

Métodos Clásicos de Optimización para Problemas No-Lineales sin Restricciones

Dr. Gonzalo Hernández Oliva

UChile - Departamento de Ingeniería Matemática

07 de Mayo 2006

Abstract

En este apunte veremos una breve descripción de la teoría y métodos que se utilizan para enfrentar problemas de optimización no-lineales sin restricciones. La no presencia de restricciones puede ser vista en un comienzo como una grave limitación, sin embargo, la razón del porque dedicamos tiempo a resolver este tipo de problemas desde el punto de vista de la optimización, es que existe toda una metodología que permite tratar problemas de optimización no-lineales con restricciones como problemas de optimización no-lineales sin restricciones: Los Métodos de Penalización.

1 Introducción

Nos interesa resolver el problema siguiente:

Dada $f : \mathbb{R}^n \rightarrow \mathbb{R}$, campo escalar 2 veces diferenciable con continuidad, encontrar $\bar{x} \in \mathbb{R}^n$, solución de:

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

Encontrar una solución de este problema corresponde a encontrar un punto \bar{x} que satisface:

$$f(\bar{x}) \leq f(x) \quad \forall x \in \mathbb{R}^n$$

Este punto \bar{x} se denomina *mínimo global de f sobre \mathbb{R}^n* .

Desde un punto de vista práctico, encontrar el mínimo o máximo global de una función no-lineal cualquiera es un problema abierto en matemáticas. Los métodos existentes sólo obtienen mínimos locales.

Definición 0: Un punto \bar{x} es un *mínimo local* de f en \mathbb{R}^n si existe $r > 0$ talque:

$$f(\bar{x}) \leq f(x) \quad \text{para } \|x - \bar{x}\| \leq r$$

El *mínimo global* de f será el *mínimo local* de menor valor de la función objetivo.

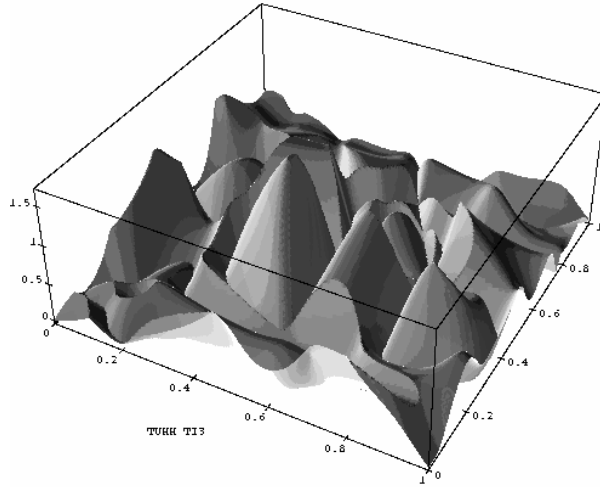


Figura 1. Mínimos locales de una función de 2 variables

Desde un punto de vista teórico, las condiciones más generales para que \bar{x} sea *mínimo local* de f son las siguientes:

Teorema 1: Un punto \bar{x} es mínimo local de f si se verifican las siguientes condiciones:

- 1) \bar{x} es un punto estacionario de f

$$\nabla f(\bar{x}) = \vec{0} \iff \begin{bmatrix} \frac{\partial f(\bar{x})}{\partial x_1} \\ \frac{\partial f(\bar{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\bar{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- 2) La matriz hessiana $\nabla^2 f(\bar{x})$ de f en \bar{x} es definida positiva:

$$\nabla^2 f(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\bar{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\bar{x})}{\partial x_n^2} \end{bmatrix}$$

Dem: Se demuestra mediante la aproximación de Taylor de 2º Orden de la función f en torno a \bar{x} . ■

Entonces, para obtener el mínimo local se debe resolver un sistema de ecuaciones no lineales, en general complejo. Surge de esta forma la necesidad de aplicar métodos numéricos.

Consideraremos métodos numéricos llamados de *descenso*, que pueden ser clasificados en 3 grupos:

$$\text{Métodos Descenso} \left\{ \begin{array}{ll} \text{Primer Orden} & \left\{ \begin{array}{l} \text{Método del Gradiente y Variedades} \\ \text{Método del Gradiente Conjugado} \\ \text{Método de Fletcher-Reeves} \end{array} \right. \\ \text{Segundo Orden} & \left\{ \begin{array}{l} \text{Métodos de Newton} \\ \text{Variedades} \end{array} \right. \\ \text{Cuasi-Newton} & \left\{ \begin{array}{l} \text{DFP: Davidon - Fletcher - Powell} \\ \text{BFGS} \end{array} \right. \end{array} \right.$$

Todos los métodos numéricos que se utilizan en la actualidad reducen este problema a una secuencia de problemas unidimensionales.

2 Método de Gradiente o de Cauchy

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable. La derivada direccional de f en la dirección $d \in \mathbb{R}^n$ está dada por:

$$Df(x; d) = \nabla f(x)^t d$$

Para obtener la dirección de *máximo descenso* de la función f en un punto $x \in \mathbb{R}^n$ tal que $\nabla f(x) \neq 0$, se debe resolver el problema:

$$\boxed{\begin{array}{l} \min_{d \in \mathbb{R}^n} \nabla f(x)^t d \\ \|d\|^2 = 1 \end{array}}$$

La solución de este problema es:

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

Y por lo tanto la dirección de máximo descenso de la función f es:

$$d = -\nabla f(x)$$

Definición 2: El vector $d \in \mathbb{R}^n$ es una dirección de descenso de f , función diferenciable, en el punto $x \in \mathbb{R}^n$ si:

$$\nabla f(x)^t \cdot d < 0 \iff (-\nabla f(x))^t \cdot d > 0$$

A través del método del gradiente, se busca definir una sucesión de puntos que tenga la dirección de máximo descenso de la función f .

Definición 3: Las curvas de nivel θ de f están definidas por:

$$\varphi(f, \theta) = \{x \in \mathbb{R}^n \text{ tal que: } f(x) = \theta\}$$

Lema 4: Sea $d \in \mathbb{R}^n$ una dirección de descenso de la función f en el punto x . Se define la función unidimensional:

$$\begin{aligned} h_{x,d} : \mathbb{R}_+ &\rightarrow \mathbb{R} \\ \alpha &\rightarrow h_{x,d}(\alpha) = f(x + \alpha d) - f(x) \quad \text{para } x, d \in \mathbb{R}^n \text{ fijos} \end{aligned}$$

Luego, si $\bar{\alpha}$ minimiza $h_{x,d}$ sobre \mathbb{R}_+ , entonces:

$$h_{x,d}(\bar{\alpha}) < h_{x,d}(\alpha = 0) = 0 \Leftrightarrow f(x + \bar{\alpha}d) < f(x)$$

donde $\bar{\alpha}$ verifica:

$$h'_{x,d}(\bar{\alpha}) = 0 \iff \nabla f(x + \bar{\alpha}d)^t d = 0$$

Dem: Directa de las definiciones hechas para $h_{x,d}$. ■

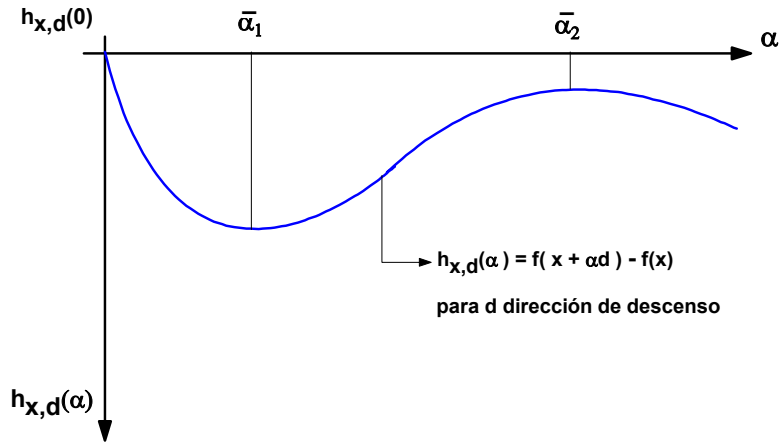


Figura 2. Función Unidimensional $h_{x,d}$

De acuerdo a los resultados anteriores, se define la iteración canónica del Método del Gradiente o Máximo Descenso, según:

METODO DEL GRADIENTE O MAXIMO DESCENSO

- Etapla 0:** Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 $k = 0$
- Etapla 1:** Calcular $\nabla f(x^k)$
 Si $\|\nabla f(x^k)\| = 0 \Rightarrow STOP$
 Si no, seguir a Etapla 2.
- Etapla 2:** $d^k = -\nabla f(x^k)$
 Calcular: α^k solución del problema uni-dimensional:
 $\min_{\alpha \geq 0} h_{x^k, d^k}(\alpha) = f(x^k + \alpha d^k) - f(x^k)$
 $x^{k+1} = x^k + \alpha^k d^k$
 $k = k + 1$ y volver a Etapla 1.

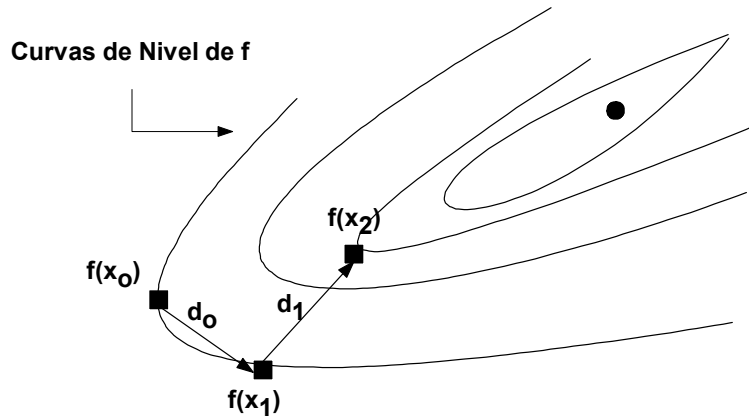


Figura 3. Esquema iteración del método del Gradiente

Ejemplo: Se quiere resolver el problema de optimización bi-dimensional, ver ref. [1]:

$$\min_{(x_1, x_2) \in \mathbb{R}^2} (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

En la siguiente tabla se muestran 7 iteraciones del método del gradiente a partir de $x^0 = (0.00, 3.00)$

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\ \nabla f(x^k)\ $	α^k
0	(0.00, 3.00)	52.00	(-44.00, 24.00)	50.12	0.062
1	(2.70, 1.51)	0.34	(0.73, 1.28)	1.47	0.24
2	(2.52, 1.20)	0.09	(0.80, -0.48)	0.93	0.11
3	(2.43, 1.25)	0.04	(0.18, 0.28)	0.33	0.31
4	(2.37, 1.16)	0.02	(0.30, -0.20)	0.36	0.12
5	(2.33, 1.18)	0.01	(0.08, 0.12)	0.14	0.36
6	(2.30, 1.14)	0.009	(0.15, -0.08)	0.17	0.13
7	(2.28, 1.15)	0.007	(0.05, 0.08)	0.09	

Observaciones:

- 1) En cada etapa α_k , llamado el paso del método en la etapa $(k + 1)$ -ésima, debe verificar:

$$\nabla f(x^k + \alpha d^k) \cdot d^k = 0$$

que en general será una ecuación difícil de resolver exactamente en α . Por esta razón se utilizan pasos α^k que serán aproximaciones del paso exacto, determinados por algún método de elección de paso, como se verá más adelante.

- 2) Dadas las condiciones de optimalidad sobre f (i.e. existencia única de la solución del problema (P)) se tiene que el método posee una convergencia global, i.e. para cualquier punto inicial la iteración del método converge, a la solución de (POSR).

- 3) La rapidez de convergencia del método es lineal: Si x_k es la sucesión de puntos determinados por el método del gradiente, se tiene que $\exists \lambda \in (0, 1)$ tal que:

$$\left| x^{k+1} - \bar{x} \right| \leq \lambda \left| x^k - \bar{x} \right|$$

donde \bar{x} es un punto estacionario de f : $\nabla f(\bar{x}) = 0$. Este punto estacionario depende el punto inicial utilizado.

- 4) Dos direcciones consecutivas generadas por el método del gradiente son ortogonales, ver figura 3. En efecto: supongamos que estamos en la iteración $k + 1$ del método; la determinación del paso exacto α^k obtiene el hecho anterior, pues:

$$\alpha^k = \arg \min_{\alpha \geq 0} f(x^k + \alpha d^k)$$

Ahora bien, si $\alpha^k = 0$, significa que estamos en el óptimo en x^k y el método finaliza. Si $\alpha^k > 0$, α^k debe verificar la ecuación:

$$\left. \frac{df(x^k + \alpha d^k)}{d\alpha} \right|_{\alpha=\alpha^k} = 0 \Leftrightarrow \nabla f(x^k + \alpha^k d^k)^t d^k = 0 \Leftrightarrow \nabla f(x^k)^t d^k = -(d^{k+1})^t d^k = 0$$

- 5) Se llama método del gradiente o de Cauchy, o de máximo descenso, a la iteración del método anterior con la elección del paso α^k exacto.
- 6) Existen otros métodos de gradiente que no consideran la elección del paso exacto, sino que un paso elegido según reglas que dan una buena aproximación del paso exacto; o que consideran un paso constante para todas las iteraciones. Luego, en general, se considera un algoritmo del gradiente con elección de paso variable según algún criterio:

$$\begin{cases} x^0 \in \mathbb{R}^n \\ x^{k+1} = x^k + \lambda^k d^k \end{cases} \quad d^k = -\nabla f(x^k), \lambda^k \geq 0$$

3 Métodos de Direcciones Conjugadas

Por tener velocidad lineal, la convergencia del método del gradiente es lenta, aún en el caso de funciones cuadráticas. Luego, se busca un método de convergencia más rápida. Una manera posible de acelerar esta convergencia es utilizar la noción de direcciones conjugadas u ortogonales.

3.1 Método del Gradiente Conjugado

Sea (PC) el problema de optimización cuadrático definido por:

$$\min_{x \in \mathbb{R}^n} q(x) = \frac{1}{2}x^t Qx - b^t x \quad (\text{PC})$$

donde Q es una matriz definida positiva

Definición 5: Un conjunto de direcciones $\{d^1, \dots, d^k\}$ de \mathbb{R}^n , no nulas; $k \leq n$ se dirá Q -conjugado 2 a 2 si:

$$(d^i)^t Q d^j = \delta_{ij} \quad i, j = 1, \dots, k \quad \text{donde } \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

Por ejemplo, si $Q = I$, las direcciones son ortonormales.

Proposición 6: Si $\{d^1, \dots, d^k\}$ es un conjunto Q -conjugado, entonces $\{d^1, \dots, d^k\}$ es de rango máximo, i.e. los vectores d^i son linealmente independientes.

Dem: En efecto, sean $\alpha^i \in \mathbb{R}$; $i = 1, \dots, k$, tal que:

$$\sum_{i=1}^k \alpha^i d^i = 0$$

premultiplicando por $(d^j)^t Q \neq 0$, se tiene que:

$$\sum_{i=1}^k \alpha^i (d^j)^t Q d^i = \sum_{i=1}^k \alpha^i \delta_{ij} = \alpha^j \quad \forall j = 1, \dots, k$$

Luego $\alpha^j = 0 \quad \forall j = 1, \dots, k$. ■

El método de direcciones conjugadas para (PC), suponiendo que no se tienen las direcciones conjugadas ya calculadas, es el siguiente:

METODO DEL GRADIENTE CONJUGADO

Etapla 0: Seleccionar un punto inicial $x_0 \in \mathbb{R}^n$
 $k = 0$

$$g^0 = \nabla q(x^0) = Qx^0 - b$$

$$d^0 = -\nabla q(x^0)$$

Etapla 1:

$$\alpha^k = -\frac{(g^k)^t d^k}{(d^k)^t Q d^k}$$

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

Si $\|x^{k+1} - x^k\| \approx 0 \Rightarrow STOP$

Si no, seguir a Etapa 2.

Etapla 2:

$$g^{k+1} = \nabla q(x^{k+1}) = Qx^{k+1} - b$$

$$\rho_k = \frac{(g^{k+1})^t Q d^k}{(d^k)^t Q d^k}$$

$$d^{k+1} = -g^{k+1} + \rho^k d^k$$

$k = k + 1$ y volver a Etapa 1.

Teorema 7: El método del gradiente conjugado es un método de direcciones conjugadas. Si no termina en x_k , entonces:

i) $\langle g^0, \dots, g^k \rangle = \langle g^0, Qg^0, \dots, Q^k g^0 \rangle$

ii) $\langle d^0, \dots, d^k \rangle = \langle g^0, Qg^0, \dots, Q^k g^0 \rangle$

iii) $(d^k)^t Q d^i = 0 \quad \forall i = 0, \dots, k-1$

iv) $\alpha^k = \frac{(g^k)^t g^k}{(d^k)^t Q d^k} = \frac{\|g^k\|^2}{\|d^k\|_Q^2}$

v) $\rho_k = \frac{(g^{k+1})^t g^{k+1}}{(g^k)^t g^k} = \frac{\|g^{k+1}\|^2}{\|g^k\|^2}$

Dem: La demostración se deja propuesta. Se puede ver en refs. [1, 2]■

Observación:

Como las direcciones conjugadas son Q -ortogonales entre si, el método del gradiente conjugado converge en a lo más n iteraciones.

3.2 Método de Fletcher-Reeves

Corresponde a la versión del método del gradiente conjugado en el caso de una función f general. La iteración se define de la siguiente manera:

METODO DE FLETCHER-REEVES

Etapla 0: Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 $k = 0$
 $d^0 = -\nabla f(x^0)$

Etapla 1: Calcular α^k solución del problema unidimensional:
 $\alpha^k = \arg \min_{\alpha \geq 0} f(x^k + \alpha d^k)$
 $x^{k+1} = x^k + \alpha^k d^k$
 Si $\|x^{k+1} - x^k\| \approx 0 \implies STOP$
 Si no, seguir a Etapla 2.

Etapla 2: Calcular $\nabla f(x^{k+1})$
 $\rho^k = \frac{\nabla f(x^{k+1})^t \nabla f(x^{k+1})}{\nabla f(x^k)^t \nabla f(x^k)}$
 $d^{k+1} = -\nabla f(x^{k+1}) + \rho^k d^k$
 $k = k + 1$ y volver a Etapla 1.

Ejemplo: Volviendo al problema de optimización bi-dimensional anterior, ver ref. [1]:

$$\min_{(x_1, x_2) \in \mathbb{R}^2} (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

En la siguiente tabla se muestran 3 iteraciones del método de Fletcher-Reeves a partir de $x^0 = (0.00, 3.00)$

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\ \nabla f(x^k)\ $	α^k
0	(0.00, 3.00)	52.00	(-44.00, 24.00)	50.12	0.062
1	(2.54, 1.21)	0.10	(0.87, -0.48)	0.99	0.11
2	(2.25, 1.10)	0.008	(0.16, -0.20)	0.32	0.10
3	(2.19, 1.09)	0.0017	(0.05, -0.04)	0.06	0.11

4 Métodos Aproximados de Elección de Paso

Su utilizan para lograr una buena elección del paso en cada iteración de un método de optimización como alternativa a la determinación del paso exacto. Sea k la iteración de un método de optimización cualquiera y se quiere determinar un paso para la sucesión minimizante:

$$x^{k+1} = x^k + t^k d^k$$

Donde $t_k \geq 0$, d_k es una dirección de descenso en x_k de la función objetivo f , i.e:

$$(-\nabla f(x^k))^t d^k > 0$$

La idea del algoritmo para determinación del paso t_k , será calcular un intervalo $[t_g, t_d]$, $0 \leq t_g \leq t_d$ en el cual se pueda escoger $t_k \in [t_g, t_d]$ y se asegure una buena elección en todo el intervalo.

ALGORITMO GENERAL DE CÁLCULO DE PASO

- Etapa 0:** Inicialmente $t_g = t_d = 0$. Sea $t > 0$ un paso tentativo inicial.
- Etapa 1:** Analizamos t :
- Si el criterio (a) se satisface: $t_k = t$
 - Si el criterio (b) se satisface entonces el paso t es muy grande. Se define $t_d = t$, y se sigue a la etapa 2.
 - Si criterio (c) se satisface entonces el paso t es muy pequeño. Se define $t_g = t$, y se sigue a la etapa 2.
- Etapa 2:** Nuevo paso tentativo:
- Si $t_d = 0$, se define un nuevo t , mayor que t_g :
 $t = 10t_g$
 Y se vuelve a la etapa 1.
 - Si $t_d \neq 0$, se define un nuevo t en el intervalo (t_g, t_d) .
 La elección más utilizada es:
 $t = \frac{t_g + t_d}{2}$
 Y se vuelve a la etapa 1.

La Regla de Goldstein determina los criterios (a), (b) y (c) en el algoritmo general de determinación de paso. Sea:

$$h(t) = f(x^k + td^k) - f(x^k) \quad \text{Para } t \geq 0$$

se tiene que:

$$\begin{aligned} h'(t) &= \nabla f(x^k + td^k)^t d^k \\ h'(0) &= \nabla f(x^k)^t d^k < 0 \\ h(t) &\leq 0 \end{aligned}$$

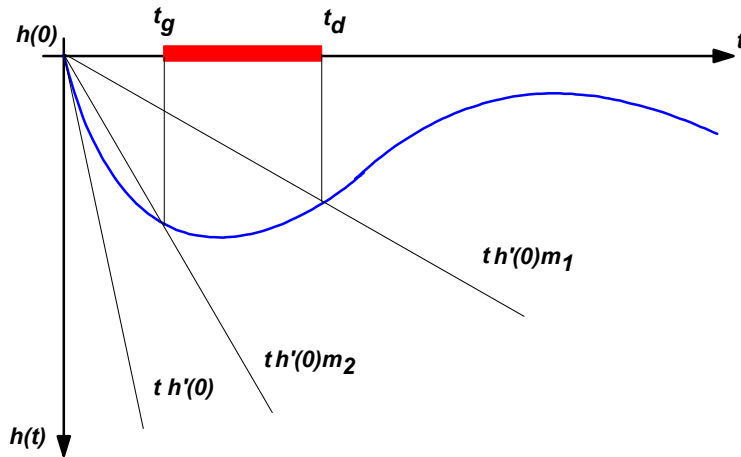


Figura 4. Regla de Elección de Paso de Goldstein

REGLA DE GOLDSTEIN

Sean $0 < m_1 < m_2 < 1$. Los criterios (a),(b) y (c) según esta regla son:

Criterio (a): $m_2 h'(0)t \leq h(t) \leq m_1 h'(0)t$

Criterio (b): $h(t) > m_1 h'(0)t$

Criterio (c): $h(t) < m_2 h'(0)t$

En la práctica, se prueba con distintos valores para m_1 y m_2 .

5 Métodos de 2º Orden: Newton - Kantorovich

El método de Newton es un método numérico que se utiliza para encontrar ceros de una función. Sea $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un campo vectorial de clase $\zeta^1(\mathbb{R}^n)$. Un punto $\bar{x} \in \mathbb{R}^n$ será un cero de F si:

$$F(\bar{x}) = 0$$

El método de Newton para este problema es el siguiente:

METODO DE NEWTON-KANTOROVICH

- Etapa 0:** Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 $k = 0$
- Etapa 1:** Calcular $F(x^k)$
Si $\|F(x^k)\| \approx 0 \implies STOP$
Si no, calcular $\nabla F(x^k)$, $[\nabla F(x^k)]^{-1}$ y seguir a Etapa 2.
- Etapa 2:** Calcular:
 $x^{k+1} = x^k - [\nabla F(x^k)]^{-1} F(x^k)$
 $k = k + 1$ y volver a Etapa 1.

Este método se puede aplicar a la búsqueda de los ceros del gradiente de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, suficientemente regular. Es decir, se buscan los puntos estacionarios de f .

METODO DE OPTIMIZACION DE NEWTON

- Etapa 0:** Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 $k = 0$
- Etapa 1:** Calcular $\nabla f(x^k)$
Si $\|\nabla f(x^k)\| \approx 0 \implies STOP$
Si no, calcular $\nabla^2 f(x^k)$, $[\nabla^2 f(x^k)]^{-1}$ y seguir a Etapa 2.
- Etapa 2:** Calcular:
 $x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$
 $k = k + 1$ y volver a Etapa 1.

Ejemplo: En la tabla siguiente se aplica el método de Newton al problema de optimización bi-dimensional a partir de $x^0 = (0.00, 3.00)$, ver ref. [1]:

$$\min_{(x_1, x_2) \in \mathbb{R}^2} (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

k	x^k	$f(x^k)$	$\nabla f(x^k)$	$\nabla^2 f(x^k)$	$-\left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k)$
0	(0.00, 3.00)	52.00	(-44.00, 24.00)	$\begin{bmatrix} 50.0 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.67, -2.67)
1	(0.67, 0.33)	3.13	(-9.39, -0.04)	$\begin{bmatrix} 23.23 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.44, 0.23)
2	(1.11, 0.56)	0.63	(-2.84, -0.04)	$\begin{bmatrix} 11.5 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.30, 0.14)
3	(1.41, 0.70)	0.12	(-0.80, -0.04)	$\begin{bmatrix} 6.18 & 4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.20, 0.10)
4	(1.61, 0.80)	0.02	(-0.22, -0.04)	$\begin{bmatrix} 3.83 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.13, 0.07)
5	(1.74, 0.87)	0.005	(-0.07, -0.00)	$\begin{bmatrix} 2.81 & -4.0 \\ -4.0 & 8.0 \end{bmatrix}$	(0.09, 0.04)

Observaciones:

- 1) La dirección del Método de Newton:

$$d^k = -\left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k)$$

es una dirección de descenso si y sólo si: $\nabla^2 f(x^k)$ es definida positiva, es decir:

$$(-\nabla f(x^k))d^k > 0 \iff \nabla^2 f(x^k) \text{ es definida positiva} \quad (1)$$

- 2) El método de Newton en cada iteración, considera una aproximación cuadrática de la función objetivo, y define el nuevo punto de la sucesión minimizante como el óptimo de la aproximación cuadrática de la función objetivo. Cerca de un óptimo local de f , la aproximación exacta.
- 3) El punto inicial x^0 no puede ser arbitrario, ya que para que el método converja, la dirección d^k debe ser de descenso. Esta corresponde a la principal desventaja

del método: su convergencia local. Sin embargo, su rapidez de convergencia es su mayor ventaja, posee convergencia de velocidad cuadrática, es decir:

$$\|x^{k+1} - \bar{x}\| \leq \lambda \|x^k - \bar{x}\|^2 \quad \text{para } \lambda < 1$$

- 4) El principal problema del Método de Newton es que la matriz $\nabla^2 f(x^k)$ debe ser definida positiva. Si se parte de un punto x^0 suficientemente cercano a un mínimo local, esto se verifica. En general se desconoce toda información acerca de la solución óptima del problema, y por lo tanto no se tiene un buen criterio para escoger x^0 .
- 5) Esto sugiere, que un buen método de minimización es una combinación del Método del Gradiente y del Método de Newton: Inicialmente se aplica el Método del Gradiente que tiene convergencia global de velocidad lineal y luego se aplica el Método del Newton que tiene convergencia local de velocidad cuadrática.

6 Métodos Cuasi-Newton

En cada iteración del método de Newton es necesario calcular la inversa de la matriz hessiana de f en x_k de manera exacta, lo que es costoso computacionalmente, $O(n^3)$ operaciones aritméticas. Por esto razón, se propone un método iterativo de la forma:

$$x^{k+1} = x^k - t^k S^k g^k \quad g^k = \nabla f(x^k) \quad (2)$$

donde S^k es una matriz que aproxima a $[\nabla^2 f(x^k)]^{-1}$ y $t^k \geq 0$ minimiza f sobre $x^k - \lambda g^k$ para $\lambda \geq 0$ (paso exacto o aproximado).

Se presentarán 2 métodos que permiten construir iterativamente la matriz S^k de manera que se verifiquen las siguientes condiciones:

- i) Definida Positiva: Si S^k es definida positiva $\implies S^{k+1}$ también lo es
- ii) Aproximación de la matriz hessiana inversa de f : $x^k \longrightarrow \bar{x}$, $S^k \longrightarrow [\nabla^2 f(\bar{x})]^{-1}$ para $k \longrightarrow \infty$.

La forma de construir estos métodos asegura:

- Convergencia Global
- Rapidez de convergencia mayor que lineal
- Convergencia a un mínimo local

METODO DFP (Davidon-Fletcher-Powell)

Etapa 0: Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 Inicializar $S_0 = I_{n \times n}$
 $k = 0$

Etapa 1: Calcular $g^k = \nabla f(x^k)$
 Si $\|g^k\| \approx 0 \implies STOP$
 Si no, calcular $x^{k+1} = x^k - t^k S^k g^k$
 donde $t_k \geq 0$ se escoge según regla de Goldstein
 y seguir a Etapa 2.

Etapa 2: Calcular:
 $p^k = x^{k+1} - x^k \quad q^k = g^{k+1} - g^k$
 $S^{k+1} = S^k + \frac{p^k (p^k)^t}{(p^k)^t q^k} - \frac{S^k q^k (q^k)^t S^k}{(q^k)^t S^k q^k}$
 $k = k + 1$ y volver a Etapa 1.

METODO BFGS (Broyden-Fletcher-Goldfarb-Shanno)

Etapa 0: Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
 Inicializar $S_0 = I_{n \times n}$
 $k = 0$

Etapa 1: Calcular $g^k = \nabla f(x^k)$
 Si $\|g^k\| \approx 0 \implies STOP$
 Si no, calcular $x^{k+1} = x^k - t^k S^k g^k$
 donde $t_k \geq 0$ se escoge según regla de Goldstein
 y seguir a Etapa 2.

Etapa 2: Calcular:
 $p^k = x^{k+1} - x^k \quad q^k = g^{k+1} - g^k$
 $S^{k+1} = S^k + \left[1 + \frac{(q^k)^t S^k q^k}{(p^k)^t q^k} \right] \frac{p^k (p^k)^t}{(p^k)^t q^k} - \frac{[p^k (q^k)^t S^t + S^k q^k (p^k)^t]}{(p^k)^t q^k}$
 $k = k + 1$ y volver a Etapa 1.

References

- [1] M. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming: Theory and Algorithms, Second Edition, John Wiley and Sons, 1993.
- [2] D. G. Luenberger, Linear and Nonlinear Programming, Second Edition, John Wiley and Sons, 1984.