Optimización no restringida

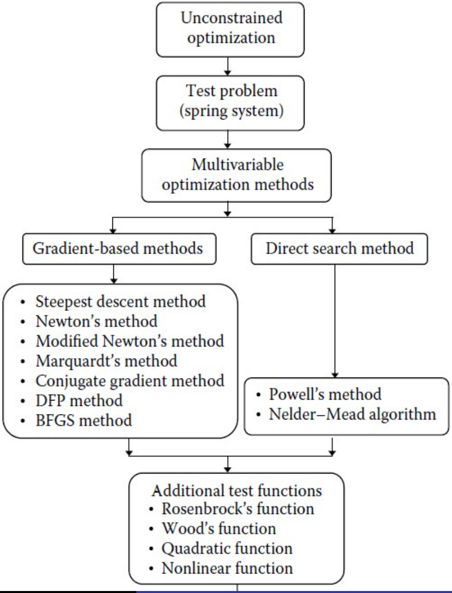
Gamaliel Moreno

Maestría en Ciencias del Procesamiento de la Información

Ene-Julio, 2021

Los métodos para los problemas de optimización no restringida pueden ser clasificados en basados en gradiente y basados en no-gradiente. Como su nombre lo indica, los métodos basados en gradiente requieren la información del gradiente para determinar la dirección de búsqueda. Los métodos basados en gradiente son: El paso más descendente, Davidon-Fletcher-Powell (DFP), Broyden-Fletcher-Goldfarb-Shanno (BFGS), Newton, y Levenberg-Marquardt.

La dirección de búsqueda por estos métodos usan la información, la Hessian o combinación de estas. Algunos métodos usan una aproximación de la matrix Hessiana. Una vez identificada la dirección, se necesita evaluar cuanto se mueve en esa dirección para minimizar la función. Este es un problema unidimensional.



Unconstrained optimization
Test problem (spring system)
Multivariable optimization methods
Gradient-based methods
Direct search method
• Steepest descent method
• Newton's method
• Modified Newton's method
• Marquardt's method
• Conjugate gradient method
• DFP method
• BFGS method
• Powell's method
• Nelder–Mead algorithm
Additional test functions
• Rosenbrock's function
• Wood's function
• Quadratic function
• Nonlinear function

Para una función de una sola variable, se sabe que la derivada de la función se desvanece en el óptimo y la segunda derivada de la función es mayor que cero en el mínimo de la función. Lo mismo se puede extender a una función multivariable. Las condiciones necesarias para que x^* sea un mínimo son que

$$\nabla f(x^*) = 0$$

y $x^T Hx$ sea positiva definida ($x^T Hx > 0$). Para asegurar esto, los eigenvalores de H debe ser positivos.

Encuentre el valor óptimo de la función, determine si la matriz hessiana es definida positiva, grafique en matlab las funciones.

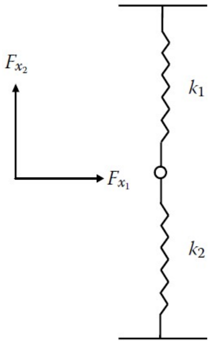
1. $f(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1$ grafiquela con $5- < x_1 < 5$ y $5- < x_2 < 5$
2. $f(\mathbf{x}) = x_1^2 - x_2^2$ grafiquela con $5- < x_1 < 5$ y $5- < x_2 < 5$

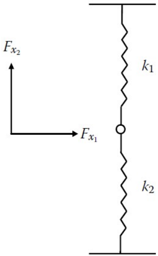
La búsqueda unidireccional se refiere a minimizar el valor de una función multivariable en una dirección específica. Por ejemplo, si x_i es el punto de inicio inicial de las variables de diseño para minimizar una función multivariable y S_i es la dirección de búsqueda, entonces necesitamos determinar una cantidad escalar α tal que la función

$$f(\alpha) = x_i + \alpha S_i$$

sea minimizada. El valor de α en el que esta función alcanza un mínimo es dado por α^*. Este es un problema de optimización unidimensional y podemos usar la técnica de la sección dorada para minimizar esta función.

Dos resortes de longitud unitaria, con rigidez k_1 y k_1, están unidos en el origen. Los otros extremos están fijados en una pared



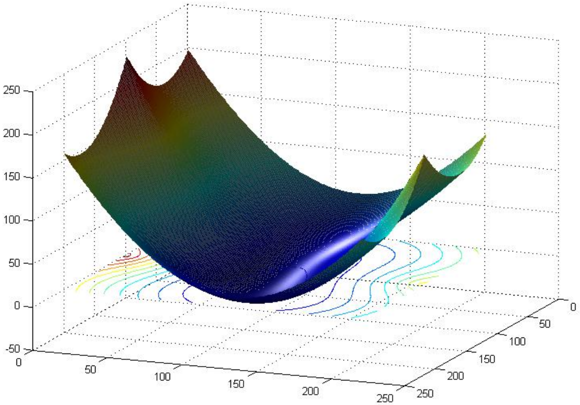


Al aplicar una fuerza, el sistema de resorte se desviará a una posición de equilibrio, que estamos interesados en determinar. El potencial del sistema de resorte está dado por

$$U = k_1\left(\sqrt{x_1^2 + (x_2+1)^2} - 1\right)^2 + k_2\left(\sqrt{x_1^2 + (x_2+1)^2} - 1\right)^2 - \left(F_{x_1}x_1 + F_{x_2}x_2\right)$$

$$U = k_1\left(\sqrt{x_1^2+(x_2+1)^2}-1\right)^2 + k_2\left(\sqrt{x_1^2+(x_2+1)^2}-1\right)^2 - (F_{x_1}x_1 + F_{x_2}x_2)$$

donde (F_{x_1}, F_{x_2}) es la fuerza aplicada en el origen debido a la cual mueve su position a (x_1, x_2). Asumiendo que $k_1 = 100N/m$, $k_2 = 90N/m$, y $(F_{x_1}, F_{x_2}) = (20, 40)$, la meta es evaluar (x_1, x_2) de tal manera que U sea minimizada.



Minimum Potential $= -7.3529$ occurs at $x1 = 0.5, x2 = 0$

La primera y segunda derivadas se pueden calcular utilizando la fórmula de diferencia central como dada a continuación.

$$\frac{\partial f}{\partial x_i} = \frac{f(x_i + \Delta x_i) - f(x_i - \Delta x_i)}{2\Delta x_i}$$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{f(x_i + \Delta x_i) - 2f(x_i) + f(x_i - \Delta x_i)}{\Delta x_i^2}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \Big[f(x_i + \Delta x_i, x_j + \Delta x_j) - f(x_i + \Delta x_i, x_j - \Delta x_j) \\ - f(x_i - \Delta x_i, x_j + \Delta x_j) + f(x_i - \Delta x_i, x_j - \Delta x_j) \Big] \Big/ (4\Delta x_i \Delta x_j)$$

Método del descenso más pronunciado, DFP, BFGS, Newton y Levenberg-Marquardt. El método de dirección conjugada de Powell es de búsqueda directa. La eficiencia de los métodos de solución se puede medir por tres criterios:

1. Número de funciones a evaluar
2. Tiempo computacional
3. Tasa de convergencia. Con esto nos referimos a la rapidez con que la secuencia $x_i, x_{i+1}, \ldots$ converge a x^*. La tasa de convergencia está dada por el parámetro n en la ecuación.

$$\frac{||x_{i+1} - x^*||}{||x_i - x^*||^n} \leq c, c \geq 0, n \geq 0$$

Para $n = 1$ y $0 \leq c \leq 1$ el método tiene una convergencia lineal.
Para n=2, el método tiene un convergencia cuadrática.
Cuanto mayor es la tasa de convergencia, el método de optimización es mejor.
Un método tiene una convergencia superlineal si

$$\lim_{i \to \infty} \left(\frac{||x_{i+1} - x^*||}{||x_i - x^*||^n} \right) \leq c, c \geq 0, n \geq 0$$

La dirección de búsqueda S_i que reduce el valor de la función es una dirección de descenso. A lo largo de la dirección del gradiente, existe el cambio máximo en el valor de la función. Así, a lo largo de la dirección negativa del gradiente, el valor de la función es el que más disminuye. La dirección negativa del gradiente se denomina la dirección de descenso más pronunciada. Es decir

$$S_i = -\nabla f(x_i)$$

En iteraciones sucesivas, las variables de diseño pueden actualizarse usando la ecuación

$$x_{i+1} = x_i - \alpha \nabla f(x_i)$$

donde α es un parámetro escalar positivo que se puede determinar utilizando el algoritmo de búsqueda de lineal, como el método de la sección dorada.

El método de descenso más pronunciado garantiza una reducción en el valor de la función en cada iteración. Si el punto de inicio está lejos del mínimo, el gradiente será mayor y la reducción de la función se maximizará en cada iteración. Debido a que el valor de gradiente de la función cambia y disminuye a un valor pequeño cerca del óptimo, la reducción de la función es desigual y el método se vuelve lento (convergencia lenta) cerca del mínimo. Por lo tanto, el método se puede utilizar como iniciador para otros algoritmos basados en gradientes.

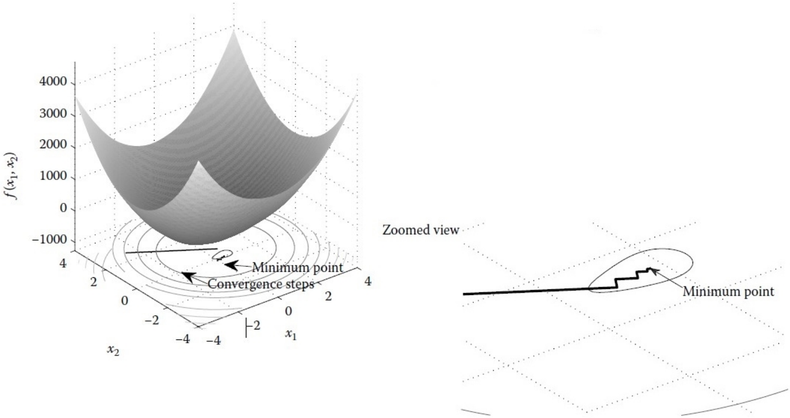
Algorithm for the Steepest Descent Method

Step 1: Given x_i (starting value of design variable)
ε_1 (tolerance of function value from previous iteration)
ε_2 (tolerance on gradient value)
Δx (required for gradient computation)

Step 2: Compute $f(x_i)$ and $\nabla f(x_i)$ (function and gradient vector)
$S_i = -\nabla f(x_i)$ (search direction)
Minimize $f(x_{i+1})$ and determine α (use golden section method)
$x_{i+1} = x_i + \alpha S_i$ (update the design vector)
If $\left| f(x_{i+1}) - f(x_i) \right| > \varepsilon_1$ or $\|\nabla f(x_i)\| > \varepsilon_2$
then goto Step 2
else goto Step 3

Step 3: Converged. Print $x^* = x_{i+1}$, $f(x^*) = f(x_{i+1})$

Convergencia del problema de prueba



Convergencia del problema de prueba

Initial function value = 1452.2619

No.	x-vector		f(x)	Deriv.
1	0.095	0.023	-2.704	1006.074
2	0.170	0.141	-5.278	37.036
3	0.318	0.048	-7.369	23.451
4	0.375	0.138	-8.773	26.656
5	0.448	0.092	-9.375	14.021
6	0.470	0.127	-9.583	10.071
7	0.491	0.114	-9.639	4.403
8	0.497	0.123	-9.652	2.509
9	0.501	0.120	-9.655	1.050
10	0.503	0.122	-9.656	0.554
11	0.504	0.122	-9.656	0.236
12	0.504	0.122	-9.656	0.125
13	0.504	0.122	-9.656	0.047
14	0.504	0.122	-9.656	0.027
15	0.504	0.122	-9.656	0.016

La dirección de búsqueda en este método se basa en la información de la primera y segunda derivada y está dada por:

$$S_i = -[H]^{-1} \nabla f(x_i)$$

donde $[H]$ es la matriz Hessian. Si la matriz es positiva definida, entonces S_i será en dirección descendente.

Características

- Si el punto inicial esta lejos de óptimo, la dirección de la búsqueda no siempre es descendente.
- Frecuentemente un reinicio es requerido con un diferente punto de inicio.
- Aunque el método de Newton es conocido por converger en una sola iteración para una función cuadrática, rara vez encontramos funciones en problemas prácticos que son cuadráticos.
- Este método es frecuentemente usado como método híbrido en conjunto con otros métodos.

El método es similar al método de Newton 1D con una modificación que se realiza una búsqueda con dirección S_i.

Step 1: Given x_i (starting value of design variable)

ε_1 (tolerance of function value from previous iteration)

ε_2 (tolerance on gradient value)

Δx (required for gradient computation)

Step 2: Compute $f(x_i)$, $\nabla f(x_i)$, and $[H]$ (function, gradient, and Hessian)

$S_i = -[H]^{-1}\nabla f(x_i)$ (search direction)

Minimize $f(x_{i+1})$ and determine α (use golden section method)

$x_{i+1} = x_i + \alpha S_i$ (update the design vector)

If $\left|f(x_{i+1}) - f(x_i)\right| > \varepsilon_1$ or $\|\nabla f(x_i)\| > \varepsilon_2$

then goto Step 2

else goto Step 3

Step 3: Converged. Print $x^* = x_{i+1}$, $f(x^*) = f(x_{i+1})$

Ejecutando el algoritmo con un punto inicial $x(-3, 2)$

Initial function value = 1452.2619

No.	x-vector		f(x)	Deriv.
1	-0.754	0.524	44.244	1006.074
2	-0.362	-0.010	8.398	116.281
3	0.094	0.125	-3.920	50.597
4	11.775	0.324	22007.14	21.420
5	1.042	0.093	14.533	4076.916
6	0.640	0.142	-8.479	102.745
7	0.524	0.122	-9.635	18.199
8	0.505	0.122	-9.656	2.213
9	0.504	0.122	-9.656	0.059
10	0.504	0.122	-9.656	0.000

Ejecutando el algoritmo con un punto inicial $x(1,1)$

Initial function value = 92.7864

No.	x-vector		f(x)	Deriv.
1	0.818	0.041	-1.428	202.492
2	0.569	0.138	-9.386	56.085
3	0.510	0.122	-9.655	8.516
4	0.504	0.122	-9.656	0.602
5	0.504	0.122	-9.656	0.004

Este método es similar al método de Newton, con la modificación que la búsqueda se realiza unidireccionalmente, en la dirección S_i del método de Newton.

Algorithm for Modified Newton's Method

Step 1: Given x_i (starting value of design variable)
ε_1 (tolerance of function value from previous iteration)
ε_2 (tolerance on gradient value)
Δx (required for gradient computation)

Step 2: Compute $f(x_i)$, $\nabla f(x_i)$, and $[H]$ (function, gradient, and Hessian)
$S_i = -[H]^{-1}\nabla f(x_i)$ (search direction)
Minimize $f(x_{i+1})$ and determine α (use golden section method)
$x_{i+1} = x_i + \alpha S_i$ (update the design vector)
If $|f(x_{i+1}) - f(x_i)| > \varepsilon_1$ or $\|\nabla f(x_i)\| > \varepsilon_2$
then goto Step 2
else goto Step 3

Step 3: Converged. Print $x^* = x_{i+1}$, $f(x^*) = f(x_{i+1})$

Ejecuntando el método de Newton modificado con un punto incial $x(-3, 2)$

Initial function value = 1452.2619

No.	x-vector		f(x)	Deriv.
1	0.006	0.025	-1.010	1006.074
2	0.498	0.026	-8.227	36.392
3	0.496	0.121	-9.653	29.839
4	0.504	0.122	-9.656	0.873
5	0.504	0.122	-9.656	0.018
6	0.504	0.122	-9.656	0.003

El método del paso más descendente se acerca al mínimo en muy pocas iteraciones cuando el punto esta lejos del mínimo, sin embargo si esta cerca el método se vuelve lento. Y por el contrario, el Mpetodo de Newton se aproxima muy rápido al mínimo cuando esta cerca de el y cuando esta lejos es lento o incluso puede no converger.

El método de Levenberg-Marquardt es un tipo de método híbrido que combina la fuerza del descenso más rápido y el método de Newton. La dirección de búsqueda en este método está dada por

$$S_i = -[H + \lambda I]^{-1} \nabla f(x_i)$$

donde I es una matriz de identidad y λ es un escalar que se establece en un valor alto al comienzo del algoritmo

Características

- El valor de λ se altera durante cada iteración dependiendo de si el valor de la función está disminuyendo o no
- Si el valor de la función disminuye en la iteración, λ disminuye en un factor (menos ponderación en la dirección de descenso más pronunciado)
- Si el valor de la función aumenta en la iteración, λ aumenta en un factor (más ponderación en la dirección de descenso más pronunciada).

Step 1: Given x_i (starting value of design variable)

ε_1 (tolerance of function value from previous iteration)

ε_2 (tolerance on gradient value)

Δx (required for gradient computation)

Step 2: Compute $f(x_i)$, $\nabla f(x_i)$, and $[H]$ (function, gradient, and Hessian)

$S_i = -[H + \lambda I]^{-1} \nabla f(x_i)$ (search direction)

$x_{i+1} = x_i + S_i$ (update the design vector)

If $f(x_{i+1}) < f(x_i)$

then change the value of λ as $\lambda/2$

else change the value of λ as 2λ

If $\left| f(x_{i+1}) - f(x_i) \right| > \varepsilon_1$ or $\|\nabla f(x_i)\| > \varepsilon_2$

then goto Step 2

else goto Step 3

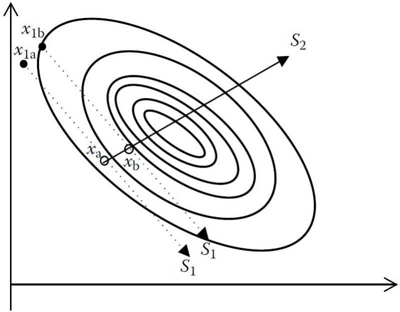
Step 3: Converged. Print $x^* = x_{i+1}$, $f(x^*) = f(x_{i+1})$

El método de gradiente conjugado es un método de primer orden, pero tiene la propiedad de la convergencia cuadrática y, por lo tanto, tiene una ventaja significativa sobre los métodos de segundo orden. Se dice que dos direcciones, S_1 y S_2, se conjugan si

$$S_1^T H S_2 = 0$$

donde H es una matriz simétrica.

Las direcciones ortogonales son direcciones conjugadas. Iniciando en el punto x_{1a}, La dirección de búsqueda S_1 resulta en el punto mínimo x_a^*. Similarmente, iniciando en el punto x_{1b}. El resultado de la búsqueda en dirección S_1 en el punto mínimo x_b^*. La linea que une x_a^* y x_b^* es la búsqueda de dirección S_2, entonces S_1 y S_2 son direcciones conjugadas.



El método del paso más descendente fue modificado por Fletcher y Reeves en el método del gradiente conjugado. Iniciando con la dirección de búsqueda

$$S_1 = -\nabla f(x_1)$$

y subsecuentemente la dirección de búsqueda es tomado como combinanción lineal de S_1 y $-\nabla f(x_2)$. Eso es

$$S_2 = -\nabla f(x_2) + \alpha S_1$$

Método gradiente conjunado Fletcher-Reeves

Usando la propiedad de $S_1^T H S_2 = 0$ de las direcciones conjugadas, α puede ser evaluada con

$$\alpha = \frac{||\nabla f(x_{i+1})||^2}{||\nabla f(x_i)||^2}$$

Iniciando con $S_1 = -\nabla f(x_1)$, la dirección de búsqueda en cada iteración es calculada usando la ecuación

$$S_{i+1} = -\nabla f(x_i) + \frac{||\nabla f(x_{i+1})||^2}{||\nabla f(x_i)||^2} S_i$$

Step 1: Given x_i (starting value of design variable)
ε_1 (tolerance of function value from previous iteration)
ε_2 (tolerance on gradient value)
Δx (required for gradient computation)

Step 2: Compute $f(x_i)$ $\nabla f(x_i)$ (function and gradient)
$S_i = -\nabla f(x_i)$ (search direction)
Minimize $f(x_{i+1})$ and determine α (use the golden section method)
$x_{i+1} = x_i + \alpha S_i$ (update the design vector)

Step 3: $S_{i+1} = -\nabla f(x_{i+1}) + \dfrac{\left\|\nabla f(x_{i+1})\right\|^2}{\left\|\nabla f(x_i)\right\|^2} S_i$
Minimize $f(x_{i+2})$ and determine α (use the golden section method)
$x_{i+2} = x_{i+1} + \alpha S_{i+1}$
Minimize $f(x_{i+2})$ and determine α (use the golden section method)
If $\left|f(x_{i+2}) - f(x_{i+1})\right| > \varepsilon_1$ or $\|\nabla f(x_{i+1})\| > \varepsilon_2$
then goto Step 3
else goto Step 4

Step 4: Converged. Print $x^* = x_{i+2}$, $f(x^*) = f(x_{i+2})$