . Vemos que si partimos de un intervalo  $[a_0, b_0]$ , de longitud  $L_0$ . Si en ese punto ocurre  $f'(x_m) \neq 0$ , en cualquiera de los casos (2 o 3), nos quedamos con un intervalo,  $[a_1, b_1]$ , de longitud  $L_1 = L_0/2$ . Al cabo de  $n$  pasos consecutivos del algoritmo obtenemos un intervalo  $[a_n, b_n]$  de longitud  $L_n = L/2^n$ . Si cogemos el punto medio de este intervalo final como solución aproximada de la ecuación  $f'(x) = 0$ , estaremos cometiendo un error menor o igual que  $L_n/2$ , i.e., si queremos calcular el punto estacionario con un error  $< \varepsilon$ , entonces el intervalo final tiene que cumplir

$$L_n/2 < \varepsilon \Rightarrow \frac{L}{2^n} < \varepsilon$$

y por tanto tenemos que repetir los pasos del algoritmo un número  $n$  de veces determinado por

$$\log_2 \frac{L}{2^n} < \log_2 \varepsilon \Leftrightarrow \log_2 L - \log_2 2^n < \log_2 \varepsilon \Leftrightarrow \log_2 L - n < \log_2 \varepsilon \Leftrightarrow \log_2 L - \log_2 \varepsilon < n$$

$$n > \log_2 \frac{L}{\varepsilon}$$

*Algoritmo bisección*

- **Paso inicial.** Sea  $[a_0, b_0]$ , y el intervalo inicial de incertidumbre donde tenemos que calcular el punto estacionario. Asumimos que  $f'(a_0) < 0$ , y  $f'(b_0) > 0$ , (en caso contrario podemos cambiar  $f$  por  $-f$ ). Sea  $\varepsilon$  la longitud final del intervalo de incertidumbre. Calculamos  $n$  como el menor entero positivo que verifica  $(\frac{1}{2})^n \leq \varepsilon/(b_0 - a_0)$ . Hacer  $k = 1$  e ir al paso principal.

- **Paso principal.**

**Paso 1.-** Hacer  $\lambda_k = \frac{1}{2}(a_k + b_k)$ , evaluamos  $f'(\lambda_k)$ . Si  $f'(\lambda_k) = 0$ , parar;  $\lambda_k$  es la solución óptima. En caso contrario, si  $f'(\lambda_k) > 0$  ir al **Paso 2**, o ir al **Paso 3** si  $f'(\lambda_k) < 0$ .

**Paso 2.-** Hacer  $a_{k+1} = a_k$  y  $b_{k+1} = \lambda_k$ . Ir al **Paso 4**.

Iteración $k$	$a_k$	$b_k$	$\lambda_k$	$\theta'(\lambda_k)$
1	-3.0000	6.0000	1.5000	5.0000
2	-3.0000	1.5000	-0.7500	0.5000
3	-3.0000	-0.7500	-1.8750	-1.7500
4	-1.8750	-0.7500	-1.3125	-0.6250
5	-1.3125	-0.7500	-1.0313	-0.0625
6	-1.0313	-0.7500	-0.8097	0.2186
7	-1.0313	-0.8097		

Tabla 2.4: Método de bisección.

**Paso 3.-** Hacer  $a_{k+1} = \lambda_k$  y  $b_{k+1} = b_k$ . Ir al **Paso 4**.

**Paso 4.-** Si  $k = n$ , parar; el mínimo cae en el intervalo  $[a_{n+1}, b_{n+1}]$ . En caso contrario, reemplazar  $k$  por  $k + 1$  y repetir el **Paso 1**.

**Ejemplo 2.39** Considerar el siguiente ejemplo

$$\begin{array}{ll} \text{Minimizar} & f(x) = x^2 + 2x \\ \text{Sujeto a} & -3 \leq x \leq 6 \end{array}$$

**Solución:** Supongamos que queremos reducir el intervalo de incertidumbre hasta un intervalo cuya longitud sea como mucho 0.2. Por tanto debemos elegir  $n$  tal que  $\frac{1}{2}^n \leq L_0/(b - a) = 0.2/9 = 0.0222$ . Esto da un valor de  $n = 6$ .

Repetimos el proceso, y presentamos un resumen en la tabla 2.4. Notar que el intervalo final de incertidumbre es  $[-1.0313, -0.8097]$ , por tanto el mínimo puede tomarse como  $-0.961$ .

### Método de la secante

El *método de la secante* combina el método de Newton con un esquema de reducción de intervalo para encontrar, si existe, la raíz de la ecuación  $f'(x) = 0$ , en el intervalo  $(a_0, b_0)$ . Supongamos que en ese intervalo se cumplen las condiciones del método de bisección discutido en la sección anterior, i.e.,  $f'(a_0) \cdot f'(b_0) < 0$ . El algoritmo de la secante aproxima la función  $f'(x)$  como una “línea secante” que une los puntos  $(a_0, f'(a_0))$  y  $(b_0, f'(b_0))$  y determina el siguiente punto en la sucesión donde esta recta se anula.

Partimos de la condición

$$f'(a_0) \cdot f'(b_0) < 0$$

La recta que une los puntos  $(a_0, f'(a_0))$  y  $(b_0, f'(b_0))$ , es

$$\frac{x - b_0}{a_0 - b_0} = \frac{y - f'(b_0)}{f'(a_0) - f'(b_0)}$$

que corta al eje  $x$ , ( $y = 0$ ) en el punto

$$\begin{aligned} \frac{x - b_0}{a_0 - b_0} &= \frac{-f'(b_0)}{f'(a_0) - f'(b_0)} \\ \lambda &= b_0 - \frac{f'(b_0)(b_0 - a_0)}{f'(b_0) - f'(a_0)} \end{aligned}$$

Siendo  $\lambda$  el siguiente punto del algoritmo. Para el siguiente paso se elige el punto  $c$ , de entre  $a$  y  $b$ , que cumpla

$$f(\lambda) \cdot f'(c) < 0$$

**Ejemplo 2.40 (Método de la Secante)** Consideremos de nuevo el problema

$$\text{Minimizar } f(x) = 2x^2 + \frac{16}{x} \quad 1 \leq x \leq 5$$

**Solución.**

$$f'(x) = 4x - \frac{16}{x^2}$$

**1ª Iteración.**

**Paso 1.**

$$\begin{array}{ll} b_0 = 5 & f'(b_0) = 19.36 \\ a_0 = 1 & f'(a_0) = -12 \end{array}$$

**Paso 2.**

$$\lambda = 5 - \frac{19.36 \cdot (5 - 1)}{(19.36 + 12)} = 2.53$$

**Paso 3.**

$$f'(\lambda) = 7.62 \Rightarrow b_1 = 2.53, \quad a_1 = 1$$

**2ª Iteración.**

**Paso 2.**

$$\lambda = 2.53 - \frac{7.62 \cdot (2.53 - 1)}{(7.62 + 12)} = 1.94$$

**Paso 3.**

$$f'(\lambda) = 3.51 \Rightarrow b_2 = 1.94, \quad a_2 = 1$$

Seguir hasta que  $|f'(c)| \leq \varepsilon$ .

## 2.4.2 Optimización sin restricciones multivariante

Describiremos a continuación algunos de los métodos numéricos básicos utilizados en la resolución de problemas no lineales sin restricciones en varias variables.

Prácticamente para todos los métodos existe una estructura común. El algoritmo comienza desde un *punto inicial*, a continuación se determina mediante una regla dada una *dirección de movimiento* y a continuación se busca el mínimo sobre esa dirección para obtener la siguiente aproximación desde la cual se repite el proceso.

Como se puede suponer la diferencia entre los métodos radica en la regla mediante la cual se selecciona la dirección de movimiento en cada paso del algoritmo. El proceso de búsqueda del mínimo sobre la recta se llama *búsqueda lineal*. Una vez fijada la dirección de movimiento, el valor de la función objetivo depende solamente de una variable, que es precisamente la longitud recorrida en la dirección elegida, y sobre esta función “univariante” se pueden utilizar los algoritmos de búsqueda descritos en el tema anterior para funciones de este tipo.

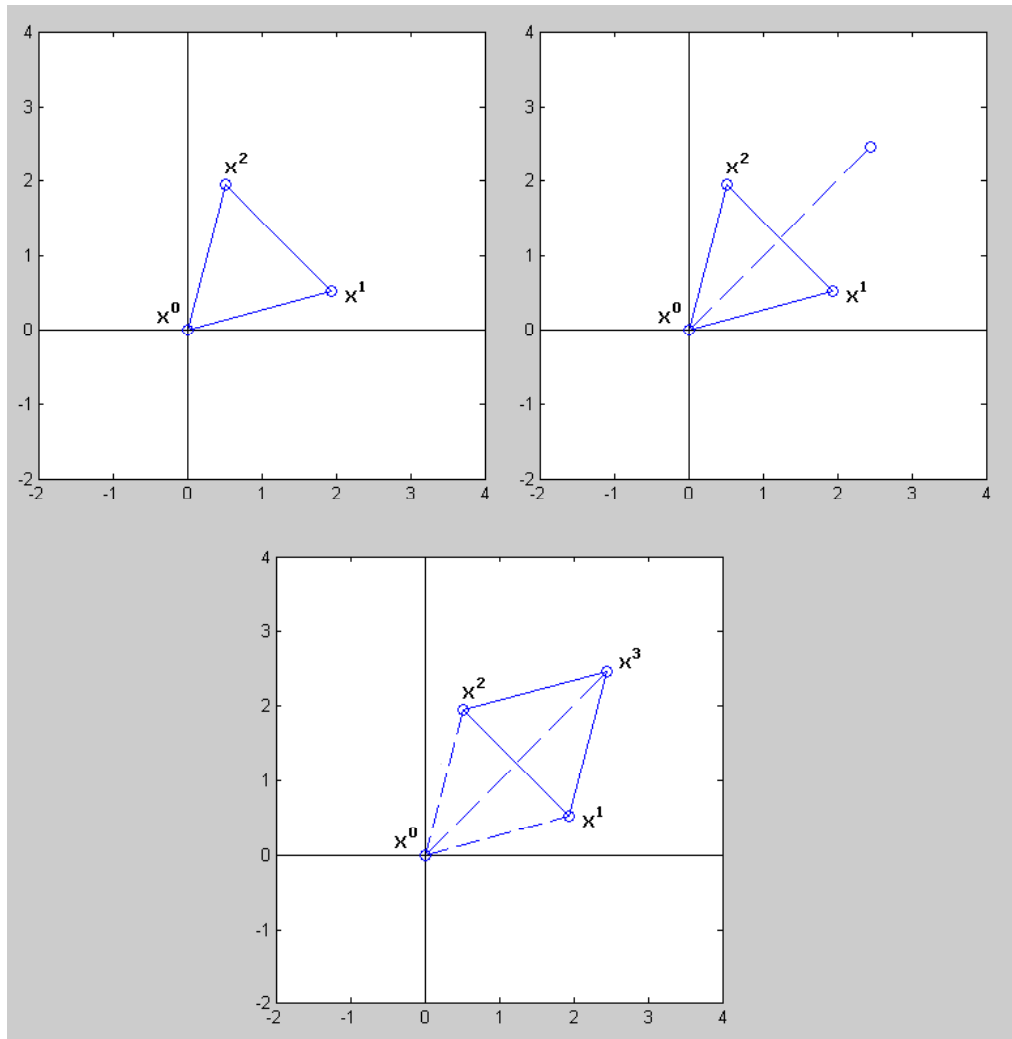
Describiremos en primer lugar un método de búsqueda directa, que solamente utilizan evaluaciones de la función en diferentes puntos. Posteriormente se describirán métodos que utilizan las primeras y segundas derivadas de la función.

### Métodos de búsqueda directa

Para la aplicación de estos métodos solamente es necesario conocer el valor de la función objetivo en cualquier punto del espacio y no es necesaria ninguna hipótesis adicional acerca de la diferenciabilidad de la función. Podemos emplear estos métodos, bien cuando el gradiente de la función,  $\nabla f(\mathbf{x})$ , no exista, no sea conocido o simplemente porque su expresión es demasiado compleja para poder manejarlo con eficacia.

Podemos clasificar los métodos que solamente utilizan evaluaciones de la función en métodos heurísticos y métodos teóricos.

**Método simplex o método  $S^2$**  Es un método de búsqueda heurística desarrollado por Spendley, Hext y Himsworth y está basado en la evaluación de la función en los vértices de un simplex regular. En  $n$  dimensiones, un *simplex regular* es un poliedro compuesto por  $(n + 1)$  puntos equidistantes, estos puntos son los *vértices* del simplex. Por ejemplo, un triángulo equilátero es un simplex en dimensión 2, mientras que un tetraedro es un simplex para  $n = 3$ . La principal propiedad del simplex explotada por este algoritmo es que se puede generar un nuevo simplex proyectando uno cualquiera de sus vértice una determinada distancia a través del centroide de los demás vértices. El nuevo simplex se construye reemplazando el vértice reflejado por el generado mediante este procedimiento de reflexión. De esta forma construimos un nuevo simplex utilizando solamente una nueva evaluación de la función objetivo. En la figura ?? se puede observar el proceso para 2 dimensiones.



**Figura 2.9:**

El método comienza construyendo un primer simplex regular en el espacio de las variables independiente y evaluando la función en cada uno de sus vértices. Se localiza el “peor” vértice, es decir, el vértice en el que la función objetivo toma el valor mayor de entre todos los vértices del Simplex. Este vértice se “refleja” a través del centroide de los demás para generar un nuevo punto. Este punto junto con los restantes constituyen los vértices del nuevo simplex. El proceso se repite hasta que se obtiene el mínimo o se entra en un ciclo entre 2 o más simplex, es decir, volvemos a obtener un simplex que ya habíamos construido previamente. Esta situación se puede resolver siguiendo tres reglas:

**Regla 1.** *Acotamiento mínimo*



Si el vértice que hay eliminar (por reflexión) ha sido generado en la iteración previa, entonces se elige en su lugar el segundo “peor” vértice.

**Regla 2. Ciclado**

Si alguno de los vértices permanece invariante después de  $M$  iteraciones, se reduce el tamaño del simplex mediante un factor determinado. El nuevo simplex se construye a partir del vértice en el que la función alcanza el valor más pequeño como punto base. Los trabajos de Spendley y colaboradores sugieren un valor para  $M$  de

$$M = 1.65n + 0.05n^2$$

donde  $n$  es la dimensión del problema, y  $M$  se obtiene por redondeo al entero más cercano. Obviamente, esta regla requiere la especificación de un factor de reducción.

**Regla 3. Criterio de finalización**

La búsqueda termina cuando el simplex se hace suficientemente pequeño o cuando la desviación estándar de los valores de la función en los vértices se hace suficientemente pequeña, es decir

$$S_f = \sqrt{\frac{\sum_{i=0}^n (f(x_i) - \bar{f})^2}{n+1}} = \sqrt{\frac{\sum_{i=0}^n f(x_i)^2}{n+1} - \bar{f}^2} < \varepsilon$$

donde

$$\bar{f} = \frac{\sum_{i=0}^n f(x_i)}{n+1}$$

Esta regla requiere la especificación de un parámetro de finalización:  $\varepsilon$

La implementación de este algoritmo requiere sólo dos tipos de cálculos: (1) la generación de un simplex regular a partir de un punto base inicial y un factor de escala apropiado, y (2) el cálculo del punto reflejado. El primer tipo de cálculo es muy fácil de realizar, puesto que puede demostrarse por geometría elemental que a partir de un punto cualquiera  $\mathbf{x}^{(0)}$  y un factor de escala  $\alpha$  (longitud del lado del simplex), se pueden generar los otros  $n$  vértices del simplex regular (en dimensión  $n$ ) utilizando las expresiones:

$$\mathbf{x}_j^{(i)} = \begin{cases} \mathbf{x}_j^{(0)} + \delta_1 & j = i \\ \mathbf{x}_j^{(0)} + \delta_2 & j \neq i \end{cases}$$

para  $i, j = 1, 2, 3, \dots, n$ . Los valores de  $\delta_1$  y  $\delta_2$  dependen solamente de  $n$  y  $\alpha$  a través de las expresiones:

$$\delta_1 = \left[ \frac{(n+1)^{1/2} + n - 1}{n\sqrt{2}} \right] \alpha$$

$$\delta_2 = \left[ \frac{(n+1)^{1/2} - 1}{n\sqrt{2}} \right] \alpha$$

El segundo tipo de cálculo, reflexión a través del centroide, es también muy sencillo de realizar. Supongamos que  $\mathbf{x}^{(j)}$  es el punto que hay que reflejar. Entonces el centroide de los restantes  $n$  puntos es

$$\mathbf{x}_c = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq j}}^n \mathbf{x}^{(i)}$$

Cualquier punto en la línea que une  $\mathbf{x}^{(j)}$  y  $\mathbf{x}_c$  viene dado por

$$\mathbf{x} = \mathbf{x}^{(j)} + \lambda (\mathbf{x}_c - \mathbf{x}^{(j)})$$

Donde  $\lambda \in \mathbb{R}$ . Por ejemplo, para  $\lambda = 0$  tenemos el punto original  $\mathbf{x}^{(j)}$ , mientras que para  $\lambda = 1$  obtenemos el centroide,  $\mathbf{x}_c$ . Para mantener la regularidad del simplex, la reflexión debe ser simétrica y por tanto el valor buscado es  $\lambda = 2$ . Así

$$\mathbf{x}_{\text{nuevo}}^{(j)} = 2\mathbf{x}_c - \mathbf{x}_{\text{antiguo}}^{(j)}$$

Veamos como ilustración del método, ambos cálculos en el siguiente ejemplo:

**Ejemplo 2.41** Utiliza el método  $S^2$  para resolver el problema

$$\text{Minimizar } f(x) = (1 - x_1)^2 + (2 - x_2)^2$$

**Solución:** La construcción del simplex inicial requiere la especificación de un punto inicial y de un factor de escala. Supongamos  $\mathbf{x}^{(0)} = (0, 0)$  y  $\alpha = 2$ . Entonces

$$\delta_1 = \left[ \frac{\sqrt{3} + 1}{2\sqrt{2}} \right] \cdot 2 = 1.9319$$

$$\delta_2 = \left[ \frac{\sqrt{3} - 1}{2\sqrt{2}} \right] \cdot 2 = 0.5176$$

Con esos valores, los dos vértices restantes serán

$$\begin{aligned} \mathbf{x}^{(2)} &= \begin{bmatrix} 0 + 0.5176 \\ 0 + 1.9319 \end{bmatrix} = \begin{bmatrix} 0.5176 \\ 1.9319 \end{bmatrix} \\ \mathbf{x}^{(1)} &= \begin{bmatrix} 0 + 1.9319 \\ 0 + 0.5176 \end{bmatrix} = \begin{bmatrix} 1.9319 \\ 0.5176 \end{bmatrix} \end{aligned}$$

La función objetivo  $f(x, y)$  en esos puntos toma los siguientes valores

$$\begin{aligned} f(x^{(0)}) &= 5 \\ f(x^{(1)}) &= 3.0658 \\ f(x^{(2)}) &= 0.2374 \end{aligned}$$

Por tanto  $\mathbf{x}^{(0)}$  es el punto que tiene que ser reflejado para formar el nuevo simplex. El punto  $\mathbf{x}^{(3)}$ , se calcula de la siguiente forma

$$\begin{aligned} \mathbf{x}_c &= \frac{1}{2} \sum_{i=1}^2 \mathbf{x}^{(i)} = \frac{1}{2} (x^{(1)} + x^{(2)}) \\ \mathbf{x}^{(3)} &= 2\mathbf{x}_c - \mathbf{x}^{(0)} = \mathbf{x}^{(1)} + \mathbf{x}^{(2)} - \mathbf{x}^{(0)} = \begin{bmatrix} 2.4495 \\ 2.4495 \end{bmatrix} \end{aligned}$$

En el nuevo punto  $f(\mathbf{x}^{(3)}) = 2.3027$  y por tanto se mejora el resultado. El nuevo simplex estaría compuesto por los puntos:  $\mathbf{x}^{(1)}$ ,  $\mathbf{x}^{(2)}$  y  $\mathbf{x}^{(3)}$ . El algoritmo continuaría ahora reflejando el punto donde la función toma el mayor valor:  $\mathbf{x}^{(1)}$ . El algoritmo se reproduce como anteriormente, salvo cuando nos encontramos con situaciones que requieren las reglas 1, 2 o 3, descritas con anterioridad.

El algoritmo  $S^2$  dado anteriormente tiene una serie de ventajas:

1. Los cálculos son muy simples y la lógica del método no es complicada, por tanto el programa será corto.
2. El almacenamiento es relativamente pequeño: un vector de dimensiones  $(n + 1, n + 2)$
3. Se necesitan pocos parámetros modificables: el factor de escala  $\alpha$ , un factor para reducir  $\alpha$  si se utiliza la regla 2, y un criterio de parada.
4. El algoritmo es efectivo cuando las evaluaciones de los errores son significativas, puesto que opera en el peor caso en lugar del mejor punto.

Por otra parte el algoritmo tiene también varias e importantes desventajas:

1. Se pueden producir problemas de escala puesto que todas las coordenadas están sujetas al mismo factor de escala  $\alpha$ . Para disminuir este efecto, se pueden escalar todas las variables de forma que tengan magnitudes comparables.
2. El algoritmo es lento, puesto que no se utiliza información ya conocida para acelerar el movimiento.
3. No hay ninguna forma simple de expandir el simplex sin volver a calcular todo el patrón. Así, una vez que se reduce  $\alpha$ , la búsqueda debe comenzar con este tamaño de paso reducido.

Para eliminar parcialmente algunas de las desventajas de este método, Nelder y Mead modificaron el procedimiento del simplex, observando que aunque es conveniente utilizar la fórmula de construcción del simplex regular en el estado inicial del método, no hay por qué mantener la regularidad del simplex cuando se sigue con el procedimiento de búsqueda. De esta forma, se pueden efectuar expansiones y contracciones en el proceso de reflexión. Esta modificación requiere la utilización del punto con el mayor valor para  $f(\mathbf{x})$ ,  $\mathbf{x}^h$ , el siguiente con mayor valor para  $f(\mathbf{x})$ ,  $\mathbf{x}^g$ , y el punto con el menor valor para  $f(\mathbf{x})$ ,  $\mathbf{x}^l$ , junto con los correspondientes valores de la función en dichos puntos  $f^h$ ,  $f^g$  y  $f^l$ . Teniendo en cuenta esta información y que el paso de reflexión viene dado mediante la fórmula

$$\mathbf{x} = \mathbf{x}^h + \lambda (\mathbf{x}_c - \mathbf{x}^h)$$

o equivalentemente por

$$\mathbf{x} = \mathbf{x}^h + (1 + \theta) (\mathbf{x}_c - \mathbf{x}^h)$$

Si  $\theta = 1$ , tenemos la reflexión normal. Cuando  $-1 \leq \theta \leq 1$ , se produce una reflexión más corta o *contracción*, mientras que si  $\theta > 1$ , se generará un paso de reflexión de mayor longitud o una *expansión* del simplex. Podemos observar esta propiedad en la figura 2.10. Los tres valores de  $\theta$  utilizados para reflexión normal, contracción y expansión se indican respectivamente como  $\alpha$ ,  $\beta$  y  $\gamma$ , y se utilizan según los criterios indicados en la figura. Los cálculos de Nelder y Mead recomiendan los valores de  $\alpha = 1$ ,  $\beta = 0.5$  y  $\gamma = 2$  respectivamente.

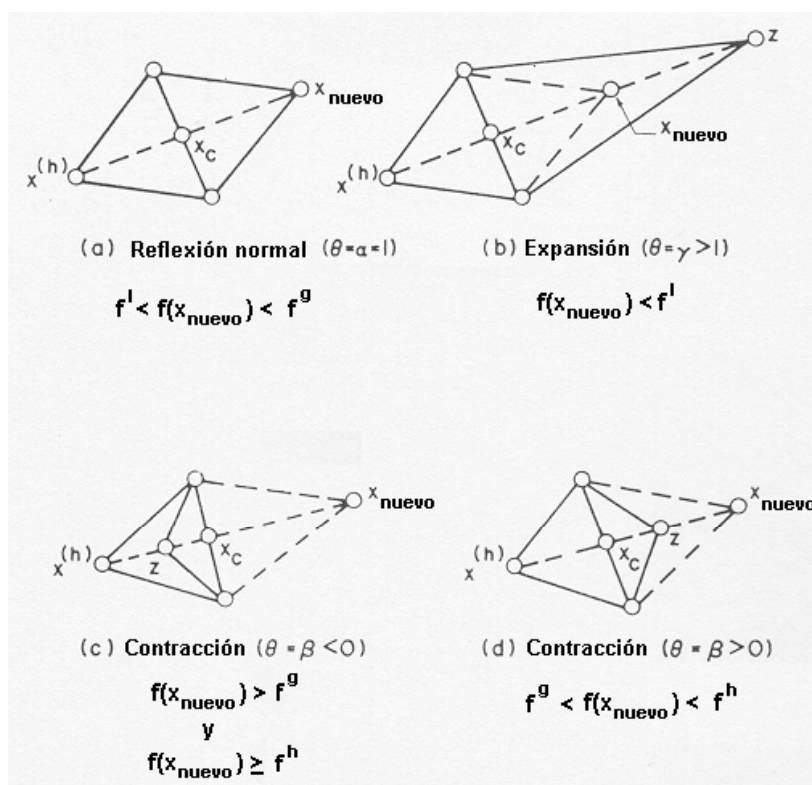


Figura 2.10: Método Simplex de Nelder-Mead

**Métodos de descenso coordenado** Como se ha comentado al principio de la sección la mayoría de los métodos (salvo el método Simplex) que vamos a ver se fundamentan en encontrar el mínimo de la función objetivo sobre una dirección determinada. Por tanto, dada la función  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , y si  $\mathbf{x}^k$  es la aproximación  $k$ -ésima al mínimo exacto  $\mathbf{x}^*$ , se elegirá una dirección  $\mathbf{d}^k \in \mathbb{R}^n$  de búsqueda a partir de esta aproximación y se buscará el mínimo de  $f(\mathbf{x})$  sobre esa dirección para conseguir la aproximación  $(k+1)$ -ésima,  $\mathbf{x}^{k+1}$ . Se trata de resolver el problema

$$\underset{\alpha}{\text{Minimizar}} f(\mathbf{x}^k + \alpha \mathbf{d}^k) = g(\alpha)$$

donde la única incógnita es  $\alpha$ , que será el desplazamiento desde  $\mathbf{x}^k$  en la dirección  $\mathbf{d}^k$  para encontrar a  $\mathbf{x}^{k+1}$ . Este problema es de una variable y podemos emplear los métodos de la sección correspondiente. Posteriormente se utilizará el punto  $\mathbf{x}^{k+1}$  como punto de partida para conseguir la siguiente aproximación sobre la dirección indicada en el método.

El método de descenso coordenado es el método más sencillo para elegir la dirección de búsqueda. Si  $\mathbf{x}^k = (x_1^k, \dots, x_i^k, \dots, x_n^k)$ , el descenso respecto a la coordenada  $x_i$ , significa resolver el problema

$$\underset{\alpha}{\text{Minimizar}} g_i(\alpha) = f(x_1^k, \dots, x_i^k + \alpha, \dots, x_n^k)$$

y por tanto este método utiliza los ejes coordenados como direcciones de búsqueda. El método busca el mínimo de  $f$  en las direcciones coordenadas

$$\mathbf{e}^i = \begin{pmatrix} 0, \dots, \overbrace{1}^i, \dots, 0 \end{pmatrix}$$

sólo hay variaciones en la coordenada  $x_i$ , mientras que las demás permanecen fijas. Si minimizamos de forma secuencial para diversas componentes de  $f(\mathbf{x})$  podríamos llegar a un mínimo relativo de  $f(\mathbf{x})$ .

En el algoritmo de los métodos de descenso coordenado una iteración completa del método termina cuando se han recorrido todas las direcciones coordenadas posibles. Una iteración, en general no acaba después de minimizar en una única dirección, sino después de hacerlo sobre un conjunto de direcciones coordenadas establecidas, de entre las que destacamos las siguientes:

1. **Descenso coordenado cíclico.** Minimiza la función  $f(\mathbf{x})$  en cada dirección de forma ordenada, desde la coordenada 1 (dirección  $\mathbf{e}^1$ ) hasta la coordenada  $n$  (dirección  $\mathbf{e}^n$ ). Posteriormente se repite el proceso comenzando de nuevo desde la primera coordenada.
2. **Doble recorrido de Aitken.** En esta variante del método de descenso coordenado, se minimiza la función en todas las direcciones coordenadas,  $\mathbf{e}^1, \dots, \mathbf{e}^n$ , para a continuación volver a minimizar hacia atrás, desde la dirección  $\mathbf{e}_{n-1}$  hasta la  $\mathbf{e}_1$ . En este método se requieren  $2(n-1)$  búsquedas lineales en cada iteración. Para  $n=2$  este método y el anterior coinciden.

A continuación se indica el procedimiento para el algoritmo del método de descenso cíclico

*Algoritmo del método de descenso coordenado cíclico*

- **Inicio.** Elegir un escalar  $\epsilon > 0$  para utilizarlo como criterio de parada del algoritmo, y  $\mathbf{e}^1, \dots, \mathbf{e}^n$  como las direcciones de búsqueda. Elegir un punto inicial  $\mathbf{x}^0$ . Hacer  $\mathbf{y}^1 = \mathbf{x}^0$ ,  $k=0$  y  $j=1$  e ir al paso principal.
- **Principal.**

**Paso 1.-** Calcular  $\alpha_j$  como la solución óptima del problema de minimizar  $f(\mathbf{y}^j + \alpha \mathbf{e}^j)$ ,  $\alpha \in \mathbb{R}$ , hacer  $\mathbf{y}^{j+1} = \mathbf{y}^j + \alpha_j \mathbf{e}^j$ . Si  $j < n$ , se reemplaza  $j$  por  $j+1$  y se repite el paso 1. En otro caso, si  $j = n$  ir al Paso 2.

**Paso 2.-** Hacer  $\mathbf{x}^{k+1} = \mathbf{y}^{n+1}$ . Si  $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| < \epsilon$ , parar. En otro caso, se define  $\mathbf{y}^1 = \mathbf{x}^{k+1}$  y  $j=1$ , se reemplaza  $k$  por  $k+1$ . Se va al Paso 1.

**Ejemplo 2.42** Considerar el siguiente problema

$$\text{Minimizar } (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

Iteración $k$	$\mathbf{x}_k^T$ $f(\mathbf{x}_k)$	$j$	$\mathbf{d}_j^T$	$\mathbf{y}_j^T$	$\alpha_j$	$\mathbf{y}_{j+1}^T$
1	(0.00, 3.00)	1	(1.0, 0.0)	(0.00, 3.00)	3.13	(3.13, 3.00)
	52.00	2	(0.0, 1.0)	(3.13, 3.00)	-1.44	(3.13, 1.56)
2	(3.13, 1.56)	1	(1.0, 0.0)	(3.13, 1.56)	-0.50	(2.63, 1.56)
	1.63	2	(0.0, 1.0)	(2.63, 1.56)	-0.25	(2.63, 1.31)
3	(2.63, 1.31)	1	(1.0, 0.0)	(2.63, 1.31)	-0.19	(2.44, 1.31)
	0.16	2	(0.0, 1.0)	(2.44, 1.31)	-0.09	(2.44, 1.22)
4	(2.44, 1.22)	1	(1.0, 0.0)	(2.44, 1.22)	-0.09	(2.35, 1.22)
	0.04	2	(0.0, 1.0)	(2.35, 1.22)	-0.05	(2.35, 1.17)
5	(2.35, 1.17)	1	(1.0, 0.0)	(2.35, 1.17)	-0.06	(2.29, 1.17)
	0.015	2	(0.0, 1.0)	(2.29, 1.17)	-0.03	(2.29, 1, 14)
6	(2.29, 1, 14)	1	(1.0, 0.0)	(2.29, 1, 14)	-0.04	(2.25, 1.14)
	0.007	2	(0.0, 1.0)	(2.25, 1.14)	-0.02	(2.25, 1.12)
7	(2.25, 1.12)	1	(1.0, 0.0)	(2.25, 1.12)	-0.03	(2.22, 1.12)
	0.004	2	(0.0, 1.0)	(2.22, 1.12)	-0.01	(2.22, 1.11)

Tabla 2.5: Método Coordinado Cíclico

**Solución:** La solución a este problema es el punto (2, 1) con un valor de la función objetivo igual a 0. El resumen de los cálculos mediante el método coordinado cíclico comenzando desde el punto (0.00, 3.00) se da en la tabla 2.5. Notar que en cada iteración, los vectores  $\mathbf{y}^2$  e  $\mathbf{y}^3$  se obtienen desarrollando una búsqueda lineal en las direcciones (1, 0) y (0, 1), respectivamente. Después de 7 iteraciones, alcanzamos el punto (2.22, 1.11), con un valor de 0.0023 para la función objetivo.

En la figura 2.11 se han dibujado los isocontornos de la función  $f(x_1, x_2) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$ , y algunos de los puntos generados mediante el método de coordenadas cíclico aplicado a esta función. Notar que el avance significativo se experimenta en las primeras iteraciones, mientras que en las posteriores iteraciones el progreso se ralentiza.

### Métodos basados en el gradiente

Los métodos desarrollados en esta sección utilizan la información proporcionada por el gradiente de la función, y en algunos casos la información dada por su matriz Hessiana, por tanto, para aplicarlos necesitamos que la función sea diferenciable e incluso dos veces diferenciable.

Discutimos en primer lugar métodos que solamente utilizan información de primer orden y asumiremos que podemos obtener los elementos del gradiente,  $\nabla f(\mathbf{x})$ , de forma directa o mediante aproximación numérica.

Todos los métodos considerados aquí emplean un procedimiento iterativo similar, según la fórmula de recursiva

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}[\mathbf{x}^{(k)}] \quad (2.13)$$

donde  $\mathbf{x}^{(k)}$  es la estimación actual de la solución  $\mathbf{x}^*$  del problema;  $\alpha_k$  es un parámetro que determina la longitud de paso y  $\mathbf{d}[\mathbf{x}^{(k)}] = \mathbf{d}^k$  es la dirección de búsqueda elegida en el espacio  $n$ -dimensional de las variables de decisión.

**Método de descenso de mayor pendiente o del gradiente** Este método es el de mayor importancia teórica debido a que es uno de los más sencillos para los que existe un análisis satisfactorio. A menudo se pueden conseguir algoritmos más avanzados modificando la técnica de descenso de mayor pendiente.

Sea  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ;  $f \in \mathcal{C}^1(\mathbb{R}^n)$ , el método de mayor pendiente o de gradiente se fundamenta en la construcción de una sucesión  $\{\mathbf{x}^k\}$  que sea convergente a la solución del problema del siguiente modo: la recta que pasa por

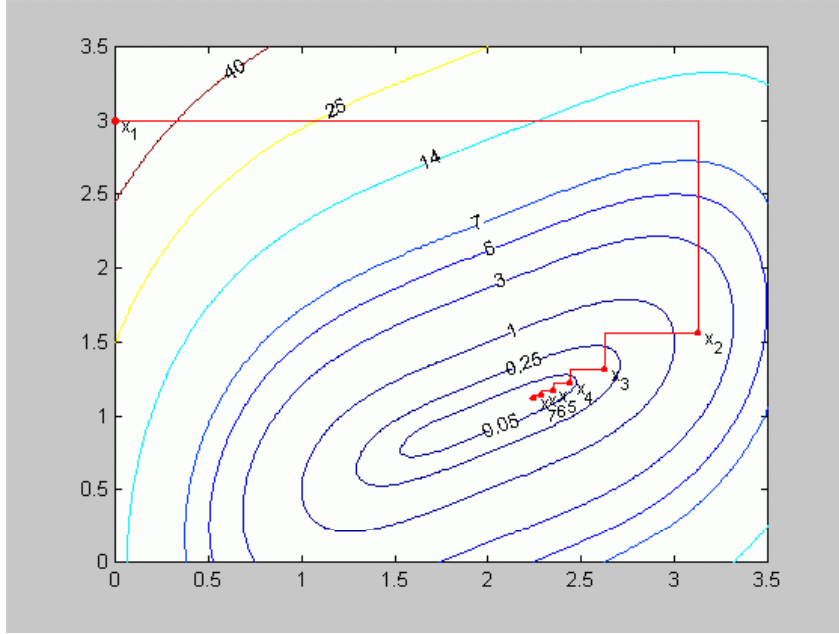


Figura 2.11: Método de coordenadas cíclico

$\mathbf{x}_k$  y  $\mathbf{x}_{k+1}$  tiene que tener la dirección de máximo descenso en  $\mathbf{x}^k$ . Si  $\mathbf{d} \in \mathbb{R}^n$  es una dirección de descenso en  $\mathbf{x}^k$ , entonces por el teorema de caracterización de las mismas (puesto que  $f(\mathbf{x})$  es diferenciable), tenemos

$$\frac{f(\mathbf{x}^k + \alpha \mathbf{d}) - f(\mathbf{x}^k)}{\alpha} \approx \nabla f(\mathbf{x}^k)^T \mathbf{d} < 0$$

y si queremos maximizar la pendiente (término izquierdo), tendremos que resolver el problema

$$\text{Minimizar } \nabla f(\mathbf{x}^k)^T \mathbf{d}$$

para todas las direcciones  $\mathbf{d}$ , de descenso.

Como estamos interesados en la dirección, tomaremos  $\mathbf{d}$  un vector unitario, es decir,  $\|\mathbf{d}\| = 1$ .

$$\nabla f(\mathbf{x}^k)^T \mathbf{d} < 0 \Rightarrow -\nabla f(\mathbf{x}^k)^T \mathbf{d} = |\nabla f(\mathbf{x}^k)^T \mathbf{d}|$$

y aplicando la desigualdad de Schwartz

$$|\nabla f(\mathbf{x}^k)^T \mathbf{d}| \leq \|\nabla f(\mathbf{x}^k)\| \cdot \|\mathbf{d}\| = \|\nabla f(\mathbf{x}^k)\|$$

y por tanto

$$-\nabla f(\mathbf{x}^k)^T \mathbf{d} \leq \|\nabla f(\mathbf{x}^k)\|$$

multiplicando por  $-1$

$$\nabla f(\mathbf{x}^k)^T \mathbf{d} \geq -\|\nabla f(\mathbf{x}^k)\|$$

y la igualdad se obtiene cuando

$$\mathbf{d} = -\frac{\nabla f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|}$$

La recta que une  $\mathbf{x}^k$  y  $\mathbf{x}^{k+1}$  es tangente a la curva de nivel de  $f(\mathbf{x})$  en  $\mathbf{x}^k$ ; se define  $\mathbf{x}^{k+1}$  como:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k \nabla f(\mathbf{x}^k) \quad \alpha_k \geq 0 \quad (2.14)$$

El problema que se plantea consiste en determinar el  $\alpha_k$  adecuado. En general para que la sucesión “siga” la curva de máxima pendiente  $\alpha_k$  debe ser pequeño, lo que ralentiza el algoritmo, mientras que si  $\alpha_k$  es grande, pueden producirse grandes desviaciones sobre dicha curva de máxima pendiente.

El valor de  $\alpha_k$  se obtiene al minimizar la función  $f(\mathbf{x})$ , en la dirección que indica el gradiente de  $f(\mathbf{x})$  en el punto  $\mathbf{x}^k$ , de modo que en el punto  $\mathbf{x}^{k+1}$  se alcance el mínimo de  $f(\mathbf{x})$  sobre la recta normal de ecuación

$$\mathbf{x} = \mathbf{x}^k - \alpha \nabla f(\mathbf{x}^k) \quad \alpha \geq 0$$

Es decir, elegimos  $\alpha_k$  de forma que minimice la curva  $g(\alpha) = f(\mathbf{x}^k - \alpha \nabla f(\mathbf{x}^k))$  respecto a  $\alpha$ .

El método se aproxima al mínimo de una forma lenta, puesto que el cambio en las variables está directamente relacionado con la magnitud del gradiente, que tiende a 0. No hay ningún mecanismo para producir una aceleración hacia el mínimo en las iteraciones finales. La mejor ventaja del método de Cauchy es la propiedad de descenso que tiene, puesto que por la naturaleza del método en cada iteración aseguramos que

$$f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$$

Este método también se denomina *método de Cauchy*.

**Ejemplo 2.43** Considera la función

$$f(\mathbf{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$

obtener el mínimo de  $f(\mathbf{x})$  utilizando el método de Cauchy

**Solución:** Para comenzar, calculamos las componentes del gradiente de  $f(\mathbf{x})$

$$\frac{\partial f}{\partial x_1} = 16x_1 + 4x_2$$

$$\frac{\partial f}{\partial x_2} = 4x_1 + 10x_2$$

Ahora para emplear el método de la máxima pendiente, utilizamos  $\mathbf{x}^{(0)}$  como aproximación inicial de la solución, donde:

$$\mathbf{x}^{(0)} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

y generamos una nueva aproximación mediante la fórmula 2.14

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)}) \\ &= \begin{bmatrix} 10 \\ 10 \end{bmatrix} - \alpha \begin{bmatrix} 200 \\ 140 \end{bmatrix} \\ &= \begin{bmatrix} 10 - 200\alpha \\ 10 - 140\alpha \end{bmatrix} \end{aligned}$$

El valor del parámetro  $\alpha_0$  se obtiene minimizando en la dirección  $\nabla f(\mathbf{x}^0)$ , de forma que:

$$f(\mathbf{x}^1) = \min_{\alpha \geq 0} f(\mathbf{x}^0 - \alpha \nabla f(\mathbf{x}^0))$$

$$\begin{aligned} f(\mathbf{x}^0 - \alpha \nabla f(\mathbf{x}^0)) &= f(10 - 200\alpha, 10 - 140\alpha) \\ &= 1700 - 59600\alpha + 530000\alpha^2 \end{aligned}$$

que es una función cuadrática unidimensional cuyo mínimo puede obtenerse fácilmente por derivación

$$f'(\alpha) = 1060000\alpha - 59600$$

$$f'(\alpha) = 0 \iff \alpha_0 = \frac{59600}{1060000} = 0.056$$

por tanto

$$\mathbf{x}^1 = \begin{bmatrix} -1.20 \\ 2.16 \end{bmatrix}$$

El paso siguiente es calcular  $\mathbf{x}^2$

$$\mathbf{x}^2 = \mathbf{x}^1 - \alpha_1 \nabla f(\mathbf{x}^1)$$

evaluando el gradiente en  $\mathbf{x}^1$  y realizando la búsqueda lineal.

Los valores obtenidos para los 4 primeras iteraciones son los siguientes:

$k$	$x_1^k$	$x_2^k$	$f(\mathbf{x}^k)$
1	-1.2403	2.1181	24.2300
2	0.1441	0.1447	0.3540
3	-0.0181	0.0309	0.0052
4	0.0021	0.0021	0.0000

Dado un punto  $\mathbf{x}$ , el algoritmo del método del gradiente o de la máxima pendiente se construye desarrollando una búsqueda lineal a lo largo de la dirección  $\frac{-\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ , o equivalentemente, a lo largo de la dirección  $-\nabla f(\mathbf{x})$ . El algoritmo asociado a este método podría realizarse como sigue:

#### *Algoritmo del método del Gradiente*

- **Inicialización.** Elegir  $\epsilon > 0$  un escalar que determina la terminación del método. Elegir un punto de partida  $\mathbf{x}^0$ , hacer  $k = 0$  e ir al paso principal.
- **Principal.** Si  $\|\nabla f(\mathbf{x}^k)\| < \epsilon$ , parar; en otro caso hacer  $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ , y calcular  $\alpha_k$  como el valor que minimiza la función univariante  $f(\mathbf{x}^k + \alpha \mathbf{d}^k)$  con  $\alpha \geq 0$ . Hacer  $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$ , reemplazar  $k$  por  $k + 1$  y repetir el paso principal.

Debido a la lentitud en la convergencia de este método, este se puede modificar de forma que limitemos el número de iteraciones realizadas, introduciendo para ello un nuevo parámetro,  $M$ , en el paso principal que controle si dicho límite es superado o no.

**Ejemplo 2.44** *Considera el siguiente problema*

$$\text{Minimizar } (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

**Solución:** Resolveremos el problema utilizando el método del gradiente, comenzando con el punto  $\mathbf{x}^0 = (0.00, 3.00)$ . La tabla 2.6 muestra un resumen de los cálculos. Después de 7 iteraciones, alcanzamos el punto  $\mathbf{x}^8 = (2.28, 1.15)$ . El algoritmo termina puesto que  $\|\nabla f(\mathbf{x}^8)\| = 0.09$  es pequeño. Podemos observar la evolución del método en la figura 2.12. Notar que la solución de este problema es  $(2.00, 1.00)$ .



Iteración $k$	$\mathbf{x}^k$	$f(\mathbf{x}^k)$	$\nabla f(\mathbf{x}^k)$	$\ \nabla f(\mathbf{x}^k)\ $	$\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$	$\alpha_k$	$\mathbf{x}^{k+1}$
1	(0.00, 3.00)	52.00	(-44.00, 24.00)	50.12	(44.00, -24.00)	0.062	(2.70, 1.51)
2	(2.70, 1.51)	0.34	(0.73, 1.28)	0.147	(-0.73, -1.28)	0.24	(2.52, 1.20)
3	(2.52, 1.20)	0.09	(0.80, -0.48)	0.93	(-0.80, 0.48)	0.11	(2.43, 1.25)
4	(2.43, 1.25)	0.04	(0.18, 0.28)	0.33	(-0.18, -0.28)	0.31	(2.37, 1.16)
5	(2.37, 1.16)	0.02	(0.30, -0.20)	0.36	(-0.30, 0.20)	0.12	(2.33, 1.18)
6	(2.33, 1.18)	0.01	(0.08, 0.12)	0.14	(-0.08, -0.12)	0.36	(2.30, 1.14)
7	(2.30, 1.14)	0.009	(0.15, -0.08)	0.17	(-0.15, 0.08)	0.13	(2.28, 1.15)
8	(2.28, 1.15)	0.007	(0.05, 0.08)	0.09			

Tabla 2.6: Método del gradiente

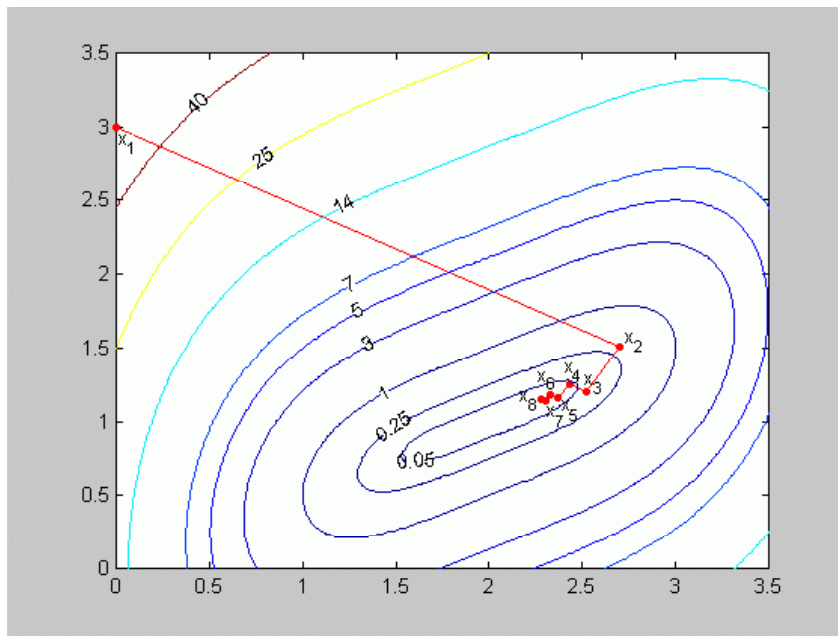


Figura 2.12: Método del Gradiente

**Método de Newton** Como en el caso univariante, el método de Newton consistirá en la aproximación a la función  $f$  mediante una función cuadrática en  $\mathbf{x}^k$  y calcular  $\mathbf{x}^{k+1}$  como el mínimo exacto de esta aproximación. Si tomamos como aproximación de  $f(\mathbf{x})$  su desarrollo en series de Taylor hasta el segundo orden en  $\mathbf{x}^k$  obtenemos la expresión:

$$f(\mathbf{x}) \simeq Q(\mathbf{x}) = f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T (\mathbf{x} - \mathbf{x}^k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^T \mathbf{H} f(\mathbf{x}^k) (\mathbf{x} - \mathbf{x}^k)$$

donde se ha supuesto que la función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $\mathcal{C}^2(\mathbb{R}^n)$ . Ahora la aproximación  $(k+1)$ -ésima es el punto estacionario de esta aproximación y por tanto

$$\nabla Q(\mathbf{x}^{k+1}) = 0 \Leftrightarrow \nabla f(\mathbf{x}^k) + \mathbf{H} f(\mathbf{x}^k) (\mathbf{x}^{k+1} - \mathbf{x}^k) = 0$$

y si  $\mathbf{H}f(\mathbf{x}_k)$  es una matriz regular ( $|\mathbf{H}f(\mathbf{x}_k)| \neq 0$ ), entonces podemos despejar  $\mathbf{x}^{k+1}$  premultiplicando por  $[\mathbf{H}f(\mathbf{x}_k)]^{-1}$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{H}f(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k) \quad (2.15)$$

Teniendo en cuenta las condiciones de suficiencia de 2º orden para que  $\mathbf{x}^*$  sea un punto de mínimo relativo de la función  $f(\mathbf{x})$ , la matriz  $\mathbf{H}f(\mathbf{x}^*)$  debe ser definida positiva  $\Rightarrow$  Si  $f \in \mathcal{C}^2(\mathbb{R}^n)$  y  $\mathbf{H}f(\mathbf{x}^*)$  es definida positiva (o al menos semidefinida positiva) el método convergerá cerca de la solución.

**Ejemplo 2.45** Consideremos la siguiente función

$$f(\mathbf{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$

donde su gradiente es:

$$\nabla f(\mathbf{x}) = (16x_1 + 4x_2, 4x_1 + 10x_2)$$

y su hessiano:

$$\mathbf{H}f(\mathbf{x}) = \begin{bmatrix} 16 & 4 \\ 4 & 10 \end{bmatrix}$$

Resuelve el problema de la minimización de  $f(\mathbf{x})$  utilizando el método de Newton

**Solución:** Utilizamos la expresión 2.15, con  $\mathbf{x}^0 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$ , obtenemos

$$\mathbf{x}^1 = \begin{bmatrix} 10 \\ 10 \end{bmatrix} - \left( \frac{1}{144} \right) \begin{bmatrix} 10 & -4 \\ -4 & 16 \end{bmatrix} \begin{bmatrix} 200 \\ 140 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \end{bmatrix} - \left( \frac{1}{144} \right) \begin{bmatrix} 1440 \\ 1440 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

que es la solución exacta del problema. Este resultado no es sorprendente, ya que, como se esperaba para el método de Newton aplicado a una función cuadrática, el método es exacto.

**Ejemplo 2.46** Considerar el siguiente problema

$$\text{Minimizar } (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

**Solución:** El resumen de los cálculos mediante el método de Newton están dados en la tabla 2.7. En cada iteración,  $\mathbf{x}^{k+1}$  está dado mediante la ecuación 2.15.

Después de 6 iteraciones, alcanzamos el punto  $\mathbf{x}_7 = (1.83, 0.91)^T$ . En este punto  $\|\nabla f(\mathbf{x}_7)\| = 0.04$ , y el procedimiento termina. En la figura 2.13 se presentan los puntos generados por el método de Newton

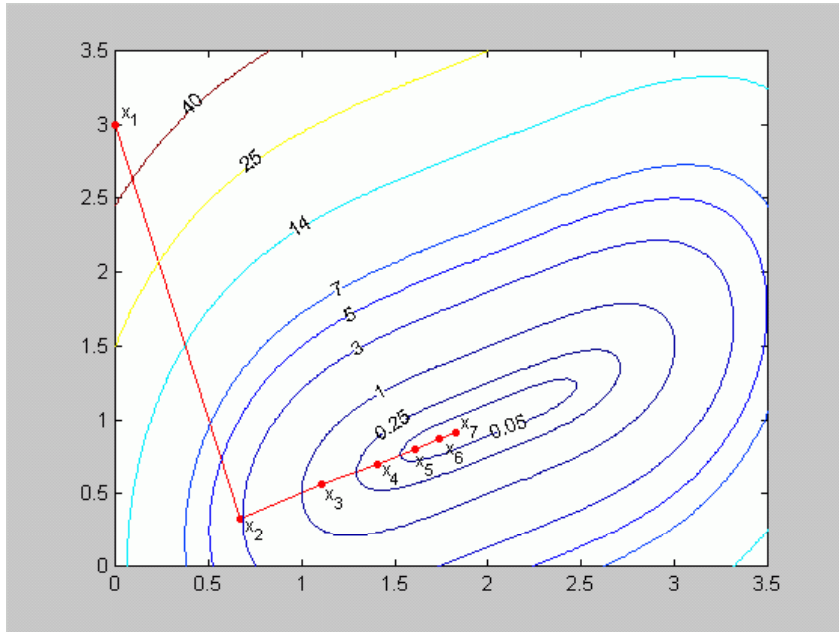


Figura 2.13: Método de Newton

$k$	$\mathbf{x}^k$	$f(\mathbf{x}^k)$	$\nabla f(\mathbf{x}^k)$	$\mathbf{H}f(\mathbf{x}^k)$	$\mathbf{H}f(\mathbf{x}^k)^{-1}$	$-\mathbf{H}f(\mathbf{x}^k)^{-1} \nabla f(\mathbf{x}^k)$	$\mathbf{x}^{k+1}$
1	$\begin{bmatrix} 0.00 \\ 3.00 \end{bmatrix}$	52.00	$\begin{bmatrix} -44.00 \\ 24.00 \end{bmatrix}$	$\begin{bmatrix} 50.0 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{384} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 50.0 \end{bmatrix}$	$\begin{bmatrix} 0.67 \\ -2.67 \end{bmatrix}$	$\begin{bmatrix} 0.67 \\ 0.33 \end{bmatrix}$
2	$\begin{bmatrix} 0.67 \\ 0.33 \end{bmatrix}$	3.13	$\begin{bmatrix} -9.39 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 23.23 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{169.84} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 23.23 \end{bmatrix}$	$\begin{bmatrix} 0.44 \\ 0.23 \end{bmatrix}$	$\begin{bmatrix} 1.11 \\ 0.56 \end{bmatrix}$
3	$\begin{bmatrix} 1.11 \\ 0.56 \end{bmatrix}$	0.63	$\begin{bmatrix} -2.84 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 11.50 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{76} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 11.50 \end{bmatrix}$	$\begin{bmatrix} 0.30 \\ 0.14 \end{bmatrix}$	$\begin{bmatrix} 1.41 \\ 0.70 \end{bmatrix}$
4	$\begin{bmatrix} 1.41 \\ 0.70 \end{bmatrix}$	0.12	$\begin{bmatrix} -0.80 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 6.18 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{33.44} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 6.18 \end{bmatrix}$	$\begin{bmatrix} 0.20 \\ 0.10 \end{bmatrix}$	$\begin{bmatrix} 1.61 \\ 0.80 \end{bmatrix}$
5	$\begin{bmatrix} 1.61 \\ 0.80 \end{bmatrix}$	0.02	$\begin{bmatrix} -0.22 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 3.83 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{14.64} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 3.83 \end{bmatrix}$	$\begin{bmatrix} 0.13 \\ 0.07 \end{bmatrix}$	$\begin{bmatrix} 1.74 \\ 0.87 \end{bmatrix}$
6	$\begin{bmatrix} 1.74 \\ 0.87 \end{bmatrix}$	0.005	$\begin{bmatrix} -0.07 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 2.81 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	$\frac{1}{6.84} \begin{bmatrix} 8.0 & 4.0 \\ 4.0 & 2.81 \end{bmatrix}$	$\begin{bmatrix} 0.09 \\ 0.04 \end{bmatrix}$	$\begin{bmatrix} 1.83 \\ 0.91 \end{bmatrix}$
7	$\begin{bmatrix} 1.83 \\ 0.91 \end{bmatrix}$	0.0009	$\begin{bmatrix} 0.0003 \\ -0.04 \end{bmatrix}$				

Tabla 2.7: Método de Newton

En el ejemplo anterior el valor de la función objetivo decrecía en cada iteración. Sin embargo, para el método de Newton este no es el comportamiento general y tiene los mismos defectos que el método equivalente para funciones de una variable: es un método local y su convergencia depende de lo cerca de la solución que esté el punto inicial.

**Método de Newton Modificado** La desventaja del método de Newton es que si la aproximación inicial del método,  $\mathbf{x}^0$ , está situada lejos del punto solución  $\mathbf{x}^*$ , puede ocurrir, como en el caso del método de Newton-Raphson para funciones de una variable, que la sucesión construida según la ecuación 2.15 sea divergente. Sin embargo, es posible modificar el algoritmo de forma simple y lógica para asegurar que sea descendente, utilizando en cada paso una búsqueda lineal como en el método de Cauchy. Si construimos la sucesión  $\{\mathbf{x}^k\}$  definida según la ecuación iterativa

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k [\mathbf{H}f(\mathbf{x}^k)]^{-1} \nabla f(\mathbf{x}^k)$$

eligiendo  $\alpha_k$  de forma que

$$f(\mathbf{x}^{k+1}) = \min f\left(\mathbf{x}^k - \alpha [\mathbf{H}f(\mathbf{x}^k)]^{-1} \nabla f(\mathbf{x}^k)\right)$$

Este es el *método de Newton modificado*, y se muestra eficiente cuando existen y se pueden calcular de forma precisa la primera y segunda derivada de  $f(\mathbf{x})$ . La mayor dificultad del método es calcular la inversa de la matriz hessiana  $\mathbf{H}f(\mathbf{x})$ , en cada iteración.

**Método de Levenberg-Marquardt** Este método combina los métodos de Cauchy y Newton. La dirección de búsqueda lineal en cada paso del algoritmo está expresada mediante la siguiente ecuación

$$\mathbf{d}^k = -[\mathbf{H}f(\mathbf{x}^k) + \lambda_k \mathbf{I}]^{-1} \nabla f(\mathbf{x}^k) \quad (2.16)$$

donde  $\lambda_k \geq 0$ . En este caso  $\alpha_k = +1$  en la ecuación 2.13, puesto que el parámetro  $\lambda_k$  se utiliza para controlar tanto la dirección de búsqueda como la longitud del paso.  $\mathbf{I}$  es la matriz identidad en el espacio  $M_{n \times n}(\mathbb{R})$ . Para comenzar la búsqueda tomamos  $\lambda_0$  con un valor constante grande y un punto  $\mathbf{x}^0$ , de forma que

$$[\mathbf{H}f(\mathbf{x}^0) + \lambda_0 \mathbf{I}]^{-1} \simeq [\lambda_0 \mathbf{I}]^{-1} = \left(\frac{1}{\lambda_0}\right) \mathbf{I}$$

Por tanto, para  $\lambda_0$  suficientemente grande,  $\mathbf{d}^k \rightarrow -\nabla f(\mathbf{x}^0)$  y tenemos el método del gradiente. Ahora bien si  $\lambda$  decrece hasta cero, entonces obtendremos el método de Newton. De acuerdo a este esquema, después del

primer paso, si obtenemos  $f(\mathbf{x}^1) < f(\mathbf{x}^0)$ , entonces elegimos  $\lambda_1 < \lambda_0$ , e iteramos de nuevo, en caso contrario ponemos  $\lambda_0 = \beta\lambda_0$ , con  $\beta > 1$ , y realizamos el paso previo de nuevo. Resumimos el algoritmo como sigue:

*Algoritmo de Marquardt*

**Paso 1.** Definimos  $\mathbf{x}^0$ , el punto inicial o primera aproximación de  $\mathbf{x}^*$ .  $M$  el máximo número de iteraciones permitido, y  $\varepsilon$  el criterio de convergencia o parada. Hacemos  $k = 0$  y  $\lambda_1 = 10^4$

**Paso 2.** Calculamos  $\nabla f(\mathbf{x}^k)$

**Paso 3.** Si  $\|\nabla f(\mathbf{x}^k)\| < \varepsilon \Rightarrow \mathbf{x}^* = \mathbf{x}^k$ . Parar

**Paso 4.** Si  $k > M \Rightarrow$  Superado el número de iteraciones. Parar.

**Paso 5.** Calcular

$$\mathbf{d}^k = -[\mathbf{H}f(\mathbf{x}^k) + \lambda_k \mathbf{I}]^{-1} \nabla f(\mathbf{x}^k)$$

**Paso 6.** Calcular

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}^k$$

**Paso 7.** Si  $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k) \Rightarrow \lambda_{k+1} = \frac{1}{2}\lambda_k$  y  $k \rightarrow k+1$ . Ir al **Paso 2**.

**Paso 8.** Si  $f(\mathbf{x}^{k+1}) \geq f(\mathbf{x}^k) \Rightarrow \lambda_k \rightarrow 2\lambda_k$ . Ir al **Paso 5**.

Las ventajas del método son su simplicidad, la propiedad de descenso, la excelente convergencia cerca de la solución,  $\mathbf{x}^*$  y la ausencia de búsqueda lineal. La mayor desventaja es la necesidad de calcular  $\mathbf{H}f(\mathbf{x}^k)$  y resolver el conjunto de ecuaciones correspondientes a 2.16. El método se utiliza concretamente en problemas en los que la función  $f(\mathbf{x})$  es una suma de cuadrados de funciones, i.e.

$$f(\mathbf{x}) = f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}) + \cdots + f_m^2(\mathbf{x})$$

**Métodos Cuasi-Newton o de métrica variable** En esta sección se presenta otro enfoque al desarrollo de métodos entre el descenso de mayor pendiente y el método de Newton. Trabajando con el hecho de que el cálculo de la matriz hessiana no es práctico o es costoso de realizar, la idea básica de todos los métodos cuasi Newton es utilizar una aproximación a la inversa del hessiano, en lugar de dicha inversa, como requiere el método de Newton.

Un proceso iterativo básico para resolver el problema

$$\text{Minimizar } f(\mathbf{x})$$

que incluye como casos especiales la mayoría de los anteriores es

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k \mathbf{D}_k \nabla f(\mathbf{x}^k) \quad (2.17)$$

donde  $\mathbf{D}_k$  es una matriz simétrica de  $n \times n$  y donde, como es usual,  $\alpha_k$  se elige para minimizar  $f(\mathbf{x}^{k+1})$ . Si  $\mathbf{D}_k$  es la inversa del Hessiano de  $f(\mathbf{x})$ , se obtiene el método de Newton; mientras que si  $\mathbf{D}_k = \mathbf{I}$  se tiene el descenso de mayor pendiente. Se elige  $\mathbf{D}_k$ , pues, como una aproximación a la inversa del hessiano.

El método elaborado aquí fue propuesto por Davidon y desarrollado posteriormente por Fletcher y Powell.

**Método de Davidon-Fletcher-Powell** El método de Davidon-Fletcher-Powell tiene la siguiente estructura:

Se parte de un punto inicial  $\mathbf{x}^0 \in \mathbb{R}^n$  y se toma una matriz simétrica definida positiva inicial  $\mathbf{D}_0$ , por ejemplo la identidad  $\mathbf{I}$  y se toma como dirección inicial de búsqueda  $\mathbf{d}^0 = -\mathbf{D}_0 \nabla f(\mathbf{x}_0)$ . A continuación, se construye la sucesión  $\{\mathbf{x}^k\}$  definida mediante la ecuación recursiva

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k$$

donde  $\alpha_k$  es la solución del problema: Minimizar  $f(\mathbf{x}^k + \alpha \mathbf{d}^k)$  con  $\alpha \geq 0$ .

La siguiente dirección conjugada,  $\mathbf{d}^{k+1}$ , se determina a partir de la expresión

$$\mathbf{d}^{k+1} = -\mathbf{D}_{k+1} \nabla f(\mathbf{x}^{k+1})$$

donde la matriz  $\mathbf{D}_{k+1}$  se obtiene según la expresión

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \frac{\mathbf{p}_k \mathbf{p}_k^T}{\mathbf{p}_k^T \mathbf{q}_k} - \frac{\mathbf{D}_k \mathbf{q}_k \mathbf{q}_k^T \mathbf{D}_k}{\mathbf{q}_k^T \mathbf{D}_k \mathbf{q}_k} \quad (2.18)$$

donde

$$\mathbf{p}_k = \alpha_k \mathbf{d}^k = \mathbf{x}^{k+1} - \mathbf{x}^k \quad (2.19)$$

$$\mathbf{q}_k = \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k) \quad (2.20)$$

Se puede demostrar que la matriz  $\mathbf{D}_{k+1}$  creada según la ecuación 2.18, es una matriz simétrica y definida positiva

Resumimos aquí el método de Davidon-Fletcher-Powell para minimizar una función diferenciable de varias variables. En particular, si la función es cuadrática, entonces, el método produce direcciones conjugadas y termina en una iteración completa.

#### *Método de Davidon-Fletcher-Powell*

- **Paso inicial.** Elegir  $\epsilon > 0$ , para la terminación del algoritmo. Elegir un punto inicial  $\mathbf{x}^0$ , y una matriz simétrica definida positiva inicial,  $\mathbf{D}_0$ , por ejemplo la matriz identidad. Hacer  $\mathbf{y}^1 = \mathbf{x}^0$  y  $\mathbf{D}_1 = \mathbf{D}_0$ ,  $k = 0$  y  $j = 1$ , e ir al paso principal.

- **Paso Principal.**

1. Si  $\|\nabla f(\mathbf{y}^j)\| < \epsilon$ , parar; en caso contrario hacer  $\mathbf{d}^j = -\mathbf{D}_j \nabla f(\mathbf{y}^j)$  y buscar  $\alpha_j$  como la solución óptima del problema de minimizar  $f(\mathbf{y}^j + \alpha \mathbf{d}^j)$  sujeto a  $\alpha \geq 0$ . Hacer  $\mathbf{y}^{j+1} = \mathbf{y}^j + \alpha_j \mathbf{d}^j$ . Si  $j < n$  ir al paso 2. Si  $j = n$ , hacer  $\mathbf{y}^1 = \mathbf{x}^{k+1} = \mathbf{y}^{n+1}$ , reemplazar  $k$  por  $k + 1$ , hacer  $j = 1$  y repetir el paso 1.
2. Construir  $\mathbf{D}_{j+1}$  como sigue

$$\mathbf{D}_{j+1} = \mathbf{D}_j + \frac{\mathbf{p}_j \mathbf{p}_j^T}{\mathbf{p}_j^T \mathbf{q}_j} - \frac{\mathbf{D}_j \mathbf{q}_j \mathbf{q}_j^T \mathbf{D}_j}{\mathbf{q}_j^T \mathbf{D}_j \mathbf{q}_j} \quad (2.21)$$

donde

$$\mathbf{p}_j = \alpha_j \mathbf{d}^j = \mathbf{y}^{j+1} - \mathbf{y}^j \quad (2.22)$$

$$\mathbf{q}_j = \nabla f(\mathbf{y}^{j+1}) - \nabla f(\mathbf{y}^j) \quad (2.23)$$

reemplazar  $j$  por  $j + 1$ , y repetir el paso 1.

De nuevo es necesario una minimización a lo largo de  $n$  direcciones antes de conseguir la siguiente aproximación de  $\mathbf{x}^*$ , el mínimo de  $f(\mathbf{x})$ .

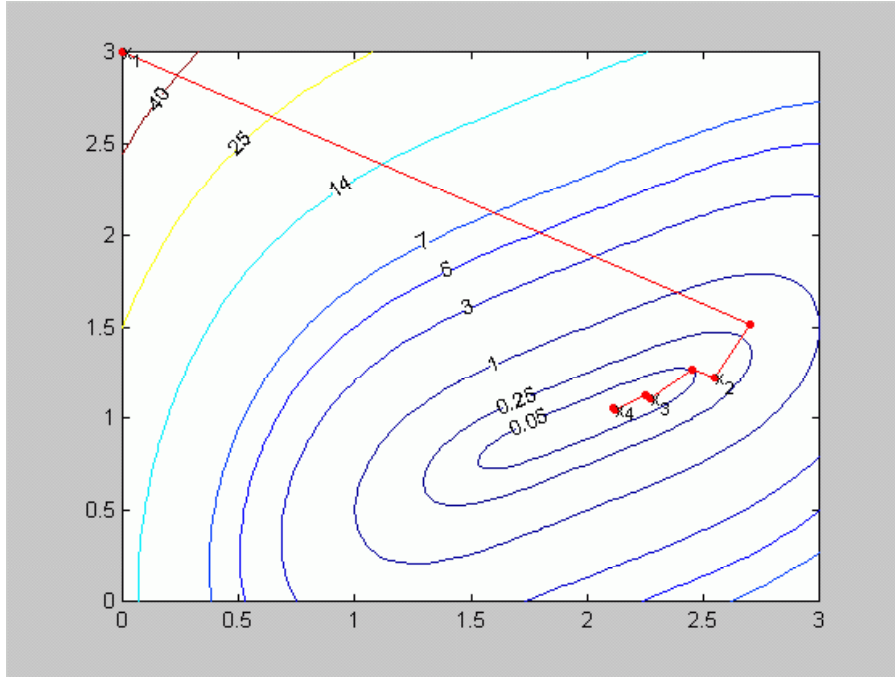


Figura 2.14: Método de Davidon-Fletcher-Powell

$k$	$\mathbf{x}^k$ $f(\mathbf{x}^k)$	$j$	$\mathbf{y}^j$ $f(\mathbf{y}^j)$	$\nabla f(\mathbf{y}^j)$	$\ \nabla f(\mathbf{y}^j)\ $	$\mathbf{D}_j$	$\mathbf{d}_j^T$	$\alpha_j$	$\mathbf{y}^{j+1}$
1	(0.00, 3.00) 52.00	1	(0.00, 3.00) 52.00	(-44.00, 24.00)	50.12	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	(44.00, -24.00)	0.062	(2.70, 1.51)
		2	(2.70, 1.51) 0.34	(0.73, 1.28)	1.47	$\begin{bmatrix} 0.25 & 0.38 \\ 0.38 & 0.81 \end{bmatrix}$	(-0.67, -1.31)	0.22	(2.55, 1.22)
2	(2.55, 1.22) 0.1036	1	(2.55, 1.22) 0.1036	(0.89, -0.44)	0.99	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	(-0.89, 0.44)	0.11	(2.45, 1.27)
		2	(2.45, 1.27) 0.0490	(0.18, 0.36)	0.40	$\begin{bmatrix} 0.65 & 0.45 \\ 0.45 & 0.46 \end{bmatrix}$	(-0.28, -0.25)	0.64	(2.27, 1.11)
3	(2.27, 1.11) 0.008	1	(2.27, 1.11) 0.008	(0.18, -0.20)	0.27	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	(-0.18, 0.20)	0.10	(2.25, 1.13)
		2	(2.25, 1.13) 0.004	(0.04, 0.04)	0.06	$\begin{bmatrix} 0.80 & 0.38 \\ 0.38 & 0.31 \end{bmatrix}$	(-0.05, -0.03)	2.64	(2.12, 1.05)
4	(2.12, 1.05) 0.0005	1	(2.12, 1.05) 0.0005	(0.05, -0.08)	0.09	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	(-0.05, 0.08)	0.10	(2.115, 1.058)
		2	(2.115, 1.058) 0.0002	(0.004, 0.004)	0.006				

Tabla 2.8: Método de Davidon-Fletcher-Powell

**Ejemplo 2.47** Considerar el siguiente problema

$$\text{Minimizar } (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

**Solución:** El resumen de los cálculos efectuados mediante el método DFP se presentan en la tabla 2.8.

En cada iteración, para  $j = 1, 2$ ,  $\mathbf{d}^j = \mathbf{D}_j - \nabla f(\mathbf{y}^1)$ , donde  $\mathbf{D}_1$  es la matriz identidad y  $\mathbf{D}_2$  se calcula a partir de 2.18, 2.19 y 2.20. En la iteración  $k = 1$ , tenemos  $\mathbf{p}_1^T = (2.7, -1.49)$  y  $\mathbf{q}_1^T = (44.73, -22.72)$ , y finalmente en la iteración 3, tenemos  $\mathbf{p}_3^T = (-0.1, 0.05)$  y  $\mathbf{q}_3^T = (-0.7, 0.8)$ . El punto  $\mathbf{y}^{j+1}$  se calcula optimizando en la dirección  $\mathbf{d}^j$ , a partir de  $\mathbf{y}^j$  para  $j = 1, 2$ . El procedimiento termina en el punto  $\mathbf{y}^2 = (2.115, 1.058)$  en la cuarta iteración, puesto que  $\nabla f(\mathbf{y}^2) = 0.006$ . El camino descrito por el método se muestra en la figura 2.14.

Otro método propuesto simultáneamente por Broyden, Fletcher, Goldfarb y Shanno (BFGS) utiliza la siguiente variación para actualizar la aproximación de la inversa del Hessiano:

$$\mathbf{D}_{k+1} = \left( \mathbf{I} - \frac{\mathbf{p}_k \mathbf{q}_k^T}{\mathbf{p}_k^T \mathbf{q}_k} \right) \mathbf{D}_k \left( \mathbf{I} - \frac{\mathbf{p}_k \mathbf{q}_k^T}{\mathbf{p}_k^T \mathbf{q}_k} \right) - \frac{\mathbf{p}_k \mathbf{p}_k^T}{\mathbf{p}_k^T \mathbf{q}_k}$$

de nuevo  $\mathbf{p}_k$  y  $\mathbf{q}_k$  están descritos por las ecuaciones 2.19 y 2.20.

### 2.4.3 Optimización con restricciones

Solamente describiremos un método de resolución numérica para problemas con restricciones: el llamado *método de transformación*, que consiste en convertir un problema con restricciones en un problema sin restricciones.

#### Métodos de transformación

A partir del problema no lineal general

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, p \end{array}$$

Asumimos en primer lugar que tenemos una estimación inicial  $\mathbf{x}^0$  de la solución del problema  $\mathbf{x}^*$  que puede o no, ser factible. Discutiremos algoritmos que generan una sucesión de puntos en  $\mathbb{R}^n$  tal que

$$\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n \rightarrow \mathbf{x}^N \simeq \mathbf{x}^*$$

Los puntos  $\mathbf{x}^k$  serán puntos estacionarios de una función sin restricciones construida a partir de las funciones del problema anterior. Con estos métodos el problema original se *transforma* en una sucesión de problemas sin restricciones a través de la llamada función de penalización. La estructura de la función de penalización junto con las reglas para actualizar los *parámetros de penalización* en cada paso de la minimización sin restricciones definen a cada método particular. La función de penalización será *exacta* si solamente necesitamos un paso de minimización sin restricciones.

Básicamente hay dos aproximaciones alternativas, que se diferencian en la forma de tratar las restricciones de desigualdad, puesto que esencialmente todos los métodos de transformación tratan las restricciones de igualdad de la misma manera. La primera de ellas es el llamado *método de la penalización* o *método de la función de penalización exterior*, en el cual se añade un término de penalización a la función objetivo para cada violación de las restricciones. Este método genera una secuencia de puntos infactibles, de ahí su nombre, cuyo límite es una solución óptima del problema original. El segundo método es llamado *métodos de barrera* o *método de la función de penalización interior*, en el que se incluye un término “barrera” que evita que los puntos generados abandonen la región factible. El método genera una sucesión de puntos factibles cuyo límite es una solución óptima del problema original. También podemos considerar métodos *mixtos* cuando la sucesión de puntos generada puede contener tanto puntos factibles como infactibles.

La estructura de los problemas sin restricciones utilizados para resolver el problema con restricciones es de la forma

$$F(\mathbf{x}, M) = f(\mathbf{x}) + MP(\mathbf{h}(\mathbf{x}), \mathbf{g}(\mathbf{x}))$$

donde  $F(\mathbf{x}, M)$  es la *función de penalización*,  $M$  es el *parámetro de penalización* y  $P$  es el *término de penalización* que como puede comprobarse es una función que depende de las restricciones del problema. Como se ha comentado en el párrafo anterior, la forma exacta en la que los parámetros de penalización y las restricciones se combinan y las reglas para actualizar los parámetros de penalización especifican el método particular. La función  $P$ , se elige de manera que tiende a favorecer la selección por parte de los métodos de búsqueda sin restricciones de puntos factibles para el problema con restricciones frente a puntos infactibles.

En general cualquier técnica de transformación tiene que tener las siguientes características:

1. Las soluciones de los problemas sin restricciones construidos tiene que aproximarse a la solución del problema con restricciones, es decir, debe ocurrir

$$\lim_{k \rightarrow N < \infty} \mathbf{x}^k = \mathbf{x}^*$$

2. El problema de minimizar  $F(\mathbf{x}, M)$  debe ser de dificultad similar al problema de minimizar  $f(\mathbf{x})$ , es decir, el método será de poca utilidad si los problemas utilizados son excesivamente difíciles de resolver, no importa como de buena sea la convergencia.
3. La construcción de los parámetros en el paso  $(k+1)$ -ésimo, debe obtenerse fácilmente a partir de los parámetros en el paso  $(k)$ -ésimo.

**Penalización exterior** Las restricciones se incluyen en la función objetivo por medio de un término de penalización de manera que se penaliza cualquier violación de las restricciones. Para motivar las funciones de penalización, se considera el siguiente problema con una única restricción de igualdad  $h(\mathbf{x}) = 0$

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & h(\mathbf{x}) = 0 \end{array}$$

Transformaremos este problema en el siguiente problema sin restricciones,

$$\text{Minimizar} \quad f(\mathbf{x}) + M [h(\mathbf{x})]^2$$

donde  $M > 0$  es un número grande.

Podemos intuitivamente observar que una solución óptima del problema anterior debe tener  $[h(\mathbf{x})]^2$  próximo a cero, en otro caso se incluye un término  $M [h(\mathbf{x})]^2$  muy grande.

Consideremos ahora el siguiente problema con una sola restricción de desigualdad  $g(\mathbf{x}) \leq 0$

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & g(\mathbf{x}) \leq 0 \end{array}$$

Está claro ahora que la forma  $f(\mathbf{x}) + M [g(\mathbf{x})]^2$  no es apropiada, puesto que ocurrirá una penalización tanto si  $g(\mathbf{x}) < 0$ , como si  $g(\mathbf{x}) > 0$ . No es necesario decir, que la penalización solamente es necesaria si el punto es infactible, es decir,  $g(\mathbf{x}) > 0$ . Un problema sin restricciones adecuado podría estar dado por ejemplo por

$$\text{Minimizar} \quad f(\mathbf{x}) + M \max \{0, g(\mathbf{x})\}$$

En este caso si  $g(\mathbf{x}) \leq 0$ , entonces  $\max \{0, g(\mathbf{x})\} = 0$  y no hay penalización. Por otra parte si  $g(\mathbf{x}) > 0$ , entonces  $\max \{0, g(\mathbf{x})\} > 0$  y el término de penalización  $Mg(\mathbf{x})$  es empleado. Sin embargo, podemos comprobar que aunque  $g$  sea diferenciable la función objetivo puede no serlo en los puntos donde  $g(\mathbf{x}) = 0$ . Si necesitamos derivabilidad podríamos considerar en su lugar un término de penalización del tipo

$$M [\max \{0, g(\mathbf{x})\}]^2$$

En general, una función de penalización adecuada debe proporcionar una penalización positiva para puntos infactibles y ninguna penalización para puntos factibles. Si las restricciones son de la forma  $g_j(\mathbf{x}) \leq 0$  para  $j = 1, \dots, p$  y  $h_i(\mathbf{x})$  para  $i = 1, \dots, m$ , una función de penalización adecuada podría ser de la forma  $MP_e(\mathbf{x})$

$$P_e(\mathbf{x}) = \sum_{i=1}^m \psi[h_i(\mathbf{x})] + \sum_{j=1}^p \phi[g_j(\mathbf{x})]$$



donde  $\phi$  y  $\psi$  son funciones continuas cumpliendo las siguientes propiedades

$$\phi(y) = \begin{cases} 0 & y \leq 0 \\ > 0 & y > 0 \end{cases}$$

$$\psi(y) = \begin{cases} 0 & y = 0 \\ > 0 & y \neq 0 \end{cases}$$

Típicamente,  $\phi$  y  $\psi$  son de la forma

$$\phi(y) = [\max\{0, y\}]^p$$

$$\psi(y) = |y|^p$$

donde  $p \in \mathbb{Z}^+$ . De esta forma, la función de penalización  $P_e(\mathbf{x})$  podría tomar la forma:

$$P_e(\mathbf{x}) = \sum_{i=1}^m |h_i(\mathbf{x})|^p + \sum_{j=1}^p [\max\{0, g_j(\mathbf{x})\}]^p$$

Por ejemplo, para restricciones de igualdad se suele considerar la *penalización parabólica*

$$\sum_{j=1}^m h_j(\mathbf{x})^2$$

mientras que para restricciones de desigualdad se utiliza el llamado *operador corchete*

$$P_e(\mathbf{x}) = \sum_{k=1}^p \langle g(\mathbf{x}) \rangle^2 = \sum_{k=1}^p \max\{0, g_k(\mathbf{x})\}^2$$

donde

$$\langle x \rangle = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

En ambos casos se construye la función sin restricciones

$$F(\mathbf{x}) = f(\mathbf{x}) + MP_e(\mathbf{x})$$

con  $M \gg 0$  y tomando límites cuando  $M \rightarrow \infty$

**Ejemplo 2.48 (Penalización parabólica)** *Resuelve el problema*

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) = (x_1 - 4)^2 + (x_2 - 4)^2 \\ \text{Sujeto a} & h(\mathbf{x}) = x_1 + x_2 - 5 = 0 \end{array}$$

*mediante un método de penalización exterior adecuado.*

**Solución:** Consideramos en primer lugar la función de penalización utilizando el término de penalización parabólica

$$F(\mathbf{x}, M) = (x_1 - 4)^2 + (x_2 - 4)^2 + M(x_1 + x_2 - 5)^2$$

Buscamos los puntos estacionarios de  $F(\mathbf{x}, M)$

$$\frac{\partial F}{\partial x_1} = 2(x_1 - 4) + 2M(x_1 + x_2 - 5) = 0$$

$$\frac{\partial F}{\partial x_2} = 2(x_2 - 4) + 2M(x_1 + x_2 - 5) = 0$$

La solución del sistema anterior es

$$x_1 = x_2 = \frac{10M + 8}{4M + 2}$$

¿Es el método convergente?

$$\lim_{M \rightarrow \infty} \left( \frac{10M + 8}{4M + 2} \right) = \frac{10}{4} = 2.5$$

La respuesta es sí, puesto que la solución óptima del problema es

$$\mathbf{x}^* = (2.5, 2.5)$$

Los valores estacionarios de  $F(\mathbf{x}, M)$  para diversos valores de  $M$  están dados en la siguiente tabla

$M$	$x_1^k = x_2^k$	$F(\mathbf{x}^{(k)}, M)$
0	4.0000	0.0000
0.1	3.7500	0.7500
1	3.0000	3.0000
10	2.5714	4.2857
100	2.5075	4.4776
$\infty$	2.5000	4.5000

donde se comprueba la convergencia a la solución, cuando  $R \rightarrow \infty$  ■

Supongamos el problema anterior con la restricción de desigualdad:

**Ejemplo 2.49 (Operador corchete)** Encuentra la solución del problema

$$\begin{aligned} \text{Minimizar} \quad & f(\mathbf{x}) = (x_1 - 4)^2 + (x_2 - 4)^2 \\ \text{Sujeto a} \quad & g(\mathbf{x}) = x_1 + x_2 - 5 \leq 0 \end{aligned}$$

mediante un método de penalización exterior adecuado.

**Solución:** Consideramos la función de penalización utilizando el operador corchete

$$F_{pc}(\mathbf{x}, M) = (x_1 - 4)^2 + (x_2 - 4)^2 + M \langle x_1 + x_2 - 5 \rangle^2$$

Buscamos los puntos estacionarios de  $F_{pc}(\mathbf{x}, M)$

$$\frac{\partial F_{pc}}{\partial x_1} = 2(x_1 - 4) + 2M \langle x_1 + x_2 - 5 \rangle = 0$$

$$\frac{\partial F_{pc}}{\partial x_2} = 2(x_2 - 4) + 2M \langle x_1 + x_2 - 5 \rangle = 0$$

Restando ambas ecuaciones se obtiene

$$x_1 = x_2$$

de donde, sustituyendo en cualquiera de las dos

$$(x_1 - 4) + M \langle 2x_1 - 5 \rangle = 0$$

Consideremos ahora que el argumento del operador corchete es  $=$ ,  $>$  o  $<$  0, respectivamente, y analicemos los resultados. Si  $2x_1 \geq 5$

$$(x_1 - 4) + M(2x_1 - 5) = 0$$

o equivalentemente

$$x_1 = \frac{5R + 4}{2R + 1}$$

Para  $2x_1 \leq 5$ , ocurre  $\langle 2x_1 - 5 \rangle = 0$  y por tanto

$$(x_1 - 4) = 0 \Rightarrow x_1 = 4$$

Lo que nos da el punto  $(4, 4)$  que es infactible porque incumple la restricción.

Por tanto para  $2x_1 \geq 5$ , tomamos  $M$  creciente

$$\lim_{M \rightarrow \infty} x_1 = \frac{5}{2} = 2.5$$

Podemos comprobar que con este método, todos los puntos estacionarios son infactibles.

**Penalización interior** De forma similar a las funciones de penalización exterior, las funciones de penalización interior o funciones barrera también se utilizan para transformar un problema con restricciones en un problema sin restricciones o en una sucesión de problemas sin restricciones. Estas funciones colocan una “barrera” para evitar que los puntos generados abandonen la región factible. Este tipo de métodos se utiliza para problemas que contienen solamente restricciones de desigualdad

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \mathbf{g}(\mathbf{x}) \leq 0 \end{array}$$

donde  $\mathbf{g} = (g_1, \dots, g_p)$ . Además las funciones  $f(\mathbf{x})$  y  $\mathbf{g}(\mathbf{x})$  son continuas.

La función sin restricciones que se construye para cada problema de la sucesión, tiene la forma

$$F(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{M} P_i(\mathbf{x})$$

donde  $M \in \mathbb{R}^+$ ,  $P_i(\mathbf{x})$  se denomina *función barrera*, y está definida por :

$$P_i(\mathbf{x}) = \sum_{j=1}^p \phi[g_j(\mathbf{x})]$$

siendo  $\phi$  es una función univariante continua sobre  $\{y | y < 0\}$  que cumple

$$\phi(y) \geq 0 \text{ si } y < 0 \quad \text{y} \quad \lim_{y \rightarrow 0^-} \phi(y) = \infty$$

De nuevo tomamos  $M \rightarrow \infty$  y por tanto  $\frac{1}{M} \rightarrow 0$ . Si definimos  $R = \frac{1}{M}$  entonces podemos escribir la función penalizada como

$$F(\mathbf{x}) = f(\mathbf{x}) + R P_i(\mathbf{x})$$

tomando  $R \rightarrow 0$  para encontrar la solución.

A continuación se indica alguna de estas funciones de barrera

#### 1. Penalización logarítmica

El primer ejemplo de penalización interior que podemos utilizar es la *penalización logarítmica*, donde

$$\phi(y) = -\ln(-y)$$

Notar que el término produce una penalización positiva para todos los puntos  $\mathbf{x}$  tales que

$$0 < -y < 1 \Leftrightarrow 0 > y > -1$$

y penalizaciones negativas para aquellos puntos  $\mathbf{x}$  tales que  $y < -1$ . De esta forma, los puntos interiores se favorecen artificialmente sobre los puntos de la frontera. Las penalizaciones negativas pueden eliminarse asignando un valor 0 cuando  $y < -1$ , es decir, podemos definir la función  $\phi$  como:

$$\phi(y) = -\ln \min[1, -y]$$

Al modificar el término de penalización de esa forma, introducimos una discontinuidad en  $\nabla F$  en el exterior de la región cerca de los bordes. La penalización logarítmica es una función barrera que no está definida para  $\mathbf{x}$  infactibles ( $g(\mathbf{x}) > 0$ ). Se necesita por tanto incluir un procedimiento para evitar estos puntos.

#### 2. Penalización inversa

Otra función de penalización interior utilizada es la llamada función de *penalización inversa*, definida como:

$$\phi(y) = \frac{-1}{y}$$

Esta función tiene problemas asociados a la evaluación de puntos infactibles, tratando de evitar este tipo de puntos.

**Ejemplo 2.50 (Penalización interior)** Resolver el problema

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) = (x_1 - 4)^2 + (x_2 - 4)^2 \\ \text{Sujeto a} & g(\mathbf{x}) = x_1 + x_2 - 5 \leq 0 \end{array}$$

utilizando una función de penalización interior adecuada.

**Solución:** Resolvamos en primer lugar utilizando una penalización logarítmica, la función penalizada es ahora

$$F_{pl}(\mathbf{x}, R) = (x_1 - 4)^2 + (x_2 - 4)^2 - R \ln(5 - x_1 - x_2)$$

Buscamos los puntos estacionarios de  $F_{pl}(\mathbf{x}, R)$

$$\frac{\partial F_{pl}}{\partial x_1} = 2(x_1 - 4) + R \frac{1}{(5 - x_1 - x_2)} = 0$$

$$\frac{\partial F_{pl}}{\partial x_2} = 2(x_2 - 4) + R \frac{1}{(5 - x_1 - x_2)} = 0$$

La solución del sistema anterior es de nuevo

$$x_1 = x_2$$

Que en la primera ecuación lleva a:

$$2x_1^2 - 13x_1 + 20 - \frac{R}{2} = 0$$

Las raíces de esta ecuación produce los valores estacionarios deseados

$$x_1 = \frac{13}{4} - \frac{1}{4}\sqrt{9 + 4R}$$

donde hemos desestimado la otra raíz puesto que es infactible. Ahora si  $R \rightarrow 0$

$$\lim_{R \rightarrow 0} x_1 = \lim_{R \rightarrow 0} \frac{13}{4} - \frac{1}{4}\sqrt{9 + 4R} = \frac{13}{4} - \frac{3}{4} = \frac{10}{4} = 2.5$$

Los valores estacionarios de  $F_{pl}(\mathbf{x}, R)$  para diversos valores de  $R$  están dados en la siguiente tabla

$R$	$x_1^k = x_2^k$	$f(\mathbf{x})$	$g(\mathbf{x})$	$-R \log(-g(\mathbf{x}))$	$F_{pl}(\mathbf{x}^k, R)$
100	-1.8059	67.4170	-8.6118	-215.3133	-147.8963
10	1.5000	12.5000	-2.0000	-6.9315	5.5685
1	2.3486	5.4542	-0.3028	1.1947	6.6489
0.1	2.4835	4.5995	-0.0034	0.3411	4.9406
0.01	2.4983	4.5100	-0.0034	0.0568	4.5668
0	2.5000	4.5000	0.0000	0.0000	4.5000

Consideremos ahora la penalización inversa

$$F_{pi}(\mathbf{x}, R) = (x_1 - 4)^2 + (x_2 - 4)^2 - R \frac{1}{(x_1 + x_2 - 5)}$$

Como anteriormente, buscamos los puntos estacionarios de  $F(\mathbf{x}, R)$

$$\frac{\partial F_{pi}}{\partial x_1} = 2(x_1 - 4) + \frac{R}{(x_1 + x_2 - 5)^2} = 0$$

$$\frac{\partial F_{pi}}{\partial x_2} = 2(x_2 - 4) + \frac{R}{(x_1 + x_2 - 5)^2} = 0$$

La solución de este sistema es de nuevo

$$x_1 = x_2$$

Sustituyendo en la primera ecuación y después de quitar denominadores

$$4x_1^3 - 36x_1^2 + 105x_1 - 100 + \frac{R}{2} = 0$$

Los puntos estacionarios están definidos por los ceros de esta función no lineal. Utilizando un algoritmo de búsqueda de raíces encontramos los valores que aparecen en la siguiente tabla

$R$	$x_1 = x_2$	$f(\mathbf{x})$	$g(\mathbf{x})$	$-R/g(\mathbf{x})$	$F_{pi}(\mathbf{x}^{(k)}, R)$
100	0.5864	23.3053	3.8272	26.1288	49.4341
10	1.7540	10.0890	1.4920	6.7024	16.7914
1	2.2340	6.2375	0.5320	1.8797	8.1172
0.1	2.4113	5.0479	0.1774	0.5637	5.6116
0.01	2.4714	4.6732	0.0572	0.1748	4.8480
0.001	2.4909	4.5548	0.0182	0.0182	4.6097
0	2.5000	4.5000	0.0000	0.0000	4.5000

De nuevo obtenemos que los valores estacionarios de la función penalizada se aproximan a la solución del problema con restricciones cuando  $R \rightarrow 0$ . ■

En un mismo problema podemos emplear penalización interior y exterior, simplemente hay que tener en cuenta que el parámetro de penalización asociado a los términos de penalización exterior crece hacia valores grandes, mientras que si la penalización es interior, dicho parámetro decrece hacia cero, de esta forma si empleamos el valor  $M$ , para una penalización exterior, podemos utilizar el valor  $R = 1/M$ , para utilizar penalización interior, de esta forma cuando  $M \rightarrow \infty$ , en este caso,  $R \rightarrow 0$ . Por otra parte es relativamente directo construir un algoritmo útil con los conceptos discutidos previamente. El algoritmo tendrá la siguiente forma

#### Algoritmo de Penalización

**Paso 1.** Definiremos  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\varepsilon_3$ ,  $\mathbf{x}^0$  y  $M^0$ . Donde

$\varepsilon_1$	Criterio de parada de la búsqueda lineal
$\varepsilon_2$	Criterio de parada del método irrestricto
$\varepsilon_3$	Criterio de parada del método Penalización
$\mathbf{x}^0$	Estimación inicial de $\mathbf{x}^*$
$M^0$	Conjunto inicial de parámetros de penalización

Hacer  $j = 0$ ;

**Paso 2.** Construir  $F(\mathbf{x}, M^j) = f(\mathbf{x}) + P(M^j, \mathbf{h}(\mathbf{x}), \mathbf{g}(\mathbf{x}))$

**Paso 3.** Calcular el punto  $\mathbf{x}^{j+1}$  como el punto que minimiza la función  $F(\mathbf{x}, M_j) = f(\mathbf{x}) + M_j P(\mathbf{h}(\mathbf{x}), \mathbf{g}(\mathbf{x}))$  partiendo desde el punto  $\mathbf{x}^j$  y utilizando  $\varepsilon_2$  como parada y  $\varepsilon_1$  para la búsqueda lineal.

**Paso 4.** Si  $|F(\mathbf{x}^{j+1}, M_j) - F(\mathbf{x}^j, M_j)| \leq \varepsilon_3$ , la solución es  $\mathbf{x}^{j+1}$  y parar. En caso contrario continuar.

**Paso 5.** Elegir  $M^{j+1} = M^j + \Delta M^j$ , según una regla descrita (que depende del término de penalización) e ir al **Paso 2**.

## 2.5 Bibliografía Básica del Tema

1. *Decisiones de Optimización*. Manuel Mocholí Arce y Ramón Sala Garrido. **Editorial Tirant lo Blanch (1996)**
2. *Engineering Optimization. Methods and Applications*. G.V. Reklaitis, A. Ravindran and K.M. Ragsdell. **Editorial John Wiley and Sons (1983)**
3. *Nonlinear Programming*. Mokhtar S. Bazaraa, Hanif D. Sherali and C.M. Shetty. **Ed. John Wiley & Sons, Inc (Wiley-Interscience) (1993)**

4. *Teoría de la Optimización*. Vicente Novo Sanjurjo. **Editorial U.N.E.D. (Aula Abierta) (1997)**
5. *Practical Optimization*. Philip E. Gill, Walter Murray & Margaret H. Wright. **Editorial Academic Press (1981)**
6. *Programación Lineal y No Lineal*. David e. Luenberger. **Editorial Addison-Wesley Iberoamericana (1989)**
7. *Programación Matemática*. Balbas, A. y Gil, J.A. **Editorial AC.**