

# Optimización en Ingeniería

**Dr. Carlos A. Coello Coello**  
**Departamento de Computación**  
**CINVESTAV-IPN**  
**Av. IPN No. 2508**  
**Col. San Pedro Zacatenco**  
**México, D.F. 07300**  
**email: [ccoello@cs.cinvestav.mx](mailto:ccoello@cs.cinvestav.mx)**

## El Método de Marquardt

### Algoritmo

Paso 1: Elegir un punto inicial  $X^{(0)}$  el número máximo de iteraciones,  $M$ , y un parámetro de terminación,  $\epsilon$

Hacer  $k = 0$  y  $\lambda^{(0)} = 1 \times 10^4$  (un valor grande)

Paso 2: Calcular  $\nabla f(X^{(k)})$

Paso 3: IF  $\|\nabla f(X^{(k)})\| \leq \epsilon$  o  $k \geq M$  THEN Terminar

ELSE GOTO Paso 4.

Paso 4: Calcular  $s(X^{(k)}) = - [\mathbf{H}^{(k)} + \lambda^{(k)} \mathbf{I}]^{-1} \nabla f(X^{(k)})$

Hacer  $X^{(k+1)} = X^{(k)} + s(X^{(k)})$

Paso 5: ¿Es  $f(X^{(k+1)}) < f(X^{(k)})$ ?

Si es así, GOTO Paso 6.

ELSE GOTO Paso 7.

Paso 6: Hacer  $\lambda^{(k+1)} = \frac{1}{2} \lambda^{(k)}$

$k = k + 1$ . GOTO Paso 2

Paso 7: Hacer  $\lambda^{(k)} = 2\lambda^{(k)}$

GOTO Paso 4.

## El Método de Marquardt

En este método suele usarse un valor grande de  $\lambda$  inicialmente. Por tanto, la matriz Hessiana tiene poco efecto en la determinación de la dirección de búsqueda (ver paso 4 del algoritmo). Inicialmente, la búsqueda es similar a la del método de Cauchy. Después de un cierto número de iteraciones (cuando se espera que la solución se encuentre ya cerca del mínimo), el valor de  $\lambda$  se vuelve pequeño y el efecto de la matriz Hessiana es similar al del método de Newton.

## El Método de Marquardt

Las mayores ventajas de este método son su simplicidad, la propiedad de descenso con que cuenta, la excelente tasa de convergencia cerca de  $x^*$  y la ausencia de una búsqueda lineal (o sea que no tienen que hacerse las búsquedas unidireccionales de los otros métodos de gradiente que hemos visto hasta ahora).

## El Método de Marquardt

La principal desventaja de este método es la necesidad de calcular la matriz Hessiana y realizar los cálculos matriciales posteriores (sumas y multiplicaciones matriciales, además del cálculo de la inversa de una matriz). Esto puede resultar computacionalmente muy costoso para problemas grandes.

## El Método de Marquardt

Es importante hacer notar que el algoritmo puede hacerse más rápido efectuando una búsqueda unidireccional al estar buscando el nuevo punto en el paso 4 del algoritmo, aunque esta modificación normalmente no se introduce.

## El Método de Marquardt

Puesto que el calcular la matriz Hessiana y su inversa es computacionalmente costoso, normalmente no se efectúa la búsqueda unidireccional a lo largo de  $s^{(k)}$ . Sin embargo, para funciones objetivo más simples, puede efectuarse una búsqueda unidireccional en el paso 4 del algoritmo, para encontrar el punto  $X^{(k+1)}$ .

## El Método de Marquardt

Este método ha sido usado extensivamente con problemas donde  $f(x)$  es una suma de cuadrados. Esto es:

$$f(x) = f_1^2(x) + f_2^2(x) + \dots + f_m^2(x) \quad (1)$$

De hecho, éste fue el problema considerado originalmente por Marquardt al desarrollar su método.

## El Método de Marquardt

Powell (1972) y Bard (1974) proporcionaron resultados numéricos que indican que el método es particularmente atractivo para este tipo de problemas.

## Métodos de Quasi-Newton

Estos métodos son similares a los métodos de gradiente conjugado en el sentido de que se basan principalmente en propiedades de las funciones cuadráticas. Sin embargo, en el método del gradiente conjugado, la principal fortaleza de la búsqueda se deriva del uso de las direcciones conjugadas de búsqueda, mientras que los métodos de Quasi-Newton están diseñados para imitar más directamente las características positivas del método de Newton pero usando sólo información de primer orden.

## Métodos de Quasi-Newton

Todos los métodos de este grupo generan las direcciones a usarse para generar el siguiente punto con:

$$s(X^{(k)}) = -\mathbf{A}^{(k)} \nabla f(X^{(k)}) \quad (2)$$

donde  $\mathbf{A}^{(k)}$  es una matriz de  $N \times N$  a la que se denomina la **métrica**.

## Métodos de Quasi-Newton

A los métodos que emplean direcciones de esta forma, se les suele llamar **métodos de métrica variable**, porque **A** cambia a cada iteración. Para ser precisos, un método de métrica variable es un método **quasi-Newton** si está diseñado de tal forma que satisfaga la siguiente propiedad cuadrática:

$$\Delta x = \mathbf{C}^{-1} \Delta g \quad (3)$$

## Métodos de Quasi-Newton

Desafortunadamente, la literatura no es precisa ni consistente en el uso de estos términos, por lo que algunos autores sugieren usarlos de manera intercambiable. Esto es apropiado, puesto que ambas expresiones son de igual importancia en el diseño y ejecución de estos métodos.

## Métodos de Quasi-Newton

Supongamos una recursión para estimar la inversa de la matriz Hessiana:

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \mathbf{A}_c^{(k)} \quad (4)$$

donde  $\mathbf{A}_c^{(k)}$  es una corrección a la métrica actual.

## Métodos de Quasi-Newton

La idea es formar  $\mathbf{A}_c^{(k)}$  tal que la secuencia  $\mathbf{A}^{(0)}$ ,  $\mathbf{A}^{(1)}$ ,  $\mathbf{A}^{(2)}$ ,  $\dots$ ,  $\mathbf{A}^{(k+1)}$  aproxime  $H^{-1} = \nabla^2 f(x^*)^{-1}$ , porque al lograr esto, una búsqueda lineal adicional producirá  $x^*$  si  $f(x)$  es cuadrática.

Obviamente, se espera que el éxito de este mecanismo en funciones cuadráticas nos hará tener éxito con funciones no lineales generales.

## Métodos de Quasi-Newton

Recordemos la importante propiedad de una función cuadrática dada en la ecuación (3), y supongamos que nuestra aproximación de  $\mathbf{C}^{-1}$  es de la forma:

$$\mathbf{C}^{-1} = \beta \mathbf{A}^{(k)} \quad (5)$$

donde  $\beta$  es un escalar.

## Métodos de Quasi-Newton

Nos gustaría, sin embargo, que la aproximación satisfaga la ecuación (3); esto es:

$$\Delta x^{(k)} = \mathbf{A}^{(k)} \Delta g^{(k)} \quad (6)$$

pero esto es claramente imposible, puesto que necesitamos  $\mathbf{A}^{(k)}$  para encontrar  $\Delta g^{(k)}$ , donde:

$$\Delta x^{(k)} = x^{(k+1)} - x^{(k)} \quad (7)$$

$$\Delta g^{(k)} = g(x^{(k+1)}) - g(x^{(k)}) \quad (8)$$

## Métodos de Quasi-Newton

Por otro lado, podemos requerir que la nueva aproximación satisfaga:

$$\Delta x^{(k)} = \beta \mathbf{A}^{(k+1)} \Delta g^{(k)} \quad (9)$$

Combinando las ecuaciones (4) y (9), tenemos:

$$\mathbf{A}_c^{(k)} \Delta g^{(k)} = \left( \frac{1}{\beta} \right) \Delta x^{(k)} - \mathbf{A}^{(k)} \Delta g^{(k)} \quad (10)$$

## Métodos de Quasi-Newton

No es obvio encontrar una solución, pero puede verificarse por sustitución directa que:

$$\mathbf{A}_c^{(k)} = \frac{1}{\beta} \left( \frac{\Delta x^{(k)} y^T}{y^T \Delta g^{(k)}} \right) - \frac{\mathbf{A}^{(k)} \Delta g^{(k)} z^T}{z^T \Delta g^{(k)}} \quad (11)$$

es una solución. Como  $y$  &  $z$  son vectores arbitrarios, esta es realmente una familia de soluciones.

## Métodos de Quasi-Newton

Si hacemos:

$$y = \Delta x^{(k)}, \quad z = \mathbf{A}^{(k)} \Delta g^{(k)} \quad (12)$$

se obtiene el **Método de Davidon-Fletcher-Powell**.

## Método de Davidon-Fletcher-Powell

Este método fue propuesto en 1959 y mejorado en 1963. La expresión que utiliza para estimar la matriz Hessiana es la siguiente:

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k-1)} + \frac{\Delta \mathbf{x}^{(k-1)} \Delta \mathbf{x}^{(k-1)T}}{\Delta \mathbf{x}^{(k-1)T} \Delta \mathbf{g}^{(k-1)}} - \frac{\mathbf{A}^{(k-1)} \Delta \mathbf{g}^{(k-1)} [\mathbf{A}^{(k-1)} \Delta \mathbf{g}^{(k-1)}]^T}{\Delta \mathbf{g}^{(k-1)T} \mathbf{A}^{(k-1)} \Delta \mathbf{g}^{(k-1)}} \quad (13)$$

Es conveniente hacer  $\mathbf{A}^{(0)} = \mathbf{I}$ .

## Método de Davidon-Fletcher-Powell

Puede demostrarse que esta fórmula preserva las propiedades de una matriz positivamente definida y simétrica, de tal manera que si  $\mathbf{A}^{(0)}$  es simétrica y positivamente definida, también lo serán  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots$  en la ausencia de errores de redondeo.

## Método de Davidon-Fletcher-Powell

Recordemos que para que la primera variación de  $f(x)$  escribimos:

$$\Delta f(x) = \nabla f(x^{(k)})^T \Delta x \quad (14)$$

Esto se puede reescribir (usando expresiones vistas anteriormente) como:

$$\Delta f(x) = -\nabla f(x^{(k)})^T \alpha^{(k)} \mathbf{A}^{(k)} \nabla f(x^{(k)}) \quad (15)$$

## Método de Davidon-Fletcher-Powell

Por lo tanto:

$$\Delta f(x) = -\alpha^{(k)} \nabla f(x^{(k)})^T \mathbf{A}^{(k)} \nabla f(x^{(k)}) \quad (16)$$

asegura que:  $f(x^{(k+1)}) < f(x^{(k)})$  para toda  $\alpha^{(k)} > 0$  y  $\mathbf{A}^{(k)}$  es positivamente definida. Por tanto, el algoritmo de Davidon-Fletcher-Powell tiene una propiedad descendente.

## Método de Davidon-Fletcher-Powell

### Algoritmo

- Paso 1: Elegir un punto inicial  $X^{(0)}$  y tolerancias  $\epsilon_1$ ,  $\epsilon_2$  y  $\epsilon_3$
- Paso 2: Encontrar  $\nabla f(X^{(0)})$  y hacer:  $s^{(0)} = -\nabla f(X^{(0)})$
- Paso 3: Encontrar  $\lambda^{(0)}$  tal que:  $f(X^{(0)} + \lambda^{(0)}s^{(0)})$  se minimice con una tolerancia  $\epsilon_1$ .  
Hacer  $X^{(1)} = X^{(0)} + \lambda^{(0)}s^{(0)}$  y  $k = 1$   
Calcular  $\nabla f(X^{(1)})$
- Paso 4: Hacer:  $s^{(k)} = -\mathbf{A}^{(k)}\nabla f(x^{(k)})$
- Paso 5: Encontrar  $\lambda^{(k)}$  tal que:  
 $f(X^{(k)} + \lambda^{(k)}s^{(k)})$  sea mínima con una tolerancia  $\epsilon_1$   
Hacer  $X^{(k+1)} = X^{(k)} + \lambda^{(k)}s^{(k)}$
- Paso 6: ¿Es  $\frac{\|X^{(k+1)} - X^{(k)}\|}{\|X^{(k)}\|} \leq \epsilon_2$  o  $\|\nabla f(X^{(k+1)})\| \leq \epsilon_3$ ?  
Si es así, terminar.  
ELSE  $k = k + 1$ . GOTO Paso 4.

## Método de Davidon-Fletcher-Powell

El método de Davidon-Fletcher-Powell ha sido y sigue siendo una técnica de gradiente ampliamente utilizada. El método tiende a ser robusto; esto es, típicamente tiende a tener un buen comportamiento en una amplia variedad de problemas prácticas.

La mayor desventaja de este tipo de métodos es su necesidad de almacenar la matriz  $\mathbf{A}$  de  $N \times N$ .

## Método de Davidon-Fletcher-Powell

Desde la publicación de la fórmula de Davidon, se han propuesto varios métodos más que resultan de usar diferentes valores de  $\beta$ ,  $y$ ,  $z$  en la ecuación (11).

Una de las dificultades prácticas comunes de estos métodos es la tendencia de  $\mathbf{A}^{(k+1)}$  a estar mal condicionada, lo que causa una mayor dependencia a un procedimiento de reinicialización.

## Método de Broyden-Fletcher-Goldfarb-Shanno

Este método fue propuesto en 1970 por Broyden y refinado posteriormente por Fletcher, Goldfarb y Shanno. Pertenece a la misma familia del método de Davidon-Fletcher-Powell, pero se le considera más poderoso. Este método ha sido muy elogiado en la literatura especializada e investigadores como Powell (1975) lo han recomendado ampliamente.