

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

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 β , 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.

Método de Broyden-Fletcher-Goldfarb-Shanno

Algoritmo

- Paso 1: Elegir un punto inicial $X^{(0)}$ y tolerancias ϵ_1 , ϵ_2 y ϵ_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 ϵ_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 ϵ_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 Broyden-Fletcher-Goldfarb-Shanno

El algoritmo del Método de Broyden-Fletcher-Goldfarb-Shanno es exactamente el mismo que el del método de Davidon-Fletcher-Powell. La única diferencia es la forma de calcular $\mathbf{A}^{(k)}$. En este caso se usa:

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

Método de Broyden-Fletcher-Goldfarb-Shanno

Nótese, sin embargo, que una diferencia entre el método de Davidon-Fletcher-Powell y éste es que en este caso lo que se actualiza iterativamente es la matriz Hessiana misma a diferencia del método de Davidon donde se actualiza la inversa de la matriz Hessiana.

Al igual que en el método de Davidon-Fletcher-Powell, la matriz $\mathbf{A}^{(0)}$ debe ser simétrica y positivamente definida. Se sugiere usar:
 $\mathbf{A}^{(0)} = \mathbf{I}$.

Método de Broyden-Fletcher-Goldfarb-Shanno

Este método puede ser considerado como quasi-Newton, de gradiente conjugado y de métrica variable.

Puesto que se aproxima la inversa de la matriz Hessiana, a este método se le puede llamar también de actualizaciones indirectas.

Se ha demostrado que este método exhibe convergencia superlineal en la proximidad del óptimo (Dennis & More, 1977).

Método de Broyden-Fletcher-Goldfarb-Shanno

Si las búsquedas lineales son exactas, la matriz \mathbf{A} retiene su propiedad de ser positivamente definida. Sin embargo, en la práctica, la matriz \mathbf{A} puede volverse indefinida e incluso singular debido a errores de redondeo acumulados. Para lidiar con este problema, se sugiere, al igual que en el método de Davidon-Fletcher-Powell, que se reinicialice periódicamente la matriz \mathbf{A} usando la matriz identidad.

Método de Broyden-Fletcher-Goldfarb-Shanno

Nótese, sin embargo, que hay bastante evidencia empírica que indica que este método es menos susceptible a las búsquedas lineales inexactas que el método de Davidon-Fletcher-Powell y que requiere también menos reinicializaciones.

Métodos de Métrica Variable (Discusión Adicional)

Existe una gran similitud entre los esfuerzos actuales por mejorar los métodos de métrica variable y aquellos que buscan mejorar los métodos de gradiente conjugado. Por tanto, mucho de lo que dijimos antes sobre los métodos de gradiente conjugado aplica también para los métodos de métrica variable.

Métodos de Métrica Variable (Discusión Adicional)

Davidon (1975) ha propuesto una clase de actualizaciones que permiten búsquedas lineales inexactas. Powell (1977) estableció posteriormente la propiedad de terminación cuadrática de estos métodos en la ausencia de búsquedas lineales.

Métodos de Métrica Variable (Discusión Adicional)

Pareciera ser que un método de métrica variable con terminación cuadrática pero sin la necesidad de búsquedas lineales costosas sería robusto y rápido en funciones generales. Goldfarb (1977) se cuenta entre los que han explorado esta promisorio línea de investigación.

Métodos de Métrica Variable (Discusión Adicional)

En general, el método de Davidon-Fletcher-Powell suele reinicializarse (es decir, hacemos $\mathbf{A} = \mathbf{I}$) después de N actualizaciones. Esto es usualmente una medida muy conservadora, ya que pueden requerirse muchos más pasos antes de que realmente se requiera una reinicialización.

Métodos de Métrica Variable (Discusión Adicional)

La necesidad de reinicializar se incrementa conforme empeora el valor de condicionamiento de la matriz:

$$K(\mathbf{A}) = \left| \frac{\lambda_h}{\lambda_l} \right| \quad (2)$$

donde: $K(\mathbf{A})$ es el número de condicionamiento de la matriz \mathbf{A} , λ_h y λ_l son los eigenvalores de mayor y menor módulo, respectivamente.

Una matriz con un valor de condicionamiento grande está **mal condicionada**. Una matriz con un valor de condicionamiento cercano a 1 está **bien condicionada**.

Métodos de Métrica Variable (Discusión Adicional)

Es importante hacer notar que aunque las reinicializaciones proporcionan un cierto grado de seguridad (o robustez) en los métodos de métrica variable, éstas típicamente hacen más lento el progreso a la solución, puesto que se desecha una estimación de segundo orden.

Métodos de Métrica Variable (Discusión Adicional)

McCormick (1972), Shanno (1978) y Shanno & Phua (1979) han investigado de manera extensiva la relación entre el gradiente conjugado y los métodos de métrica variable.

Shanno ve a los métodos de gradiente conjugado como métodos de métrica variable “sin memoria”.

La mayor parte de los investigadores coinciden en la actualidad en afirmar que estas 2 clases de métodos comparten mucho más de lo que originalmente se creía.

Métodos de Métrica Variable (Discusión Adicional)

Además, resulta claro que las muchas variantes de los métodos de métrica variable tienen mucho en común en la práctica (Dixon, 1972), lo que hace que se tenga que sopesar cuidadosamente el costo computacional adicional de los métodos más complejos.

Shanno & Phua (1978) proporcionan numerosos resultados que favorecen el uso del método de Broyden-Fletcher-Goldfarb-Shanno.

Métodos de Métrica Variable (Discusión Adicional)

Los métodos de métrica variable han sido utilizados de manera extensiva para el desarrollo de técnicas de optimización para problemas con restricciones.

También hay una cantidad importante de trabajo en torno a hacer los métodos de métrica variable más atractivos para problemas grandes.

Métodos de Métrica Variable (Discusión Adicional)

Los métodos de gradiente conjugado son los que suelen considerarse como los mejores para problemas de alta dimensionalidad (o sea, aquellos con muchas variables de decisión).

Sin embargo, se han logrado muchos avances en lo que a los métodos de métrica variable se refiere, sobre todo en aquellos casos en los que la matriz Hessiana es dispersa.

Regiones de Confianza

Todos los métodos de gradiente que hemos visto emplean la misma estrategia básica:

1. Elegir una dirección $s(\vec{x})$ que se designa para producir un descenso.
2. Explotar esta dirección al máximo; es decir, realizar una búsqueda unidireccional.

Regiones de Confianza

Cuando ya no es posible encontrar una dirección de descenso hemos localizado un mínimo, al menos local. Aún los métodos de búsqueda directa como Hooke-Jeeves y Nelder-Mead usan esta estrategia básica, aunque no esto no resulte del todo obvio, ni se plantee explícitamente en la descripción del algoritmo.

Regiones de Confianza

Los métodos basados en regiones de confianza usan una estrategia ligeramente diferente:

1. Formar una aproximación “confiable” de $f(\vec{x})$.
2. Encontrar el mínimo de la aproximación.

Regiones de Confianza

La “región de confianza” se define como un vecindario alrededor de la solución actual ($x^{(t)}$) en la cual la aproximación actual de $f(\vec{x})$ produce un descenso. El énfasis cambia ahora de encontrar direcciones de descenso a formar aproximaciones “seguras” a $f(\vec{x})$.

Regiones de Confianza

El método de Marquardt es un método basado en regiones de confianza. Nótese que este método no requiere una búsqueda unidireccional para converger. Asimismo, la elección del parámetro λ determina la aproximación a emplearse en cada iteración. El método de Marquardt también exhibe el principal problema de los métodos basados en regiones de confianza: el costo computacional de formar la función de aproximación.

Comparaciones de Métodos

No ha sido sino hasta recientemente que se ha podido decir algo respecto a las tasas de convergencia de los métodos de gradiente desde una perspectiva puramente teórica. La mayor parte de lo publicado al respecto se deriva de estudios empíricos. A continuación revisaremos rápidamente algunos de los trabajos más representativos que se han publicado a este respecto.

Comparaciones de Métodos

Himmelblau (1972) compara una gran cantidad de métodos de optimización (tanto directos como de gradiente) con base en su robustez, número de evaluaciones de la función objetivo y tiempo de cómputo empleado para terminar, al aplicarse a una familia de problemas de prueba. La robustez es, en este caso, una medida del éxito para obtener una solución óptima de una precisión dada para una amplia gama de problemas. Himmelblau concluye que los mejores métodos son: Broyden-Fletcher-Goldfarb-Shanno, Davidon-Fletcher-Powell y el método de búsqueda directa de Powell.

Comparaciones de Métodos

Sargent y Sebastian (1971) realizaron un estudio similar, pero enfocado sólo a métodos de gradiente. En él estudian el efecto de los criterios de terminación de la búsqueda lineal, la frecuencia de las reinicializaciones, y la precisión del gradiente. Los resultados indican nuevamente la superioridad de los métodos quasi-Newton para funciones generales (en particular, Broyden-Fletcher-Goldfarb-Shanno).

Comparaciones de Métodos

Carpenter y Smith (1975) reportan la eficiencia computacional relativa del método de búsqueda directa de Powell, de Davidon-Fletcher-Powell y del método de Newton para aplicaciones estructurales. Concluyen que para los problemas de prueba considerados, Davidon-Fletcher-Powell y Newton son métodos superiores al método de búsqueda directa de Powell, y que Davidon-Fletcher-Powell es superior al método de Newton para problemas grandes.

Comparaciones de Métodos

Shanno y Pua (1978) realizaron un estudio similar, muy exhaustivo, en el cual evalúan una amplia variedad de métodos de gradiente conjugado y de métrica variable. Los resultados sugieren que el método de Broyden-Fletcher-Goldfarb-Shanno es superior a los otros.

Optimización en Espacios de Búsqueda Restringidos

Previamente discutimos los criterios de optimalidad necesarios y suficientes para problemas de optimización sin restricciones. Sin embargo, la mayor parte de los problemas de ingeniería involucran optimizaciones sujetas a varias restricciones.

Optimización en Espacios de Búsqueda Restringidos

La presencia de restricciones esencialmente reduce la región en la cual buscaremos el óptimo. A simple vista, pudiera pensarse que la reducción del tamaño de la región factible debiera simplificar la búsqueda. No obstante, ocurre exactamente lo contrario, ya que el proceso de optimización se vuelve más complicado puesto que los criterios de optimización presentados previamente no necesariamente se cumplen en la presencia de restricciones.

Optimización en Espacios de Búsqueda Restringidos

Aún la condición básica de que un óptimo debe estar en un punto estacionario, donde el gradiente es cero, puede ser violada. Por ejemplo, el **mínimo sin restricciones** de $f(x) = (x - 2)^2$ ocurre en el punto estacionario $x = 2$.

Optimización en Espacios de Búsqueda Restringidos

Sin embargo, si al problema se le agrega la restricción de que $x \geq 4$, entonces el mínimo restringido ocurre en el punto $x = 4$. Este NO es un punto estacionario de f , puesto que $f'(4) = 4$. Debemos entonces revisar las condiciones necesarias y suficientes de optimalidad para problemas restringidos.

Problema con Restricciones de Igualdad

Consideremos un problema de optimización que involucre varias restricciones de igualdad:

$$\text{Min } f(x_1, x_2, \dots, x_N) \quad (3)$$

sujeta a: $h_k(x_1, x_2, \dots, x_N) = 0, k = 1, \dots, K$

En principio, este problema puede resolverse como un problema de optimización sin restricciones si se eliminan explícitamente K variables independientes usando las restricciones de igualdad.

Problema con Restricciones de Igualdad

En efecto, la presencia de restricciones de igualdad reduce la **dimensionalidad** del problema original de N a $N - K$. Una vez que el problema ha sido reducido a uno sin restricciones, podemos usar los métodos vistos anteriormente.

Problema con Restricciones de Igualdad

El método de eliminación de variables es aplicable siempre y cuando las restricciones de igualdad puedan resolverse explícitamente para un conjunto dado de variables independientes. En la presencia de varias restricciones de igualdad, el proceso de eliminación puede volverse engorroso. Además, en ciertas situaciones, puede no ser posible resolver explícitamente las restricciones para eliminar una variable.

Problema con Restricciones de Igualdad

Por ejemplo, si en el problema anterior la restricción $h_1(x) = 0$ estuviera dada por:

$$h_1(x) = x_1^2 x_3 + x_2 x_3^2 + x_2^{-1} x_1 = 0 \quad (4)$$

entonces no sería posible una solución explícita de una variable en términos de las demás. Para esos casos, es mejor usar el método de los **multiplicadores de Lagrange** que veremos a continuación.

Multiplicadores de Lagrange

El método de los multiplicadores de Lagrange esencialmente proporciona un conjunto de condiciones necesarias para identificar candidatos al óptimo en problemas de optimización con restricciones de igualdad. Esto se hace convirtiendo el problema con restricciones en un problema equivalente sin restricciones con la ayuda de ciertos parámetros no especificados conocidos como **multiplicadores de Lagrange**.

Multiplicadores de Lagrange

Consideremos la minimización de una función de n variables sujeta a una restricción de igualdad:

$$\text{Min } f(x_1, x_2, \dots, x_N) \quad (5)$$

sujeta a:

$$h_1(x_1, x_2, \dots, x_N) = 0 \quad (6)$$

Multiplicadores de Lagrange

El método de los multiplicadores de Lagrange consiste en convertir este problema en el siguiente (sin restricciones):

$$\text{Min } L(x; v) = f(x) - v h_1(x) \quad (7)$$

La función sin restricciones $L(x; v)$ es llamada la **función Lagrangiana** y v es una constante no especificada llamada el **multiplicador Lagrangiano**. No hay restricciones de signo para el valor de v .

Multiplicadores de Lagrange

Supongamos que para un valor fijo de $v = v^0$, el mínimo sin restricciones de $L(x; v)$ con respecto a x ocurre en $x = x^0$ y x^0 satisface $h_1(x^0) = 0$. Entonces, es claro que x^0 minimiza a la ecuación (3) sujeta a (4), porque para todos los valores de x que satisfacen (4), se cumple que $h_1(x) = 0$ y $\min L(x; v) = \min f(x)$.

Multiplicadores de Lagrange

Obviamente el reto es determinar el valor apropiado para $v = v^0$ de manera que el mínimo restringido x^0 satisfaga (4). Pero esto puede lograrse tratando a v como una variable, encontrando el mínimo sin restricciones de (5) como una función de v y ajustando v de manera que se satisfaga (4).

Multiplicadores de Lagrange

En la solución del ejemplo anterior, tratamos a $L(x; v)$ como una función de 2 variables (x_1 y x_2) y consideramos a v como un parámetro cuyo valor es “ajustado” para satisfacer la restricción. En problemas donde es difícil obtener una solución explícita a:

$$\frac{\partial L}{\partial x_j} = 0 \quad \text{para } j = 1, 2, \dots, N \quad (8)$$

como una función de v , pueden determinarse los valores de x y v simultáneamente.

Multiplicadores de Lagrange

Para ello, se resuelve el siguiente sistema de $N + 1$ ecuaciones con $N + 1$ incógnitas:

$$\frac{\partial L}{\partial x_j} = 0 \quad \text{para } j = 1, 2, \dots, N \quad (9)$$

$$h_1(x) = 0 \quad (10)$$

Multiplicadores de Lagrange

Cualquier técnica de búsqueda para funciones no restringidas de varias variables, de las vistas anteriormente, puede usarse para determinar todas las soluciones posibles.

Por cada solución $(x^0; v^0)$, debe evaluarse la matriz Hessiana de L con respecto a x , a fin de determinar si ésta es positivamente definida para un mínimo local (o negativamente definida para un máximo local).

Multiplicadores de Lagrange

Es importante hacer notar que si consideramos a L como una función de 3 variables en el ejemplo anterior (o sea x_1 , x_2 y v), entonces los puntos $(x^{(1)}; v_1)$ y $(x^{(2)}; v_2)$ no corresponden a un mínimo o a un máximo de L con respecto a x y v . De hecho, se vuelven puntos de paso de la función $L(x, v)$. Más adelante hablaremos en más detalle de los puntos de paso.

Multiplicadores de Lagrange

El método de los multiplicadores de Lagrange puede extenderse a varias restricciones de igualdad.

Consideremos el problema general:

$$\text{Min } f(x) \quad (11)$$

sujeta a: $h_k(x) = 0, k = 1, 2, \dots, K$

Multiplicadores de Lagrange

La función Lagrangiana será entonces:

$$L(x; v) = f(x) - \sum_{k=1}^K v_k h_k \quad (12)$$

donde v_1, v_2, \dots, v_k son los multiplicadores de Lagrange, los cuales son parámetros no especificados cuyos valores serán determinados posteriormente.

Multiplicadores de Lagrange

Planteando las derivadas parciales de L con respecto a x como iguales a cero, obtenemos el siguiente sistema de N ecuaciones con N incógnitas:

$$\begin{aligned}\frac{\partial L(x; v)}{\partial x_1} &= 0 \\ \frac{\partial L(x; v)}{\partial x_2} &= 0 \\ &\vdots \\ \frac{\partial L(x; v)}{\partial x_N} &= 0\end{aligned}\tag{13}$$

Multiplicadores de Lagrange

Es difícil resolver el sistema anterior explícitamente como una función del vector v . Sin embargo, podemos aumentar el sistema con las ecuaciones de las restricciones:

$$h_1(x) = 0$$

$$h_2(x) = 0$$

$$\vdots$$

$$h_k(x) = 0$$

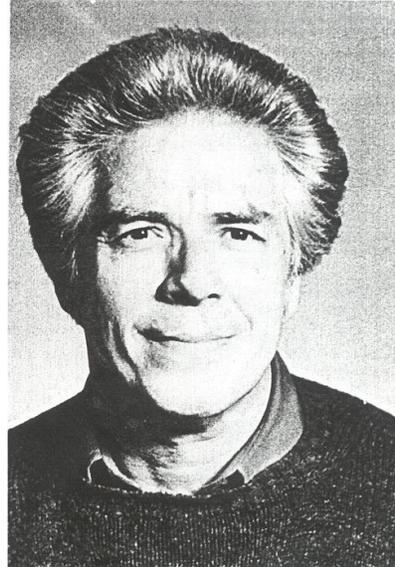
Multiplicadores de Lagrange

Una solución al sistema aumentado de $N + K$ ecuaciones en $N + K$ variables proporciona un punto estacionario de L . Puede verificarse si este punto es un mínimo o un máximo para calcular la matriz Hessiana de L con respecto a x tal y como hicimos en el ejemplo anterior en el que se usó una sola restricción.

Multiplicadores de Lagrange

Pueden existir algunos problemas para los cuales el sistema aumentado de $N + K$ ecuaciones con $N + K$ incógnitas pudiera no tener una solución. En tales casos, el método de los multiplicadores de Lagrange fallaría. Sin embargo, tales casos son raros en la práctica.

Condiciones de Kuhn-Tucker



Harold W. Kuhn y Albert W. Tucker desarrollaron las condiciones de optimalidad para el problema general de programación no lineal tanto con restricciones de igualdad como con restricciones de desigualdad.

Condiciones de Kuhn-Tucker

Consideremos el problema general de optimización no lineal:

$$\text{Min } f(x) \quad (14)$$

sujeta a:

$$g_j(x) \leq 0 \quad \text{para } j = 1, 2, \dots, J \quad (15)$$

$$h_k(x) = 0 \quad \text{para } k = 1, 2, \dots, K \quad (16)$$

$$x = (x_1, x_2, \dots, x_N)$$

Condiciones de Kuhn-Tucker

Definición:

La restricción de desigualdad $g_j(x) \geq 0$ se dice que está **activa** en el punto \bar{x} si $g_j(\bar{x}) = 0$; se dice que está **inactiva** si $g_j(\bar{x}) > 0$.

Condiciones de Kuhn-Tucker

Si podemos identificar las restricciones inactivas en el óptimo antes de resolver el problema, entonces podemos eliminar esas restricciones del modelo y reducir el tamaño del problema. La principal dificultad yace, sin embargo, en identificar las restricciones inactivas antes de resolver el problema.

Condiciones de Kuhn-Tucker

Kuhn y Tucker desarrollaron las condiciones de optimalidad necesarias y suficientes para el problema de programación no lineal, suponiendo que las funciones f , g_j y h_k son diferenciables.