

# Optimización Multidimensional no restringida 2023

Prof.: Dr. Juan Ignacio Manassaldi

JTP: Ing. Amalia Rueda

- Decimos que  $\underline{x}^*$  es un mínimo local si:

$$\underline{x}^* \in S \text{ tal que } f(\underline{x}^*) < f(\underline{x}^* + \underline{\delta})$$

- $\underline{x}^*$  es un mínimo global si:

$$\underline{x}^* \in S \text{ tal que } f(\underline{x}^*) < f(\underline{x}) \forall \underline{x} \in S$$

- Decimos que  $\underline{x}^*$  es un máximo local si:

$$\underline{x}^* \in S \text{ tal que } f(\underline{x}^*) > f(\underline{x}^* + \underline{\delta})$$

- $\underline{x}^*$  es un máximo global si:

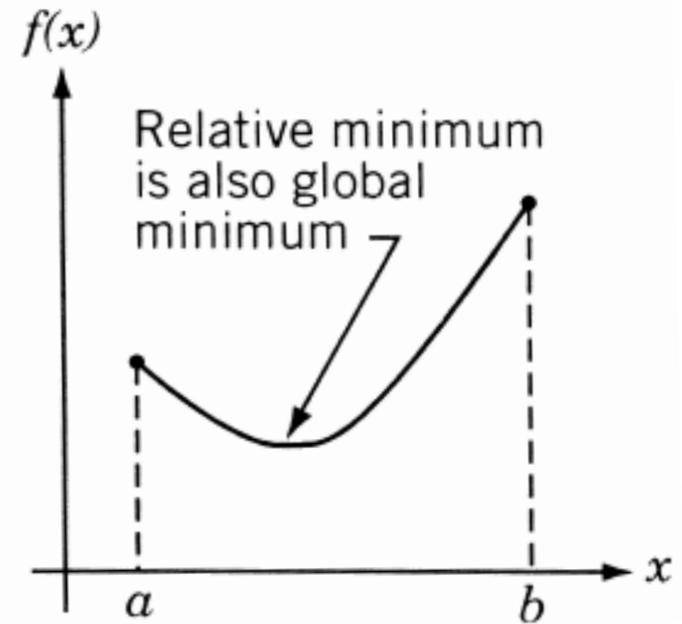
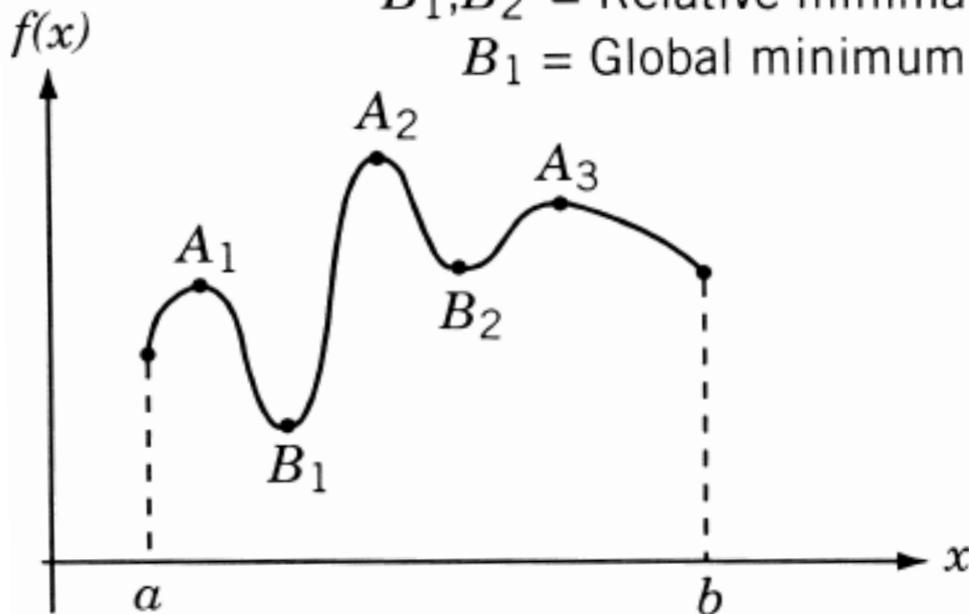
$$\underline{x}^* \in S \text{ tal que } f(\underline{x}^*) > f(\underline{x}) \forall \underline{x} \in S$$

➤ Si tenemos algún punto  $\underline{x}^*$  del dominio tal que:  $\nabla f(\underline{x}^*) = \underline{0}$

Entonces  $\underline{x}^*$  *es un punto crítico*, esto es, puede ser:

- 1) Un mínimo local
- 2) Un máximo local
- 3) Un punto de ensilladura

$A_1, A_2, A_3 =$  Relative maxima  
 $A_2 =$  Global maximum  
 $B_1, B_2 =$  Relative minima  
 $B_1 =$  Global minimum



Aproximación de primer orden:

$$f(\underline{x}, \underline{x}_0) = f(\underline{x}_0) + (\underline{x} - \underline{x}_0)^T \nabla f(\underline{x}_0)$$

Aproximación de segundo orden:

$$f(\underline{x}, \underline{x}_0) = f(\underline{x}_0) + (\underline{x} - \underline{x}_0)^T \nabla f(\underline{x}_0) + \frac{1}{2} (\underline{x} - \underline{x}_0)^T \nabla^2 f(\underline{x}_0) (\underline{x} - \underline{x}_0)$$

Los métodos que se han ideado se pueden clasificar en tres amplias categorías según el tipo de información que debe proporcionar el usuario:

- Métodos de búsqueda directa, que utilizan solo valores de función.

## Método de una variable a la vez

- Métodos de gradiente (primer orden), que requieren estimaciones de la primera derivada de  $f(x)$ .

## Descenso mas pronunciado

- Métodos de segundo orden, que requieren estimaciones de la primera y segunda derivada de  $f(x)$ .

## Método de Newton

- La búsqueda de línea nos permite encontrar el valor mínimo (o máximo) que toma una función sobre una dirección determinada de búsqueda.
- Para realizar la búsqueda de línea sobre  $f(\underline{x})$ , necesitamos un valor perteneciente al dominio de la función  $\underline{x}^0$  y una dirección de búsqueda  $\underline{d}$ .
- Partiendo del punto  $\underline{x}^0$  y utilizando la dirección en el espacio  $\underline{d}$  podemos encontrar el valor mínimo que toma la función mediante una optimización unidimensional.

$$\underline{x} = \underline{x}^0 + \alpha \underline{d}$$

Ecuación de todos los puntos sobre la dirección  $\underline{d}$  desde  $\underline{x}^0$

Ecuación de todos los puntos sobre la dirección  $\underline{d}$  desde  $\underline{x}^0$ :

$$\underline{x} = \underline{x}^0 + \alpha \underline{d}$$

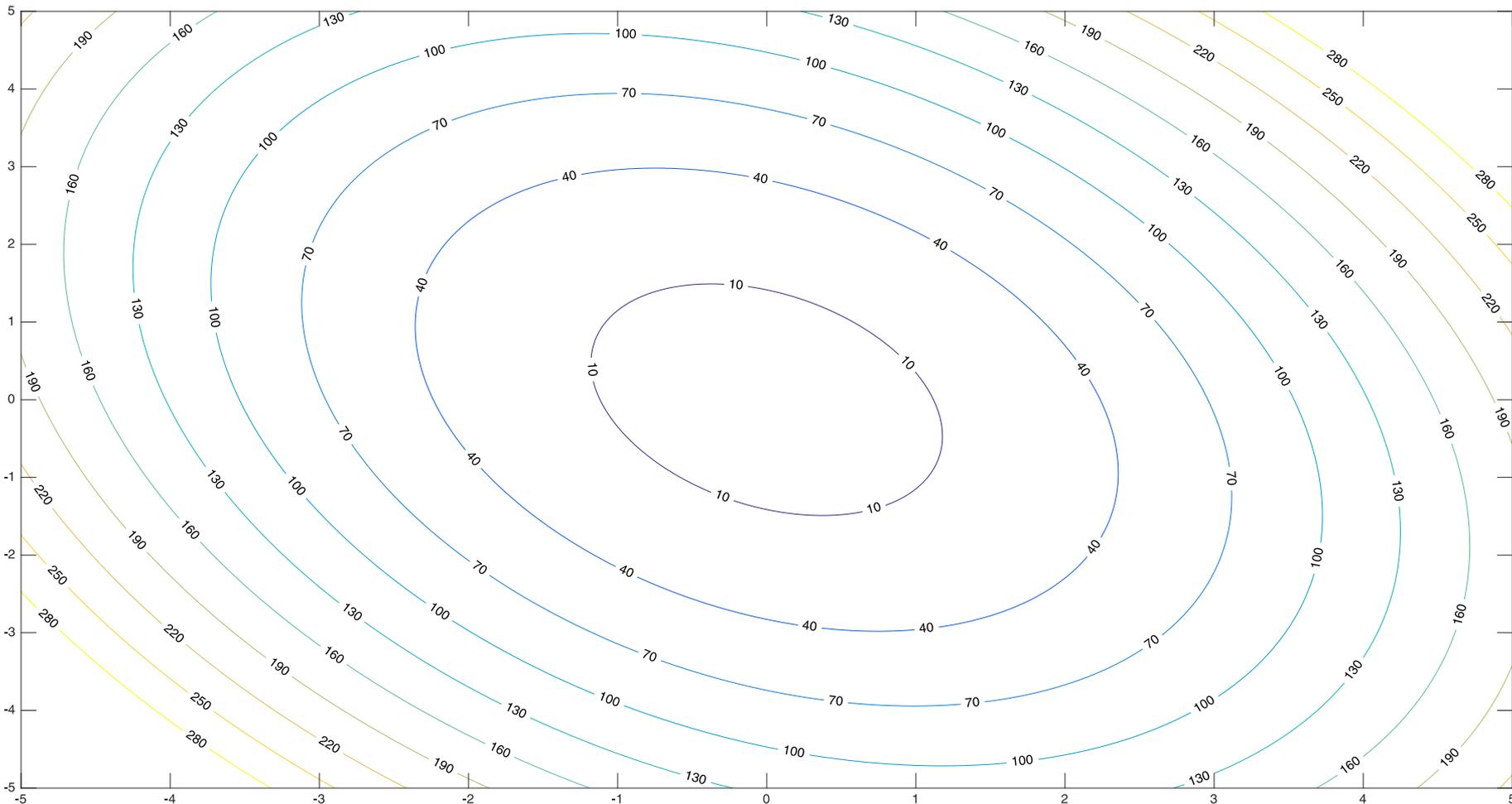
Para encontrar el valor mínimo sobre esta dirección se debe resolver el siguiente problema:

$$\text{Min. } f\left(\underline{x}^0 + \alpha \underline{d}\right)$$

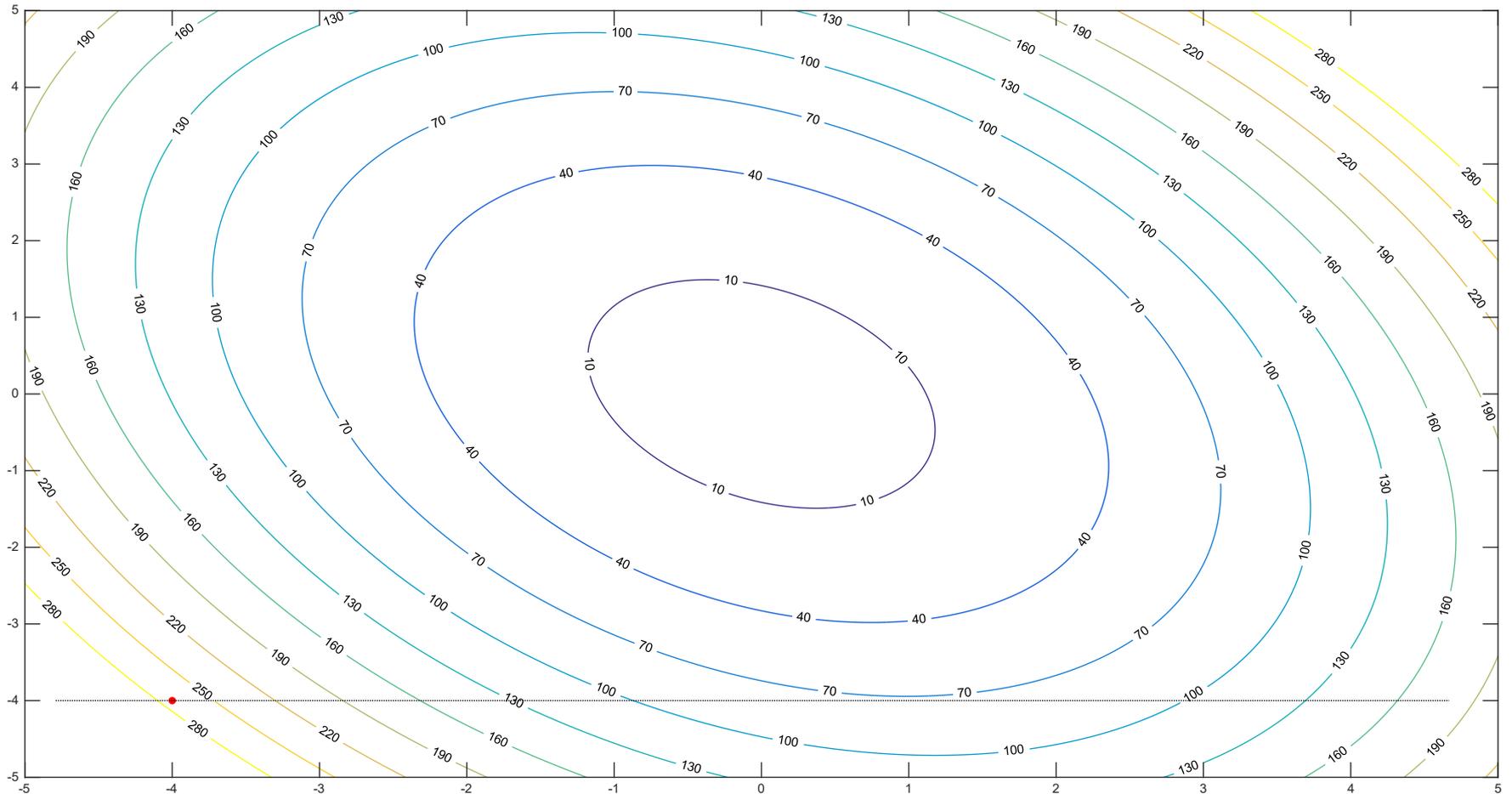
¡Es un problema de optimización unidimensional!

Solo debo encontrar el  $\alpha$  que minimiza  $f$

$$f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$



$$f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2 \quad \underline{x}^0 = \begin{bmatrix} -4 \\ -4 \end{bmatrix} \quad \underline{d} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



$$\underline{x}^0 = \begin{bmatrix} -4 \\ -4 \end{bmatrix} \quad \underline{d} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \rightarrow \quad \underline{x} = \underline{x}^0 + \alpha \underline{d}$$

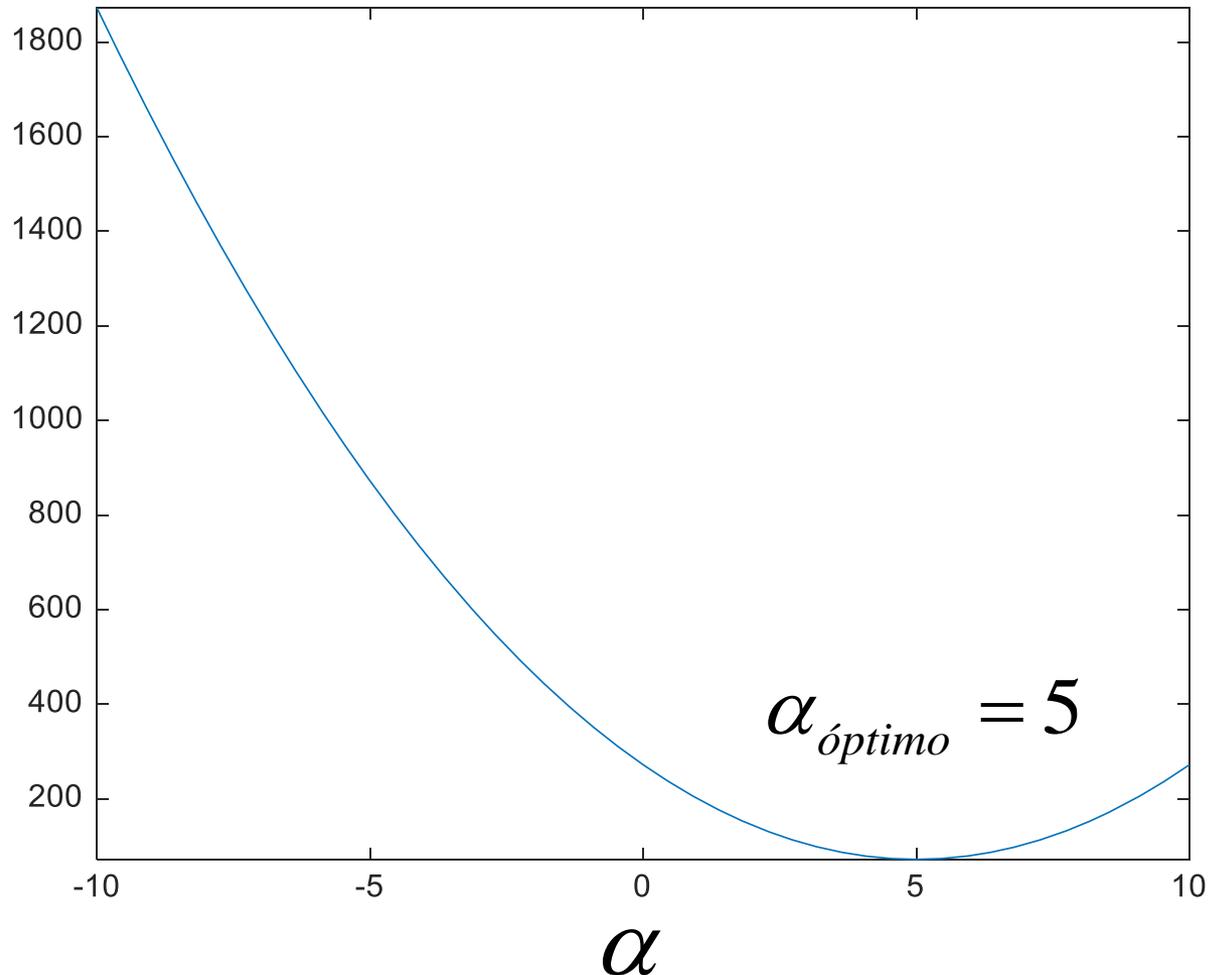
$$\underline{x} = \begin{bmatrix} -4 \\ -4 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \rightarrow \quad \underline{x} = \begin{bmatrix} -4 + \alpha \\ -4 \end{bmatrix}$$

$$\text{Min. } f(\underline{x}^0 + \alpha \underline{d}) \quad f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$

$$\text{Min } 8(-4 + \alpha)^2 + 4(-4 + \alpha)(-4) + 5(-4)^2$$

$$\text{Min } 8(-4 + \alpha)^2 + 4(-4 + \alpha)(-4) + 5(-4)^2$$

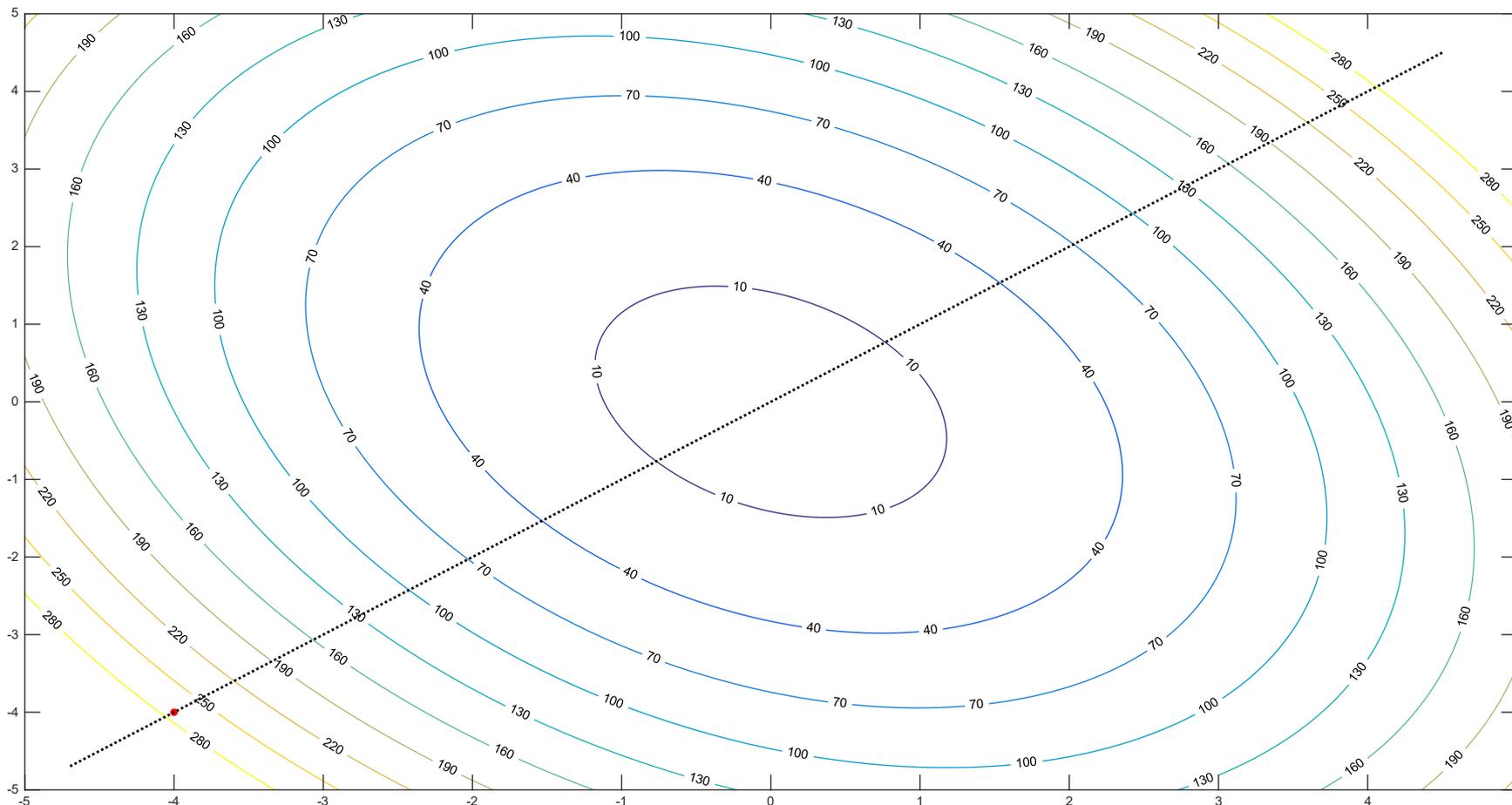
$$f(\underline{x}^0 + \alpha \underline{d})$$



$$\underline{x}^0 = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$$

$$\underline{d} = \begin{bmatrix} \cos(\pi/4) \\ \text{sen}(\pi/4) \end{bmatrix}$$

$$\text{Min. } f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$



- En el método de una variable a la vez, buscamos el mínimo a lo largo de las direcciones paralelas a los ejes.
- Este proceso se realiza de manera iterativa hasta que no sea posible una mejora en la FO.
- El método de optimización unidimensional elegido influye en el desempeño del método.
- Este método puede no converger y en algunos casos su convergencia resultar muy lenta.
- Estos problemas se pueden evitar cambiando las direcciones de búsqueda de una manera favorable en lugar de mantenerlas siempre paralelas a los ejes de coordenadas.

$$\underline{x}^{(0)} \in D \subseteq \mathbb{R}^n; \underline{d}_1, \underline{d}_2, \dots, \underline{d}_n; k = 1$$

$$\underline{x}^{(k)} = \underline{x}^{(k-1)}$$

$i = 1$  a  $n$

$$\alpha^* = \text{Min. } f(\underline{x}^{(k)} + \alpha \underline{d}_i)$$

$$\underline{x}^{(k)} = \underline{x}^{(k)} + \alpha^* \underline{d}_i$$



$$\|\underline{x}^{(k)} - \underline{x}^{(k-1)}\| \leq tol$$

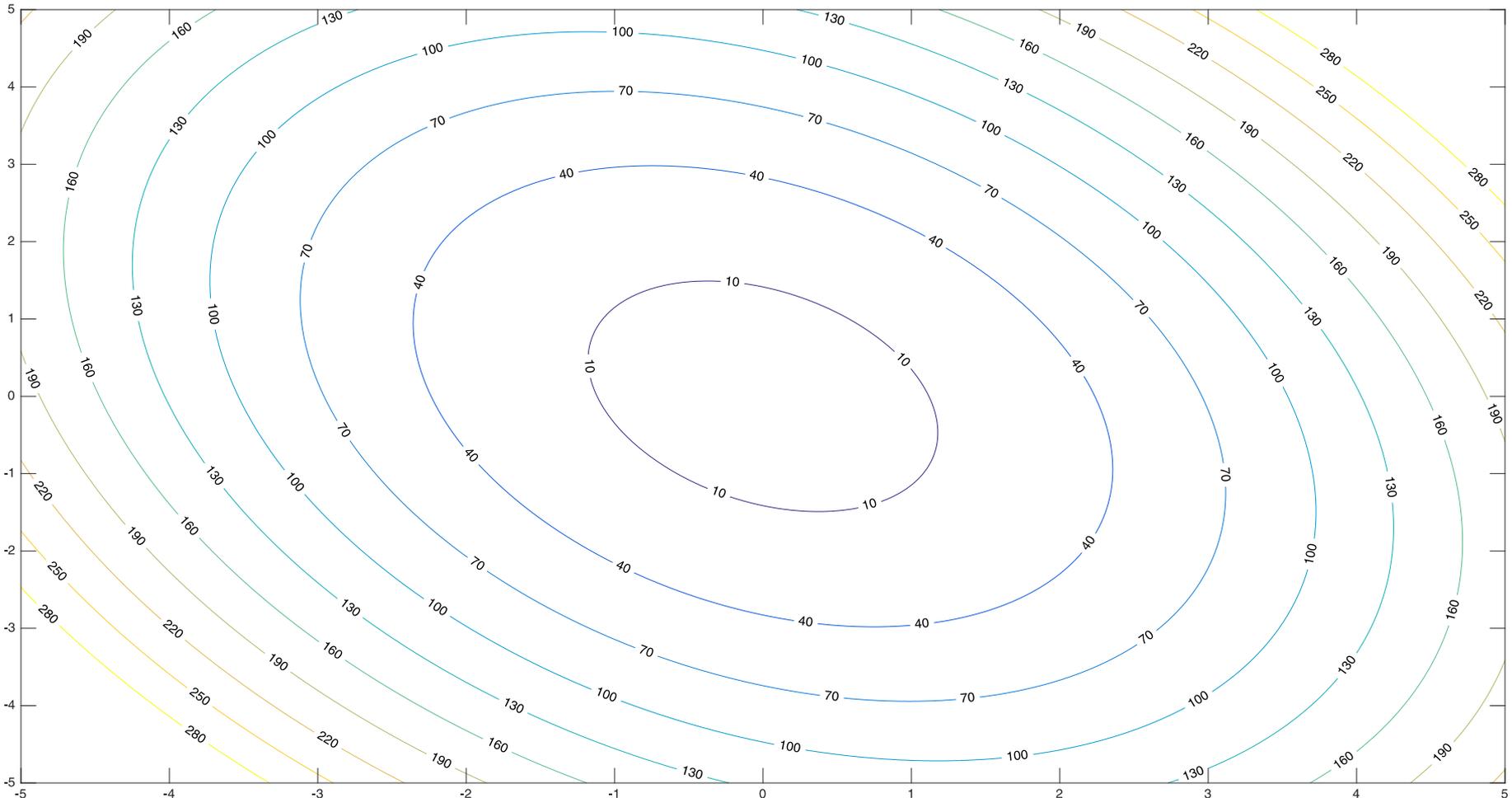
no

$$k = k + 1$$

si

$$\underline{x}^* = \underline{x}^{(k)}$$

$$\text{Min. } f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2 \quad \underline{x}^{(0)} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$$



$$\text{Min. } f(\underline{x}) = 8x_1^2 + 4x_1x_2 + 5x_2^2$$

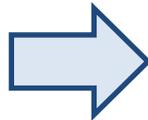
$$k = 1$$

$$\underline{x}^{(0)} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$$

$$\underline{d}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underline{d}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\underline{x}^{(1)} = \underline{x}^{(0)} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$$



$$i = 1$$

$$\alpha = \text{Min. } f(\underline{x}^{(1)} + \alpha \underline{d}_1) \rightarrow \alpha = 5$$

$$\underline{x}^{(1)} = \underline{x}^{(0)} + \alpha \underline{d}_1 \rightarrow \underline{x}^{(1)} = \begin{bmatrix} -4 \\ -4 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$$

$$i = 2$$

$$\alpha = \text{Min. } f(\underline{x}^{(1)} + \alpha \underline{d}_2) \rightarrow \alpha = 3.6$$

$$\underline{x}^{(1)} = \underline{x}^{(1)} + \alpha \underline{d}_2 \rightarrow \underline{x}^{(1)} = \begin{bmatrix} 1 \\ -4 \end{bmatrix} + 3.6 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix}$$

$$\underline{x}^{(0)} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}; \underline{x}^{(1)} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix} \rightarrow \begin{cases} f(\underline{x}^{(0)}) = 272 \\ f(\underline{x}^{(1)}) = 7.2 \end{cases}$$

¿? ¡Seguimos!

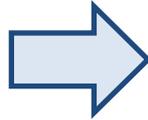
$$k = 2$$

$$\underline{x}^{(1)} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix}$$

$$\underline{d}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underline{d}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\underline{x}^{(2)} = \underline{x}^{(1)} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix}$$



$$i = 1$$

$$\alpha = \text{Min. } f(\underline{x}^{(2)} + \alpha \underline{d}_1) \rightarrow \alpha = -0.9$$

$$\underline{x}^{(2)} = \underline{x}^{(2)} + \alpha \underline{d}_1 \rightarrow \underline{x}^{(2)} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix} - 0.9 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.1 \\ -0.4 \end{bmatrix}$$

$$i = 2$$

$$\alpha = \text{Min. } f(\underline{x}^{(2)} + \alpha \underline{d}_2) \rightarrow \alpha = 0.36$$

$$\underline{x}^{(2)} = \underline{x}^{(2)} + \alpha \underline{d}_2 \rightarrow \underline{x}^{(2)} = \begin{bmatrix} 0.1 \\ -0.4 \end{bmatrix} + 0.36 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix}$$

$$\underline{x}^{(1)} = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix}; \underline{x}^{(2)} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix} \rightarrow \begin{cases} f(\underline{x}^{(1)}) = 7.2 \\ f(\underline{x}^{(2)}) = 0.072 \end{cases} \quad \text{¿?} \quad \text{iSeguimos!}$$

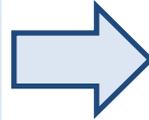
$$k = 3$$

$$\underline{x}^{(2)} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix}$$

$$\underline{d}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underline{d}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\underline{x}^{(3)} = \underline{x}^{(2)} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix}$$



$$i = 1$$

$$\alpha = \text{Min. } f(\underline{x}^{(3)} + \alpha \underline{d}_1) \rightarrow \alpha = -0.09$$

$$\underline{x}^{(3)} = \underline{x}^{(3)} + \alpha \underline{d}_1 \rightarrow \underline{x}^{(3)} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix} - 0.09 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.01 \\ -0.04 \end{bmatrix}$$

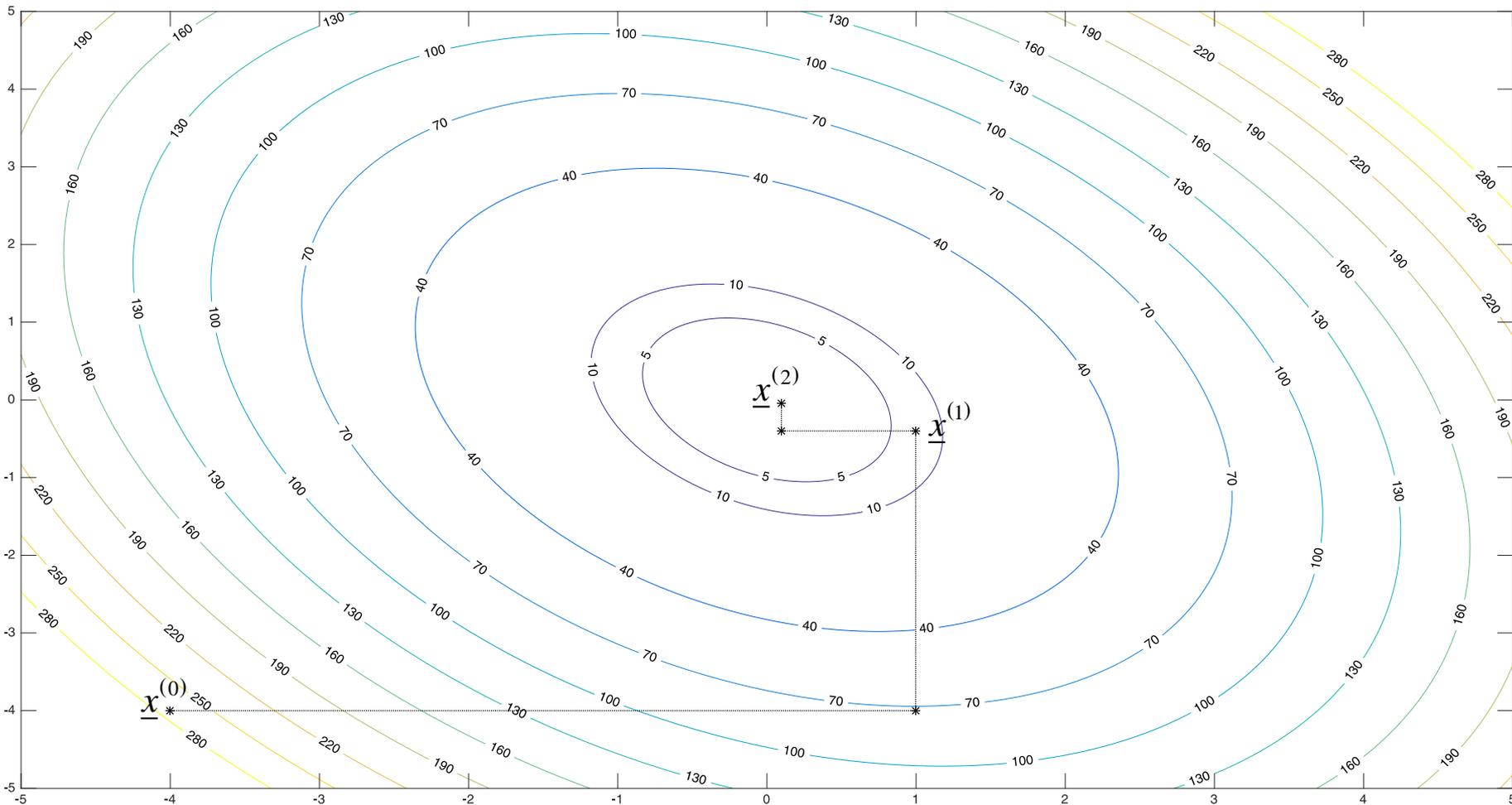
$$i = 2$$

$$\alpha = \text{Min. } f(\underline{x}^{(3)} + \alpha \underline{d}_2) \rightarrow \alpha = 0.036$$

$$\underline{x}^{(3)} = \underline{x}^{(3)} + \alpha \underline{d}_2 \rightarrow \underline{x}^{(3)} = \begin{bmatrix} 0.01 \\ -0.04 \end{bmatrix} + 0.036 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.01 \\ -0.004 \end{bmatrix}$$

$$\underline{x}^{(2)} = \begin{bmatrix} 0.1 \\ -0.04 \end{bmatrix}; \underline{x}^{(3)} = \begin{bmatrix} 0.01 \\ -0.004 \end{bmatrix} \rightarrow \begin{cases} f(\underline{x}^{(2)}) = 0.072 \\ f(\underline{x}^{(3)}) = 0.0072 \end{cases}$$

¿?  
¡Sigan!



- Los métodos directo requieren solo valores de la función objetivo para avanzar hacia la solución.
- Los métodos directos son importantes, porque muy a menudo, en un problema práctico de ingeniería, esta es la única información confiable disponible.
- Por otro lado, incluso los mejores métodos directos pueden requerir un número excesivo de evaluaciones de la función para localizar la solución.
- El ítem anterior, combinado con el deseo natural de buscar puntos críticos, motiva a considerar métodos que emplean información del gradiente.
- Se supone en todo momento que  $f(\underline{x})$ ,  $\nabla f(\underline{x})$  y  $\nabla^2 f(\underline{x})$  existen y son continuas.

- Todos los métodos que se presentaremos emplean un procedimiento de iteración similar, el nuevo punto se obtiene a partir de una dirección de búsqueda:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^k \underline{s}(\underline{x}^{(k)})$$

$\underline{x}^{(k)}$  : *Estimación actual de  $\underline{x}^*$*

$\alpha^k$  : *Tamaño del salto*

$\underline{s}(\underline{x}^{(k)})$  : *Dirección de búsqueda en el espacio  $\mathbb{R}^n$*

- En general se busca entonces encontrar el tamaño del salto en la dirección de  $\underline{s}$  que minimiza el valor de la función objetivo (búsqueda de línea).

- Supongamos que estamos en el punto  $\underline{x}^{(k)}$  y deseamos determinar la dirección de “el mayor descenso local”.
- A partir de la serie de Taylor, formamos una aproximación lineal a  $f(x)$  a partir  $\underline{x}^{(k)}$  eliminando los términos de orden 2 y superiores:

$$f(\underline{x}, \underline{x}^{(k)}) = f(\underline{x}^{(k)}) + (\underline{x} - \underline{x}^{(k)})^T \nabla f(\underline{x}^{(k)})$$

- Evaluando la aproximación de la función objetivo en el siguiente punto  $(k+1)$  del proceso iterativo se llega a:

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

- Suponemos que la aproximación lineal representa a la función, por lo que:

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

- Por lo tanto, el cambio en la función objetivo corresponde a:

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

- Siguiendo la estrategia de resolución propuesta:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^k \underline{s}(\underline{x}^{(k)}) \Rightarrow (\underline{x}^{(k+1)} - \underline{x}^{(k)})^T = \alpha^k \underline{s}(\underline{x}^{(k)})^T$$

- Reemplazando en la expresión anterior:

$$f(\underline{x}^{(k+1)}) - f(\underline{x}^{(k)}) = \alpha^k \underline{s}(\underline{x}^{(k)})^T \nabla f(\underline{x}^{(k)}) \rightarrow \|\underline{s}(\underline{x}^{(k)})\| \|\nabla f(\underline{x}^{(k)})\| \cos \alpha$$

- Analizando, proponemos la siguiente dirección:

$$\underline{s}(\underline{x}^{(k)}) = -\nabla f(\underline{x}^{(k)})$$

- Luego, reemplazando la dirección propuesta:

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

$$f(\underline{x}^{(k+1)}) - f(\underline{x}^{(k)}) = -\alpha^k \left\| \nabla f(\underline{x}^{(k)}) \right\|^2$$

- El método mas simple basado en el gradiente consiste en tomar la dirección opuesta del gradiente y fijar  $\alpha$  como un parámetro fijo positivo:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha \nabla f(\underline{x}^{(k)})$$

- Esto tiene dos desventajas: la necesidad de hacer una elección adecuada para  $\alpha$  y la lentitud inherente cerca del mínimo debido a que en esta zona el gradiente tiende a cero.

- El Método del descenso mas pronunciado propone realizar una optimización unidimensional para la elección de  $\alpha$  en cada iteración:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha^{(k)} \nabla f \left( \underline{x}^{(k)} \right)$$

- Por lo tanto, en cada iteración se debe resolver el siguiente problema de optimización unidimensional:

$$\alpha^* = \min_{\alpha} f \left( \underline{x}^{(k)} - \alpha \nabla f \left( \underline{x}^{(k)} \right) \right) \Rightarrow \underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha^* \nabla f \left( \underline{x}^{(k)} \right)$$

- Este método es más confiable que el método de gradiente simple, pero la convergencia sigue siendo lenta para la mayoría de los problemas prácticos.

**Ejemplo:**  $\text{Min. } f(\underline{x}) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha \nabla f(\underline{x}^{(k)}) \quad \nabla f(\underline{x}) = \begin{bmatrix} 2(x_1^2 + x_2 - 11)(2x_1) + 2(x_1 + x_2^2 - 7) \\ 2(x_1^2 + x_2 - 11) + 2(x_1 + x_2^2 - 7)(2x_2) \end{bmatrix}$$

$$\left. \begin{array}{l} \alpha = 1 \\ \underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \end{array} \right\}$$

$$\Rightarrow \underline{x}^{(1)} = \begin{bmatrix} 44 \\ 20 \end{bmatrix}$$

$$\Rightarrow \underline{x}^{(2)} = \begin{bmatrix} -343150 \\ -38830 \end{bmatrix}$$

$$\left. \begin{array}{l} \alpha = 10^{-1} \\ \underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \end{array} \right\}$$

$$\Rightarrow \underline{x}^{(1)} = \begin{bmatrix} 6.2 \\ 3.8 \end{bmatrix}$$

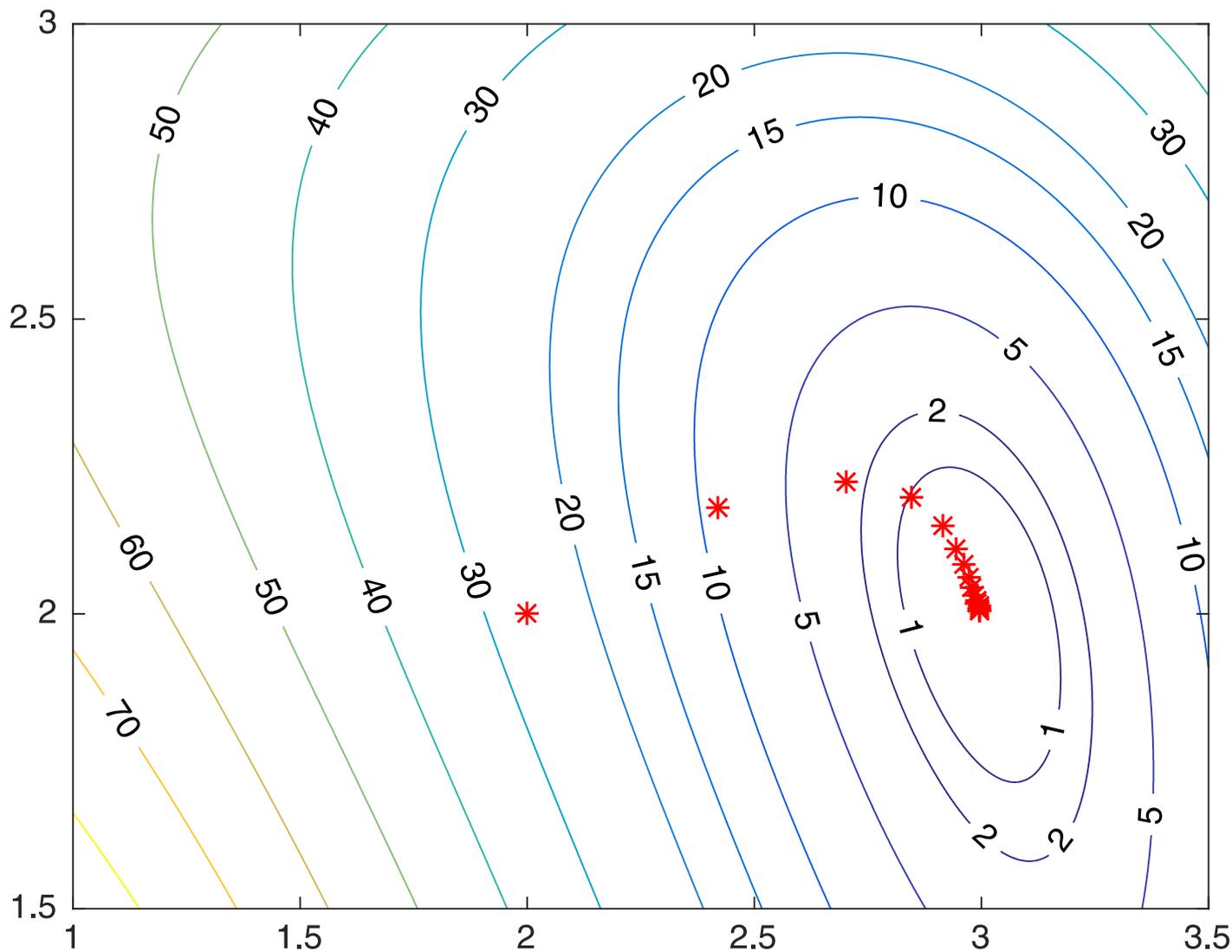
$$\Rightarrow \underline{x}^{(2)} = \begin{bmatrix} -74.003 \\ -23.181 \end{bmatrix}$$

$$\left. \begin{array}{l} \alpha = 10^{-2} \\ \underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \end{array} \right\}$$

$$\Rightarrow \underline{x}^{(1)} = \begin{bmatrix} 2.42 \\ 2.18 \end{bmatrix}$$

$$\Rightarrow \underline{x}^{(2)} = \begin{bmatrix} 2.703 \\ 2.224 \end{bmatrix}$$

$$\text{Min. } f(\underline{x}) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$



$$\alpha = 10^{-2}$$

$$\underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\underline{x}_1^{(33)} = \begin{bmatrix} 2.999989 \\ 2.000025 \end{bmatrix}$$

$$\text{tol} = 10^{-5}$$

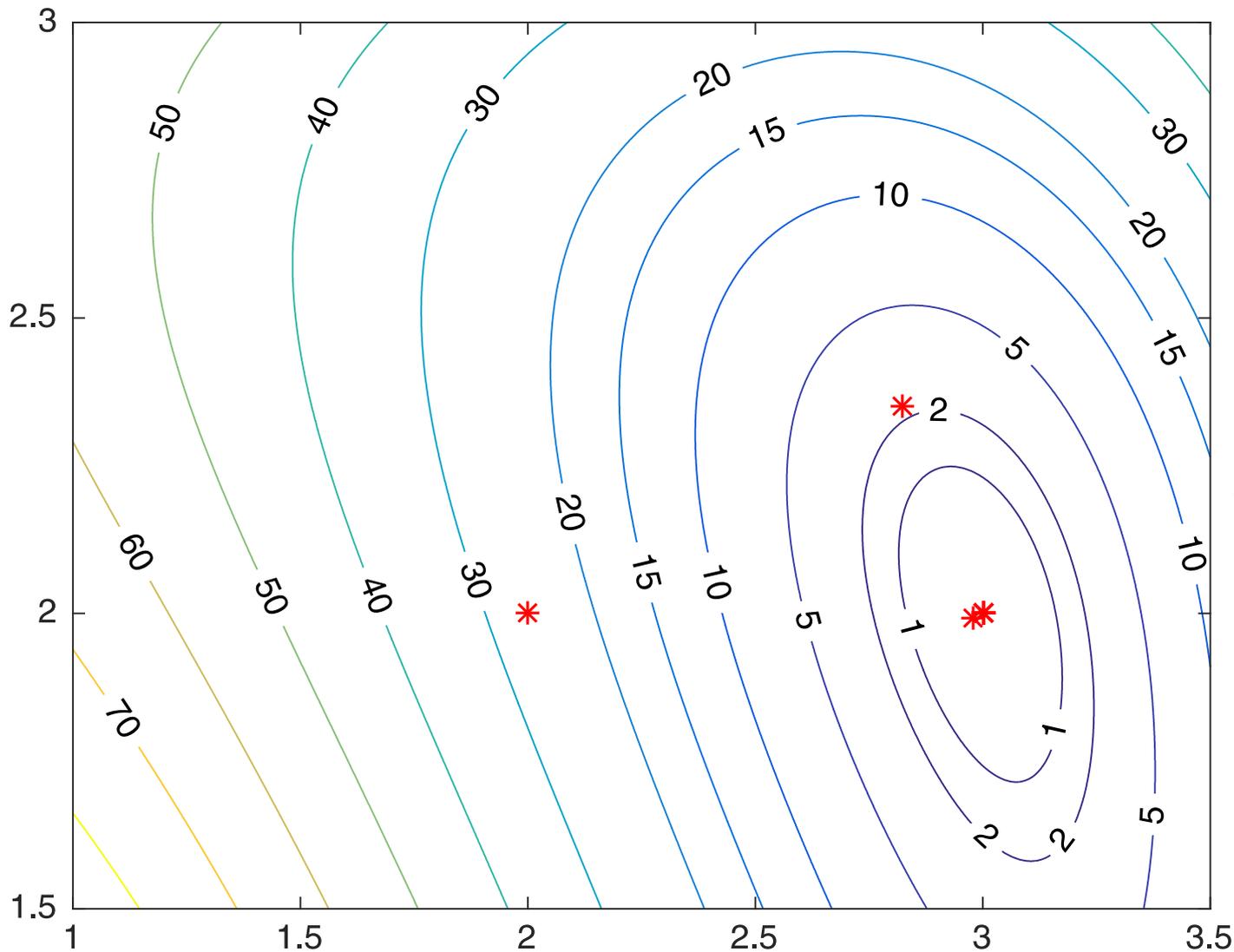
$$\text{Min. } f(\underline{x}) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2(x_1^2 + x_2 - 11)(2x_1) + 2(x_1 + x_2^2 - 7) \\ 2(x_1^2 + x_2 - 11) + 2(x_1 + x_2^2 - 7)(2x_2) \end{bmatrix}$$

$$\underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\alpha^* = \min f(\underline{x}^{(0)} - \alpha \nabla f(\underline{x}^{(0)}))$$

$$\underline{x}^{(1)} = \underline{x}^{(0)} - \alpha^* \nabla f(\underline{x}^{(0)})$$



$$x^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$x^{(13)} = \begin{bmatrix} 2.999999 \\ 2.000002 \end{bmatrix}$$

$$tol = 10^{-5}$$

$$\underline{x}^{(0)} \in D \subseteq \mathbb{R}^n; \nabla f(\underline{x}); k = 0$$

$$\underline{d} = -\nabla f(\underline{x}^{(k)})$$

$$\alpha^* = \text{Min. } f(\underline{x}^{(k)} + \alpha \underline{d})$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^* \underline{d}$$



$$\|\alpha^* \underline{d}\| \leq tol$$

no

$$k = k + 1$$

si

$$\underline{x}^* = \underline{x}^{(k+1)}$$

- El gradiente negativo apunta directamente hacia el mínimo solo cuando los contornos de  $f(x)$  son circulares, por lo tanto el gradiente negativo no es una buena dirección global (en general) para funciones no lineales.
- El método del descenso mas pronunciado emplea aproximaciones lineales sucesivas al objetivo y requiere valores de la función objetivo y su gradiente en cada iteración.
- Un avance es considerar el uso de información de orden superior (derivadas segundas) en un esfuerzo por construir una estrategia de iteración más global.

- A partir de la serie de Taylor, formamos una aproximación cuadrática de  $f(\underline{x})$  y suponemos que esta representa a la función.

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

$$\nabla f(\underline{x}) = \nabla f(\underline{x}^{(k)}) + \nabla^2 f(\underline{x}^{(k)}) (\underline{x} - \underline{x}^{(k)})$$

- Se propone una secuencia de iteración en la que el siguiente punto, es un punto en donde el gradiente de la aproximación es cero.

$$\nabla f(\underline{x}^{(k+1)}) = \nabla f(\underline{x}^{(k)}) + \nabla^2 f(\underline{x}^{(k)}) (\underline{x}^{(k+1)} - \underline{x}^{(k)}) = 0$$

- Siguiendo la estrategia de resolución propuesta:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^k s(\underline{x}^{(k)}) \Rightarrow (\underline{x}^{(k+1)} - \underline{x}^{(k)}) = \alpha^k s(\underline{x}^{(k)})$$

- Reemplazando y asumimos  $\alpha=1$ :

$$\nabla f(\underline{x}^{(k)}) + \nabla^2 f(\underline{x}^{(k)}) \alpha^k s(\underline{x}^{(k)}) = 0 \Rightarrow s(\underline{x}^{(k)}) = -\left(\nabla^2 f(\underline{x}^{(k)})\right)^{-1} \nabla f(\underline{x}^{(k)})$$

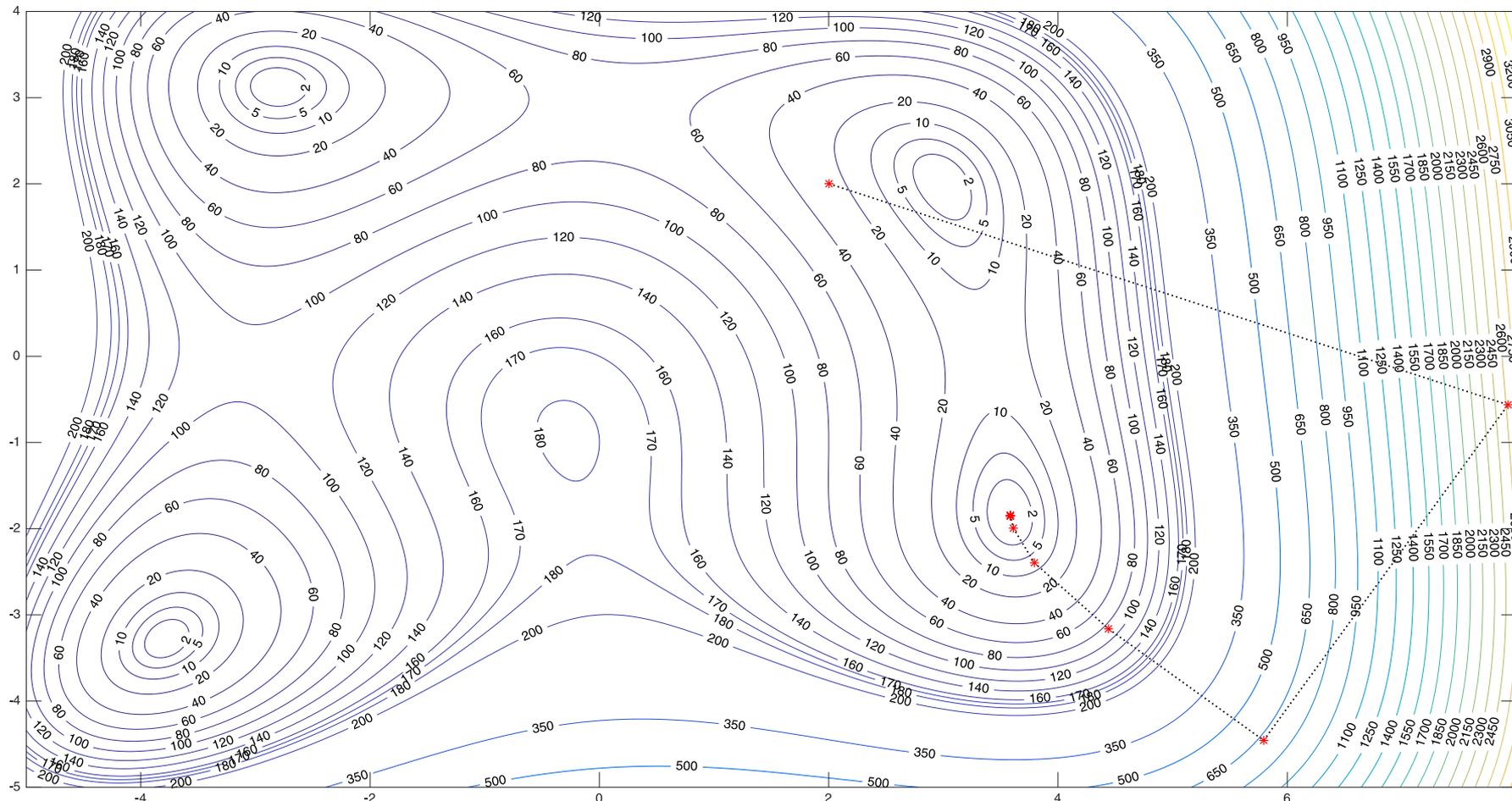
- Por lo que finalmente:

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \left(\nabla^2 f(\underline{x}^{(k)})\right)^{-1} \nabla f(\underline{x}^{(k)})$$

**Ejemplo:**  $Min. f(\underline{x}) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2(x_1^2 + x_2 - 11)2x_1 + 2(x_1 + x_2^2 - 7) \\ 2(x_1^2 + x_2 - 11) + 2(x_1 + x_2^2 - 7)2x_2 \end{bmatrix}$$

$$\nabla^2 f(\underline{x}) = \begin{bmatrix} 8x_1^2 + 4(x_1^2 + x_2 - 11) + 2 & 4x_1 + 4x_2 \\ 4x_1 + 4x_2 & 2 + 8x_2^2 + 4(x_1 + x_2^2 - 7) \end{bmatrix}$$



$$\underline{x}^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \Rightarrow \underline{x}^{(9)} = [3.584428 \quad -1.848126]^T$$

$tol = 10^{-5}$

- El método de Newton minimizará una función cuadrática (desde cualquier punto de partida) en exactamente un paso.
- La experiencia ha demostrado que el método de Newton puede ser poco confiable para funciones no cuadráticas.
- Es decir, el paso de Newton a menudo será grande cuando  $\underline{x}^{(0)}$  está lejos de  $\underline{x}^*$ , y existe la posibilidad real de divergencia.
- Es posible modificar el método de una manera lógica y simple para asegurar el descenso agregando una búsqueda de línea como en el método del descenso mas pronunciado.

$$\alpha^* = \min f \left( \underline{x}^{(k)} - \alpha^k \left( \nabla^2 f \left( \underline{x}^{(k)} \right) \right)^{-1} \nabla f \left( \underline{x}^{(k)} \right) \right)$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha^* \left( \nabla^2 f \left( \underline{x}^{(k)} \right) \right)^{-1} \nabla f \left( \underline{x}^{(k)} \right)$$

- El método de Newton II (o modificado) es confiable y eficiente cuando la primera y la segunda derivada se calculan de manera precisa y económica (desde el punto de vista computacional).
- La mayor dificultad es la necesidad de calcular y resolver en cada iteración el conjunto de ecuaciones lineales que involucra el término de la matriz Hessiana.

$$\underline{x}^{(0)} \in D \subseteq \mathbb{R}^n; \nabla f(\underline{x}); \nabla^2 f(\underline{x}^{(k)}); k = 0$$

$$\underline{d} = -\left(\nabla^2 f(\underline{x}^{(k)})\right)^{-1} \nabla f(\underline{x}^{(k)})$$

También puede resolverse como un SEL

$$\alpha^* = \text{Min. } f(\underline{x}^{(k)} + \alpha \underline{d})$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^* \underline{d}$$



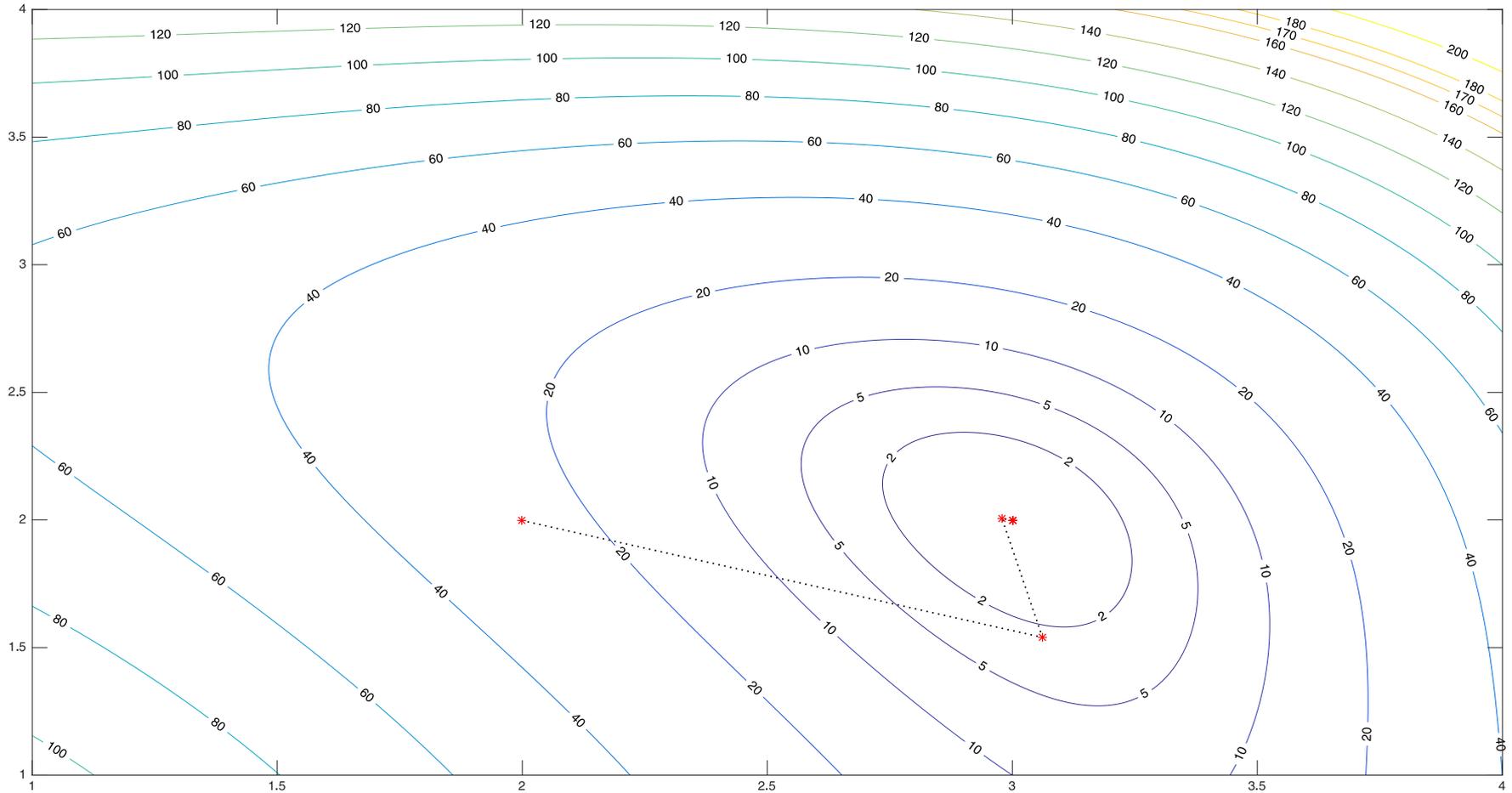
$$\|\alpha^* \underline{d}\| \leq tol$$

no

$$k = k + 1$$

si

$$\underline{x}^* = \underline{x}^{(k+1)}$$

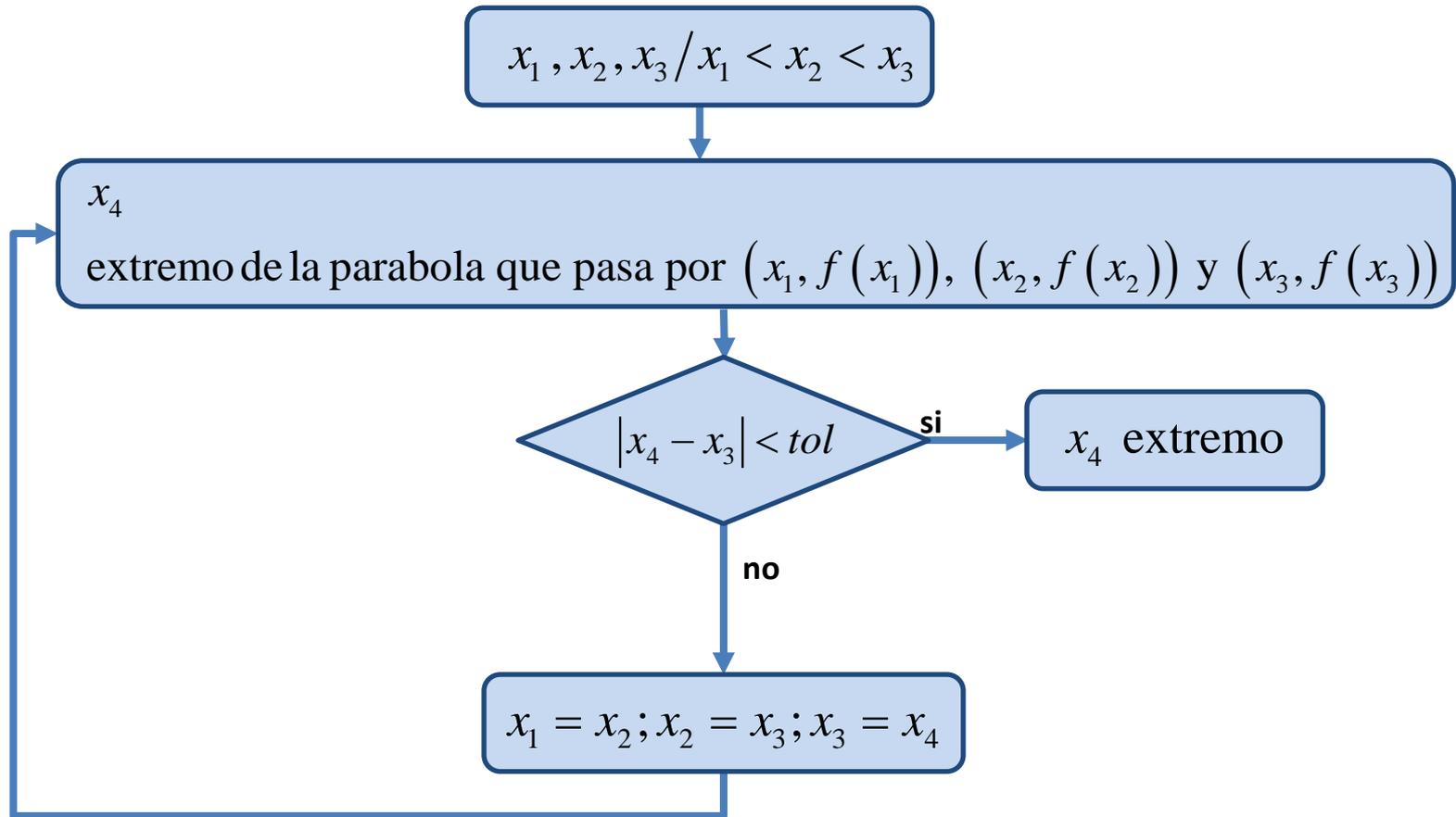


$$\underline{x}^{(0)} = [2 \quad 2]^T \Rightarrow \underline{x}^{(5)} = [3 \quad 2]^T$$

$$tol = 10^{-5}$$

- Programar las “function” necesarias en las dimensiones correctas.
- Manejar las incógnitas como un único vector.
- Tener una buena implementación de los algoritmos de búsqueda unidimensional.
- Una buena herramienta es programar una “function” que realice la búsqueda de línea.

Volvemos a Optimización Unidimensional...



$$\text{Min. } f \left( \underline{x}^{(i)} + \alpha \underline{d} \right)$$

Función a minimizar

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Solución actual

$$\underline{x}^{(i)} \in \mathbb{R}^n$$

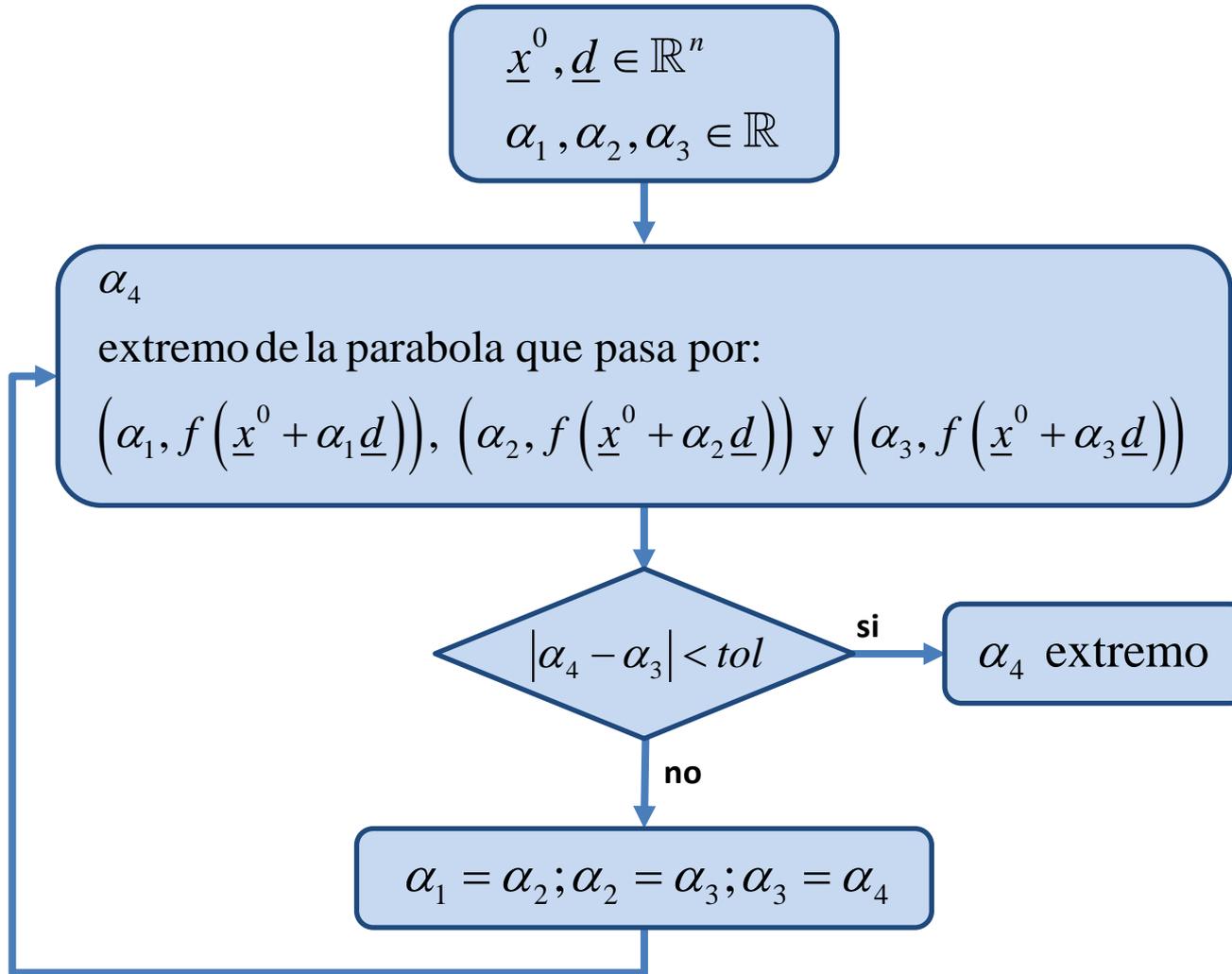
Tamaño del salto

$$\alpha \in \mathbb{R}$$

Dirección de búsqueda

$$\underline{d} \in \mathbb{R}^n$$

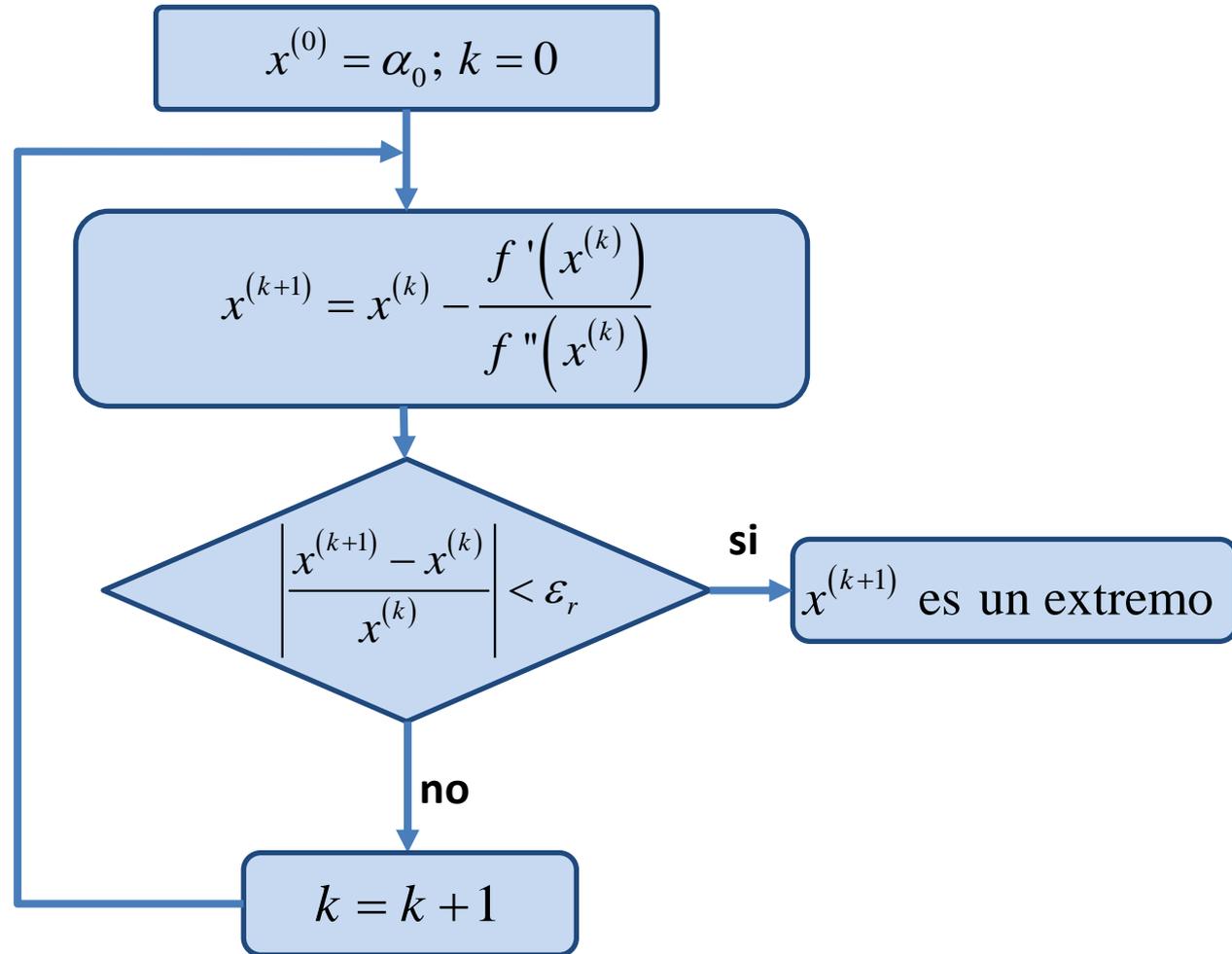
Objetivo de la búsqueda de línea



```

function alfa=linesearchMIPS(fun, xk,d,a_0)
x1 = a_0*0.98;
x2 = a_0;
x3 = a_0*1.02;
for i= 1:100
    A=[x1^2 x1 1
        x2^2 x2 1
        x3^2 x3 1 ];
    b=[fun(xk + x1*d)
        fun(xk + x2*d)
        fun(xk + x3*d)];
    xx=A\b;
    x4=-xx(2)/(2*xx(1));
    if norm(x4-x3) < 1e-6
        alfa=x4;
        break
    end
    x1=x2; x2=x3; x3=x4;
end
if i==100
    alfa=[];
end
endfunction

```



$$\text{Min. } f \left( \underline{x}^{(k)} + \alpha \underline{d} \right)$$

Función a minimizar

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Solución actual

$$\underline{x}^{(i)} \in \mathbb{R}^n$$

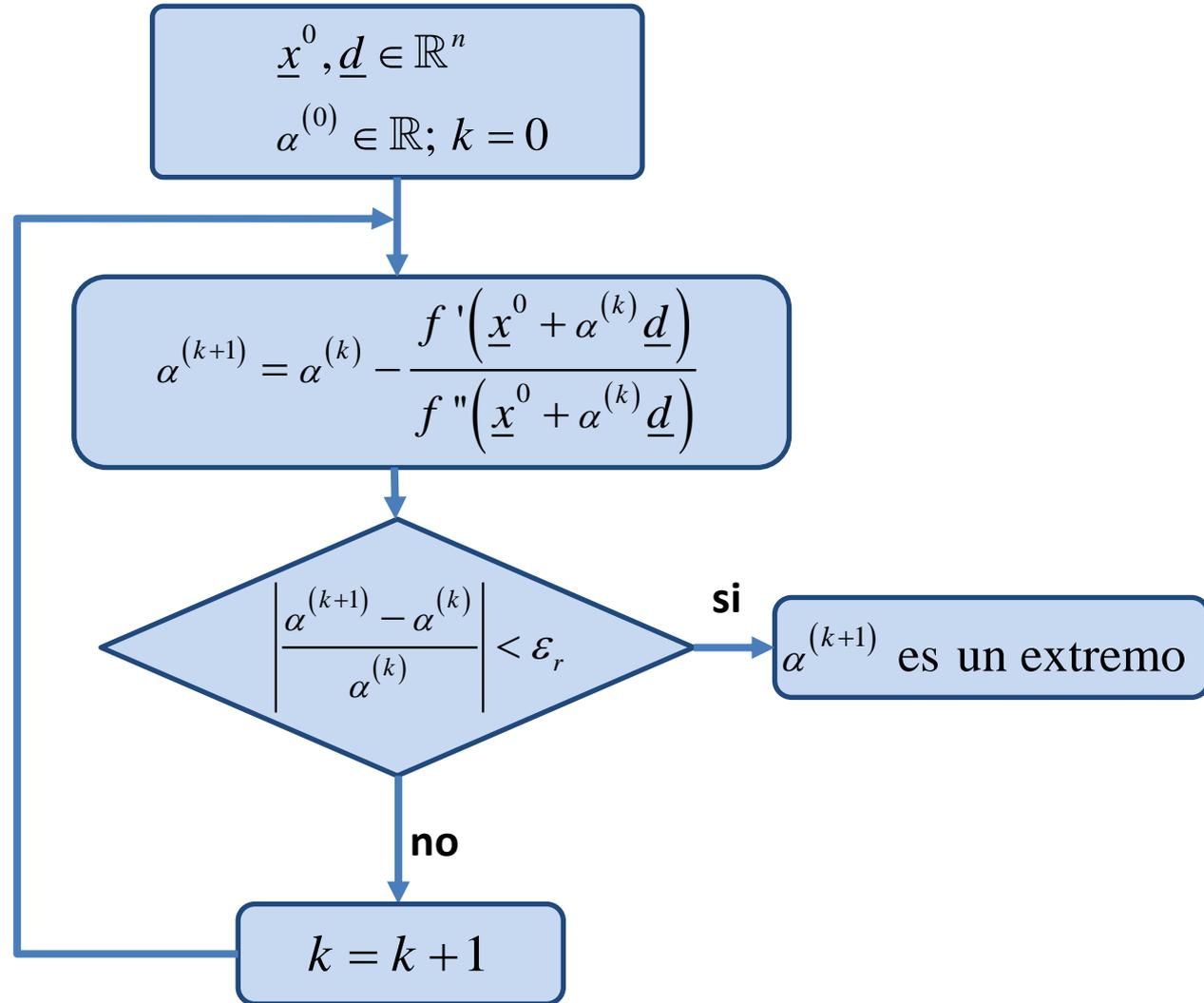
Tamaño del salto

$$\alpha \in \mathbb{R}$$

Dirección de búsqueda

$$\underline{d} \in \mathbb{R}^n$$

Objetivo de la búsqueda de línea



$$f'(\underline{x}^{(k)} + \alpha \underline{d}) = \frac{df(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha}$$

$$\frac{df(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha} = \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial \alpha} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial \alpha} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial \alpha}$$

$$\frac{df(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha} = \frac{\partial f}{\partial x_1} d_1 + \frac{\partial f}{\partial x_2} d_2 + \dots + \frac{\partial f}{\partial x_n} d_n$$

$$\frac{df(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha} = \nabla f(\underline{x}^{(k)} + \alpha \underline{d})^T \underline{d}$$

$$\frac{df(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha} = \frac{\partial f}{\partial x_1} d_1 + \frac{\partial f}{\partial x_2} d_2 + \dots + \frac{\partial f}{\partial x_n} d_n$$

$$\begin{aligned} \frac{d^2 f(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha^2} &= \left( \frac{\partial^2 f}{\partial x_1^2} \frac{\partial x_1}{\partial \alpha} + \frac{\partial^2 f}{\partial x_1 \partial x_2} \frac{\partial x_2}{\partial \alpha} + \dots + \frac{\partial^2 f}{\partial x_1 \partial x_n} \frac{\partial x_n}{\partial \alpha} \right) d_1 \\ &+ \left( \frac{\partial^2 f}{\partial x_2 \partial x_1} \frac{\partial x_1}{\partial \alpha} + \frac{\partial^2 f}{\partial x_2^2} \frac{\partial x_2}{\partial \alpha} + \dots + \frac{\partial^2 f}{\partial x_2 \partial x_n} \frac{\partial x_n}{\partial \alpha} \right) d_2 \\ &+ \dots \\ &+ \left( \frac{\partial^2 f}{\partial x_n \partial x_1} \frac{\partial x_1}{\partial \alpha} + \frac{\partial^2 f}{\partial x_n \partial x_2} \frac{\partial x_2}{\partial \alpha} + \dots + \frac{\partial^2 f}{\partial x_n^2} \frac{\partial x_n}{\partial \alpha} \right) d_n \end{aligned}$$

$$\begin{aligned}
 \frac{d^2 f(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha^2} &= \left( \frac{\partial^2 f}{\partial x_1^2} d_1 + \frac{\partial^2 f}{\partial x_1 \partial x_2} d_2 + \dots + \frac{\partial^2 f}{\partial x_1 \partial x_n} d_n \right) d_1 \\
 &+ \left( \frac{\partial^2 f}{\partial x_2 \partial x_1} d_1 + \frac{\partial^2 f}{\partial x_2^2} d_2 + \dots + \frac{\partial^2 f}{\partial x_2 \partial x_n} d_n \right) d_2 \\
 &+ \dots \\
 &+ \left( \frac{\partial^2 f}{\partial x_n \partial x_1} d_1 + \frac{\partial^2 f}{\partial x_n \partial x_2} d_2 + \dots + \frac{\partial^2 f}{\partial x_n^2} d_n \right) d_n
 \end{aligned}$$

$$\frac{d^2 f(\underline{x}^{(k)} + \alpha \underline{d})}{d\alpha^2} = \left( \frac{\partial^2 f}{\partial x_1^2} d_1 + \frac{\partial^2 f}{\partial x_1 \partial x_2} d_2 + \dots + \frac{\partial^2 f}{\partial x_1 \partial x_n} d_n \right) d_1 + \left( \frac{\partial^2 f}{\partial x_2 \partial x_1} d_1 + \frac{\partial^2 f}{\partial x_2^2} d_2 + \dots + \frac{\partial^2 f}{\partial x_2 \partial x_n} d_n \right) d_2$$

$$+ \dots + \left( \frac{\partial^2 f}{\partial x_n \partial x_1} d_1 + \frac{\partial^2 f}{\partial x_n \partial x_2} d_2 + \dots + \frac{\partial^2 f}{\partial x_n^2} d_n \right) d_n$$

$\frac{\partial^2 f(\underline{x}_0)}{\partial x_1 \partial x_1}$	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_1 \partial x_2}$	...	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_1 \partial x_n}$	$d_1$
$\frac{\partial^2 f(\underline{x}_0)}{\partial x_2 \partial x_1}$	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_2 \partial x_2}$	...	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_2 \partial x_n}$	$d_2$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\frac{\partial^2 f(\underline{x}_0)}{\partial x_n \partial x_1}$	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_n \partial x_2}$	...	$\frac{\partial^2 f(\underline{x}_0)}{\partial x_n \partial x_n}$	$d_n$

$d_1 \quad d_2 \quad \dots \quad d_n$

$\underline{d}^T \nabla^2 f(\underline{x}^{(k)} + \alpha \underline{d}) \underline{d}$

```
function alfa=linearsearchMN(Fun, Grad, Hess, xk, d, a_0)
    a=a_0;
    for i=1:100
        fp = Grad(xk + a*d)'*d;
        fpp= d'*Hess(xk + a*d)*d;
        a=a-fp/fpp;
        if norm(fp/fpp) < 1e-5
            alfa=a;
            break
        end
    end
    if i==100
        alfa=[];
    end
endfunction
```

$$\text{Min. } f(\underline{x}) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2(x_1^2 + x_2 - 11)2x_1 + 2(x_1 + x_2^2 - 7) \\ 2(x_1^2 + x_2 - 11) + 2(x_1 + x_2^2 - 7)2x_2 \end{bmatrix}$$

$$\nabla^2 f(\underline{x}) = \begin{bmatrix} 8x_1^2 + 4(x_1^2 + x_2 - 11) + 2 & 4x_1 + 4x_2 \\ 4x_1 + 4x_2 & 2 + 8x_2^2 + 4(x_1 + x_2^2 - 7) \end{bmatrix}$$

function **f=fun(x)**

$$f = (x(1)^2 + x(2) - 11)^2 + (x(1) + x(2)^2 - 7)^2$$

endfunction

function **Gf=gradf(x)**

$$Gf(1,1) = 2 * (x(1)^2 + x(2) - 11) * (2 * x(1)) + 2 * (x(1) + x(2)^2 - 7)$$

$$Gf(2,1) = 2 * (x(1)^2 + x(2) - 11) + 2 * (x(1) + x(2)^2 - 7) * (2 * x(2))$$

endfunction

function **Hf=hessf(x)**

$$Hf(1,1) = 2 * (2 * x(1)) * (2 * x(1)) + 2 * (x(1)^2 + x(2) - 11) * (2) + 2$$

$$Hf(1,2) = 2 * (2 * x(1)) + 2 * (2 * x(2))$$

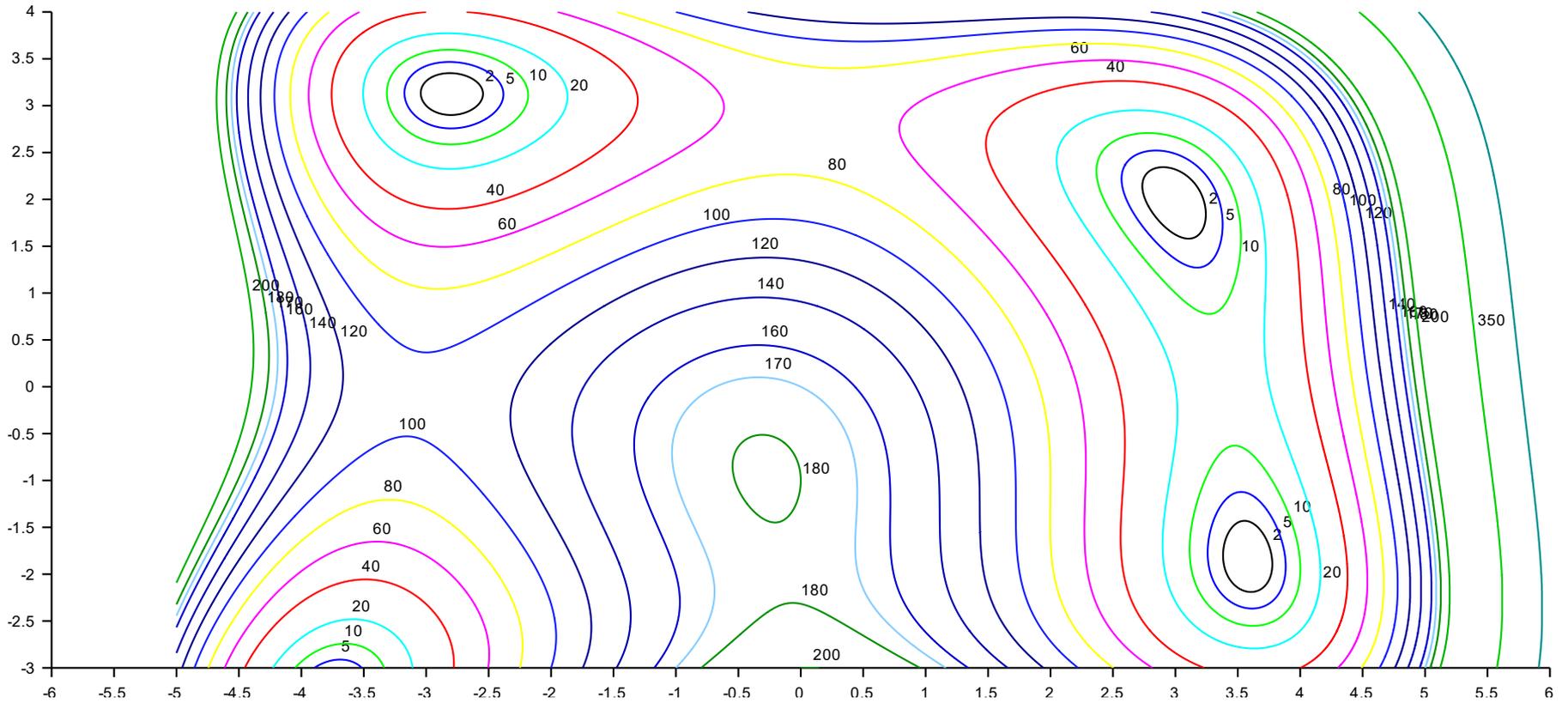
$$Hf(2,1) = 2 * (2 * x(1)) + 2 * (2 * x(2))$$

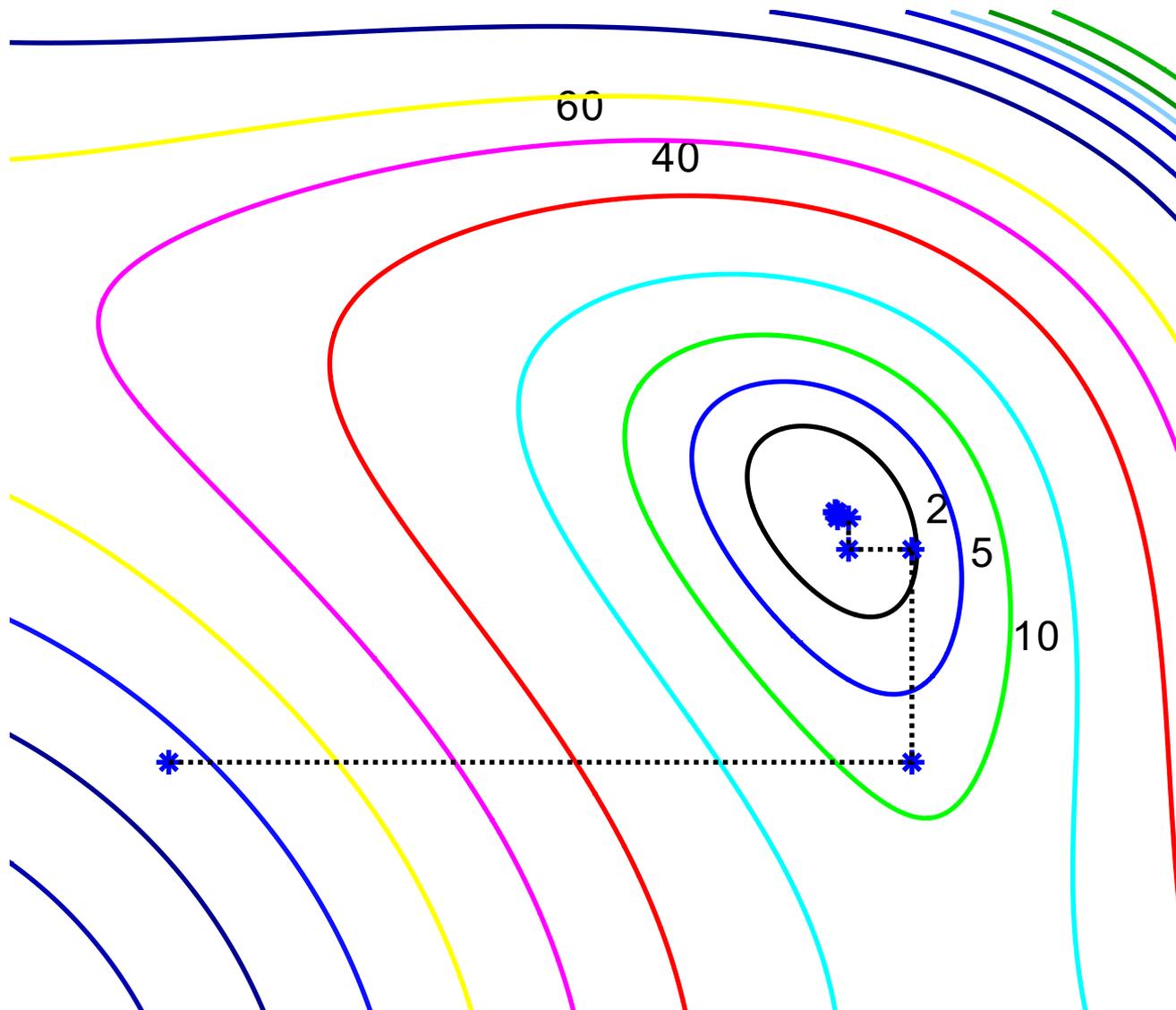
$$Hf(2,2) = 2 + 2 * (2 * x(2)) * (2 * x(2)) + 2 * (x(1) + x(2)^2 - 7) * (2)$$

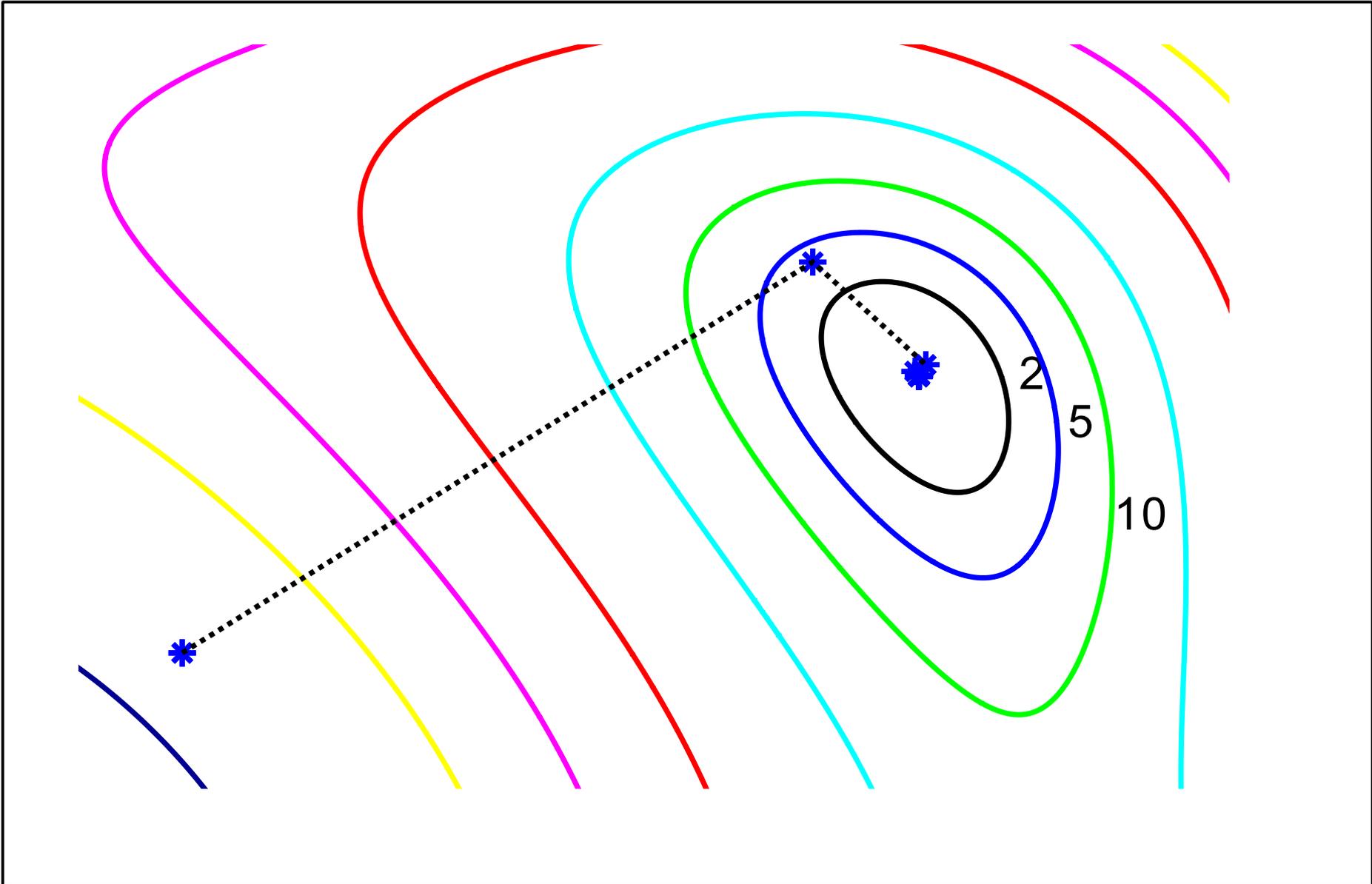
endfunction

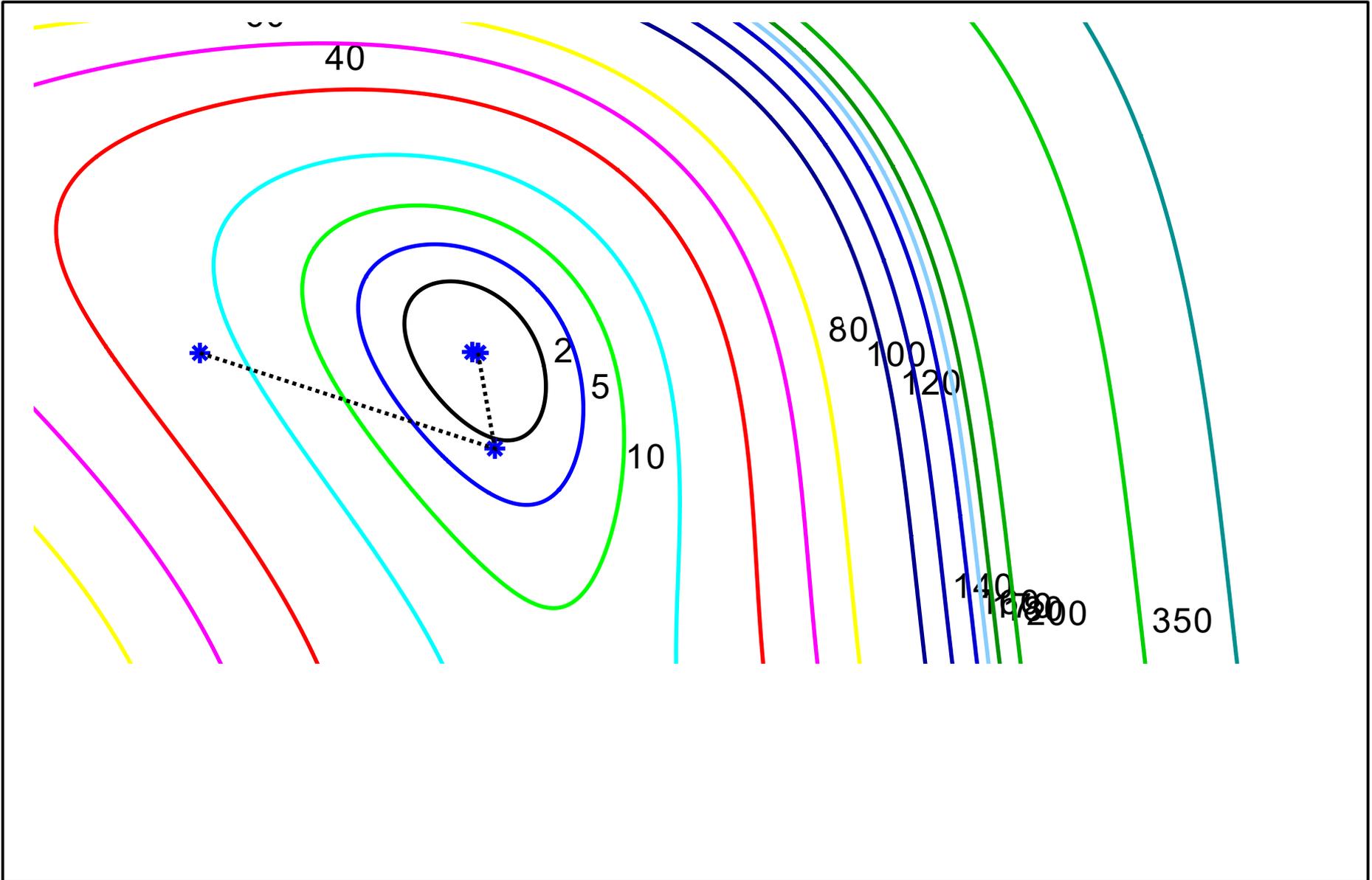
```
x1=linspace(-5,6,1000);
x2=linspace(-3,4,1000);
for i=1:length(x1)
    for j=1:length(x2)
        z(i,j)=fun([x1(i);x2(j)]);
    end
end
```

```
contour(x1,x2,z,[2 5 10 20:20:160 170 180 200 350 500])
```









$$\text{Min. } f(\underline{x}) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$$