



Unidad n°4-b

Métodos matemáticos de optimización no restringida Búsqueda multivariable

La optimización de funciones objetivo multivariables no lineales necesita el empleo de técnicas robustas y eficientes. La eficiencia es importante debido a que estos problemas requieren procedimientos de solución iterativos, la robustez es una propiedad deseable debido a que predecir el comportamiento de funciones no lineales y de varias variables es imposible.

Estos métodos tienen como objetivo determinar puntos $x^* = [x_1, x_2, x_3 \dots x_n]$ que minimicen a $f(x_1, x_2, x_3 \dots) \equiv f(x)$

Prácticamente para todos los métodos existe una estructura fundamental. El algoritmo comienza desde un punto inicial (x^0), a continuación se determina, mediante una regla, una dirección de movimiento (s^k), y se sigue en esa dirección hasta llegar a un mínimo (relativo) de la función objetivo sobre esa recta. En ese nuevo punto se determina una nueva dirección utilizando la misma regla anterior y se repite el proceso, la búsqueda finaliza cuando se cumple con un criterio de convergencia. 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.

Los métodos pueden ser clasificados en tres categorías, basándonos en la información que debe ser suministrada por el usuario:

- * Métodos de búsqueda directa: los cuales utilizan solo valores de la función objetivo.
- * Métodos de gradiente, aquellos que requieren valores exactos de la primera derivada de la función objetivo.
- * Métodos de segundo orden, utilizan la segunda derivada de la función objetivo.

Los dos últimos métodos son los denominados métodos de búsqueda indirecta.

Métodos directos	1. Búsqueda random	
	2. Búsqueda en una grilla	
	3. Método simplex	
	4. Método de las direcciones conjugadas	
	5. Método de Powell	
Métodos indirectos	Métodos de 1 ^{er} orden	1. Método del gradiente
		2. Método del gradiente conjugado
	Métodos de 2 ^{do} orden	1. Método de Newton
		2. Método de la secante

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 necesitamos ninguna hipótesis adicional acerca de la diferenciabilidad de la función. Podemos emplear estos métodos, bien cuando el gradiente de la función, $\nabla f(x)$, no exista, no sea conocido o simplemente porque su expresión es demasiado compleja para poder manejarlo con eficacia. Para el desarrollo de estos métodos supondremos que $f(x)$ es continua y además unimodal.



1. Búsqueda random

Tal como el nombre lo indica, este método evalúa la función repetidamente en los valores de las variables independientes seleccionados al azar. Si se realizan un número suficiente de evaluaciones de la función objetivo, eventualmente, se localizará el óptimo.

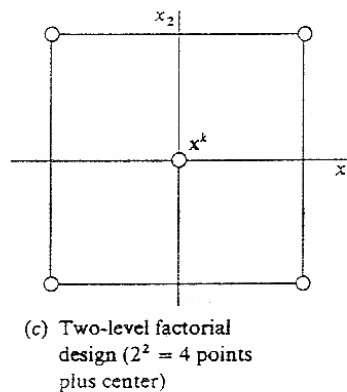
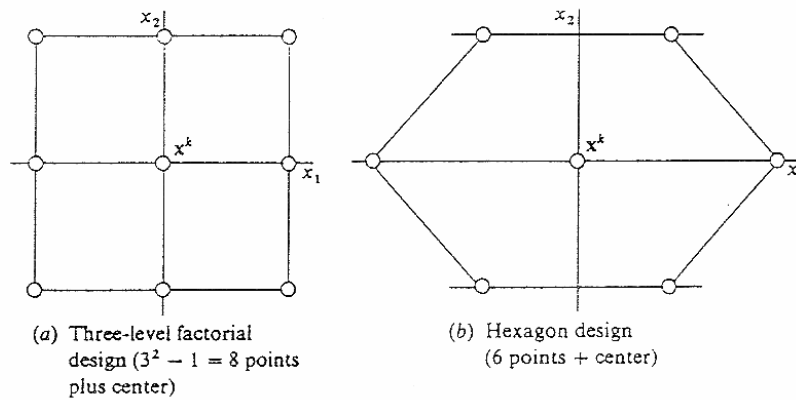
Este método selecciona un vector inicial, se evalúa luego la función en dicho punto, y se determina aleatoriamente un nuevo punto. En efecto en este método se eligen simultáneamente una dirección de búsqueda y un tamaño de paso.

Después de una o más etapas el valor de $f(x^k)$ se compara con el mejor valor de $f(x)$ y se decide continuar o finalizar la búsqueda.

Es un método ineficiente, pero puede ser útil para proveer un punto de partida para otros métodos.

2. Búsqueda en una grilla

En este método se evalúan una serie de puntos alrededor de un punto de referencia, según un diseño determinado. El punto de referencia para la segunda iteración de este método será aquel que da el mejor valor de función objetivo, repitiendo de esta manera el procedimiento de búsqueda.



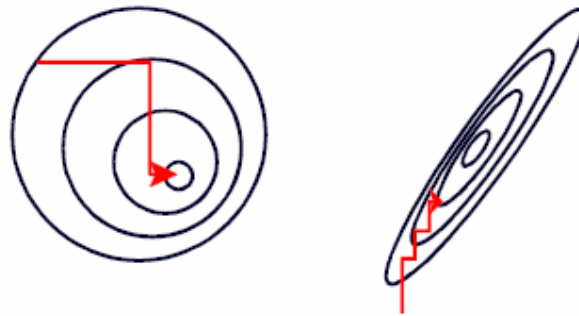
Various grid search designs to select vectors x to evaluate $f(x)$.



3. Búsqueda univariable

Este método consiste en elegir n direcciones fijas de búsqueda, usualmente los ejes coordenados, para una función objetivo de n variables. Luego $f(x)$ es minimizada en cada dirección secuencialmente, utilizando una búsqueda usando unidimensional.

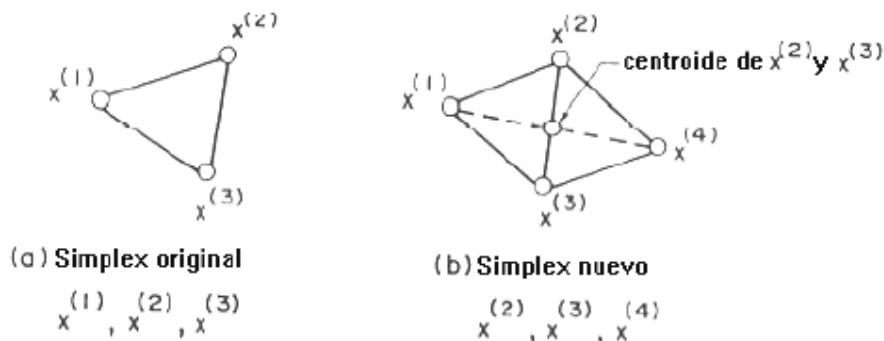
Este método funciona bien cuando las curvas de nivel son circulares, siempre llega al óptimo en dos etapas; pero fracasa cuando se encuentra con curvas de nivel elípticas que originan aristas.



Método de una variable a la vez.

4. Método Simplex o s^2

Es un método de búsqueda heurística desarrollado por Spendley, Hext y Himsworth. 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 empleada por este algoritmo es que se puede generar un nuevo simplex proyectando uno cualquiera de sus vértices a 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.



Reflexión

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

• Regla 2. Ciclado

Si alguno de los vértices permanece invariante después de M iteraciones, reducimos el tamaño del simplex mediante un factor determinado. Construimos el nuevo simplex a partir del vértice en el que la función alcanza el valor más pequeño como punto base. El valor de M puede predecirse a través de la siguiente expresión: $M = 1.65n + 0.05n^2$, 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.

La implementación de este algoritmo requiere sólo dos tipos de cálculos: la generación de un simplex regular a partir de un punto base inicial y un factor de escala apropiado, y 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 x^0 y un factor de escala, α (longitud del lado del simplex), pueden generarse los otros n vértices del simplex regular (en dimensión n) utilizando las expresiones:

$$x^{(i)} = \begin{cases} x_j^{(0)} + \delta_1 & \text{si } j \neq i \\ x_j^{(0)} + \delta_2 & \text{si } j = i \end{cases} \quad \text{para } i \text{ y } j = 1, 2, 3, \dots, N$$

Los valores de δ_1 , y δ_2 dependen solamente de n y α , mediante 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 $x^{(i)}$ es el punto que hay que reflejar. Entonces el centroide de los restantes n puntos es:

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



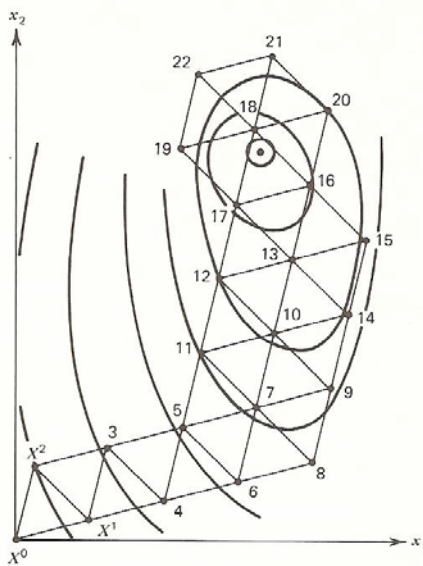
Cualquier punto en la línea que une $x^{(j)}$ y x_c viene dado por:

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

Donde $\lambda \in \mathbb{R}$. Por ejemplo, para $\lambda = 0$ tenemos el punto original $x^{(j)}$, mientras que para $\lambda = 1$ obtenemos el centroide, x_c . Para mantener la regularidad del simplex, la reflexión debe ser simétrica. Por tanto, $\lambda = 2$ conduce al nuevo vértice buscado. Así:

$$x_{new}^{(j)} = 2x_c - x^{(j)}$$

El método prosigue descartando un vértice por vez hasta que el simplex rodee al óptimo.



Sequential simplex search.

Este algoritmo tiene una serie de ventajas:

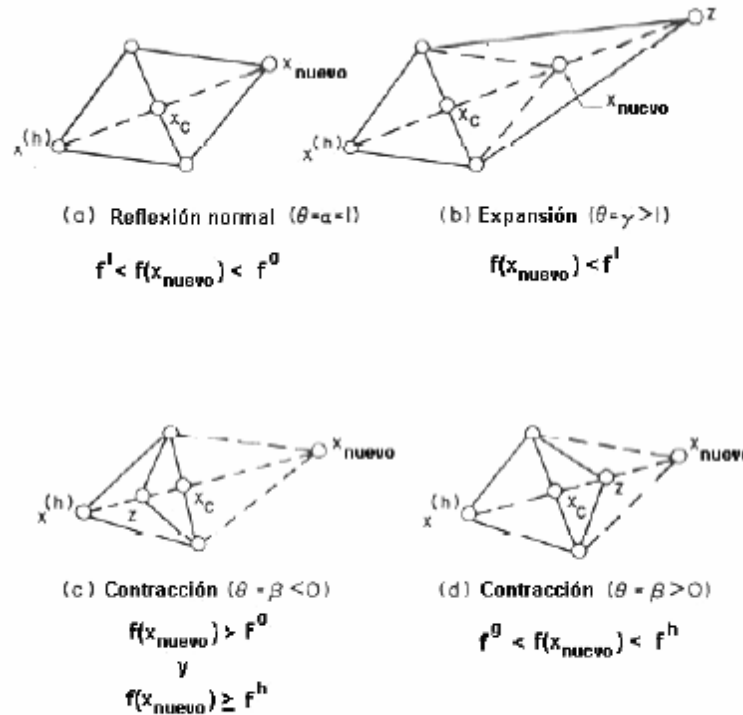
- * Los cálculos son muy simples y la lógica del método no es complicada.
- * El almacenamiento es relativamente pequeño: un vector de dimensiones $(n + 1, n + 2)$
- * Se necesitan pocos parámetros modificables: el factor de escala α y un criterio de parada.
- * El algoritmo es efectivo cuando las evaluaciones 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 desventajas importantes:

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



Para eliminar parcialmente algunas de las desventajas de este método, Nelder y Mead modifican 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.



Método Simplex de Nelder-Mead

5. Direcciones de búsqueda conjugadas

La experiencia ha demostrado que la exploración a través de direcciones conjugadas es mucho más efectiva que la búsqueda con direcciones que se determinan arbitrariamente. Dos direcciones s^i y s^j se dice que son conjugadas si:

$$(s^i)^T Q(s^j) = 0$$

En general, un conjunto de n direcciones linealmente independientes ($s^0, s^1, s^2, \dots, s^{n-1}$), se dice que son conjugadas con respecto a la matriz cuadrada definida positiva Q si:

$$(s^i)^T Q(s^j) = 0 \quad 0 \leq i \neq j \leq n-1$$

En optimización la matriz Q es el Hessiano de la función objetivo. Para una función cuadrática de n variables, para la cual H es una matriz constante, el método garantiza que se puede alcanzar el óptimo en n etapas, si en cada etapa se optimiza en la correspondiente dirección conjugada. Las direcciones conjugadas no son únicas, se debe especificar una dirección inicial para que el resto queden determinadas.



Si la función fuese cuadrática, el desarrollo de Taylor sería:

$$f(x) = f(x^k + \lambda s^k) = f(x^k) + \nabla^T f(x^k) \lambda s^k + \frac{1}{2} (\lambda s^k)^T H(x^k) (\lambda s^k)$$

Para obtener el mínimo de $f(x^k + \lambda s^k)$, derivamos con respecto a λ , y luego igualamos la derivada a cero:

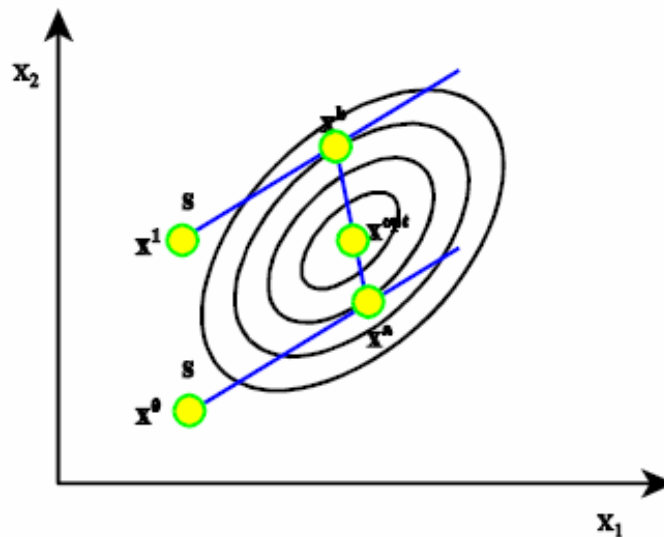
$$\frac{df(x^k + \lambda s^k)}{d\lambda} = 0 = \nabla^T f(x^k) s^k + (s^k)^T H(x^k) (\lambda s^k)$$

Despejando se obtiene:

$$\lambda^{opt} = -\frac{\nabla^T f(x^k) s^k}{(s^k)^T H(x^k) (s^k)}$$

Cálculo de la dirección conjugada sin usar derivadas

Partiendo del punto x^0 , se localiza el punto x^a que minimiza la función en la dirección s . Luego, a partir de otro punto x^1 distinto del primero se localiza el punto x^b que minimiza la función en la misma dirección s . La dirección $(x^b - x^a)$ es conjugada a s . La figura siguiente muestra esta situación.

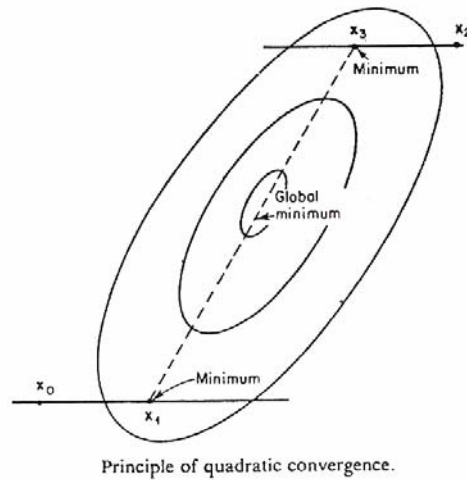


Determinación de direcciones conjugadas.

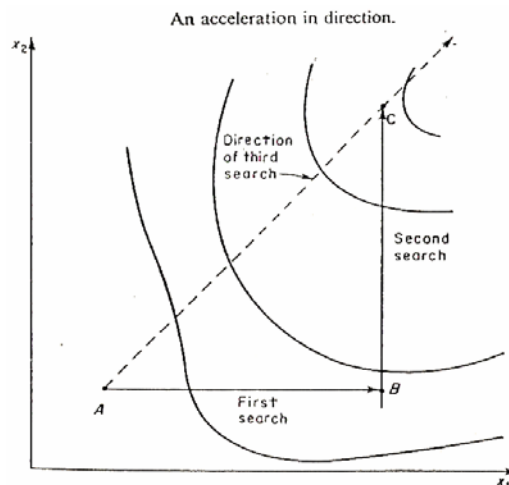


6. Método de Powell

El método de Powell localiza el mínimo de una función f mediante búsquedas secuenciales unidimensionales. Las bases de este método pueden verse en la construcción geométrica de la siguiente figura.



Un posible método de búsqueda podría comenzar con dos direcciones no necesariamente conjugadas $\xi_1^{(0)}$ y $\xi_2^{(0)}$. En la primer etapa estas podrían ser las direcciones de los ejes coordenados. Llevamos a cabo nuestra búsqueda unidimensional partiendo desde un punto base elegido arbitrariamente $x_0^{(0)}$ y encontrando un punto mejorado $x_2^{(0)}$. Luego definimos una dirección de búsqueda favorable $\mu^{(0)} = x_2^{(0)} - x_0^{(0)}$. Donde $\mu^{(0)}$ es la dirección desde el punto base al punto final de la búsqueda en una dimensión. Esta búsqueda es similar a la observada en la siguiente figura, donde $x_0^{(0)}$ sería el punto A. En este caso A es el punto óptimo obtenido en un paso anterior de la búsqueda univariable, y los puntos B y C son los respectivos óptimos de búsquedas siguientes en las direcciones de los ejes coordenados, la aceleración del método se obtiene realizando la próxima búsqueda en la dirección de la línea que une A y C. En el método de Powell $x_0^{(0)}$ no es un punto óptimo obtenido de una búsqueda anterior y tampoco tenemos la seguridad de que de que μ sea una dirección que pase a través del óptimo incluso si la función fuese cuadrática.





La segunda etapa de búsqueda se realiza a lo largo de $\mu^{(0)}$ alcanzando un punto mejorado $x_0^{(1)}$, el punto base para esta etapa. Usamos ahora las direcciones $\xi_1^{(1)} = \xi_2^{(0)}$ y $\xi_2^{(1)} = \mu^{(0)}$. Si $x_2^{(1)}$ es el óptimo encontrado luego de pasar por estas dos nuevas direcciones, entonces la dirección $\mu^{(1)} = x_2^{(1)} - x_0^{(1)}$ pasará a través del óptimo de una función cuadrática. Por supuesto, si la función no es cuadrática, $\mu^{(1)}$ no necesariamente pasará por el óptimo, pero probablemente sea una dirección de búsqueda favorable.

Para una búsqueda unidimensional el método puede resumirse de la siguiente manera:

• Etapa 1:

Comenzando con el mejor valor previo $x_0^{(k)}$ y una serie de direcciones de búsqueda linealmente independientes, $\xi_1^{(k)}, \xi_2^{(k)}, \dots, \xi_n^{(k)}$, iniciar buscando la posición del óptimo a lo largo de la línea que pasa a través de $x_0^{(k)}$, que es paralela a $\xi_1^{(k)}$. Sea este punto óptimo $x_1^{(k)}$, comenzamos una segunda búsqueda desde este nuevo punto en la dirección $\xi_2^{(k)}$. El procedimiento continúa hasta que las n direcciones de búsqueda hayan sido exploradas

• Etapa 2:

Ahora buscamos el punto particular $x_m^{(k)}$ correspondiente al mejor valor de la función objetivo respecto de sus valores previos. Entonces este punto produce el mayor cambio Δ en cualquiera de los n movimientos, donde $\Delta = |y(x_m^{(k)}) - y(x_{m-1}^{(k)})|$, también determinamos el vector $\mu = x_n^{(k)} - x_0^{(k)}$.

• Etapa 3:

Establecer $y(2x_n^{(k)} - x_0^{(k)}) = y_t^{(k)}$

• Etapa 4:

Si en la búsqueda de un mínimo $y_t^{(k)} \geq y_0^{(k)}$ y/ó

$(y_0^{(k)} - 2y_n^{(k)} + y_t^{(k)})(y_0^{(k)} - y_n^{(k)} - \Delta)^2 \geq \frac{\Delta(y_0^{(k)} - y_t^{(k)})^2}{2}$, entonces μ no es una buena dirección

para introducir en nuestra búsqueda y continuamos con la exploración comenzando desde el último punto y usando las mismas direcciones, esto es $x_0^{(k+1)} = x_n^{(k)}$ y $\xi_i^{(k+1)} = \xi_i^{(k)}$ para $i = 1, \dots, n$. Se repite entonces la etapa 1 hasta que se encuentre el mínimo.

Si ninguna de estas inecuaciones se verifica, buscamos a lo largo de la dirección μ hasta que se encuentre un mínimo. Este punto se define como $x_0^{(k+1)}$ y las nuevas direcciones de búsqueda para la etapa $(k+1)$ son $\xi_i^{(k+1)} = \xi_i^{(k)}$, $i=1, 2, \dots, m-1$; $\xi_i^{(k+1)} = \xi_{i+1}^{(k)}$, $i = m, \dots, n-1$; y $\xi_n^{(k+1)} = \mu$

Hay dos cuestiones que indaga el algoritmo antes de determinar una nueva dirección de búsqueda. La primera considera el valor de una función objetivo tentativa y_t , encontrada explorando a lo largo de la nueva dirección de búsqueda. Powell estudia la función objetivo en el punto $2x_n - x_0$ el cual es un punto a lo largo de la dirección μ ubicado a la misma distancia de x_n como el punto original x_0 . Si este sondeo tentativo no produce una mejoría en la función objetivo respecto del punto base original, la nueva dirección se rechaza.

La segunda desigualdad se usa para determinar si la función puede tener picos (en la búsqueda de un mínimo) en el movimiento desde el punto x_n al punto tentativo $2x_n - x_0$, en este caso se rechaza como dirección favorable.

Métodos indirectos de primer orden

Los métodos indirectos hacen uso de derivadas en la determinación de las direcciones de búsqueda. Una buena dirección de búsqueda debería reducir la función objetivo, entonces si x_0 es el punto inicial y x_1 es el nuevo punto: $f(x_1) < f(x_0)$. Entonces la dirección de búsqueda s es llamada dirección descendiente y satisface que $\nabla^T f(x) s < 0$.



1. Método del Gradiente

Este método utiliza sólo las derivadas primeras de la función objetivo. El gradiente es el vector en el punto x que da la dirección del mayor aumento en $f(x)$ y es el ortogonal al contorno de la función en el punto x . Para *maximización* la dirección de búsqueda es simplemente el gradiente, para *minimización*, la dirección de búsqueda es el valor negativo del gradiente. $s^k = -\nabla f(x^k)$.

En minimización, en la iteración k^{th} , la transición desde el punto x^k a otro punto x^{k+1} se obtiene con la siguiente expresión:

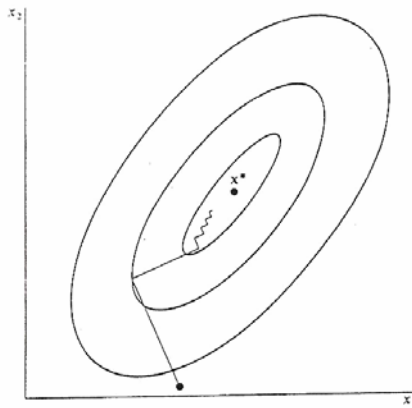
$$x^{k+1} = x^k + \Delta x^k = x^k + \lambda^k s^k = x^k - \lambda^k \nabla f(x^k)$$

donde Δx^k es el vector desde x^k hasta x^{k+1} , s^k es la dirección de búsqueda y λ^k es un escalar que determina la longitud del paso en la dirección s^k .

El método puede resumirse en las siguientes etapas:

- Elección de un punto inicial x_0 .
- Cálculo de las derivadas parciales, analítica o numéricamente.
- Cálculo del vector de búsqueda, $s^k = -\nabla f(x^k)$
- Cálculo del nuevo punto x^{k+1} , $x^{k+1} = x^k + \lambda^k s^k$
- Comparar $f(x^{k+1})$ con $f(x^k)$, si el cambio en $f(x)$ es menor que la tolerancia especificada se termina la búsqueda, sino se vuelve a la segunda etapa y se continúa con la búsqueda.

La siguiente figura muestra el avance del método del gradiente hacia el óptimo para una función cuadrática.



Este método tiene el inconveniente que a medida que nos acercamos al óptimo (o a un punto crítico) su velocidad de convergencia disminuye notoriamente (el gradiente tiende a anularse). Sin embargo, tiene una buena velocidad en las primeras etapas, cuando estamos lejos del óptimo. Siempre se debe verificar que el punto encontrado no sea un punto de ensilladura.

Otra desventaja del método es su elevada sensibilidad al escalado de la función objetivo provocando una baja velocidad de convergencia y fuertes oscilaciones.

2. Método del Gradiente Conjugado

Este método utiliza como dirección de búsqueda una combinación lineal del gradiente actual con el gradiente de la iteración anterior. Es una notable mejora con respecto al método original, porque para funciones cuadráticas se puede demostrar que las sucesivas direcciones de búsqueda



son conjugadas, por lo tanto el óptimo se encuentra en n pasos, donde n es el número de variables. Las direcciones no se especifican a priori, sino que se determinan en cada paso de la iteración. Las direcciones Q-ortogonales en cada punto se eligen de manera sencilla en función de ∇f . Esta dependencia hace que el método converja más uniformemente. Tiene tres ventajas

- * A no ser que se alcance la solución en n pasos, el gradiente siempre es $\neq 0$ y linealmente independiente de todos los vectores dirección anteriores.
- * Utiliza una fórmula sencilla para determinar el nuevo vector de dirección.
- * Como las direcciones se basan en el gradiente, el proceso, tiene en cada paso un buen avance uniforme hacia la solución.

El método general del gradiente conjugado tiene la siguiente estructura:

- * En el punto inicial x^0 se calcula el valor de la función, $f(x^0)$, luego $s^0 = -\nabla f(x^0)$
- * Luego se calcula un nuevo punto $x^1 = x^0 + \lambda^0 s^0$, mediante una minimización de $f(x)$ con respecto a λ , en la dirección de s^0 .
- * El próximo paso consiste en calcular $f(x^1)$ y $\nabla f(x^1)$. La nueva dirección de búsqueda es una combinación lineal de s^0 y $\nabla f(x^1)$:

$$s^1 = -\nabla f(x^1) + s^0 \frac{\nabla^T f(x^1) \nabla f(x^1)}{\nabla^T f(x^0) \nabla f(x^0)}$$

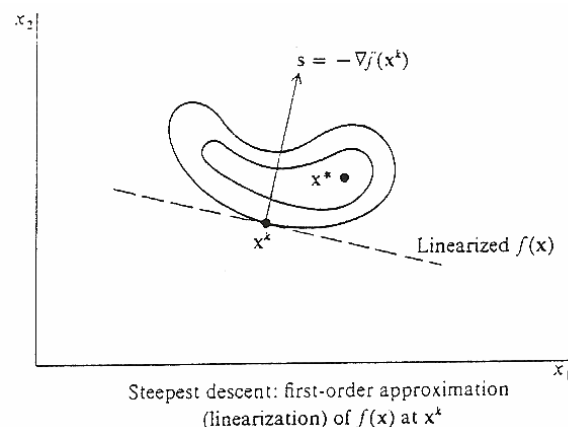
para la iteración k la relación es:

$$s^{k+1} = -\nabla f(x^{k+1}) + s^k \frac{\nabla^T f(x^{k+1}) \nabla f(x^{k+1})}{\nabla^T f(x^k) \nabla f(x^k)}$$

- * El próximo paso es verificar la convergencia del método, si esto no sucede debe regresarse a la tercer etapa y continuar con la búsqueda
- * El algoritmo finaliza cuando $\|s^k\|$ es tan pequeña como permita la tolerancia especificada.

Métodos indirectos de segundo orden

En el método del gradiente puede interpretarse que la dirección de búsqueda es ortogonal a una aproximación lineal de la función objetivo en el punto x^k , como se muestra en la siguiente figura.

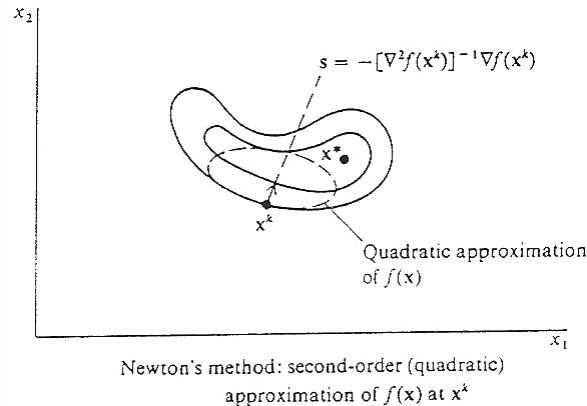




Si hacemos una aproximación cuadrática de la función en x^k :

$$f(x) = f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

donde $H(x)$ es la matriz Hessiana de $f(x)$, de esta forma es posible tener en cuenta la curvatura de $f(x)$ en x^k en la determinación de la dirección de búsqueda, como se describe para el método de Newton.



Método de Newton

En este caso, la dirección de búsqueda se determina utilizando la segunda derivada de la función objetivo. El método aproxima la función objetivo f en la vecindad de un mínimo con una serie de Taylor truncada hasta el término de segundo orden, es decir:

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

El mínimo de $f(x)$ en x^k es obtenido diferenciando la aproximación cuadrática de $f(x)$ con respecto de cada uno de los componentes de x e igualando la expresión resultante a cero.

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0 \quad \text{ó} \quad x^{k+1} - x^k = \Delta x^k = -H^{-1}(x^k) \nabla f(x^k)$$

Si la función es cuadrática, se requiere solo un paso para alcanzar el óptimo.

La desventaja del método de Newton es que si la aproximación inicial del método, $x^{(0)}$, está situada lejos del punto solución x^* , puede ocurrir que la sucesión construida 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:

$$x^{k+1} - x^k = -\lambda^k H^{-1}(x^k) \nabla f(x^k), \quad \text{con} \quad s^k = -H^{-1}(x^k) \nabla f(x^k)$$

El método de Newton considera $\lambda=1$ en cada etapa.

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 $f(x)$. La mayor dificultad consiste en calcular la inversa de la matriz Hessiana, $Hf(x)$, en cada iteración.



Método de la Secante

El cálculo de la matriz Hessiana no es práctico de realizar, la idea básica del método de la Secante es utilizar una aproximación a la inversa del Hessiano, en lugar de dicha inversa, como requiere el método de Newton.

De igual manera que el método de la Secante para una variable, el óptimo se obtiene utilizando solo valores de la función objetivo y valores de la primera derivada, la matriz Hessiana de la función se obtiene por combinación de ambos valores. El valor de la primera derivada puede ser obtenida por diferencias finitas, de esta manera el método solo utiliza valores de la función objetivo.

El método comienza planteando:

$$\nabla f(x^k) + H(x^k)\Delta x^k = 0$$

La matriz Hessiana de la función objetivo, $H(x)$, se obtiene mediante a aproximación cuadrática de la función, $\hat{H}(x)$, pero utilizando solo las primeras derivadas parciales de la función en la evaluación de $\hat{H}(x)$.

Si utilizamos una aproximación de la matriz inversa del Hessiano, $\hat{H}^{-1}(x)$, el nuevo punto se obtiene mediante la siguiente expresión:

$$x^{k+1} - x^k = \Delta x^k = -\lambda^k \hat{H}^{-1}(x^k) \nabla f(x^k)$$

donde $\hat{H}^{-1}(x^k)$ es llamada matriz de dirección.

Suponiendo que $f(x)$ es aproximada por una función cuadrática, el desarrollo del método tomando dos puntos de referencia resulta:

$$\left. \begin{aligned} \nabla f(x^{k+1}) &= \nabla f(x^p) + \hat{H}(x^{k+1} - x^p) \\ \nabla f(x^k) &= \nabla f(x^p) + \hat{H}(x^k - x^p) \end{aligned} \right\} \nabla f(x^{k+1}) - \nabla f(x^k) = \hat{H}(x^{k+1} - x^k)$$

Para una función no cuadrática, la matriz \hat{H} deberá ser obtenida utilizando esta última ecuación, utilizando los puntos x^k y x^{k+1} .

$$\hat{H}(x^k)\Delta x^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

Una relación equivalente sería:

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



Ejemplos

Métodos directos

1. Método Simplex

Minimizar la siguiente función:

$$f(x) = 10 + (5 - x_1)^2 + (1 - x_2)^2$$

La construcción inicial del simplex requiere la especificación de un punto inicial y un factor de escala. Suponiendo $x^0 = [0, 0]^T$ y $\alpha = 2$, entonces:

$$\delta_1 = \left[\frac{(2+1)^{1/2} + 2 - 1}{2\sqrt{2}} \right] 2 = 1.9318$$

$$\delta_2 = \left[\frac{(2+1)^{1/2} - 1}{2\sqrt{2}} \right] 2 = 0.5176$$

Con esos valores, los dos vértices restantes serán

$$x^1 = \begin{bmatrix} 0 + 0.5176 \\ 0 + 1.9318 \end{bmatrix} = \begin{bmatrix} 0.5176 \\ 1.9318 \end{bmatrix}$$

$$x^2 = \begin{bmatrix} 0 + 1.9318 \\ 0 + 0.5176 \end{bmatrix} = \begin{bmatrix} 1.9318 \\ 0.5176 \end{bmatrix}$$

Con valores:

$$f(x^0) = 36$$

$f(x^1) = 30.960$, por lo tanto x^0 es el punto reflejado para formar el nuevo simplex.

$$f(x^2) = 19.647$$

El punto x^3 , se calcula de la siguiente forma: $x_c = \frac{1}{2} \sum_{i=1}^2 x^i = \frac{1}{2}(x^1 + x^2)$,

$$x^3 = 2x_c - x^0 = x^1 + x^2 - x^0, \quad x^3 = \begin{bmatrix} 0.5176 \\ 1.9318 \end{bmatrix} + \begin{bmatrix} 1.9318 \\ 0.5176 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.4494 \\ 2.4494 \end{bmatrix}$$

El nuevo simplex, formado por los puntos x^1 , x^2 y x^3 , presenta los siguientes valores de función:

$$f(x^1) = 30.960$$

$f(x^2) = 19.647$ por lo tanto x^1 es el punto reflejado para formar el nuevo simplex.

$$f(x^3) = 18.609$$

$x^4 = x^2 + x^3 - x^1$, $x^4 = \begin{bmatrix} 3.8636 \\ 1.0352 \end{bmatrix}$, $f(x^4) = 11.292$, en el simplex formado por los puntos x^2 , x^3 y

x^4 , el punto que debe reflejarse es x^2 .

$x^5 = x^3 + x^4 - x^2$, $x^5 = \begin{bmatrix} 4.3812 \\ 2.967 \end{bmatrix}$, $f(x^5) = 14.252$ en el simplex formado por los puntos x^3 , x^4 y

x^5 , el punto que debe reflejarse es x^3 .

$x^6 = x^4 + x^5 - x^3$, $x^6 = \begin{bmatrix} 5.7954 \\ 1.5528 \end{bmatrix}$, $f(x^6) = 10.938$ en el simplex formado por los puntos x^4 , x^5 y

x^6 , el punto que debe reflejarse es x^5 .



$x^7 = x^4 + x^6 - x^5$, $x^7 = \begin{bmatrix} 5.2778 \\ -0.379 \end{bmatrix}$, $f(x^7) = 11.978$ en el simplex formado por los puntos x^4 , x^6 y x^7 , el punto que deber reflejarse es x^7 , de esta manera se vuelve a obtener el simplex formado por los puntos x^4 , x^5 y x^6 , para evitar esto reflejamos el segundo “peor” valor de función, es decir x^4 .

$x^8 = x^7 + x^6 - x^4$, $x^8 = \begin{bmatrix} 7.2096 \\ 0.1386 \end{bmatrix}$, $f(x^8) = 15.6243$ en esta etapa el punto que debe reflejarse es x^8 , pero se presenta el mismo inconveniente que en la iteración anterior. El método se continúa reflejando el segundo peor valor de la función. Los resultados de las próximas iteraciones son:

$$x^9 = x^6 + x^8 - x^7, x^9 = \begin{bmatrix} 7.7272 \\ 2.0704 \end{bmatrix}, f(x^9) = 18.5834 \text{ reflejo } x^8$$

$$x^{10} = x^9 + x^6 - x^8, x^{10} = \begin{bmatrix} 6.3130 \\ 6.4846 \end{bmatrix}, f(x^{10}) = 17.8972$$

Si observamos detenidamente las iteraciones realizadas podemos percibir que en las últimas cuatro ($M = 4$) permanece invariante uno de los vértices del simplex, punto x^6 , en este momento se debe reducir el tamaño del simplex, de acuerdo a un factor de reducción, es la etapa de “ciclado”. Se comienza nuevamente con las iteraciones tomando como punto base aquel en el cual se obtuvo el menor valor de función, en nuestro caso el punto de partida será x^6 . Reducimos ahora el tamaño de nuestro simplex original, $x^0 = [5.7954, 1.5528]^T$ y $\alpha = 0.5$, entonces:

$$\delta_1 = \left[\frac{(2+1)^{1/2} + 2 - 1}{2\sqrt{2}} \right] 0.5 = 0.4829, \quad \delta_2 = \left[\frac{(2+1)^{1/2} - 1}{2\sqrt{2}} \right] 0.5 = 0.1294$$

Con esos valores, los dos vértices restantes serán

$$x^1 = \begin{bmatrix} 5.9248 \\ 2.0357 \end{bmatrix}, f(x^1) = 11.9280$$

x^2 es el punto reflejado para formar el nuevo simplex.

$$x^2 = \begin{bmatrix} 6.2783 \\ 1.6822 \end{bmatrix}, f(x^2) = 12.0996$$

$$x^3 = x^1 + x^0 - x^2 \quad x^3 = \begin{bmatrix} 5.4418 \\ 1.9063 \end{bmatrix} \quad f(x^3) = 11.016 \quad \text{Reflejo } x^1$$

$$x^4 = x^3 + x^0 - x^1 \quad x^4 = \begin{bmatrix} 5.3124 \\ 1.4233 \end{bmatrix} \quad f(x^4) = 10.277 \quad \text{Reflejo } x^3$$

$$x^5 = x^4 + x^0 - x^3 \quad x^5 = \begin{bmatrix} 5.6659 \\ 1.0698 \end{bmatrix} \quad f(x^5) = 10.4484 \quad \text{Reflejo } x^0$$

El método se continua hasta que el tamaño del simplex o la desviación estándar entre los valores de función en los vértices sea lo suficientemente pequeña.

2. Direcciones de búsquedas conjugadas

Minimizar la siguiente función:

$$f(x) = 10 + (5 - x_1)^2 + (1 - x_2)^2$$



El punto y la dirección de búsqueda inicial son: $x^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $s^0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

El primer paso es obtener la dirección s^1 conjugada a s^0 :

$$(s^0)^T H(x)(s^1) = 0 \Rightarrow [2 \ 2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} s_1^1 \\ s_2^1 \end{bmatrix} = 0 \Rightarrow s^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Ahora debemos calcular el valor de λ^0 óptimo:

$$\lambda_0^{opt} = -\frac{\nabla^T f(x^0) s^0}{(s^0)^T H(x^0) s^0} = -\frac{[-10 \ -2] \begin{bmatrix} 2 \\ 2 \end{bmatrix}}{[2 \ 2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}} = \frac{3}{2}$$

Calculamos x^1

$$x^1 = x^0 + \lambda_0^{opt} s^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Con la dirección de búsqueda s^1 calculamos el próximo punto.

$$\lambda_1^{opt} = -\frac{\nabla^T f(x^1) s^1}{(s^1)^T H(x^1) s^1} = -\frac{[-4 \ 4] \begin{bmatrix} 1 \\ -1 \end{bmatrix}}{[1 \ -1] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}} = 2$$

El punto x^2 se calcula como:

$$x^2 = x^1 + \lambda_1^{opt} s^1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

Al ser la función objetivo cuadrática el óptimo se obtiene utilizando primero la dirección s^0 y luego s^1 :

3. Método de Powell

Minimizar la siguiente función:

$$f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

El método necesita un punto de partida y direcciones iniciales de búsqueda:

$$x_0^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad s_1^0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad s_2^0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

En la siguiente etapa realizamos el cálculo de λ óptimo:

$$x_1^0 = x_0^0 + \lambda_1 s_1^0 \Rightarrow x_1^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix}$$

$$f(x_1^0) = (\lambda_1 - 2)^4 + (\lambda_1 - 2 * 0)^2, \quad \frac{df(x_1^0)}{d\lambda_1} = 4(\lambda_1 - 2)^3 + 2\lambda_1 = 0 \Rightarrow \lambda_1 = 1.1648$$



$$x_1^0 = \begin{bmatrix} 1.1648 \\ 0 \end{bmatrix}, \quad f(x_1^0) = 1.84334$$

El próximo punto se calcula utilizando la segunda dirección de búsqueda.

$$x_2^0 = x_1^0 + \lambda_2 s_2^0 \Rightarrow x_2^0 = \begin{bmatrix} 1.1648 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1648 \\ \lambda_2 \end{bmatrix}$$

$$f(x_2^0) = (1.1648 - 2)^4 + (1.1648 - 2\lambda_2)^2, \quad \frac{df(x_2^0)}{d\lambda_2} = 2(1.1648 - 2\lambda_2)^2(-2) = 0 \Rightarrow \lambda_2 = 0.5824$$

$$x_2^0 = \begin{bmatrix} 1.1648 \\ 0.5824 \end{bmatrix}, \quad f(x_2^0) = 0.4865$$

En el siguiente paso se calcula x^3 como:

$$x_3^0 = 2 \begin{bmatrix} 1.1648 \\ 0.5824 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.3293 \\ 1.1648 \end{bmatrix}, \quad f(x_3^0) = 0.01180$$

Ahora debemos determinar la nueva dirección de búsqueda:

$$s_1^1 = s_2^0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad s_2^1 = x_2^0 - x_0^0 = \begin{bmatrix} 1.1648 \\ 0.5824 \end{bmatrix}, \quad \text{la nueva dirección de búsqueda debe normalizarse:}$$

$$s_2^1 = \frac{\begin{bmatrix} 1.1648 \\ 0.5824 \end{bmatrix}}{1.30} = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix}$$

$f(x_3^0) \leq f(x_0^0)$ se verifica, entonces la dirección de búsqueda es buena.

El próximo paso comienza a partir del último punto con las siguientes direcciones de búsqueda:

$$x_0^1 = \begin{bmatrix} 2.3293 \\ 1.1648 \end{bmatrix}, \quad s_1^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad s_2^1 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix}$$

s_1^1	s_2^1	x^1	$f(x^1)$
		$x_0^1 = [2.3293 \quad 1.1648]^T$	$f(x_0^1) = 0.01180$
$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix}$	$x_1^1 = [2.3293 \quad 1.16465]^T$	$f(x_1^1) = 0.01180191$
		$x_2^1 = [2.0003 \quad 0.9999]^T$	$f(x_2^1) = 0.9 \cdot 10^{-7}$
		$x_3^1 = [1.6796 \quad 0.83518]^T$	$f(x_3^1) = 0.1172$

$f(x_3^1) \leq f(x_0^1)$ se verifica, entonces la dirección de búsqueda es buena. La próxima iteración se realiza de igual manera, volviendo a calcular la nueva dirección de búsqueda como $\mu = x_n^{(k)} - x_0^{(k)}$.

**Métodos indirectos**

Minimizar la siguiente función:

$$f(x) = 10 + (5 - x_1)^2 + (1 - x_2)^2$$

1. Método del gradiente

La primer etapa de éste método consiste en la elección de un punto inicial, en este caso $x^0 = [0, 0]^T$.

Cálculo de las derivadas de la función:

$$\frac{\partial f(x)}{\partial x_1} = -2(5 - x_1)$$

$$\frac{\partial f(x)}{\partial x_2} = -2(1 - x_2)$$

El vector de búsqueda resulta:

$$s^0 = -\nabla f(x^0) = \begin{bmatrix} 10 \\ 2 \end{bmatrix}$$

Para calcular el nuevo punto es necesario conocer el valor de λ^0 , el cálculo de este parámetro se realiza de igual manera que en el método de las direcciones conjugadas.

$$\lambda_0 = -\frac{\nabla^T f(x^0) s^0}{(s^0)^T H(x^0) s^0} = -\frac{[-10 \quad -2] \begin{bmatrix} 10 \\ 2 \end{bmatrix}}{[10 \quad 2] \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 10 \\ 2 \end{bmatrix}} = \frac{1}{2}$$

$$x^1 = x^0 + \lambda_0 s^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 10 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

Al ser la función objetivo cuadrática el valor del óptimo se obtiene con una sola iteración.

2. Método de Newton

Para implementar este método es necesario conocer las primeras y segundas derivadas de la función.

$$\nabla f(x) = \begin{bmatrix} -2(5 - x_1) \\ -2(1 - x_2) \end{bmatrix} \quad Hf(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\Delta x^0 = -[Hf(x^0)]^{-1} \nabla f(x^0) = -\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} -10 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

El punto inicial, de igual manera que en los ejemplos anteriores es $x^0 = [0, 0]^T$, entonces el nuevo punto es: $x^1 = [5, 1]^T$.

El método de Newton trabaja con una aproximación de segundo orden de la función, al ser la función objetivo cuadrática el valor óptimo se obtiene en una sola iteración, y este valor es el único punto crítico de la función.



Bibliografía

- * Beveridge G., Schechter (1970) “Optimization: Theory and Practice” Ed. McGraw-Hill.
- * Biles William, Swain James (1980). “Optimization and industrial experimentation”, ed. John Wiley and Sons Inc.
- * Chapra Steven, Canale Raymond (2006). “Numerical methods for engineers”, 5th edition. Ed. McGraw-Hill.
- * Edgar T., Himmelblau. (1988) “Optimization of chemical processes” ed. McGraw-Hill.
- * Reklaitis G., Ravindran A., Ragsdell K. (1983), “Engineering optimization. Methods and applications”, ed. John Wiley and Sons Inc.
- * Tarifa E. (2006) “Optimización y simulación de procesos. Métodos numéricos” Facultad de Ingeniería - Universidad Nacional de Jujuy
- * Zerpa L., Colmenares J. (2004), “Optimización para ingenieros, optimización sin restricciones”. Notas de clase, Universidad del Zulia, Facultad de Ingeniería, División de Estudios para Graduados, Instituto de Cálculo Aplicado. República Bolivariana de Venezuela.