

PROGRAMACIÓN NO LINEAL SIN RESTRICCIONES

Pedro Miranda Menéndez

Resumen

En este tema estudiaremos varios algoritmos de optimización de problemas de Programación No Lineal sin restricciones. En definitiva, lo que se plantea es hallar un máximo o un mínimo local de una función (tal vez restringida a un intervalo compacto de \mathbb{R}^n); y supondremos que las funciones que tenemos que optimizar son suficientemente complejas como para no poder obtener la solución óptima con exactitud; en otras palabras, supondremos que aunque la función pueda ser derivable, su derivada da lugar a una ecuación que no puede resolverse analíticamente. El tema se divide en dos partes: en la primera parte estudiaremos métodos de optimización de una función unidimensional; en la segunda parte, aplicaremos lo visto en la primera parte para resolver problemas de optimización de funciones en varias variables.

1. El caso unidimensional

Empezaremos con el caso más sencillo, el de una función $f : \mathbb{R} \rightarrow \mathbb{R}$. Para simplificar la notación, supondremos que nuestro problema consiste en minimizar la función f y supondremos que esta función tiene un solo mínimo. Los algoritmos de resolución se clasifican en dos familias: métodos de búsqueda y métodos de aproximación.

En los métodos de búsqueda se supone que conocemos un intervalo $[a, b]$ en el cual está este mínimo. La idea de estos métodos es reducir paulatinamente la amplitud del intervalo, de forma que finalmente se llegue a un intervalo de amplitud menor a una tolerancia dada ϵ . Y se toma como solución el punto medio del último intervalo.

En los métodos de aproximación la idea subyacente es aproximar la función original f por otra función g_1 , que sea parecida a f y que sea fácil de optimizar. Este óptimo nos da un nuevo punto desde el que realizar otra aproximación g_2 , y así sucesivamente. Por lo tanto, estos métodos necesitan partir de un punto inicial x_0 sobre el que se realice la primera aproximación; según este punto esté más o menos cerca del óptimo buscado, se necesitará un mayor o menor número de iteraciones. Nótese que en estos métodos de resolución no se obtiene una sucesión de intervalos, sino una sucesión de puntos. Por ello, ya no podemos considerar una condición de parada basada en la amplitud. Las condiciones de parada en estos métodos se basan en que es de esperar que la sucesión de puntos obtenidos converja hacia la solución del problema. Por ello, cuando los puntos obtenidos en dos iteraciones sucesivas están muy próximos esto se toma como un indicio de que se está muy próximo al límite (condición de Cauchy). Así, una condición de parada habitual es considerar una tolerancia ϵ y parar en el momento en que $|x_{n+1} - x_n| \leq \epsilon$. Es posible sin embargo que la función atraviese una zona de “valle” y que esto haga que los puntos obtenidos sean muy próximos aunque no se esté cerca del óptimo; por ello, en ocasiones se pone como condición de parada el primer n para el que $|x_m - x_n| \leq \epsilon, m = n + 1, \dots, n + r$, donde

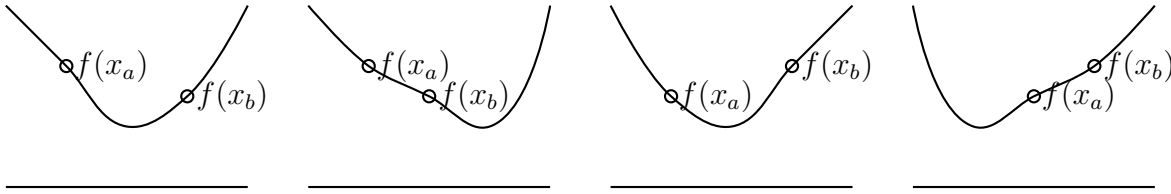
r está prefijado. Otra condición de parada es que efectivamente el punto obtenido esté próximo a cumplir las condiciones de óptimo: en concreto, se toma como condición de parada el primer n para el que $|f'(x_n)| \leq \epsilon$. Nótese que para poder aplicar este criterio necesitamos que f sea diferenciable. En todos los casos se toma como solución el último punto obtenido.

Como veremos más adelante, los métodos de resolución en el caso unidimensional serán la base de los métodos de resolución en el caso multidimensional. Los algoritmos que estudiaremos tratan de hallar aproximadamente este mínimo local.

En lo que sigue supondremos que la función f es continua en el intervalo $[a, b]$.

1.1. Algoritmos de la sección áurea y de Fibonacci

Estos métodos son métodos de búsqueda. En estos dos métodos se evalúa la función f en dos puntos intermedios $x_{a,1}, x_{b,1}$ del intervalo $[a, b]$.



Tenemos entonces dos posibilidades:

- Si $f(x_{a,1}) \geq f(x_{b,1})$ entonces necesariamente el mínimo local es mayor que $x_{a,1}$, pudiendo estar en $[x_{a,1}, x_{b,1}]$ o bien que esté en $[x_{b,1}, b]$. En definitiva, el mínimo local está en el intervalo $[x_{a,1}, b]$.
- Si $f(x_{a,1}) \leq f(x_{b,1})$ entonces se tiene que el mínimo local está en el intervalo $[a, x_{b,1}]$.

Supongamos que estamos en el primero de los casos; ahora debemos volver a evaluar la función en otros dos puntos del intervalo $[x_{a,1}, b]$ y, como puede ser costoso realizar estas evaluaciones, vamos a hacer que uno de los puntos que se evalúan sea $x_{b,1}$, que ya conocemos.

Como a priori no sabemos si elegiremos el intervalo $[x_{a,1}, b]$ o el intervalo $[a, x_{b,1}]$, no sabemos si el punto intermedio de la siguiente iteración será $x_{a,1}$ ó $x_{b,1}$; por ello, debemos suponer que los dos intervalos $[x_{a,1}, b]$ y $[a, x_{b,1}]$ tienen la misma amplitud, que denotaremos I_1 ; entonces $x_{a,1}$ y $x_{b,1}$ son simétricos respecto del punto medio del intervalo $[a, b]$. Por otra parte, queremos que la reducción en la amplitud sea lo más grande posible, por lo que procuraremos que $x_{a,1}$ y $x_{b,1}$ estén lo más cerca posible del punto medio de $[a, b]$ (es decir, $x_{b,1}$ estará más cerca de $x_{a,1}$ que de b). Ahora, suponiendo que se haya escogido el intervalo $[x_{a,1}, b]$, como aprovechamos el punto $x_{b,1}$, hacemos que $x_{a,2} = x_{b,1}$ y $x_{b,2}$ se obtiene por simetría. Si ahora seleccionamos en esta iteración el intervalo $[x_{a,2}, b] = [x_{b,1}, b]$, tendremos un nuevo intervalo de amplitud I_2 . Nótese que

$$I_0 = b - a = b - x_{b,1} + x_{b,1} - a = I_2 + I_1.$$

Y en general tendremos que

$$I_k = I_{k+1} + I_{k+2}, k \geq 0.$$

El problema ahora se reduce a la forma es que son elegidos los puntos intermedios del intervalo y aquí están las diferencias entre los dos métodos.

1.1.1. El método de la sección áurea

En este algoritmo imponemos además de las condiciones anteriores que la amplitud disminuya siempre en la misma proporción, i.e.

$$\frac{I_k}{I_{k+1}} = C.$$

En consecuencia, $\frac{I_k}{I_{k+2}} = C^2$. Ahora bien,

$$I_k = I_{k+1} + I_{k+2} \Leftrightarrow \frac{I_k}{I_{k+2}} = \frac{I_{k+1}}{I_{k+2}} + \frac{I_{k+2}}{I_{k+2}} \Leftrightarrow C^2 = C + 1.$$

Resolviendo esta ecuación se obtiene que $C = \frac{1 + \sqrt{5}}{2} \approx 1,618$, el valor de la sección áurea.

De esta manera, podemos deducir que la amplitud del intervalo resultante de la iteración k -ésima es $\frac{I_0}{C^k}$; en consecuencia, dada una tolerancia ϵ , el número de iteraciones necesarias para conseguir un intervalo de amplitud menor que ϵ será el primer n tal que

$$\frac{I_0}{C^n} \leq \epsilon \Leftrightarrow \frac{I_0}{\epsilon} \leq C^n \Leftrightarrow \log I_0 - \log \epsilon \leq n \log C.$$

Por otra parte, como sabemos la amplitud del nuevo intervalo, esto nos permite determinar directamente los valores x_a y x_b en los que tiene que evaluarse la función f en cada iteración. En efecto, como el intervalo en el que comienza la siguiente iteración es $[a, x_b]$ ó $[x_a, b]$, entonces si denotamos $I_k := \frac{I_{k-1}}{C}$, se tiene que $x_b = a + I_k$, $x_a = b - I_k$.

Ejemplo 1. Consideremos la función $f(x) = x^3 - 4x^2 + 5x - 10$. Para esta función vamos a buscar un mínimo local en el intervalo $[1, 5]$ con una tolerancia $\epsilon = 0,5$. Si aplicamos el método de la sección áurea obtenemos los siguientes resultados:

Iteración	$\frac{I_0}{C^k}$	Ext. inferior	Ext. superior	x_a	$f(x_a)$	x_b	$f(x_b)$
1	2.472	1	5	2.528	-6.767	3.472	0.795
2	1.528	1	3.472	1.944	-8.050	2.528	-6.767
3	0.944	1	2.528	1.584	-8.142	1.944	-8.050
4	0.584	1	1.944	1.360	-8.083	1.584	-8.142
5	0.360	1.36	1.944	1.584	-8.142	1.720	-8.146

El intervalo final que se toma como solución es $[1,584, 1,944]$ y el punto solución sería 1.764.

1.1.2. El método de Fibonacci

En este método no se impone ninguna restricción sobre la forma en que disminuye la amplitud de los intervalos. De esta forma, lo que buscamos es resolver el problema en el menor número posible de iteraciones.

Supongamos que hemos terminado la penúltima iteración, digamos la iteración $n - 1$. Entonces, la menor amplitud para el último intervalo se conseguiría si x_a y x_b se encuentran muy cerca del punto medio del intervalo. En el límite, hacemos coincidir x_a y x_b con ese punto medio, ya que no es necesario tener dos puntos de comparación al no realizarse la iteración $n + 1$. Entonces,

$$I_{n-1} = 2I_n.$$

Pero entonces,

$$I_{n-2} = I_{n-1} + I_n = 3I_n, \quad I_{n-3} = I_{n-2} + I_{n-1} = 5I_n, \dots$$

Y en general $I_k = F_{n-k}I_n$, donde F_i representa el i -ésimo término de la sucesión de Fibonacci, que viene dada por

$$F_i = \begin{cases} 1 & \text{si } i = 0, -1 \\ F_{i-1} + F_{i-2} & \text{si } i \geq 1 \end{cases}$$

En particular, $I_0 = F_n I_n$; de esta forma, el número de iteraciones será el menor valor n tal que $\frac{I_0}{F_n} \leq \epsilon$. Por otra parte, podemos calcular a priori la amplitud de los intervalos de las siguientes iteraciones teniendo en cuenta que conocemos la amplitud del intervalo final, que viene dada por $I_n = \frac{I_0}{F_n}$ y que conocemos las demás amplitudes en función de la de I_n .

Al resolver el problema por el algoritmo de Fibonacci tenemos el problema de que si igualamos $x_a = x_b$ en la última iteración no sabríamos si el intervalo final es $[a, x_b]$ ó $[x_a, b]$. Para evitar este problema, tomamos x_a y x_b ligeramente por debajo y por encima respectivamente del punto medio del intervalo.

Ejemplo 2. Consideremos el ejemplo anterior. Entonces se tiene la siguiente tabla:

Iteración	$\frac{I_0}{F_k}$	I_k	Ext. inferior	Ext. superior	x_a	$f(x_a)$	x_b	$f(x_b)$
1	4	4	1	5	2.5	-6.875	3.5	1.375
2	2	2.5	1	3.5	2	-8	2.5	-6.875
3	$\frac{4}{3}$	1.5	1	2.5	1.5	-8.125	2	-8
4	0.8	1	1	2	1.4	-8.096	1.6	-8.144
5	0.5	0.5						

El intervalo final que se toma como solución es $[1.5, 2]$ y el punto solución sería 1.75.

1.2. El método de la bisección

Este método también es un método de búsqueda. Para aplicar el método de la bisección tenemos que suponer que la función f es derivable en el intervalo $[a, b]$ y que en este intervalo sólo hay un punto en el que se anula la derivada, que es el mínimo buscado. Este método consiste en calcular la derivada en el punto medio del intervalo, que denotaremos por c . Tenemos ahora tres casos:

- Si $f'(c) = 0$, entonces c es el mínimo local que estamos buscando.
- Si $f'(c) > 0$, entonces la función es creciente en c , lo que implica que el mínimo se encuentra en el intervalo $[a, c]$.
- Si $f'(c) < 0$, entonces la función es decreciente en c , lo que implica que el mínimo se encuentra en el intervalo $[c, b]$.

Con este método la amplitud del intervalo se reduce a la mitad en cada iteración, es decir, $I_k = \frac{I_0}{2^k}$; en consecuencia, dada una tolerancia ϵ , el número de iteraciones necesarias para conseguir un intervalo de amplitud menor que ϵ será el primer n tal que

$$\frac{I_0}{2^n} \leq \epsilon \Leftrightarrow \frac{I_0}{\epsilon} \leq 2^n \Leftrightarrow \log I_0 - \log \epsilon \leq n \log 2.$$

Ejemplo 3. Consideremos el ejemplo anterior. La derivada de la función f es

$$f'(x) = 3x^2 - 8x + 5.$$

Entonces se tiene la siguiente tabla:

Iteración	I_k	Ext. inferior	Ext. superior	c	$f'(c)$
1	4	1	5	3	8
2	2	1	3	2	1
3	1	1	2	1.5	-0.25
4	0.5	1.5	2		

El intervalo final que se toma como solución es $[1.5, 2]$ y el punto solución sería 1.75.

El método de la bisección necesita un menor número de iteraciones que los métodos de la sección áurea y Fibonacci para obtener la tolerancia requerida; sin embargo, también necesita la diferenciabilidad de la función, condición que no era necesaria en los otros métodos.

1.3. El método de Newton unidimensional

Veamos ahora un método de aproximación. El método de Newton consiste en aproximar la función f por su desarrollo de Taylor de orden 2 centrado en x_0 . De esta forma, se tiene

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 =: g(x).$$

Optimizamos ahora la función g . Derivando e igualando a cero se obtiene la ecuación

$$f'(x_0) + f''(x_0)(x - x_0) = 0 \Leftrightarrow x = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

De esta forma tomamos como nuevo punto el punto $x_1 := x_0 - \frac{f'(x_0)}{f''(x_0)}$, y en general, si tenemos el punto x_k , el siguiente punto es

$$x_{k+1} := x_k - \frac{f'(x_k)}{f''(x_k)}.$$

Nótese que en este método se necesita que la función f original sea dos veces diferenciable. Por otra parte, es posible que este método nos lleve a un máximo o a un punto de inflexión; incluso es posible que la sucesión no converja o sea oscilante. Por ello, es muy importante que el punto inicial no esté muy alejado del mínimo o imponer condiciones sobre la función de forma que haya un único punto en que se anule la derivada y sea un mínimo (lo que por ejemplo pasa con funciones convexas).

Ejemplo 4. Consideremos nuevamente el ejemplo anterior comenzando en $x_0 = 3,5$ y consideremos como condición de parada el primer n para el que $|x_{n+1} - x_n| \leq 0,1$. Para nuestra función se tiene

$$f'(x) = 3x^2 - 8x + 5, \quad f''(x) = 6x - 8.$$

Entonces se tiene:

Iteración	x_{k-1}	$f'(x_{k-1})$	$f''(x_{k-1})$	x_k
1	3.5	13.75	13	2.442
2	2.442	3.352	6.652	1.938
3	1.938	0.763	3.628	1.728
4	1.728	0.134	2.368	1.671

Y tomamos como mínimo de la función el punto 1.671.

2. El caso multidimensional

En los tres primeros algoritmos del caso multidimensional vamos a aprovechar lo visto para el caso unidimensional. Para ello, tenemos que transformar un problema en varias dimensiones en uno en una dimensión. La idea de estos algoritmos consiste en determinar una dirección \vec{d} y desde un punto dado \vec{x} se busca optimizar la función f en la dirección \vec{d} ; así, ya tenemos un problema en una dirección, puesto que se trata de minimizar a partir de un punto \vec{x}_0 la función $g(\lambda)$ dada por

$$g(\lambda) := f(\vec{x}_0 + \lambda \vec{d}),$$

es decir, una función que depende solamente de la variable λ . Y éste es un problema que se puede resolver siguiendo alguno de los algoritmos de la sección anterior; en caso de que utilicemos algoritmos que necesitan un intervalo de búsqueda, podemos tomar un intervalo $[a, b]$ más o menos amplio.

Ahora los distintos algoritmos se diferencian por la forma de determinar la dirección \vec{d} .

Las condiciones de parada son adaptaciones de las que se tenían para el caso unidimensional: dada una tolerancia prefijada $\epsilon > 0$, casi siempre se toma como condición de parada o bien que $\|x_{n+1} - x_n\| \leq \epsilon$, o bien que $\|\nabla f(x_n)\| \leq \epsilon$.

2.1. El método de las coordenadas cíclicas

En este método se toman sucesivamente como direcciones de mejora los ejes coordenados.

Así, si empezamos en un punto \vec{x}_0 , se busca primero optimizar en $\vec{d}_1 = \vec{e}_1$. Obtenemos entonces el punto \vec{x}_0^1 . A partir de \vec{x}_0^1 , optimizamos en la dirección $\vec{d}_2 = \vec{e}_2$, obteniendo el punto \vec{x}_0^2 ; y así sucesivamente. El punto final de la iteración es $\vec{x}_1 = \vec{x}_0^n$, es decir, el punto que se obtiene al optimizar en la dirección del último eje de coordenadas.

Ejemplo 5. Consideremos la función $f(x, y) = x^2 + 2y^2 - 2y - 2xy$ y tomemos como punto inicial el punto $\vec{x}_0 = (0, 0)$.

Si aplicamos el algoritmo de las coordenadas cíclicas se obtienen los siguientes puntos en las tres primeras iteraciones:

■ Iteración 1

- Sea $\vec{d} = (1, 0)$. Entonces,

$$f(\vec{x}_0 + \lambda(1, 0)) = \lambda^2.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \lambda^2$ que se obtiene en $\lambda = 0$. Se obtiene nuevamente así el punto $\vec{x}_0^1 = (0, 0)$.

- Sea $\vec{d} = (0, 1)$. Entonces,

$$f(\vec{x}_0^1 + \lambda(0, 1)) = 2\lambda^2 - 2\lambda.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = 2\lambda^2 - 2\lambda$ que se obtiene en $\lambda = 1/2$. Se obtiene así el punto $\vec{x}_0^2 = (0, 1/2)$.

En consecuencia, $\vec{x}_1 = (0, 1/2)$ y $f(\vec{x}_1) = -1/2$.

■ Iteración 2

- Sea $\vec{d} = (1, 0)$. Entonces,

$$f(\vec{x}_1 + \lambda(1, 0)) = \lambda^2 - \frac{1}{2} + \lambda.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \lambda^2 - \frac{1}{2} + \lambda$ que se obtiene en $\lambda = 1/2$. Se obtiene así el punto $\vec{x}_1^1 = (1/2, 1/2)$.

- Sea $\vec{d} = (0, 1)$. Entonces,

$$f(\vec{x}_1^1 + \lambda(0, 1)) = \frac{1}{4} + 2\left(\frac{1}{2} + \lambda\right)^2 - 2\left(\frac{1}{2} + \lambda\right).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \frac{1}{4} + 2\left(\frac{1}{2} + \lambda\right)^2 - 2\left(\frac{1}{2} + \lambda\right)$ que se obtiene en $\lambda = 1/4$. Se obtiene así el punto $\vec{x}_1^2 = (1/2, 3/4)$.

En consecuencia, $\vec{x}_2 = (1/2, 3/4)$ y $f(\vec{x}_2) = -7/8$.

■ Iteración 3

- Sea $\vec{d} = (1, 0)$. Entonces,

$$f(\vec{x}_2 + \lambda(1, 0)) = \left(\frac{1}{2} + \lambda\right)^2 - 3\left(\frac{1}{2} + \lambda\right) - \frac{3}{8}.$$

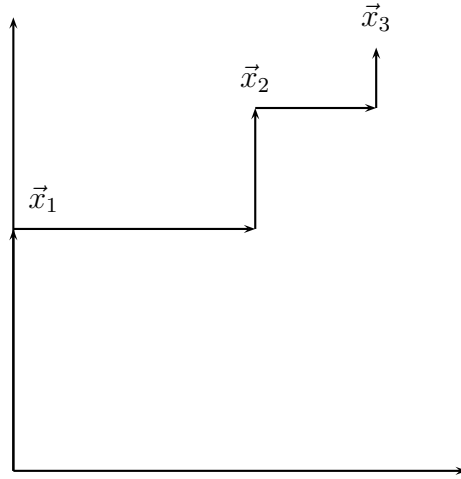
Tenemos entonces que hallar el mínimo de $g(\lambda) = \left(\frac{1}{2} + \lambda\right)^2 - 3\left(\frac{1}{2} + \lambda\right) - \frac{3}{8}$ que se obtiene en $\lambda = 1/4$. Se obtiene así el punto $\vec{x}_2^1 = (3/4, 3/4)$.

- Sea $\vec{d} = (0, 1)$. Entonces,

$$f(\vec{x}_2^1 + \lambda(0, 1)) = \frac{9}{8} + 2\left(\frac{3}{4} + \lambda\right)^2 - \frac{7}{2}\left(\frac{3}{4} + \lambda\right).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \frac{9}{8} + 2\left(\frac{3}{4} + \lambda\right)^2 - \frac{7}{2}\left(\frac{3}{4} + \lambda\right)$ que se obtiene en $\lambda = 1/8$. Se obtiene así el punto $\vec{x}_2^2 = (3/4, 7/8)$.

En consecuencia, $\vec{x}_1 = (3/4, 7/8)$ y $f(\vec{x}_3) = -31/32$.



2.2. El método de Hooke and Jeeves

Este método se basa en el método de las coordenadas cíclicas, pero añadiendo un paso adicional en cada iteración. Así, una vez obtenido el punto \vec{x}_i^n , se toma la dirección $\vec{d} = \vec{x}_i^n - \vec{x}_i$ y se optimiza en esta dirección, obteniéndose el punto \vec{x}_{i+1} , desde el que comienza la siguiente iteración.

Ejemplo 6. Vamos a aplicar este algoritmo al ejemplo anterior tomando nuevamente como punto inicial el punto $\vec{x}_0 = (0, 0)$.

■ Iteración 1

- **Paso 1.** Obtención de la dirección de mejora.

Como hemos visto al aplicar el método de las coordenadas cíclicas

$$\left. \begin{array}{l} \vec{x}_0^1 = (0, 0) \\ \vec{x}_0^2 = (0, 1/2) \end{array} \right\} \Rightarrow \vec{d} = \vec{x}_0^2 - \vec{x}_0 = (0, 1/2).$$

- **Paso 2.** Óptimo en la dirección de mejora

Ahora se tiene

$$f(\vec{x}_0 + \lambda(0, 1/2)) = \frac{\lambda^2}{2} - \lambda.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \frac{\lambda^2}{2} - \lambda$ que se obtiene en $\lambda = 1$. Se obtiene así el punto $\vec{x}_1 = (0, 1/2)$ y $f(\vec{x}_1) = -1/2$.

■ Iteración 2

- **Paso 1.** Obtención de la dirección de mejora.

Como hemos visto al aplicar el método de las coordenadas cíclicas

$$\left. \begin{array}{l} \vec{x}_1^1 = (1/2, 0) \\ \vec{x}_1^2 = (1/2, 3/4) \end{array} \right\} \Rightarrow \vec{d} = \vec{x}_1^2 - \vec{x}_1^1 = (1/2, 1/4).$$

- **Paso 2.** Óptimo en la dirección de mejora

Ahora se tiene

$$f(\vec{x}_1 + \lambda(1/2, 1/4)) = \left(\frac{\lambda}{2}\right)^2 + 2\left(\frac{1}{2} + \frac{1}{4}\lambda\right)^2 - 2\left(\frac{1}{2} + \frac{1}{4}\lambda\right) - \lambda\left(\frac{1}{2} + \frac{1}{4}\lambda\right).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \left(\frac{\lambda}{2}\right)^2 + 2\left(\frac{1}{2} + \frac{1}{4}\lambda\right)^2 - 2\left(\frac{1}{2} + \frac{1}{4}\lambda\right) - \lambda\left(\frac{1}{2} + \frac{1}{4}\lambda\right)$ que se obtiene en $\lambda = 2$. Se obtiene así el punto $\vec{x}_2 = (1, 1)$ y $f(\vec{x}_2) = -1$.

■ Iteración 3

- **Paso 1.** Obtención de la dirección de mejora.

- Sea $\vec{d} = (1, 0)$. Entonces,

$$f(\vec{x}_2 + \lambda(1, 0)) = (\lambda + 1)^2 - 2(1 + \lambda).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = (\lambda + 1)^2 - 2(1 + \lambda)$ que se obtiene en $\lambda = 0$. Se obtiene nuevamente así el punto $\vec{x}_2^1 = (1, 1)$.

- Sea $\vec{d} = (0, 1)$. Entonces,

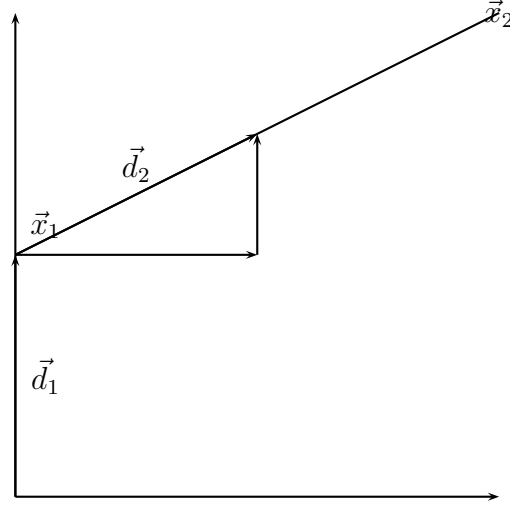
$$f(\vec{x}_2^1 + \lambda(0, 1)) = 1 + 2(1 + \lambda)^2 - 4(1 + \lambda).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = 1 + 2(1 + \lambda)^2 - 4(1 + \lambda)$ que se obtiene en $\lambda = 0$. Se obtiene así el punto $\vec{x}_2^2 = (1, 1)$.

$$\left. \begin{array}{l} \vec{x}_2^1 = (1, 1) \\ \vec{x}_2^2 = (1, 1) \end{array} \right\} \Rightarrow \vec{d} = \vec{x}_2^2 - \vec{x}_2^1 = (0, 0).$$

- **Paso 2.** Óptimo en la dirección de mejora.

Como $\vec{d} = (0, 0)$, entonces $\vec{x}_3 = \vec{x}_2 = (1, 1)$. FIN.



Es importante tener en cuenta que el método de Hooke and Jeeves no se acerca al mínimo local necesariamente más rápido que el método de coordenadas cíclicas; además, cada iteración es más costosa que en el método de las coordenadas cíclicas. Sin embargo, desde un punto de vista intuitivo es de esperar que la dirección en la que busca este algoritmo sea una buena elección para buscar, ya que es una combinación de direcciones de mejora.

2.3. El método de la pendiente más pronunciada (steepest descend)

En el método de las coordenadas cíclicas, las direcciones en las que se optimizaba estaban determinadas a priori. Es decir, no dependían de la función f ni del punto \vec{x}_i . En otras palabras, no tenemos ninguna razón para suponer que los ejes coordenados sean buenas direcciones para optimizar y la única ventaja que tenemos es que son direcciones sencillas intuitivamente, que no computacionalmente. Veamos ahora otro método que en principio parece más lógico, aunque exige la diferenciabilidad de la función.

En el método de la pendiente más pronunciada la dirección en la que buscamos es aquella en la que la función disminuye más rápidamente. Esta dirección ya sabemos que viene dada por $-\nabla f(\vec{x})$. En definitiva, el método de la pendiente más pronunciada parte de un punto inicial \vec{x}_0 y busca el óptimo de la función en la dirección $-\nabla f(\vec{x}_0)$. Esto nos dará el punto \vec{x}_1 a partir del cual se repite el proceso.

En este algoritmo el intervalo de búsqueda es del tipo $[0, b]$, puesto que la dirección $\nabla f(\vec{x})$ es la dirección de máximo aumento y, en consecuencia, la que menos nos interesa a priori. Es decir, no sirven valores negativos para λ .

Como nos sucedía al aplicar el simplex, el escoger la dirección más promisoría no nos garantiza una convergencia más rápida, pero tampoco nos la garantiza cualquier otro algoritmo y el algoritmo de la pendiente más pronunciada nos da una razón lógica para escoger esa dirección.

Ejemplo 7. Vamos a aplicar este algoritmo al ejemplo anterior tomando nuevamente como punto inicial el punto $\vec{x}_0 = (0, 0)$.

Para esta función se tiene que

$$\nabla f(x, y) = \begin{pmatrix} 2x - 2y \\ 4y - 2 - 2x \end{pmatrix}.$$

- **Iteración 1.** Para \vec{x}_0 , se tiene que $\nabla f(x, y) = (0, -2)$, luego tomamos $\vec{d} = (0, 2)$. Entonces,

$$f(\vec{x}_0 + \lambda(0, 2)) = 8\lambda^2 - 4\lambda.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = 8\lambda^2 - 4\lambda$ que se obtiene en $\lambda = 1/4$. Se obtiene así el punto $\vec{x}_1 = (0, 1/2)$ y $f(\vec{x}_1) = -1/2$.

- **Iteración 2.** Para \vec{x}_1 , se tiene que $\nabla f(x, y) = (-1, 0)$, luego tomamos $\vec{d} = (1, 0)$. Entonces,

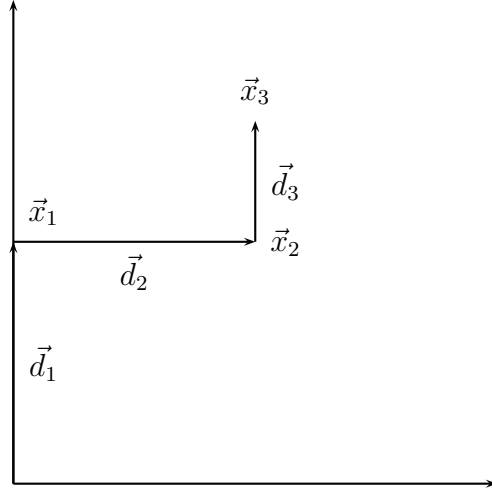
$$f(\vec{x}_1 + \lambda(1, 0)) = \lambda^2 - \lambda - \frac{1}{2}.$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \lambda^2 - \lambda - \frac{1}{2}$ que se obtiene en $\lambda = 1/2$. Se obtiene así el punto $\vec{x}_2 = (1/2, 1/2)$ y $f(\vec{x}_2) = -3/4$.

- **Iteración 3.** Para \vec{x}_2 , se tiene que $\nabla f(x, y) = (0, -1)$, luego tomamos $\vec{d} = (0, 1)$. Entonces,

$$f(\vec{x}_2 + \lambda(0, 2)) = \frac{1}{4} + 2\left(\frac{1}{2} + \lambda\right)^2 - 3\left(\frac{1}{2} + \lambda\right).$$

Tenemos entonces que hallar el mínimo de $g(\lambda) = \frac{1}{4} + 2\left(\frac{1}{2} + \lambda\right)^2 - 3\left(\frac{1}{2} + \lambda\right)$ que se obtiene en $\lambda = 1/4$. Se obtiene así el punto $\vec{x}_3 = (1/2, 3/4)$ y $f(\vec{x}_3) = -7/8$.



2.4. El método de Newton multidimensional

Este método es una extensión del método visto en el caso unidimensional. Se parte de un punto inicial \vec{x}_0 . Si aproximamos la función f por su desarrollo de Taylor de orden 2 centrado en \vec{x}_0 , se obtiene

$$f(\vec{x}) \approx f(\vec{x}_0) + \nabla f(\vec{x}_0)^t(\vec{x} - \vec{x}_0) + \frac{1}{2}(\vec{x} - \vec{x}_0)^t H(\vec{x}_0)(\vec{x} - \vec{x}_0) = g(\vec{x}).$$

Optimizamos ahora la función g . Derivando respecto a cada variable e igualando a cero se obtiene el sistema de ecuaciones

$$\nabla f(\vec{x}_0) + H(\vec{x}_0)(\vec{x} - \vec{x}_0) = \vec{0} \Leftrightarrow \vec{x} = \vec{x}_0 - H(\vec{x}_0)^{-1} \nabla f(\vec{x}_0).$$

De esta forma tomamos como nuevo punto el punto $\vec{x}_1 = \vec{x}_0 - H(\vec{x}_0)^{-1} \nabla f(\vec{x}_0)$, y en general, si tenemos el punto \vec{x}_k , el siguiente punto es

$$\vec{x}_{k+1} = \vec{x}_k - H(\vec{x}_k)^{-1} \nabla f(\vec{x}_k).$$

Nótese que en este método se necesita que la función f original sea dos veces diferenciable. Por otra parte, tiene los mismos inconvenientes que el método de Newton unidimensional. Es decir, es posible que este método nos lleve a un máximo, a un punto de silla, ...

Ejemplo 8. Consideremos la función $f(x, y) = x^4 - x^2y^2 + y^4$. Vamos a buscar el mínimo de esta función comenzando en $\vec{x}_0^t = (1, 1)$, para el que $f(1, 1) = 1$.

Nótese que

$$\frac{\partial f}{\partial x}(x, y) = 4x^3 - 2xy^2, \quad \frac{\partial f}{\partial y}(x, y) = 4y^3 - 2yx^2.$$

$$\frac{\partial^2 f}{\partial^2 x}(x, y) = 12x^2 - 2y^2, \quad \frac{\partial^2 f}{\partial^2 y}(x, y) = 12y^2 - 2x^2, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = -2xy.$$

Por lo tanto,

$$H = \begin{pmatrix} 12x^2 - 2y^2 & -4xy \\ -4xy & 12y^2 - 2x^2 \end{pmatrix}, \quad \nabla f(x, y) = \begin{pmatrix} 4x^3 - 2xy^2 \\ 4y^3 - 2yx^2 \end{pmatrix}.$$

■ **Iteración 1.** Para \vec{x}_0 se tiene

$$\nabla f(1, 1) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad H(1, 1) = \begin{pmatrix} 10 & -4 \\ -4 & 10 \end{pmatrix} \Rightarrow H^{-1}(1, 1) = \frac{1}{84} \begin{pmatrix} 10 & 4 \\ 4 & 10 \end{pmatrix}.$$

Entonces,

$$\vec{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{84} \begin{pmatrix} 10 & 4 \\ 4 & 10 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{84} \begin{pmatrix} 28 \\ 28 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix}, \quad f(2/3, 2/3) = \frac{16}{81}.$$

■ **Iteración 2.** Para \vec{x}_1 se tiene

$$\nabla f(2/3, 2/3) = \begin{pmatrix} 16/27 \\ 16/27 \end{pmatrix}, \quad H(2/3, 2/3) = \begin{pmatrix} 4 & -16/9 \\ -16/9 & 4 \end{pmatrix}$$

$$H^{-1}(2/3, 2/3) = \frac{81}{1040} \begin{pmatrix} 4 & 16/9 \\ 16/9 & 4 \end{pmatrix}.$$

Entonces,

$$\vec{x}_2 = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix} - \frac{81}{1040} \begin{pmatrix} 4 & 16/9 \\ 16/9 & 4 \end{pmatrix} \begin{pmatrix} 16/27 \\ 16/27 \end{pmatrix} = \begin{pmatrix} 2/5 \\ 2/5 \end{pmatrix}, \quad f(2/5, 2/5) = \frac{16}{625}.$$

■ **Iteración 3.** Para \vec{x}_2 se tiene

$$\nabla f(2/5, 2/5) = \begin{pmatrix} 16/125 \\ 16/125 \end{pmatrix}, \quad H(2/5, 2/5) = \begin{pmatrix} 28/25 & -16/25 \\ -16/25 & 28/25 \end{pmatrix}$$

$$H^{-1}(2/5, 2/5) = \frac{3 \cdot 11 \cdot 4}{5^4} \begin{pmatrix} 28/25 & 16/25 \\ 16/25 & 28/25 \end{pmatrix}.$$

Entonces,

$$\vec{x}_3 = \begin{pmatrix} 2/5 \\ 2/5 \end{pmatrix} - \frac{3 \cdot 11 \cdot 4}{5^4} \begin{pmatrix} 28/25 & 16/25 \\ 16/25 & 28/25 \end{pmatrix} \begin{pmatrix} 16/125 \\ 16/125 \end{pmatrix} = \begin{pmatrix} 2/15 \\ 2/15 \end{pmatrix}, \quad f(2/15, 2/15) = \frac{16}{625 \cdot 81}.$$

Puede comprobarse que el mínimo se encuentra en $(0,0)$.

