Tema 7. Optimización

7.1 Minimización de una función de una variable
7.2 Eliminación de Regiones
7.3 Método de la Sección Dorada
7.4 Optimización multidimensional. Método Nelder-Mead (Simplex)
7.5 Método del Gradiente (Método del máximo descenso)
7.6 Método de Newton generalizado
7.7 Algoritmos genéticos

Minimización de una función de una variable

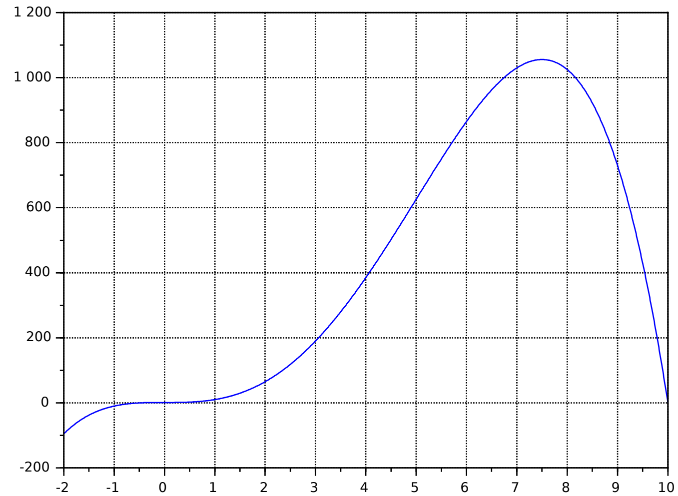
Términos: Optimizacíon=Minimizacíon - es busquieda de los extremos.

La función $f(x)$ **tiene** un **mínimo local** en $x=p$, si existe un intervalo I que contiene p tal que $f(p)\leq f(x)$ para todo $x\in I$.

Del mismo modo, $f(x)$ **tiene** un **máximo local** en $x=p$ si $f(x)\leq f(p)$ para todos los $x\in I$.

Si $f(x)$ tiene **un mínimo o máximo local** en $x=p$, se dice que han un **extremo local** en $x=p$.

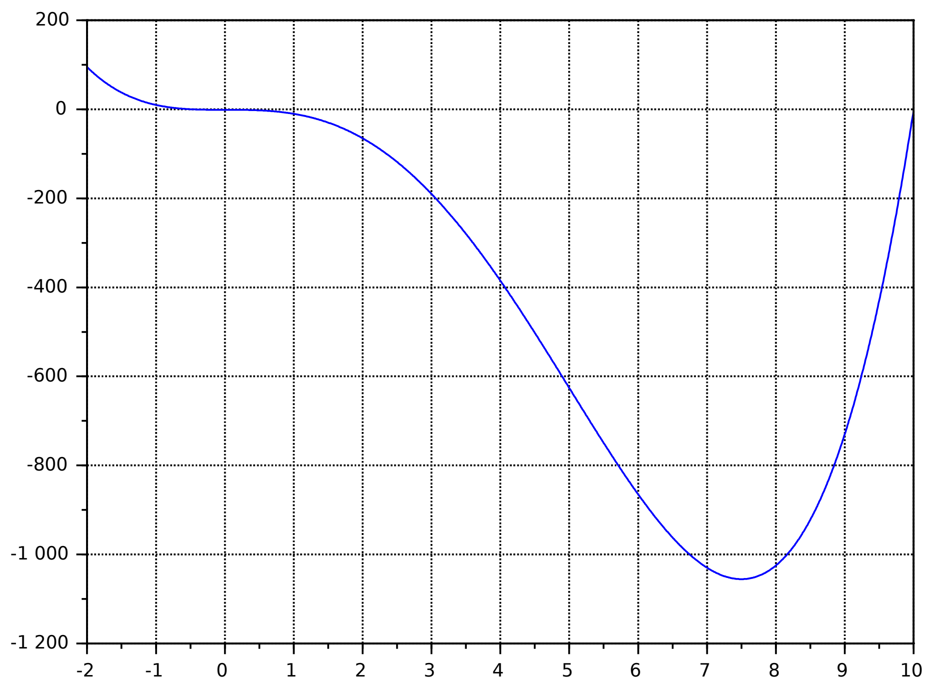
Suponga que $f(x)$ es definida en $I=[a,b]$ y tiene un extremo local en un punto interior $p\in[a,b]$. Si $f(x)$ es diferenciable en $x=p$, entonces $f'(p)=0$.



Minimización de una función de una variable

Prueba de la primera derivada. Suponga que $f(x)$ es **continua** en $I=[a,b]$. Supongamos que $f'(x)$ está definida para todo $x\in[a,b]$, excepto posiblemente en $x=p$.

(1) Si $f'(x)<0$ en (a,p) y $f'(x)>0$ en (p,b) , entonces $f(p)$ es un **mínimo local**.



(2) Si $f'(x)>0$ en (a,p) y $f'(x)<0$ <0 en (p,b) , entonces $f(p)$ es un **máximo local**.

Minimización de una función de una variable

Prueba de la segunda derivada. Suponga que f es continua en $[a,b]$ y f' y f'' se definen en (a,b). Además, supongamos que $p\in(a,b)$ es un **punto crítico** donde $f'(p)=0$.

(1) Si $f''(p)>0$, entonces $f(p)$ es un mínimo local de f.

(2) Si $f''(p)<0$, entonces $f(p)$ es un máximo local de f.

(3) Si $f''(p)=0$, entonces esta prueba no es concluyente.

Ejemplo Use la prueba de segunda derivada para clasificar los extremos locales de $f(x)=x^3+x^2-x+1$ en el intervalo $[-2,2]$.

La primera derivada es $f'(x)=3x^2+2x-1=(3x-1)(x+1)$, y el segunda derivada es $f''(x)=6x+2$. Hay dos puntos donde $f'=0$ (es decir, $x=1/3,-1$).

Caso (1): En $x=1/3$ nos encontramos con que $f'(1/3)=0$ y $f''(1/3)=4>0$, por lo que $f(x)$ tiene un mínimo local en $x=1/3$.

Caso (2): En $x=-1$ encontramos que $f'(-1)=0$ y $f''(-1)=-4<0$, por lo que $f(x)$ tiene un máximo local en $x=-1$.

Eliminación de Regiones

Se llama así a los métodos de búsqueda que localizan el optimo mediante la eliminación sucesiva de subintervalos que reduce el intervalo de búsqueda restante.

Nótese que prácticamente todos los métodos de búsqueda unidimensionales presuponen **unimodalidad (función tiene solo un extremo)** de la función a resolverse, al menos en el dominio de interés de la misma.

Este tipo de métodos se basan en la **propiedad de eliminación**.

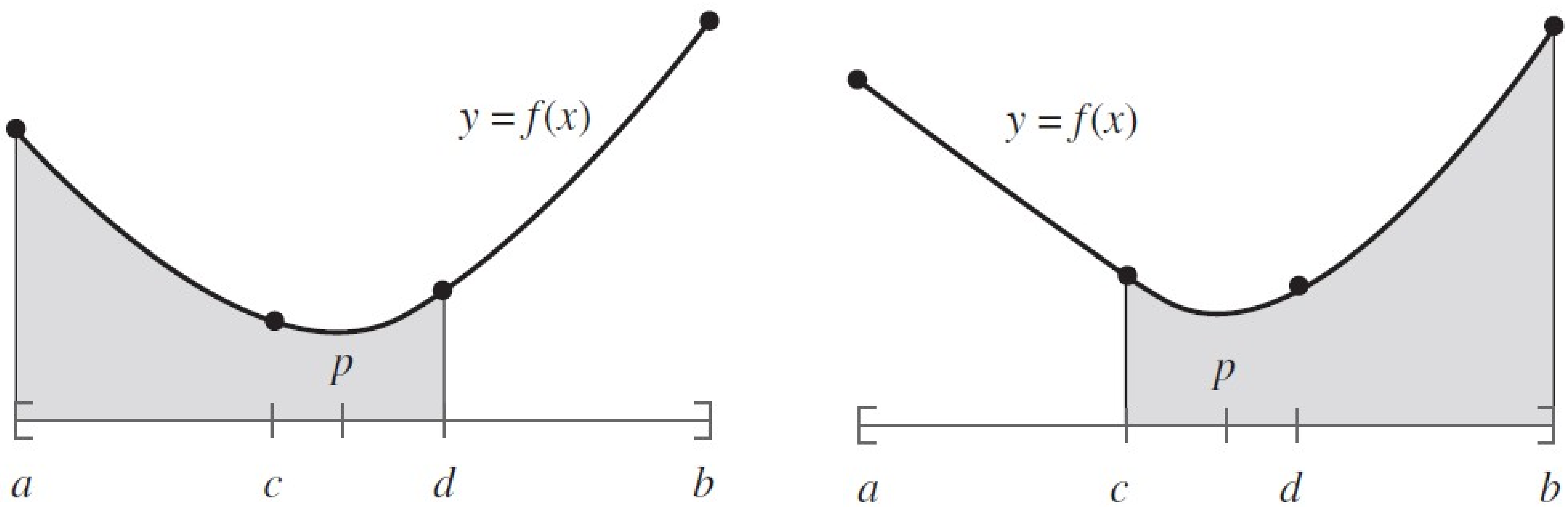
Supongamos que f es **estrictamente unimodal** (o sea, es unimodal y no tiene intervalos de longitud finita en los cuales la función sea de valor constante) en el intervalo $a \leqslant x \leqslant b$ con un mínimo en x*. Hagamos que c y d sean 2 puntos en el intervalo, de forma que $a < c < d < b$.

Eliminación de Regiones

Tenemos entonces que:

1. Si $f(d) > f(c)$, entonces el mínimo de $f(x)$ no se encuentra en el intervalo $[d, b]$. En otras palabras, $x^* \in [a, d]$.

2. Si $f(c) > f(d)$, entonces el mínimo de $f(x)$ no se encuentra en el intervalo $[a, c]$ En otras palabras, $x^* \in (c, b)$.



3. Cuando $f(c) = f(d)$ podemos eliminar los 2 extremos y $x^* \in (c, d)$.

Eliminación de Regiones

Usando la propiedad de eliminación, puede organizarse la búsqueda de manera que se encuentre el optimo eliminando recursivamente secciones del intervalo acotado originalmente. La **búsqueda termina** cuando el **subintervalo restante** es reducido a una longitud suficientemente **pequeña**.

Nótese que **la mayor ventaja de este tipo de métodos es que no requieren que la función sea diferenciable** (solo requieren evaluarla).

Métodos de minimización basados en técnica de eliminación de regiones.

1. Método de División de Intervalos por la Mitad
2. Método de la Sección Dorada
3. Método de Búsqueda de Fibonacci

Método de Búsqueda de Fibonacci.

La propiedad de los números de Fibonacci es que, dados dos números consecutivos F_{n-2} y F_{n-1}, el tercer numero se calcula usando:

$F_n = F_{n-1} + F_{n-2}$ donde $n = 2, 3, 4, ...$

Los primeros números de Fibonacci son:

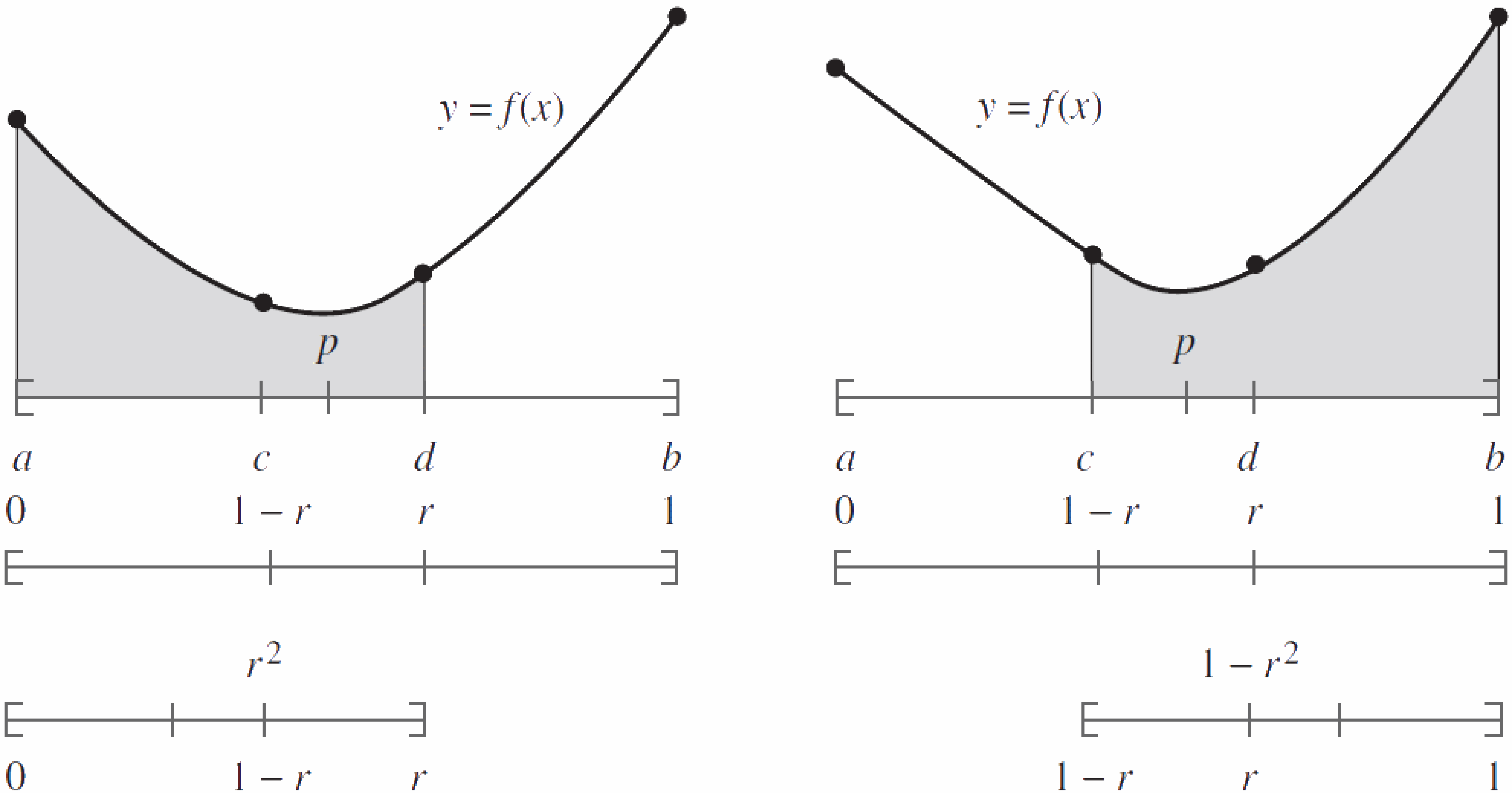
$F_0 = 0$, $F_1 = 1$, $F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$, $F_6 = 8$ ...

Los números de Fibonacci pueden usarse para crear un algoritmo de búsqueda que requiera **solo una evaluación** de la función objetivo **a cada iteración**.

El principio de la búsqueda de Fibonacci es que de dos puntos requeridos para el uso de la propiedad de eliminación de regiones, **uno es siempre el punto previo** y el otro punto es nuevo. Por ende, solo se requiere una evaluación de la función a cada iteración.

Método de la Sección Dorada

El principio del método de la Sección Dorada es que de dos puntos requeridos para el uso de la propiedad de eliminación de regiones, **uno es siempre el punto previo** y el otro punto es nuevo. Por ende, solo se requiere una evaluación de la función a cada iteración.



Método de la Sección Dorada

En el caso de eliminación de intervalo $[d, b]$ tenemos

$a_k = a_{k-1}$; $b_k = d_{k-1}$; $d_k = c_{k-1}$.

Para calcular nuevo c_k se usan $c_k = a_k + (1-r)(b_k - a_k)$.

En el caso de eliminación de intervalo $[a, c]$ tenemos

$a_k = c_{k-1}$; $b_k = b_{k-1}$; $c_k = d_{k-1}$.

Para calcular nuevo d_k se usan $d_k = b_k - (1-r)(b_k - a_k)$

En este algoritmo, el espacio de búsqueda (a, b) se mapea linealmente a un intervalo unitario [0, 1]. Posteriormente, dos puntos en r desde cualquier extremo del espacio de búsqueda se eligen en forma que a cada iteración la región eliminada sea de $(1-r)$ con respecto a la iteración previa. Esto se puede lograr igualando $\frac{r}{1} = \frac{1-r}{r}$ => $r^2 + r - 1 = 0$ este ecuación cuadrático tiene raíces $r = \frac{-1 \pm \sqrt{5}}{2}$. Con restricción $0.5 < r < 1$ tenemos

$r = (-1 + \sqrt{5})/2 \approx 0.618$. (La sol. de ecu. cuadrática es $x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$).

Método de la Sección Dorada. Ejemplo

Minimizar la función $f(x)=x^2-\sin(x)$ en intervalo $[0,1]$

1. Con un método de búsqueda de zeros de derivada.

$f'(x)=2x-\cos(x)=0$ => con un método de solución de los ecuaciones $f(x)=0$ podemos encontrar el mínimo de función. El mínimo de función va en el punto $x_{min}=0.4501836$.

$f''(0.4501836)=2.435131>0$ => el punto $x_{min}=0.4501836$ es mínimo con valor $f(0.4501836)=-0.2324656$ ..

2. Con sección dorada.

Usamos formulas

$c=a+(1-r)(b-a)$

$d=b-(1-r)(b-a)$

para los puntos c y d.

Calculamos $r=(-1+\sqrt{5})/2$.

Tenemos $a_0=0$, $b_0=1$. Calculamos c_0 y d_0

$c_0=a_0+(1-r)(b_0-a_0)\approx 0.38919660$ y $d_0=b_0-(1-r)(b_0-a_0)\approx 0.6180340$.

Método de la Sección Dorada. Ejemplo

Los valores de función en puntos c_0 y d_0 son

$f(c_0) = -0.22684748$ y $f(d_0) = -0.19746793$.

Como $f(c_0) < f(d_0)$, el nuevo intervalo es $[a_0, d_0] = [0.00000000, \;\; 0.6180340]$.
Tenemos $a_1 = a_0$, $b_1 = d_0$, $d_1 = c_0$

Calculamos c_1 . $c_1 = a_1 + (1-r)(b_1 - a_1) \approx 0.2360680$

Los valores de función en puntos c_1 y d_1 son

$f(c_1) = -0.178153$ y $f(d_1) = -0.226847$

Como $f(c_1) > f(d_1)$, el nuevo intervalo es $[c_1, b_1] = [0.2360680, \;\; 0.6180340]$.
Tenemos $a_2 = c_1$; $b_2 = b_1$; $c_2 = d_1$.

Calculamos d_2 . $d_2 = b_2 - (1-r)(b_2 - a_2) \approx 0.472136$

....

Continuamos el proceso hasta $|b_k - a_k| < tol$ o $|f(b_k) - f(a_k)| < tol$

Compare los resultados con la tabla en siguiente pagina.

Método de la Sección Dorada. Ejemplo

k	a_k	c_k	d_k	b_k	$f(c_k)$	$f(d_k)$
0	0.0000000	0.3819660	0.6180340	1	−0.22684748	−0.19746793
1	0.0000000	0.2360680	0.3819660	0.6180340	−0.17815339	−0.22684748
2	0.2360680	0.3819660	0.4721360	0.6180340	−0.22684748	−0.23187724
3	0.3819660	0.4721360	0.5278640	0.6180340	−0.23187724	−0.22504882
4	0.3819660	0.4376941	0.4721360	0.5278640	−0.23227594	−0.23187724
5	0.3819660	0.4164079	0.4376941	0.4721360	−0.23108238	−0.23227594
6	0.4164079	0.4376941	0.4508497	0.4721360	−0.23227594	−0.23246503
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	0.4501574	0.4501730	0.4501827	0.4501983	−0.23246558	−0.23246558
22	0.4501730	0.4501827	0.4501886	0.4501983	−0.23246558	−0.23246558
23	0.4501827	0.4501886	0.4501923	0.4501983	−0.23246558	−0.23246558

Método Nelder-Mead (Simplex)

El método Nelder-Mead es un método numérico para **minimizar** una función objetiva **en un espacio multidimensional**.

El método busca **de modo aproximado** una **solución óptima local** a un problema **con N variables** cuando **la función** a minimizar **varía suavemente**.

El método utiliza el concepto de un **simplex**, que es un **triangulo generalizado** de **N+1 vértices en N dimensiones**: un segmento de línea en un línea, un triángulo en un plano, un tetraedro en un espacio tridimensional y así sucesivamente.

Para dos variables el método compara valores de la función en los tres vértices de un triángulo. Lo peor de vértices, donde $f(x, y)$ es el más grande, es rechazada y sustituida por una nueva vértice. Un nuevo triángulo se forma y la búsqueda continúa. **El proceso genera una secuencia de triángulos** para lo cual la función los valores en los vértices se vuelven más y más pequeños. El tamaño de los triángulos se reduce y las coordenadas del punto mínimo se encuentran.

Método Nelder-Mead (Simplex)

Triángulo inicial BGW

Sea $f(x, y)$ la función que se quiere minimizar. Para empezar, se nos da tres vértices de un triángulo: $\boldsymbol{V_k}=(x_k, y_k)$, $k=1,2,3$. La función $f(x, y)$ es evaluado en cada de los tres puntos: $z_k=f(x_k, y_k)$ para $k=1,2,3$. Los subíndices luego se reordenan de manera que $z_1 \leq z_2 \leq z_3$. Usamos la notación $\boldsymbol{B}=(x_1, y_1)$, $\boldsymbol{G}=(x_2, y_2)$, y $\boldsymbol{W}=(x_3, y_3)$ para recordar que $\boldsymbol{B}$ es el mejor vértice, $\boldsymbol{G}$ es bueno, y $\boldsymbol{W}$ es el peor vértice.

Punto medio del lado bueno

El proceso de construcción de nuevo simplex utiliza el punto medio del segmento que une ***B*** y ***G***. Es que han encontrado un promedio de las coordenadas:

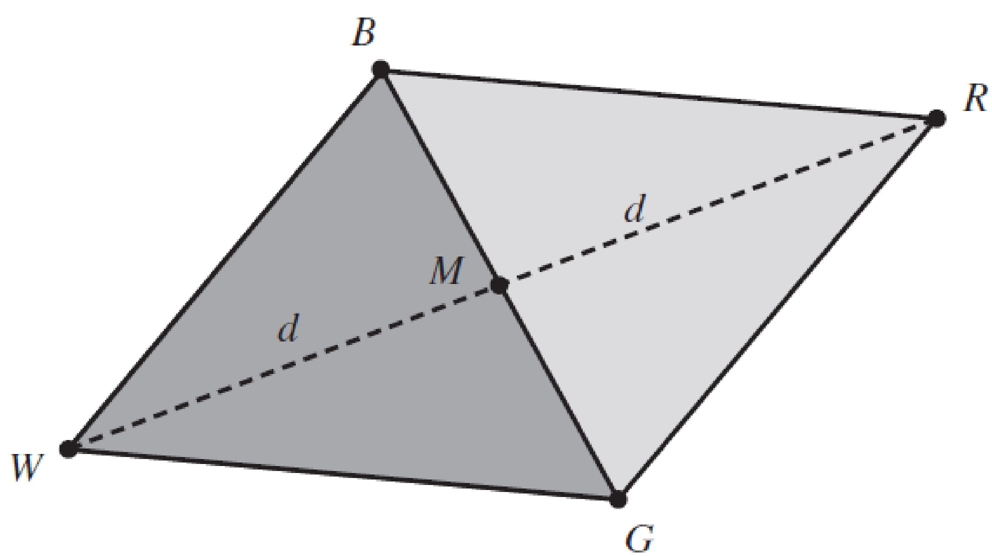
$$\boldsymbol{M}=\frac{\boldsymbol{B}+\boldsymbol{G}}{2}=\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$$

Método Nelder-Mead (Simplex)

Reflexión usando punto *R*

La función se disminuye a medida que nos movemos a lo largo del lado del triángulo de ***W*** a ***B***, y también a medida que nos movemos a lo largo de la parte de ***W*** a ***G***. Por lo tanto es factible que $f(x, y)$ toma valores más pequeños en los puntos que se encuentran lejos de ***W*** en el lado opuesto de la línea entre ***B*** y ***G***. Elegimos un punto ***R*** de prueba que se obtiene "que refleja" el triángulo a través de la lado $\overline{BG}$. Para determinar ***R***, primero encontrar el punto medio ***M*** de la lado $\overline{BG}$.

A continuación, dibuje el segmento de línea de ***W*** a ***M*** y llame a su longitud *d*. Este último se extendió una distancia *d* a ***M*** para localizar el punto ***R*.** La formula par el vector R es $\boldsymbol{R} = \boldsymbol{M} + (\boldsymbol{M} - \boldsymbol{W}) = 2\boldsymbol{M} - \boldsymbol{W}$.

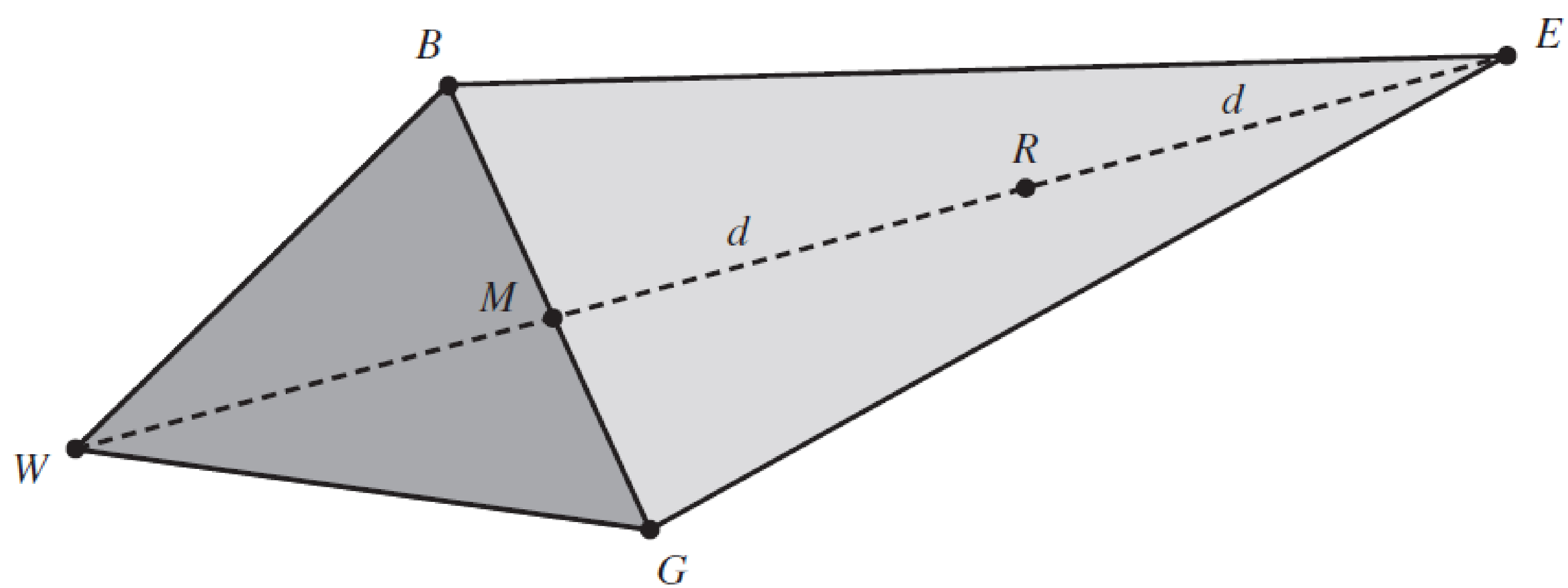


Método Nelder-Mead (Simplex)

Expansión usando el punto *E*

Si el valor de la función en ***R*** es menor que el valor de la función en ***W***, entonces nos hemos **movido en la dirección correcta** a el mínimo. Tal vez el mínimo es un poco más lejos que el punto ***R***. Por lo tanto, ampliar el segmento de recta que pasa por ***M*** y ***R*** hasta el punto ***E***.

Esto forma un triángulo ***BGE*** ampliado. El punto ***E*** se encuentra moviendo un adicional distancia *d* en dirección de $\overline{\boldsymbol{MR}}$. Si el valor de la función en ***E*** es menor que el valor de la función en ***R***, entonces encontramos un vértice mejor que el ***R***. Fórmula para vector de ***E*** es $\boldsymbol{E} = \boldsymbol{R} + (\boldsymbol{R} - \boldsymbol{M}) = 2\boldsymbol{R} - \boldsymbol{M}$.

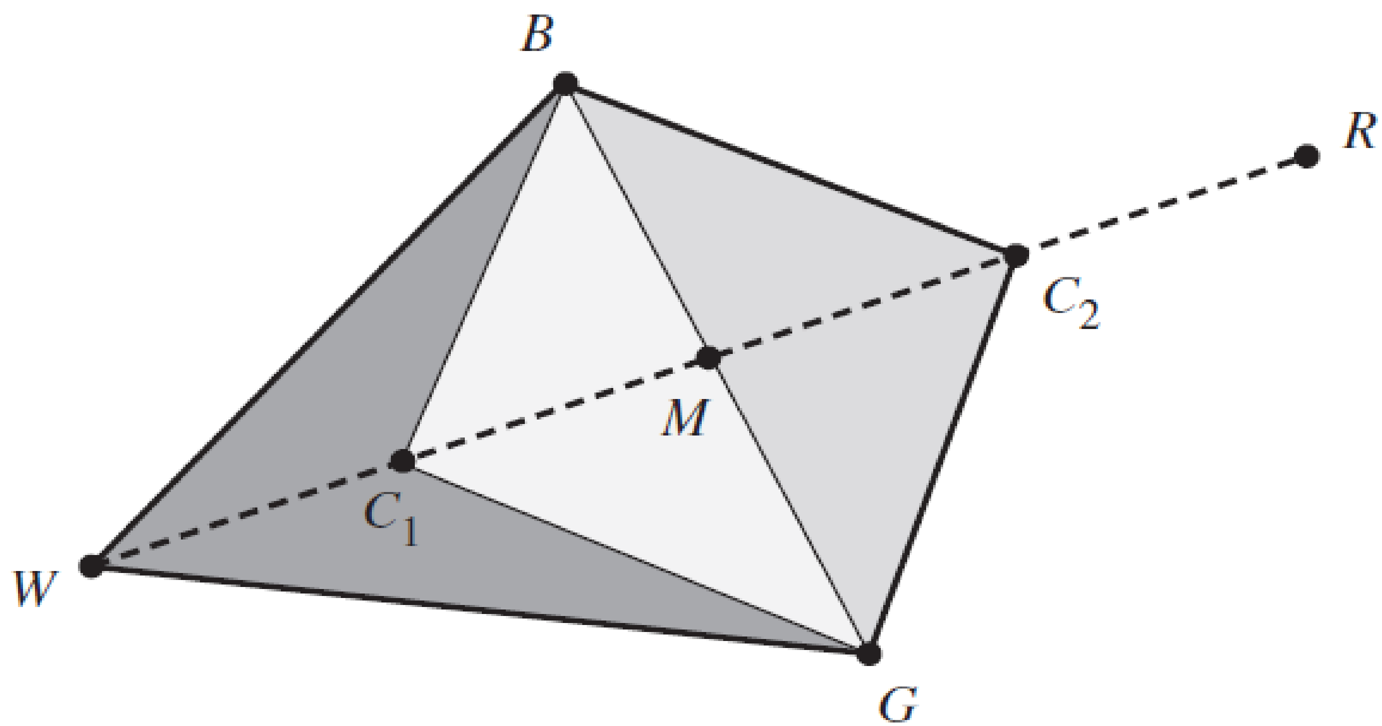


Método Nelder-Mead (Simplex)

La contracción usando el punto *C*

Si los valores de la función de ***R*** y ***W*** son los mismos, otro punto debe ser probado. Tal vez la función es más pequeño en el ***M***, pero **no puede sustituir a *W* con *M* porque debemos tener un triángulo**. Considere los dos puntos medios C_1 y C_2 de los segmentos de línea $\overline{WM}$ y $\overline{RM}$, respectivamente. $C_1=(W+M)/2$, $C_2=(M+R)/2$

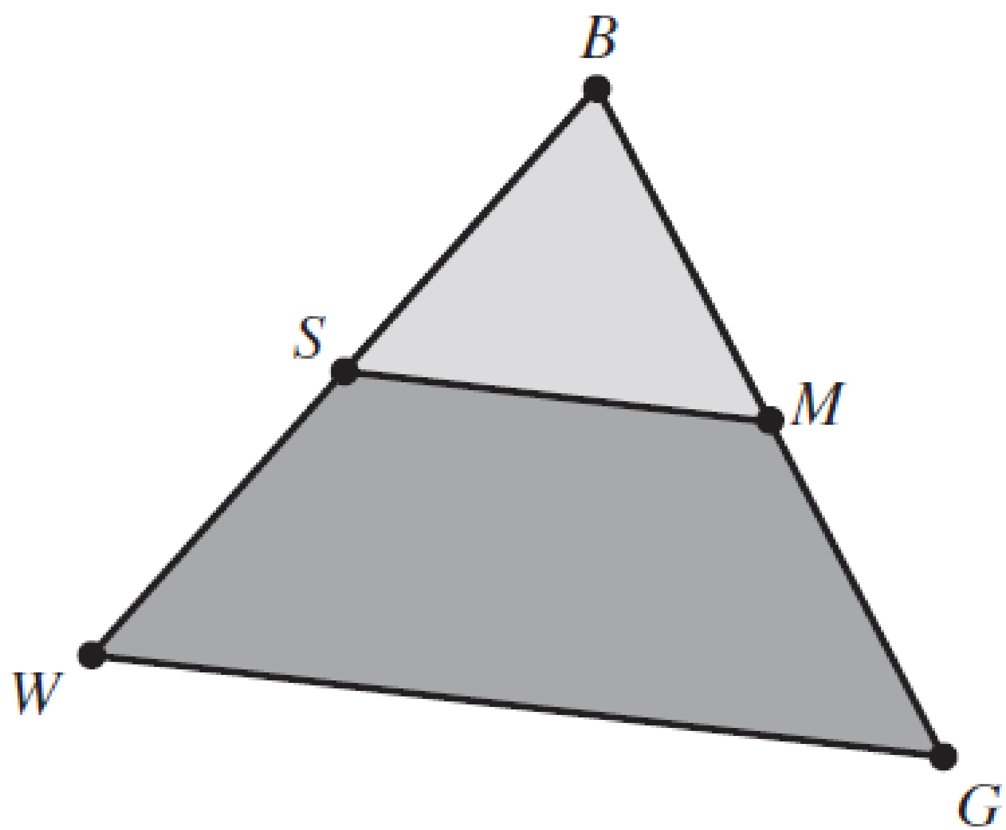
El punto con el valor de la función **más pequeño se llama *C***, y el nuevo triángulo es ***BGC***.



Método Nelder-Mead (Simplex)

Reducción hacia *B*

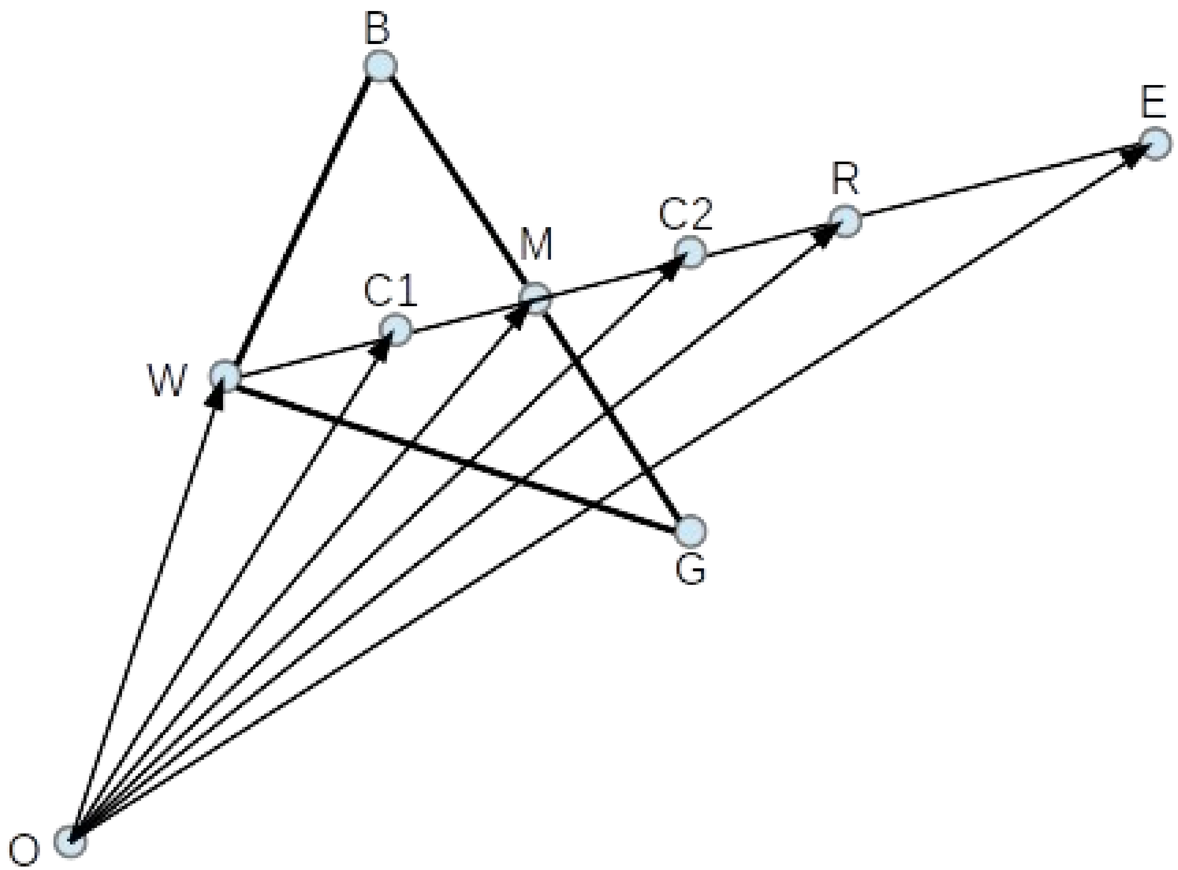
Si el valor de la función en ***C*** no es menor que el valor en el ***W***, los puntos ***G*** y ***W*** deben ser moviendo hacia ***B***. El punto ***G*** se sustituirá por ***M*** y ***W*** se sustituye con ***S***, que es el punto medio del segmento que une ***B*** con ***W***.



$$\textit{if } f(C) \geq f(W) \textit{ then}$$

$$S = \left(\frac{x_W + x_B}{2}, \frac{y_W + y_B}{2} \right)$$

Método Nelder-Mead (Simplex)



$$\boldsymbol{M} = \frac{\boldsymbol{B}+\boldsymbol{G}}{2} = \left(\frac{x_B + x_G}{2}, \frac{y_B + y_G}{2} \right)$$

$$OW + WM = OM$$

$$OM + WM = OR$$

$$\boxed{\Rightarrow OR = 2\,OM - OW}$$

Método Nelder-Mead (Simplex)

$$OM + WM = OR$$

$$OR + WM = OE$$

$$\boxed{\Rightarrow OE = 2\,OR - OM}$$

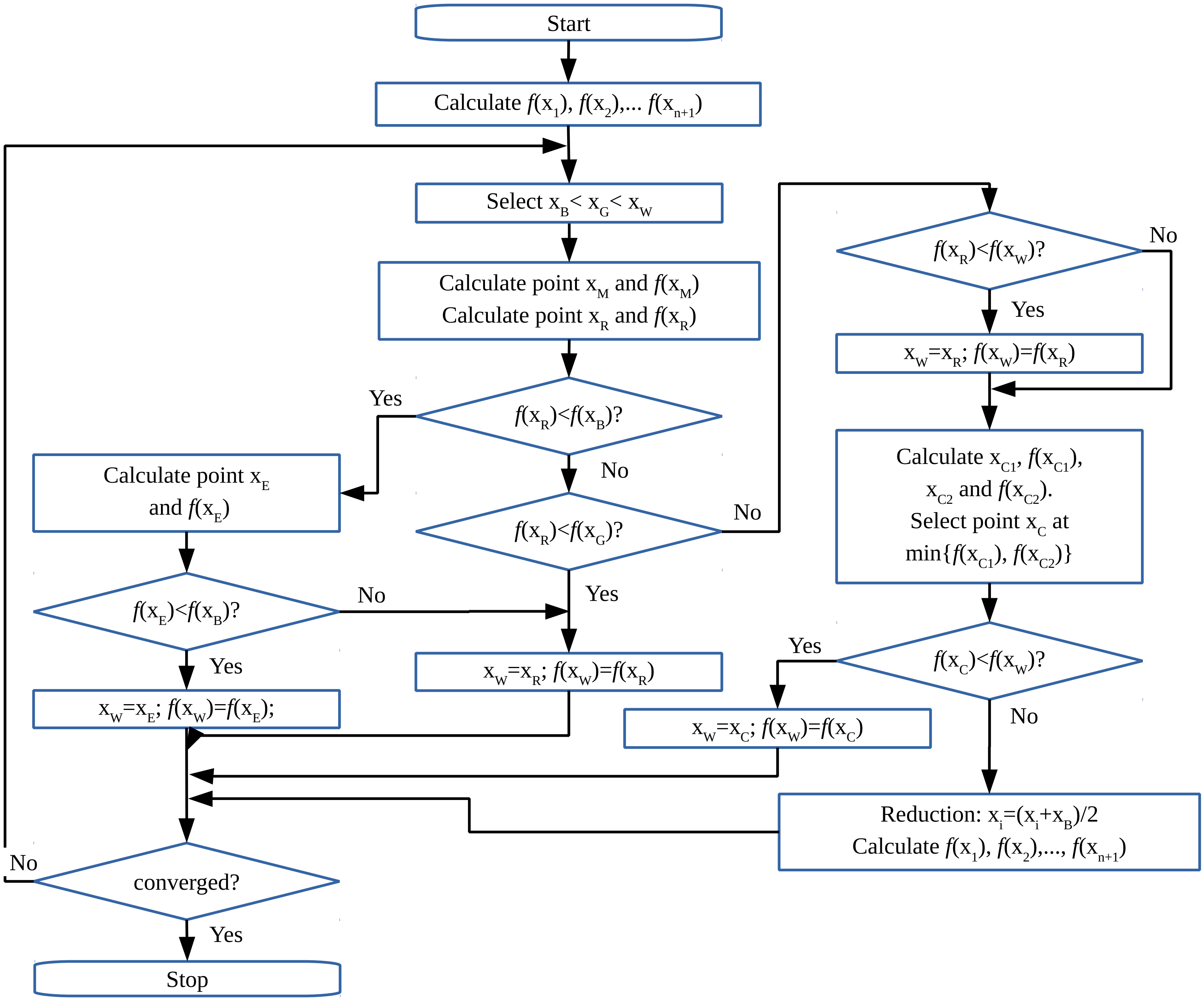
$$\boxed{OC_1 = \frac{OW + OM}{2}}$$

$$\boxed{OC_2 = \frac{OR + OM}{2}}$$

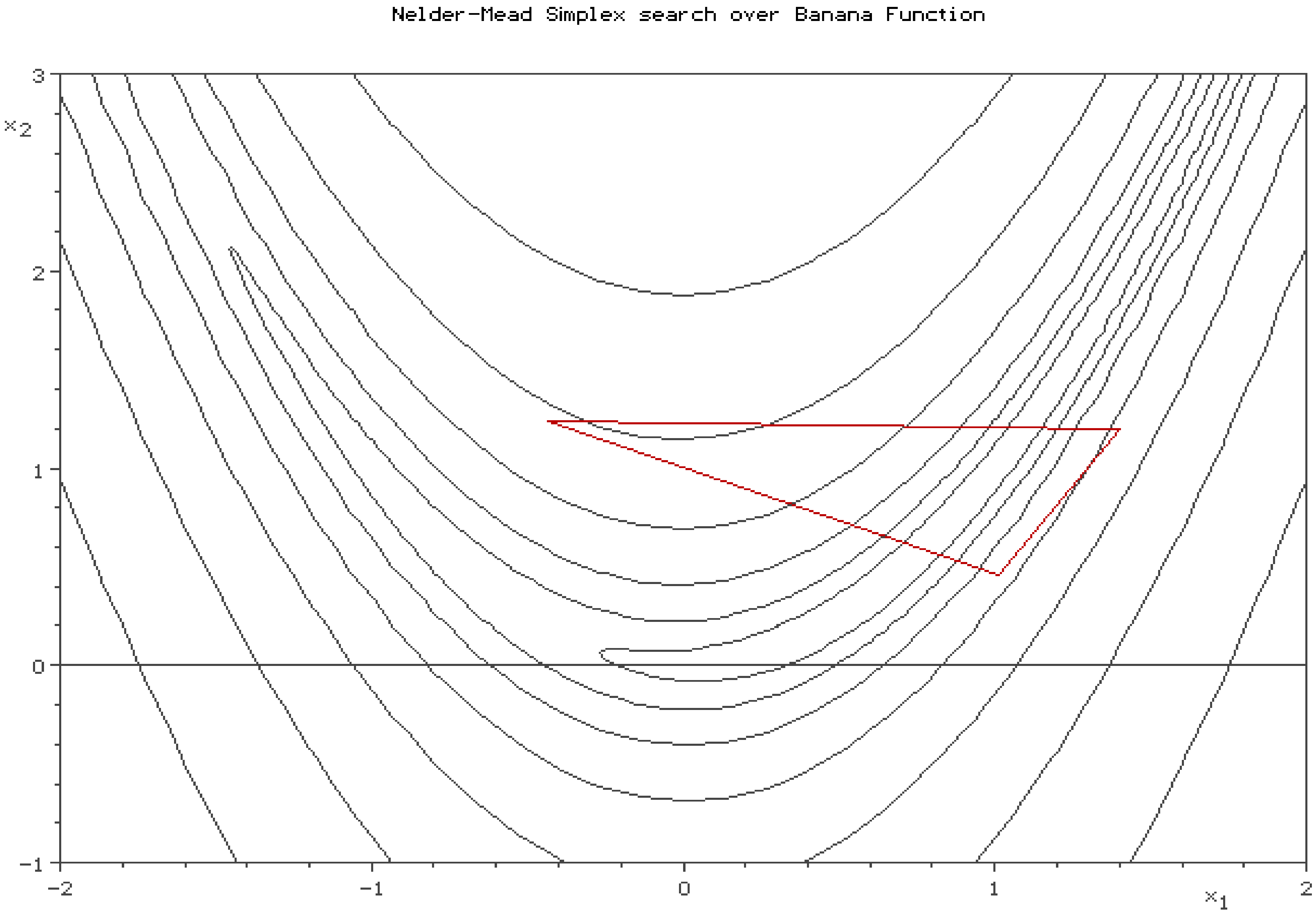
$$C = min\left(C_1, C_2\right)$$

$$S = \left(\frac{x_G + x_B}{2}, \frac{y_G + y_B}{2}\right)$$

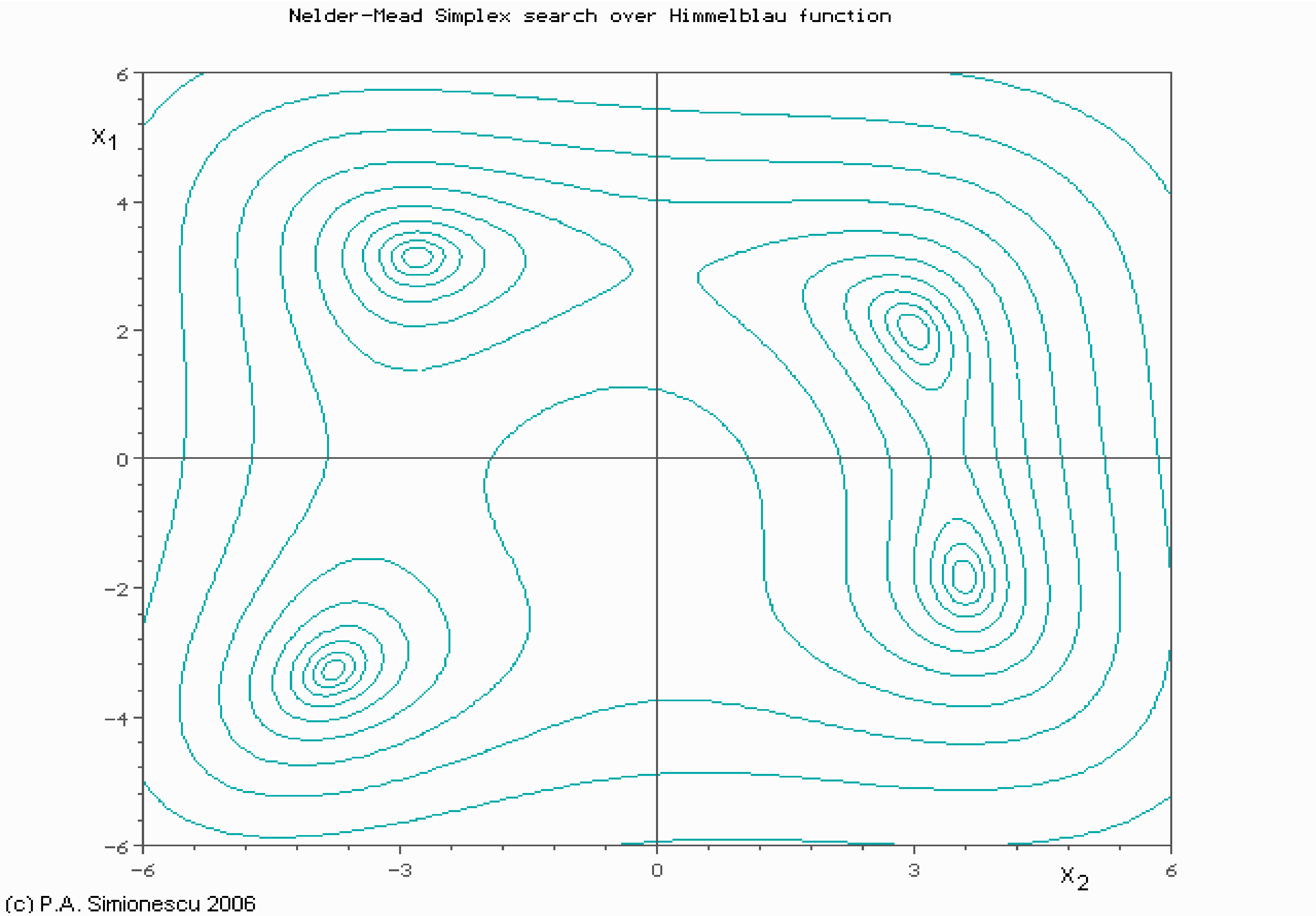
Método Nelder-Mead (Simplex). Algoritmo



Método Nelder-Mead (Simplex). Demos



Método Nelder-Mead (Simplex). Demos



Método Nelder-Mead (Simplex). Ejemplo

Utilice el algoritmo de Nelder-Mead para encontrar el mínimo de $f(x,y)=x^2-4x+y^2-y-xy$. Comience con los tres vértices

$V_1=(0,0)$, $V_2=(1.2,0.0)$, $V_3=(0.0,0.8)$.

La función $f(x,y)$ toma los valores

$f(0,0)=0.0$, $f(1.2,0.0)=-3.36$, $f(0.0,0.8)=-0.16$.

Con ayuda de Scilab calcula 10 primeros pasos. Los resultados de los cálculos escribe a la tabla en la forma

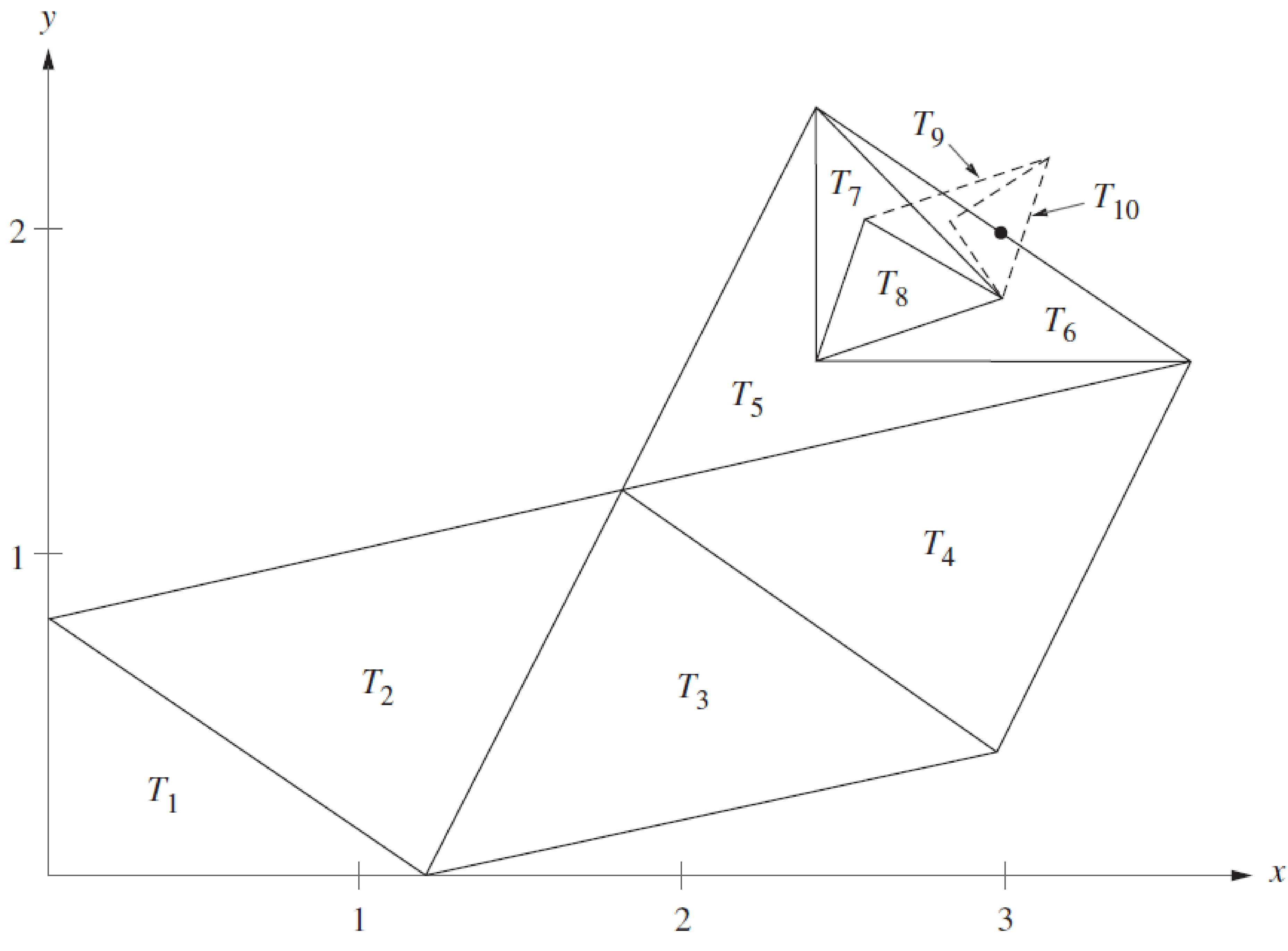
k	Best point	Good point	Worst point

Compare los resultados recibidos con los table de próxima pagina.

Método Nelder-Mead (Simplex). Ejemplo

k	Best point	Good point	Worst point
1	$f(1.2, 0.0) = -3.36$	$f(0.0, 0.8) = -0.16$	$f(0.0, 0.0) = 0.00$
2	$f(1.8, 1.2) = -5.88$	$f(1.2, 0.0) = -3.36$	$f(0.0, 0.8) = -0.16$
3	$f(1.8, 1.2) = -5.88$	$f(3.0, 0.4) = -4.44$	$f(1.2, 0.0) = -3.36$
4	$f(3.6, 1.6) = -6.24$	$f(I.8, 1.2) = -5.88$	$f(3.0, 0.4) = -4.44$
5	$f(3.6, 1.6) = -6.24$	$f(2.4, 2.4) = -6.24$	$f(1.8, 1.2) = -5.88$
6	$f(2.4, 1.6) = -6.72$	$f(3.6, 1.6) = -6.24$	$f(2.4, 2.4) = -6.24$
7	$f(3.0, 1.8) = -6.96$	$f(2.4, 1.6) = -6.72$	$f(2.4, 2.4) = -6.24$
8	$f(3.0, 1.8) = -6.96$	$f(2.55, 2.05) = -6.7725$	$f(2.4, 1.6) = -6.72$
9	$f(3.0, 1.8) = -6.96$	$f(3.15, 2.25) = -6.9525$	$f(2.55, 2.05) = -6.7725$
10	$f(3.0, 1.8) = -6.96$	$f(2.8125, 2.0375) = -6.95640625$	$f(3.15, 2.25) = -6.9525$

Método Nelder-Mead (Simplex). Ejemplo



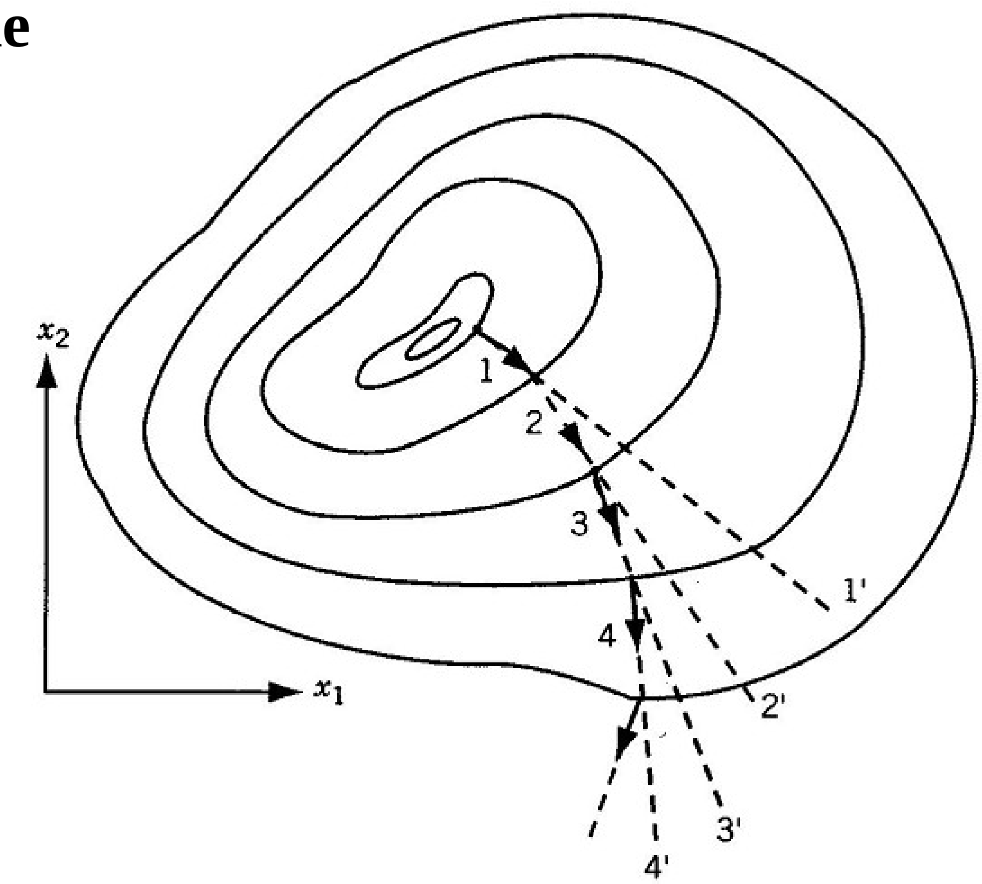
Método del Gradiente (Método del máximo descenso)

Sea $z = f(\boldsymbol{X})$ una función escalar de un vector $\boldsymbol{X}$ tal que $\partial f(\boldsymbol{X}) / \partial x_k$ **existe** para $k = 1, 2, \ldots, N$. El **gradiente** de f, denotado por $\nabla f(\boldsymbol{X})$**, es el vector**

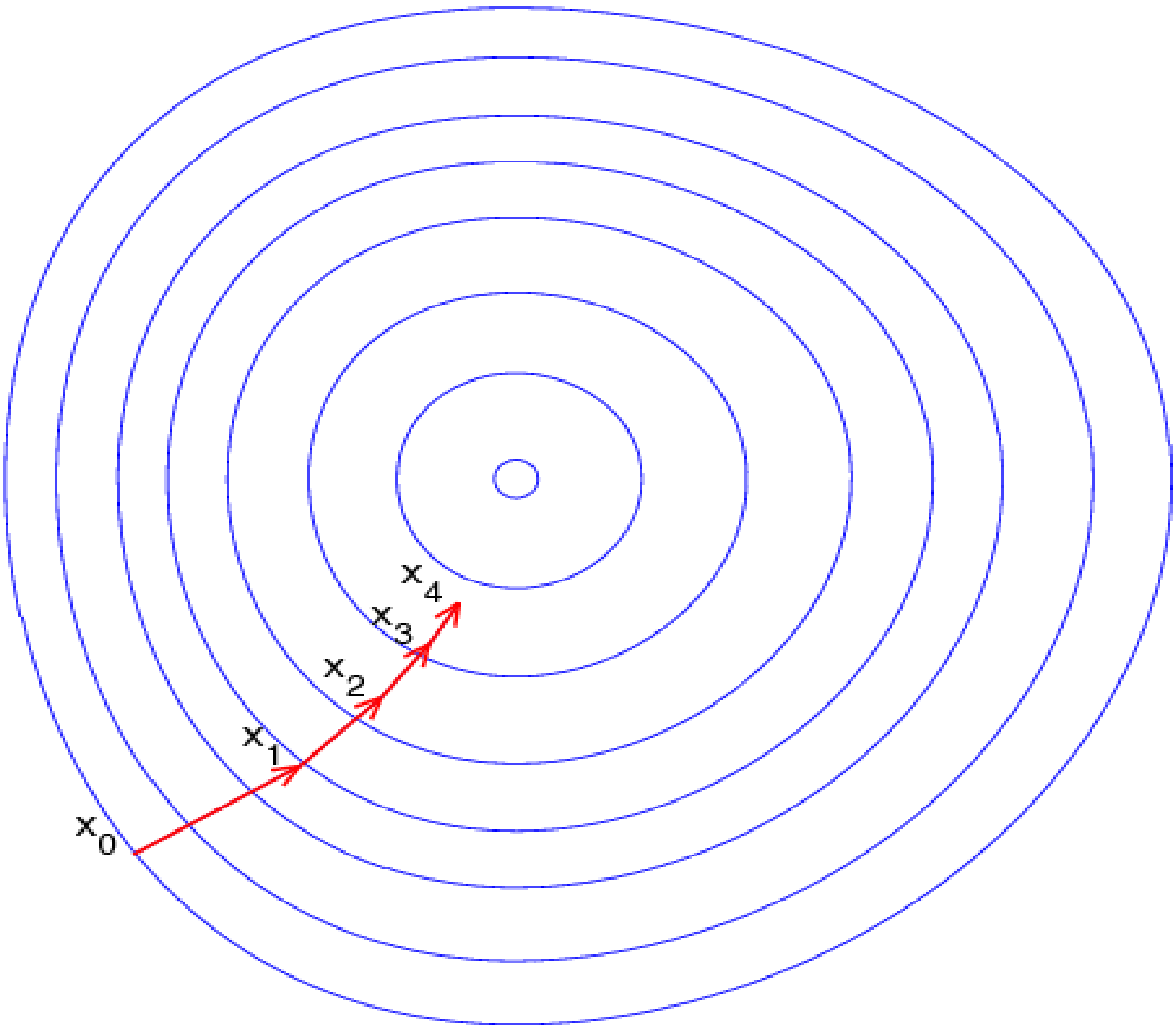
$$\nabla f(\boldsymbol{X}) = \left(\frac{\partial f(\boldsymbol{X})}{\partial x_1}, \frac{\partial f(\boldsymbol{X})}{\partial x_2}, \ldots, \frac{\partial f(\boldsymbol{X})}{\partial x_N} \right)$$

Recordemos que el **gradiente** indica a nivel local en la **dirección de tasa de crecimiento de** $f(\boldsymbol{X})$ **más grande**. Por lo tanto $-\nabla f(\boldsymbol{X})$ indica en la dirección de **mayor disminución de función**.

La dirección del **máximo descenso es una propiedad local y no una global**. Esto se ilustra con la figura siguiente:



Método del Gradiente



Método del Gradiente

Por esto, después de computación de la dirección de **mayor disminución de función** en un punto inicial $\boldsymbol{P_0}$

$$\boldsymbol{S_0} = \frac{-\nabla f(\boldsymbol{P_0})}{\|\nabla f(\boldsymbol{P_0})\|}$$

buscamos **el paso** h **en la dirección** $\boldsymbol{S_0}$ que nos da **nuevo punto.**

$$\boldsymbol{P_1} = \boldsymbol{P_0} + h_0 \boldsymbol{S_0}$$

El paso h calculamos en tal manera que función $\Phi(h) = f(\boldsymbol{P_0} + h\boldsymbol{S_0})$ tiene su mínimo local. En cuanto $\boldsymbol{P_0}$ y $\boldsymbol{S_0}$ son fijos **la función** $\Phi(h)$ **es una función de una solo variable** (es posible aplicar un método de minimización las funciones de una variable, por ejemplo el método de la sección dorada). El resultado de este minimización de $\Phi(h) = f(\boldsymbol{P_0} + h\boldsymbol{S_0})$ nos da nuestro paso $h_0 = h_{min}$. Calculamos nuevo punto $\boldsymbol{P_1} = \boldsymbol{P_0} + h_{min}\boldsymbol{S_0}$

A continuación calculamos $-\nabla f(\boldsymbol{P_1})$ y se mueven en la dirección de búsqueda $\boldsymbol{S_1} = \frac{-\nabla f(\boldsymbol{P_1})}{\|\nabla f(\boldsymbol{P_1})\|}$ con nuevo paso h_1 . Siguiente aproximación es $\boldsymbol{P_2} = \boldsymbol{P_1} + h_2 \boldsymbol{S_1}$. Iteraciónes producen una secuencia, $\{\boldsymbol{P_k}\}_{k=0}^{\infty}$ de puntos con la propiedad $f(\boldsymbol{P_0}) > f(\boldsymbol{P_1}) > \ldots > f(\boldsymbol{P_k}) > \ldots$.

Método del Gradiente

Algoritmo

Supongamos que tenemos $z = f(\boldsymbol{X})$, $\boldsymbol{P_k}$ y *tol*.

1. Evaluar el gradiente de $\nabla f(\boldsymbol{P_k})$.

2. Calcule la dirección de máximo descenso $\boldsymbol{S_k} = \dfrac{-\nabla f(\boldsymbol{P_k})}{\|\nabla f(\boldsymbol{P_k})\|}$.

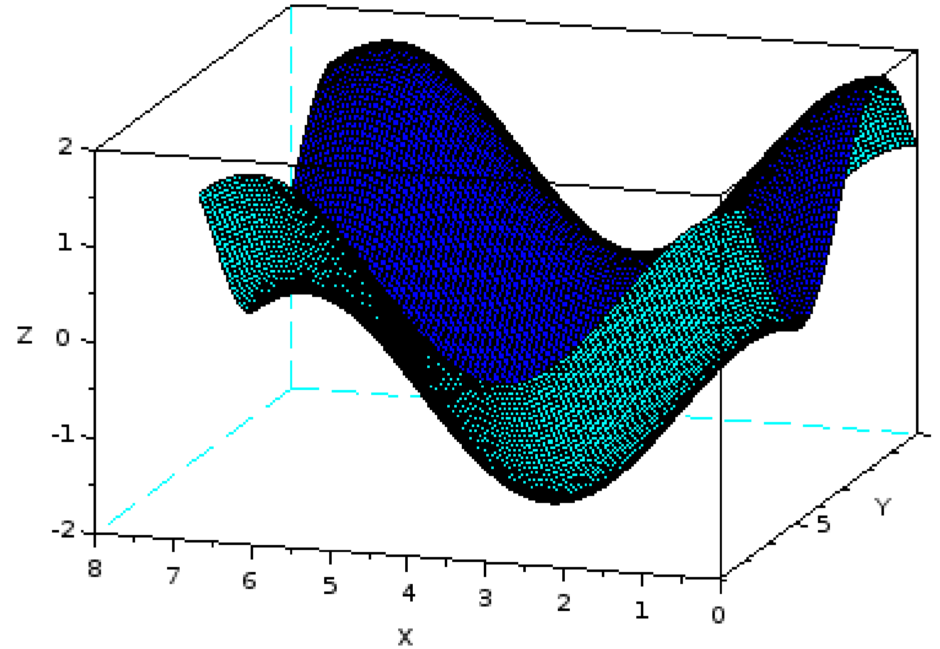
3. La búsqueda del mínimo de función $\Phi(h) = f(\boldsymbol{P_k} + h_k \boldsymbol{S_k})$ de una solo variable *h*. Esto producirá un valor de $h_k = h_{min}$ donde se ocurre un mínimo local de $\Phi(h)$. La relación $\Phi(h_{min}) = f(\boldsymbol{P_k} + h_{min} \boldsymbol{S_k})$ muestra que se trata de un mínimo de $f(\boldsymbol{X})$ a lo largo de la línea de búsqueda $\boldsymbol{X} = \boldsymbol{P_k} + h_{min} \boldsymbol{S_k}$.

4. Calcular el siguiente punto $\boldsymbol{P_{k+1}} = \boldsymbol{P_k} + h_{min} \boldsymbol{S_k}$

5. Calcular el error de convergencia $Err = max[|f(\boldsymbol{P_k}) - f(\boldsymbol{P_{k+1}})|, \|\boldsymbol{P_{k+1}} - \boldsymbol{P_k}\|]$

6. Si $Err > tol$ repetir pasos 1-6. Realice la prueba de la terminación de minimización, es decir, termina el proceso cuando los valores de la función $f(\boldsymbol{P_k})$ y $f(\boldsymbol{P_{k+1}})$ son suficientemente cerca y la distancia $\|\boldsymbol{P_{k+1}} - \boldsymbol{P_k}\|$ es suficiente pequeña.

Método del Gradiente. Ejemplo



Minimizar la función $f(x,y)=\cos(x)+\sin(y)$ con el método del Gradiente.

El punto inicial es $\boldsymbol{P_0}=[5,3]$.

El gradiente de función es

$$\nabla f(\boldsymbol{x})=\left[\frac{\partial f(x,y)}{\partial x},\frac{\partial f(x,y)}{\partial y}\right] .$$

Donde derivadas parciales de función son

$$\frac{\partial f(x,y)}{\partial x}=-\sin(x) \quad \text{y} \quad \frac{\partial f(x,y)}{\partial y}=\cos(x) .$$

Para el punto inicial $\boldsymbol{P_0}=[5,3]$ tenemos

$$\nabla f(\boldsymbol{P_0})=[0.958924,-0.989993] , \quad \|\nabla f(\boldsymbol{P_0})\|=1.378267$$

y $$\boldsymbol{S_0}=\frac{-\nabla f(\boldsymbol{P_0})}{\|\nabla f(\boldsymbol{P_0})\|}=[-0.695746,\ 0.718288]$$

Método del Gradiente. Ejemplo

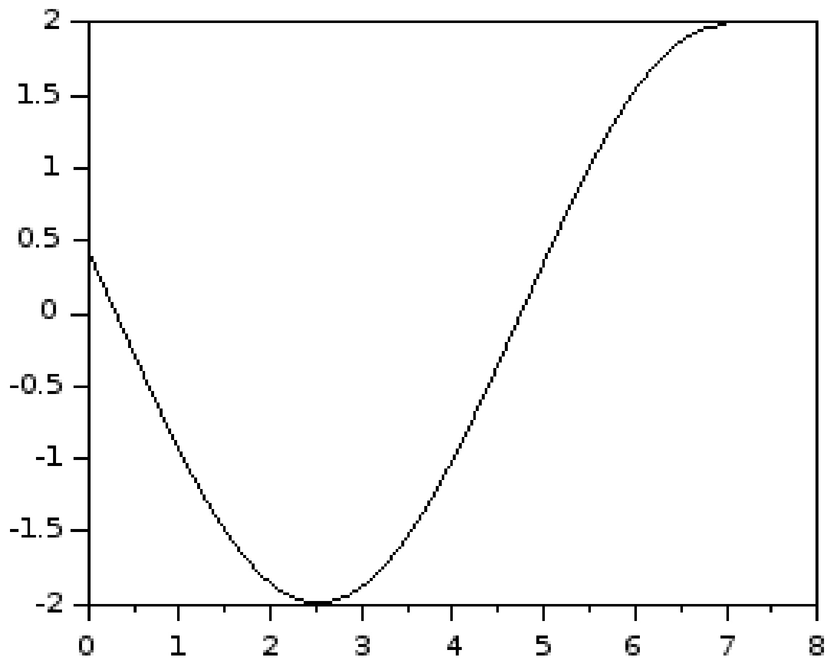
La función de una solo variable

$$\Phi(h)=f(\boldsymbol{P_0}+h\boldsymbol{S_0})=f((5,3)+h(-0.695746,\ 0.718288))=$$

$$=f(5-0.695746h,\ 3+0.718288h)$$

$$=>\ \ \Phi(h)=\cos(x)+\sin(y)=\cos(5-0.695746h)+\sin(3+0.718288h)$$

tiene un mínimo en el punto $h_1=h_{min}=2.522798$. (Recibido con el uso del programa de sección dorada).



Gráfica de la función $\Phi(k)$

Método del Gradiente. Ejemplo

Nuevo punto es $\boldsymbol{P_1}=\boldsymbol{P_0}+h\boldsymbol{S_0}=(5,3)+2.522798(-0.695746,\ 0.718288)$

$\boldsymbol{P_1}=[3.244773,\ 4.812095]$.

Siguientes iteraciónes nos dan

$\nabla f(\boldsymbol{P_1})=[0.102997,\ 0.099541]$, $\|\nabla f(\boldsymbol{P_1})\|=0.143237$,

$\boldsymbol{S_1}=[0.9037378,\ 0.4280864]$

La función

$$\Phi(h)=f(\boldsymbol{P_1}+h\boldsymbol{S_1})=f((3.244773,\ 4.812095)+h(-0.719069,\ -0.694939))$$

$$=f(3.244773-0.719069h,\ 4.812095-0.694939h)$$

$$\Phi(h)=\cos(x)+\sin(y)=\cos(3.244773-0.719069h)+\sin(4.812095-0.694939h)$$

tiene un mínimo en punto $h_2=0.143409$

Con ayuda de Scilab calcula primeros iteraciónes y compare los resultados con la tabla de próxima pagina.

Método del Gradiente. Ejemplo

iter	x_k	y_k	$f(x, y)$	h_{opt}
1	5.000000	3.000000	0.4247822	2.522798
2	3.244773	4.812095	-1.989715	0.143409
3	3.141652	4.712434	-2.000000	0.000463

Método de Newton

Para la minimizacíon de las funciones diferenciábles se puede utilizar el método de Newton. En contraste a búsqueda de los ceros de la función, aquí nos buscamos a cero de la primera derivada de la función. Obtenemos la fórmula iterativa para minimizacíon de la función de una variable mediante la diferenciación de la serie de Taylor.

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots$$

Como $f(x_0)$, $f'(x_0)$, y $f''(x_0)$ son constantes entonces

$\frac{d}{dx}f(x_0)=\frac{d}{dx}f'(x_0)=\frac{d}{dx}f''(x_0)=0$. Por esto tenemos

$$f'(x)=f'(x_0)+\frac{f''(x_0)}{2}2(x-x_0)+\ldots$$

Usamos solo primeros dos términos y recordamos que en puntos extremos (mínimos o máximos) primera derivada es igual a cero.

$$0=f'(x_0)+f''(x_0)(x-x_0) \quad => \boxed{x_{k+1}=x_k-\frac{f'(x_k)}{f''(x_k)}}$$

Método de Newton. Caso multidimencional.

En caso de las funciones multivariables, el método de Newton se basa en un desarrollo en serie de Taylor multidimensional de la función alrededor del punto $\boldsymbol{x_0}$.

$$f(\boldsymbol{x}) = f(\boldsymbol{x_0}) + \nabla f(\boldsymbol{x_0})(\boldsymbol{x}-\boldsymbol{x_0})^T + \frac{(\boldsymbol{x}-\boldsymbol{x_0})}{2!}\mathbf{H}(\boldsymbol{x}-\boldsymbol{x_0})^T + \ldots$$ donde

$\boldsymbol{x_0}$ = punto sobre el que se desarrolla la serie de Taylor

$\boldsymbol{x}$ = punto cercano $\boldsymbol{x_0}$

$\boldsymbol{x}^{\boldsymbol{T}}$ = vector transpuesto de $\boldsymbol{x}$ (en este caso se convierte el vector renglón en el vector columna)

$\mathbf{H}$ = matriz Hessiana con elementos $H_{mn} = \dfrac{\partial^2 f}{\partial x_m \partial x_n}$

Tomando el gradiente de serie de Taylor y igualando de los primeros términos a cero tenemos

$$\nabla f(\boldsymbol{x}) = \nabla f(\boldsymbol{x_0}) + (\boldsymbol{x}-\boldsymbol{x_0})\mathbf{H} = 0$$

Entonces la formula iterativa para minimización de función multidimensional es $\boxed{\boldsymbol{x_{k+1}} = \boldsymbol{x_k} - \nabla f(\boldsymbol{x_k})\mathbf{H}^{-1}}$

Método de Newton. Caso multidimencional.

En la práctica, en lugar de calcular la matriz inversa $\mathbf{H}^{-1}$, se resuelve un sistema de ecuaciones lineales

$$\mathbf{H}\,\boldsymbol{\delta}\,\boldsymbol{x} = -\nabla f(\boldsymbol{x}_{\boldsymbol{k}})$$

para $\boldsymbol{\delta}\,\boldsymbol{x} = (\boldsymbol{x}_{\boldsymbol{k+1}} - \boldsymbol{x}_{\boldsymbol{k}})^T$

Se calcula el nuevo punto

$$\boldsymbol{x}_{\boldsymbol{k+1}} = \boldsymbol{x}_{\boldsymbol{k}} + \boldsymbol{\delta}\,\boldsymbol{x}$$

Método de Newton. Ejemplo

Minimizar la función $f(x,y)=\dfrac{(x-y)}{(x^2+y^2+2)}$ con el método de Newton. El punto inicial es $\boldsymbol{P_0}=[-0.3,0.2]$.

El gradiente de función es $\nabla f(\boldsymbol{x})=[f_x, f_y]$, $f_x=\dfrac{\partial f(x,y)}{\partial x}$, $f_y=\dfrac{\partial f(x,y)}{\partial y}$

$$f_x=\frac{-x^2+2xy+y^2+2}{(x^2+y^2+2)^2} \quad f_y=\frac{-x^2-2xy+y^2-2}{(x^2+y^2+2)^2} \; .$$

La matriz Hessian es $\boldsymbol{H}=\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$, donde

$$f_{xx}=\frac{\partial^2 f(x,y)}{\partial x^2}=\frac{2(x^3-3x^2y-3xy^2+2y-6x+y^3)}{(x^2+y^2+2)^3} \; ,$$

$$f_{xy}=\frac{\partial^2 f(x,y)}{\partial x \partial y}=\frac{-2(y-x)(x^2+4xy+y^2+2)}{(x^2+y^2+2)^3}$$

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

$$f_{yy}=\frac{\partial^2 f(x,y)}{\partial y^2}=\frac{-2(x^3-3x^2y-3xy^2-6y+2x+y^3)}{(x^2+y^2+2)^3}$$

Método de Newton. Ejemplo

Para el punto inicial $\boldsymbol{P}_0 = [-0.3, 0.2]$ tenemos

$$\nabla f(\boldsymbol{P}_0) = [0.4033591, -0.4254006]$$

$$\boldsymbol{H}(\boldsymbol{P}_0) = \begin{bmatrix} 0.4476594 & -0.1955793 \\ -0.1955793 & 0.3801897 \end{bmatrix}$$

$$\boldsymbol{H}^{-1}(\boldsymbol{P}_0) = \begin{bmatrix} 2.8814429 & 1.4822882 \\ 1.4822882 & 3.3927931 \end{bmatrix}$$

$$\boldsymbol{P}_1 = \boldsymbol{P}_0 - \nabla f(\boldsymbol{P}_0)\mathbf{H}^{-1} =$$

$$= [-0.3, 0.2] - [0.4033591, -0.4254006] \begin{bmatrix} 2.8814429 & 1.4822882 \\ 1.4822882 & 3.3927931 \end{bmatrix} =$$

$$= [-0.8316899, 1.0454017]$$

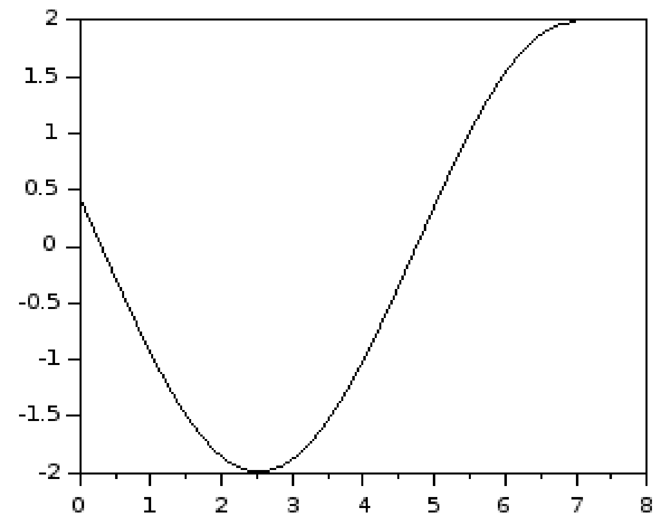
Continuamos con iteraciónes:

$$\boldsymbol{P}_2 = [-0.9880713, 0.9932699]$$

$$\boldsymbol{P}_3 = [-0.9998446, 0.9998919]$$

$$\boldsymbol{P}_4 = [-1.0000000, 1.0000000]$$

Optimización de funciones complejas.



Si está buscando un mínimo de funciones simples, puede usar **cualquiera de los métodos** que estudiamos anteriormente.

Sin embargo, estos métodos no funcionan si la función tiene más de un mínimo.

El proceso puede estar atascado en un mínimo local.

Para buscar un mínimo de funciones no unimodales, se utilizan algoritmos evolutivos, como algoritmos genéticos o evolución diferencial.

La búsqueda aleatoria

La **búsqueda aleatoria** es una familia de métodos de optimización numérica que no requieren el gradiente y, por lo tanto, se puede utilizar para las funciones que no son continuas o diferenciables. Dichos métodos de optimización también se conocen como métodos de búsqueda directa, búsqueda sin derivados o el metodo de caja negra.

Algoritmo

Sea f: $\mathbb{R}^n \rightarrow \mathbb{R}$ la función de aptitud o costo que debe minimizarse. Deje $x_k \in \mathbb{R}^n$ es una posición o solución candidata en el espacio de búsqueda. El algoritmo básico se puede describir como:

1. Inicialice x_k con una posición aleatoria en el espacio de búsqueda. $x_{min} \leq x_k \leq x_{max}$

2. Hasta que se cumpla un criterio de terminación (por ejemplo, el número de iteraciones o la aptitud adecuada alcanzada), repita lo siguiente:

3. Genera aleatoriamente una nueva posición x_{k+1} en el espacio de búsqueda.

4. Si $f(x_{k+1}) < f(x_k)$, muévase a la nueva posición estableciendo $x_k = x_{k+1}$

Algoritmos genéticos

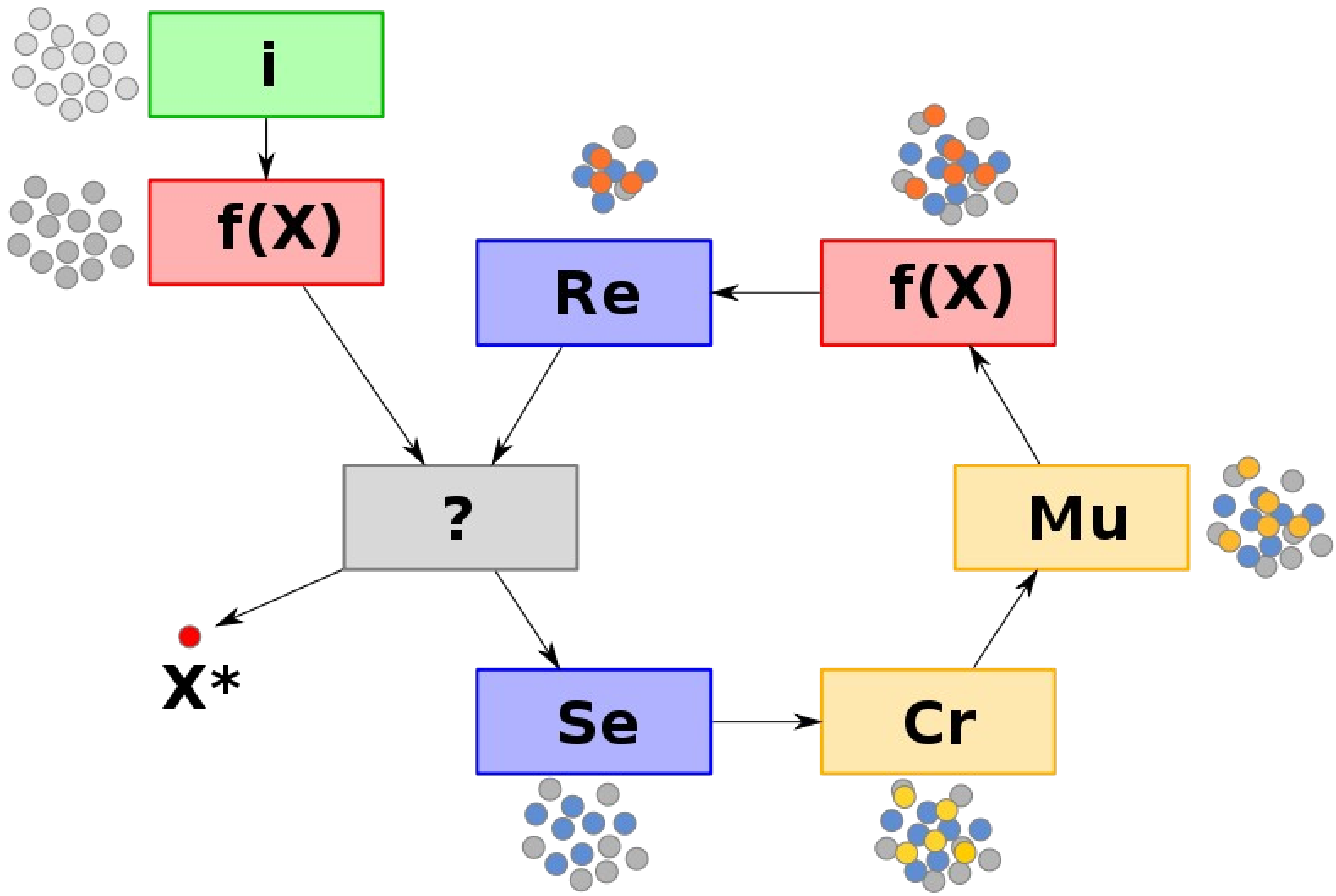
Un **algoritmo genético** (AG) es una **técnica de búsqueda/optimización** que **imita a la evolución biológica como estrategia para resolver problemas**.

Estos algoritmos hacen evolucionar una **población de individuos** sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (**mutaciones** y **recombinaciones** genéticas), así como también a una **selección** de acuerdo con algún criterio, en función del cual se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados.

Un algoritmo genético es un método de búsqueda dirigida basada en probabilidad. Bajo **una condición muy débil** (que el algoritmo mantenga **elitismo**, es decir, guarde siempre al mejor elemento de la población sin hacerle ningún cambio) se puede demostrar **que el algoritmo converge en probabilidad al óptimo**. En otras palabras, al aumentar el número de iteraciones, la probabilidad de tener el óptimo en la población tiende a 1 (uno).

Cuándo usar. Los algoritmos genéticos son de probada eficacia en caso de querer calcular funciones no derivables (o de derivación muy compleja) aunque su uso es posible con cualquier función.

Algoritmos genéticos



Algoritmo genético **i**: inicialización**, f(X)**: evaluación**, ?**: condición de término, **Se**: selección**, Cr**: cruzamiento**, Mu**: mutación**, Re**: reemplazo**, X***: mejor solución.

Algoritmos genéticos. Funcionamiento.

Inicialización: Se genera aleatoriamente la población inicial, que está constituida por un conjunto de cromosomas los cuales representan las posibles soluciones del problema.

Evaluación: A cada uno de los cromosomas de esta población se aplicará la función de aptitud para saber qué tan "buena" es la solución que se está codificando.

Condición de término. El AG se deberá detener cuando se alcance la solución óptima, pero ésta generalmente se desconoce, por lo que se deben utilizar otros criterios de detención. Normalmente se usan dos criterios: correr el AG un número máximo de iteraciones (generaciones) o detenerlo cuando no haya cambios en la población.

Selección. Después de saber la aptitud de cada cromosoma se procede a elegir los cromosomas que serán cruzados en la siguiente generación. Los cromosomas con mejor aptitud tienen mayor probabilidad de ser seleccionados.

Algoritmos genéticos. Funcionamiento.

Sobrecruzamiento. El cruzamiento es el principal operador genético, representa la reproducción sexual, opera sobre dos cromosomas a la vez para generar dos descendientes donde se combinan las características de ambos cromosomas padres.

Mutación. Modifica al azar parte del cromosoma de los individuos, y permite alcanzar zonas del espacio de búsqueda que no estaban cubiertas por los individuos de la población actual.

Reemplazo. Una vez aplicados los operadores genéticos, se seleccionan los mejores individuos para conformar la población de la generación siguiente

Preguntas de autoevaluacion

1. Definición de extremos locales
2. Prueba de la primera derivada.
3. Prueba de la segunda derivada.
4. Búsqueda numérica. Eliminación de Regiones
5. Método de la Sección Dorada. Ventaja.
6. Método Nelder-Mead (Simplex)
7. Método del Gradiente o del Descenso
8. Método de Newton
9. Algoritmos genéticos