

Método de Nelder-Mead

Por: Ricardo Rodríguez

Ian Castellanos

Método de Nelder-Mead

- Permite resolver problemas de optimización sin restricciones a partir de valores de la propia función, en vez de la derivada de la misma.
- Se usa la función objetivo no lineal $f: R^2 \rightarrow R$

$$\min f(x), x \in R^n$$

Donde n es el número de parámetros a optimizar.

- Tiene la ventaja respecto a otros métodos es que por lo regular solo requiere una o dos evaluaciones de la función para cada paso.

Estructura simplex

- El algoritmo hace uso de esta estructura que está compuesta por $n + 1$ vértices, $f_1, \dots, f_{n+1} \in R^n$ de manera que n es la dimensión de la función no degenerada.
- Esta función se irá modificando mediante una serie de transformaciones de modo que aumentan los valores obtenidos de la función objetivo.

Estructura simplex

- Ejemplo:
- $f \in R^1$: el simplex será un segmento que contiene 2 puntos.
- $f \in R^2$: El simplex será un triángulo.

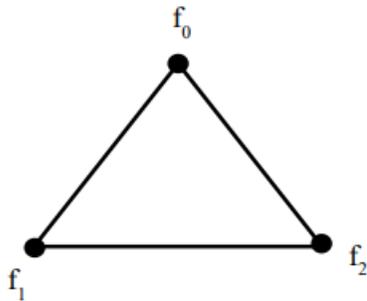


Figura 1. Simplex en dos dimensiones.

Estructura simplex

- $f \in R^3$: El simplex es un tetraedro.

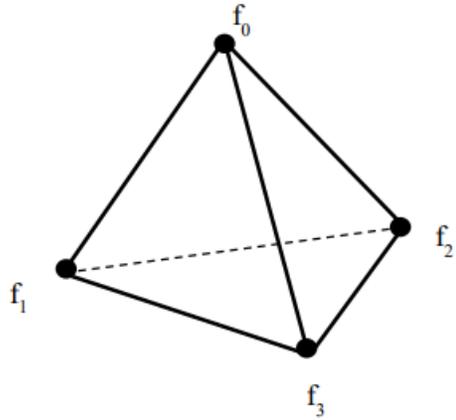


Figura 2. Simplex en 3D.

Estructura simplex

- Geométricamente, el objetivo del algoritmo es hacer que los puntos del simplex se encuentren lo suficientemente cercanos entre sí; cuando esto sucede, el algoritmo termina.
- Esto se traduce en que los valores de f_i hayan alcanzado una sensibilidad prefijada.
- El simplex inicial tiene todos los valores ubicados a la misma distancia entre sí.

Simplex inicial

Existen dos maneras de elegir el primer valor x_0 y los n-puntos:

- Se elige un valor aleatorio dentro del espacio R^n
- Se inicia a partir del final de un algoritmo previo que ha permitido acercarse a la optimización sin llegar a la misma.
- Tras elegir el valor x_0 el resto de valores se calculan como sigue:

$$x_i = x_1 + h_i e_i$$

- Donde h_i es el tamaño de paso en la dirección del vector $e_i \in R^n$

Simplex inicial

- Los vértices se ordenan en valores crecientes de manera que el mejor valor f_1 es el mejor vértice y el vértice f_{n+1} es el peor de los vértices del simplex. Este en particular es el caso para la minimización.

$$f_1 \leq f_2 \leq \dots \leq f_n \leq f_{n+1}$$

Centroide

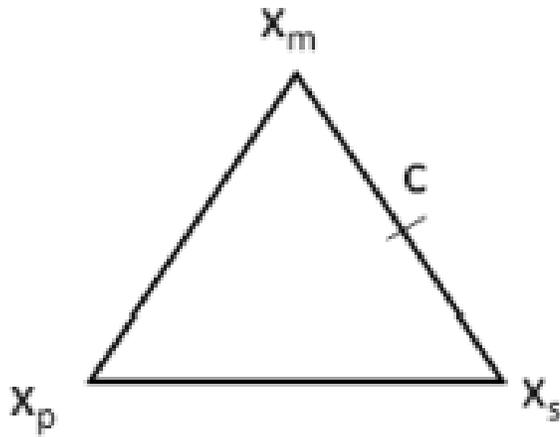


Figura 3. Centroide

- Este cálculo en particular es importante en las ecuaciones de actualización del método.

$$c = \frac{1}{n} \sum_{i \neq p}^n x_i$$

Transformaciones aplicadas

Se aplican las siguientes transformaciones, en el orden que son listadas, con el propósito de ver si es posible mejorar el valor de la función objetivo; estas transformaciones dependen de los parámetros escritos al lado:

- Reflexión: $\rho > 0$. Por convención se usa $\rho = 1$
- Expansión: $\chi > 1, \chi > \rho$. Por convención se usa $\chi = 2$
- Contracción: $0 < \gamma < 1$. Por convención se usa $\gamma = \frac{1}{2}$
- Encogimiento: $0 < \sigma < 1$. Por convención se usa $\sigma = \frac{1}{2}$

Reflexión

- Se calcula el punto de reflexión del “peor” vértice elegido:

$$x_{ref} = c + \rho(c - x_{n-1})$$

Se evalúa $f(x_{ref})$:

- $f(x_0) \geq f(x_{ref}) > f(x_j)$, $j = 1, \dots, n$: Se elimina el peor punto, sustituyéndolo por x_{ref} . Finaliza la iteración y se retorna al inicio de la misma.
- $f(x_{ref}) > f(x_{n+1})$: No se produce mejora alguna respecto a algún punto.

Reflexión

- $x_{ref} = c + \rho(c - x_{n-1})$

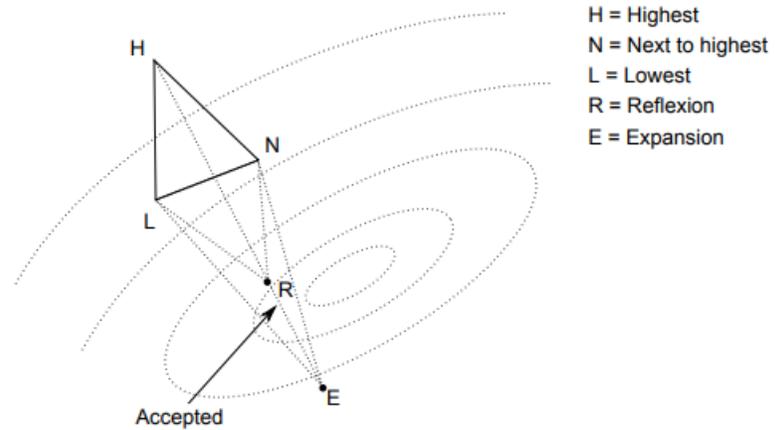


Figura 4. Transformación de reflexión

Expansión

- Se aplica expansión si $f(x_{ref}) > f(x_0)$
- Se calcula para comprobar si este nuevo valor mejora más al valor calculado previamente.

$$x_{expan} = c + \chi(x_{ref} - c)$$

- $f(x_{expan}) > f(x_{ref})$: el punto de expansión se considera el nuevo punto para la estructura simplex y se elimina el “peor” punto.
- Si no se cumple la desigualdad anterior, entonces el nuevo punto será $f(x_{ref})$. Se quita el peor punto formando el nuevo simplex. Se finaliza la iteración y se vuelve al inicio de la misma, de manera que se realiza la clasificación, el cálculo del centroide y finalmente aplicando la transformada adecuada.

Expansión

- $x_{expan} = c + \chi(x_{ref} - c)$

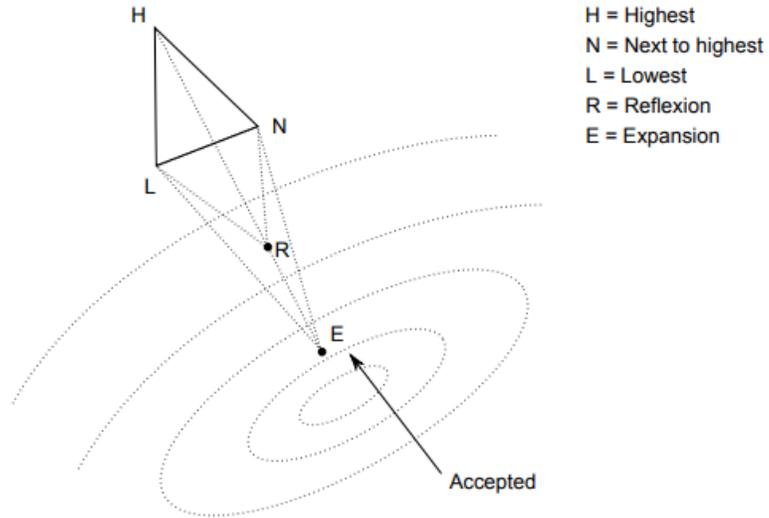


Figura 5. Transformación de expansión.

Contracción

- La contracción se realiza en la dirección del vector que va del punto peor al reflejado. *“Se aplica siempre que el valor de punto reflejado no sea de un valor mayor que el mejor o que alguno de los buenos, pudiendo ser mejor, igual o peor que el punto peor”* (Prados, 2022).

$$f(x_{ref}) < f(x_{n+1})$$

- Existen dos formas de contracción: Contracción externa y contracción interna. Para realizar la contracción se usa la mayor de las dos.

Contracción

- Contracción externa: $x_{con-e} = c + \gamma(x_{ref} - c)$
- Contracción interna: $x_{con-i} = c - \gamma(x_{ref} - c)$

Si $f(x_{con-e}) > f(x_p)$ y $f(x_{con-e}) > f(x_{ref})$ o $f(x_{con-i}) > f(x_p)$ y $f(x_{con-i}) > f(x_{ref})$, se procede a aceptar ese punto (depende de la contracción elegida) y este pasará a formar el nuevo punto del simplex.

Contracción

- Contracción Interna

$$x_{con-i} = c - \gamma(x_{ref} - c)$$

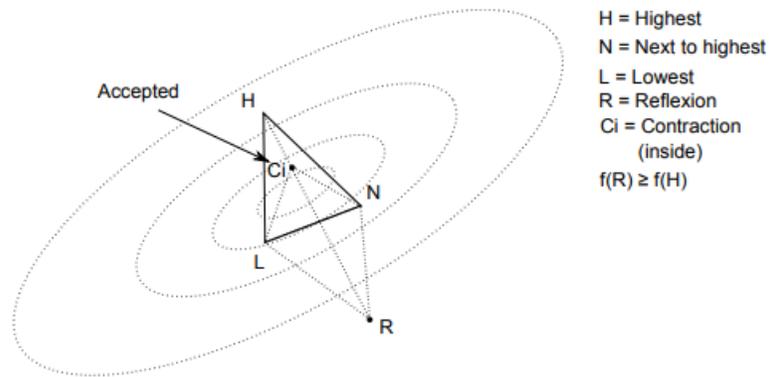


Figura 6. Contracción interna

- Contracción Externa

$$x_{con-e} = c + \gamma(x_{ref} - c)$$

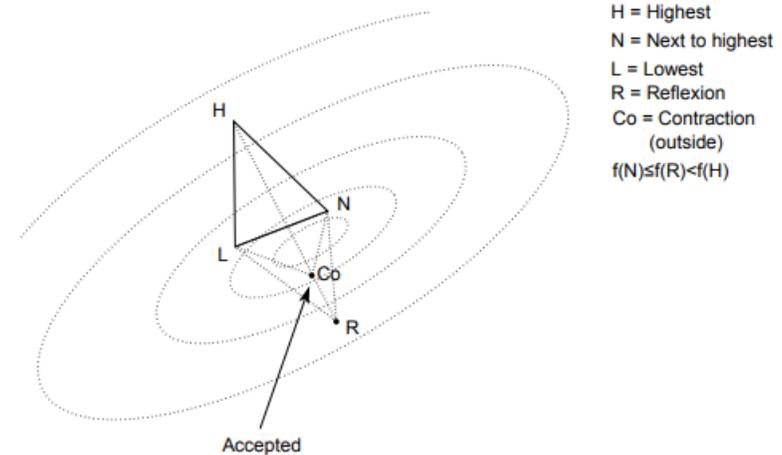


Figura 7. Contracción externa

Encogimiento

- Si no se ha conseguido mejorar el punto con la contracción, se aplica esta transformación. Se eligen n puntos nuevos como sigue:

$$v_i = x_1 - \sigma(x_i - x_1) \quad i = 2, \dots, n + 1$$

- Estos forman la nueva estructura simplex, con vértices

$$x_1, v_2, \dots, v_{n+1}$$

- Termina la iteración y se vuelve al inicio de la misma.

Encogimiento

- $v_i = x_1 - \sigma(x_i - x_1) \quad i = 2, \dots, n + 1$

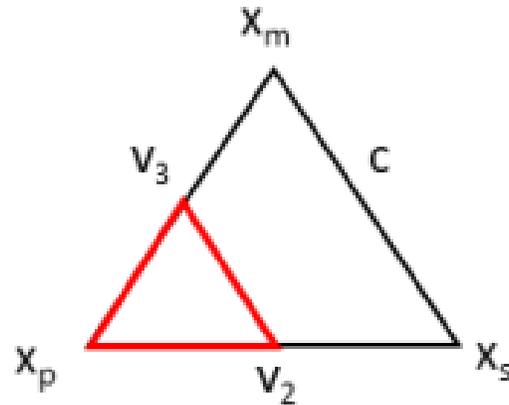


Figura 8. Transformación de encogimiento.

Análisis geométrico

- Expansión: Es realizada cuando el simplex está considerablemente lejos del óptimo. El peor vértice es movido en la dirección de descenso.
- Reflexión: Cuando es aplicada esta transformación, el simplex se acerca a un valle, puesto que el punto de expansión no se acercó lo suficiente al óptimo.
- Contracción: Cuando el simplex está cercano al óptimo, la contracción, ya sea externa o interna, decrece el tamaño del simplex.
- El encogimiento no se aplica con frecuencia.

Análisis geométrico

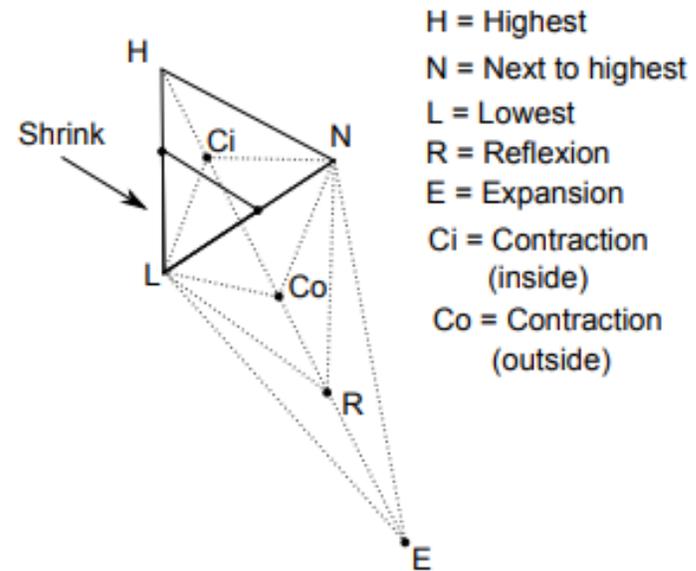


Figura 9. Transformaciones del simplex

Criterios de convergencia

Lagarias, Reeds, Wright y Wright demostraron que para funciones estrictamente convexas de 1 o 2 dimensiones, el método converge. Para 2 dimensiones, se cumple:

- 1) Los valores de la función en los vértices de los símplex con el método Nelder-Mead convergen al mismo valor.
- 2) Los símplex en el método de Nelder-Mead tienen diámetros que convergen a cero.

Criterios de convergencia

Han y Neumann demostraron que el método de Nelder-Mead se deteriora a medida que la dimensión aumenta.

Lagarias, Poonen, Wright demostraron que en el método de Nelder-Mead restringido de expansiones, si F es la clase de funciones dos veces continuamente diferenciables de $R^2 \rightarrow R$ con conjuntos de niveles acotados y el Hessiano positivo definido en todas partes, si $f \in F$ y S^0 es no degenerado, converge al mínimo único de f .

Criterios de convergencia

Galántai demostró, al hacer uso de una forma matricial del método de Nelder-Mead, que la convergencia está relacionada con la convergencia de productos matriciales infinitos

Assumption (A): *There is an induced matrix norm $\|A\|_{\vartheta}$ such that if $T_i P^{(i)} \in \mathcal{W}_1$, then $\|C_i\|_{\vartheta} < 1$.*

$$\mathcal{W}_1 = \left\{ T \begin{pmatrix} 1 \\ 2 \end{pmatrix} P_j, T \begin{pmatrix} -1 \\ 2 \end{pmatrix} P_j : j = 1, 2 \right\} \cup \{T_{shr} P : P \in \mathcal{P}_{n+1}\}$$

The entries of submatrix C are only 0, $\frac{1}{2}$ and $-\frac{1}{2}$

$$S^{(k)} = S^{(k-1)} T_k P^{(k)} = S^{(0)} \prod_{i=1}^k T_i P^{(i)}$$

Criterios de convergencia

McKinnon construyó una función estrictamente convexa $f: R^2 \rightarrow R$ con derivadas continuas en la que el método Nelder-Mead converge a un punto no estacionario.

El método de Nelder-Mead puede entrar a un ciclo de contracciones internas o externas que no llevan a una solución óptima. Esto suele suceder si la función objetivo tiene puntos de silla o regiones planas.

Reinicio del método Nelder-Mead

Se utiliza para encontrar que se ha encontrado un mínimo falso. Para estos reinicios se establece un número máximo. Luego de realizar una búsqueda se hace uso de una condición para saber si debe reiniciarse.

- Prueba factorial de O'Neill,
- Condición de estancamiento de Kelley.

Reinicio del método Nelder- Mead

```
restartnb ← 0
reached ← FALSE
for i = 1 to restartmax + 1 do
  search()
  istorestart = istorestart()
  if NOT(istorestart) then
    reached ← TRUE {Convergence}
    BREAK
  end if
  if i < restartmax then
    restartnb ← restartnb + 1 {A restart is going to be performed}
  end if
end for
if reached then
  printf ( "Convergence reached after
else
  printf ( "Convergence not reached after maximum
  status ← "maxrestart"
end if
```

Prueba Factorial de O'Neill

Se requieren a lo más $2n$ funciones de evaluación por cada vez que se realiza el test.

```
x ← x*  
istorestart = FALSE  
for i = 1 to n do  
   $\delta = \text{step}(i) * \epsilon$   
   $x(i) = x(i) + \delta$   
   $fv = f(x)$   
  if  $fv < f_{opt}$  then  
    istorestart = TRUE  
    break  
  end if  
   $x(i) = x(i) - \delta - \delta$   
   $fv = f(x)$   
  if  $fv < f_{opt}$  then  
    istorestart = TRUE  
    break  
  end if  
   $x(i) = x(i) + \delta$   
end for
```

Condición de estancamiento de Kelley

Se requieren $k + 1$ iteraciones para esta condición.

Se toman los supuestos:

- El simplex S_k es no degenerado.
- Los vértices se ordenan por valor de función creciente.

$$f_1^k \leq f_2^k \leq \dots \leq f_{n+1}^k$$

- El mejor valor de función es estrictamente decreciente.

$$f_1^{k+1} < f_1^k$$

Si no se produce ninguna contracción, el mejor valor es decreciente.

$$f_1^{k+1} - f_1^k < -c \|\bar{\mathbf{g}}(S_k)\|^2$$

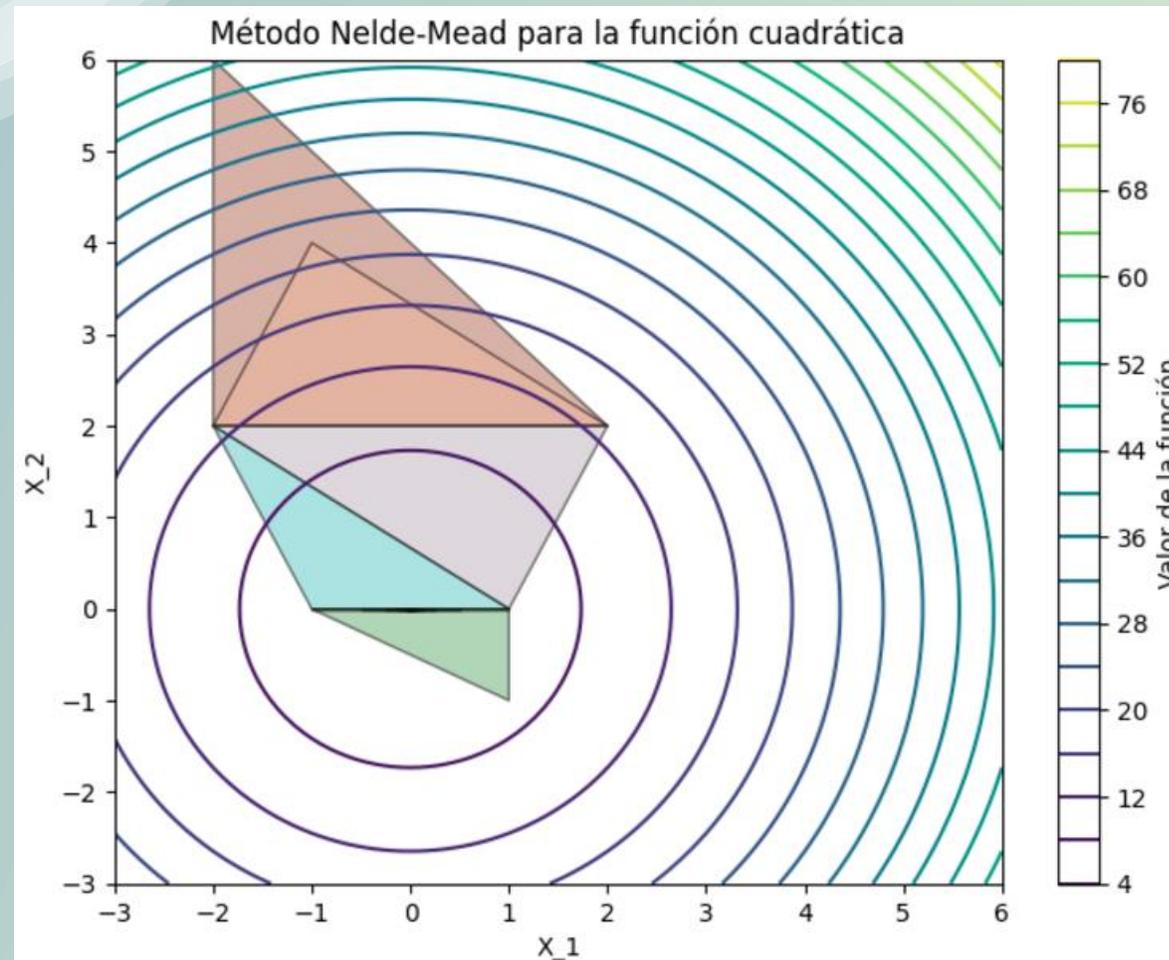
Función cuadrática

$$x_1^2 + x_2^2 + 5$$

Punto inicial	[-2, 2]
Delta	4
Punto obtenido	[0, 0]
Valor de la función (final)	5
Iteraciones	16

	V1	V2	V3	f	Error
Iteración					
0	[-2.0, 2.0]	[2.0, 2.0]	[-2.0, 6.0]	13.0	1.508494e+01
1	[-2.0, 2.0]	[2.0, 2.0]	[-1.0, 4.0]	13.0	4.242641e+00
2	[1.0, 0.0]	[-2.0, 2.0]	[2.0, 2.0]	6.0	3.299832e+00
3	[1.0, 0.0]	[-1.0, 0.0]	[-2.0, 2.0]	6.0	3.299832e+00
4	[1.0, 0.0]	[-1.0, 0.0]	[1.0, -1.0]	6.0	4.714045e-01
5	[0.0, 0.0]	[1.0, 0.0]	[-1.0, 0.0]	5.0	4.714045e-01
6	[0.0, 0.0]	[0.5, 0.0]	[-0.5, 0.0]	5.0	1.178511e-01
7	[0.0, 0.0]	[0.25, 0.0]	[-0.25, 0.0]	5.0	2.946278e-02
8	[0.0, 0.0]	[0.125, 0.0]	[-0.125, 0.0]	5.0	7.365696e-03
9	[0.0, 0.0]	[0.0625, 0.0]	[-0.0625, 0.0]	5.0	1.841424e-03
10	[0.0, 0.0]	[0.03125, 0.0]	[-0.03125, 0.0]	5.0	4.603560e-04
11	[0.0, 0.0]	[0.015625, 0.0]	[-0.015625, 0.0]	5.0	1.150890e-04
12	[0.0, 0.0]	[0.0078125, 0.0]	[-0.0078125, 0.0]	5.0	2.877225e-05
13	[0.0, 0.0]	[0.00390625, 0.0]	[-0.00390625, 0.0]	5.0	7.193062e-06
14	[0.0, 0.0]	[0.001953125, 0.0]	[-0.001953125, 0.0]	5.0	1.798266e-06
15	[0.0, 0.0]	[0.0009765625, 0.0]	[-0.0009765625, 0.0]	5.0	4.495664e-07
16	[0.0, 0.0]	[0.0009765625, 0.0]	[-0.0009765625, 0.0]	5.0	4.495664e-07

Función cuadrática

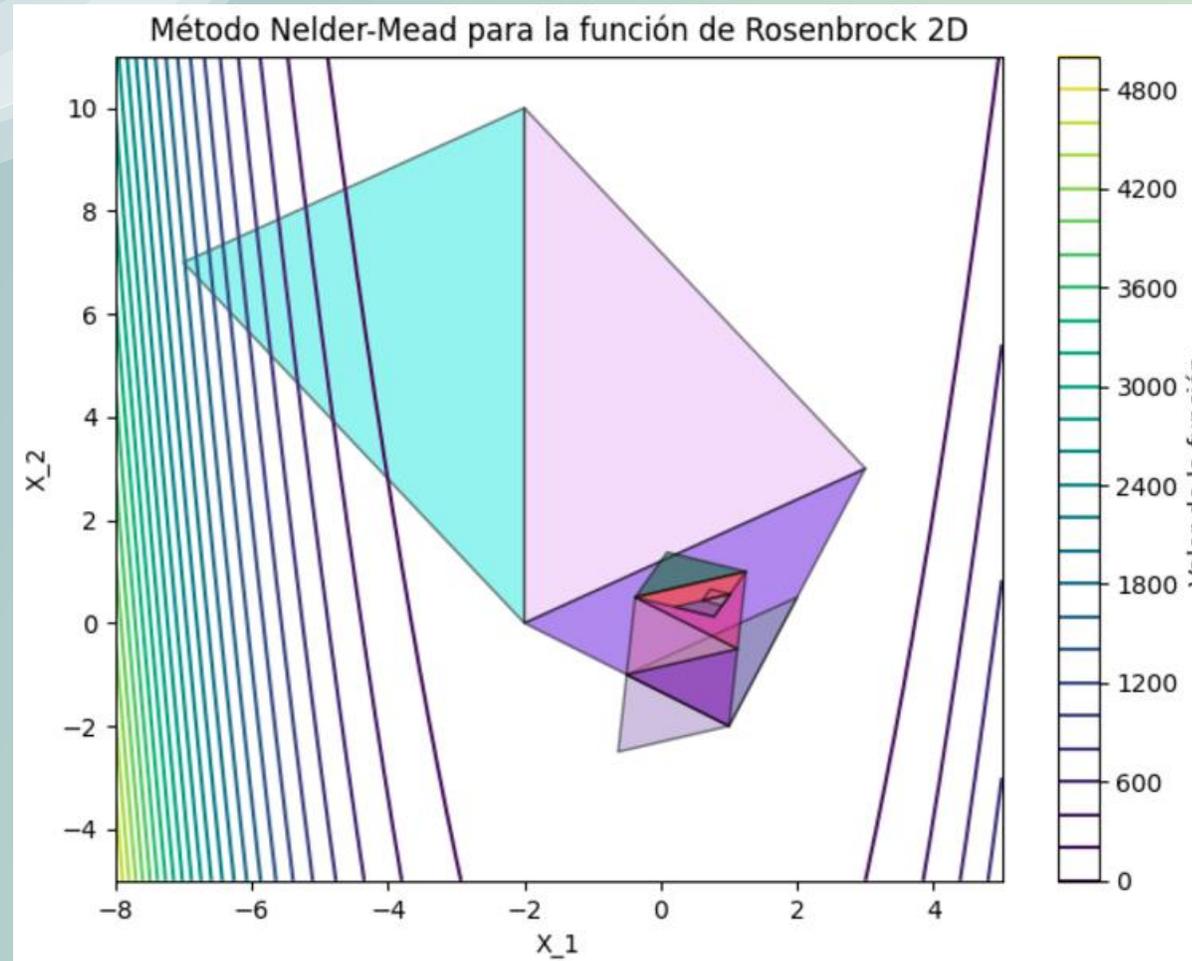
$$x_1^2 + x_2^2 + 5$$


Función de Rosenbrock 2D

Punto inicial	$[-2, 0]$
Delta	10
Punto obtenido	$[0.99933434, 0.99910355]$
Valor de la función (final)	$6.318409e-07$
Iteraciones	47

Iteración	V1	V2	V3	f	Error
40	[1.00146484375, 1.0048828125]	[0.99237060546875, 0.9908447265625]	[0.98828125, 0.9765625]	5.952087e-06	5.473659e-05
41	[1.00146484375, 1.0048828125]	[0.9925994873046875, 0.987213134765625]	[0.99237060546875, 0.9908447265625]	5.952087e-06	3.646120e-05
42	[1.00146484375, 1.0048828125]	[1.0016937255859375, 1.001251220703125]	[0.9925994873046875, 0.987213134765625]	5.952087e-06	2.447750e-05
43	[1.00146484375, 1.0048828125]	[1.0016937255859375, 1.001251220703125]	[1.0060691833496094, 1.0109939575195312]	5.952087e-06	1.487683e-05
44	[1.00146484375, 1.0048828125]	[1.0016937255859375, 1.001251220703125]	[1.003824234008789, 1.0070304870605469]	5.952087e-06	3.972239e-06
45	[0.9993343353271484, 0.9991035461425781]	[1.00146484375, 1.0048828125]	[1.0016937255859375, 1.001251220703125]	6.318409e-07	2.923921e-06
46	[0.9993343353271484, 0.9991035461425781]	[1.000514030456543, 1.0001773834228516]	[1.0003995895385742, 1.001993179321289]	6.318409e-07	3.931901e-07
47	[0.9993343353271484, 0.9991035461425781]	[1.000514030456543, 1.0001773834228516]	[1.0003995895385742, 1.001993179321289]	6.318409e-07	3.931901e-07

Función de Rosenbrock 2D



Fuentes de consulta

- Baudin, M. (2010) *Nelder-Mead User's Manual*.
- Garrido Prados, D.J. (2022) *Aplicación del método Nelder-Mead para optimización del intervalo de mantenimiento preventivo*. Dpto. Organización Industrial y Gestión de Empresas. Escuela Técnica Superior de Ingeniería. Universidad de Sevilla.
- Galántai, A. (2021). Convergence of the Nelder-Mead method. *Numerical Algorithms* 90(2022):1043–1072.
<https://doi.org/10.1007/s11075-021-01221-7>