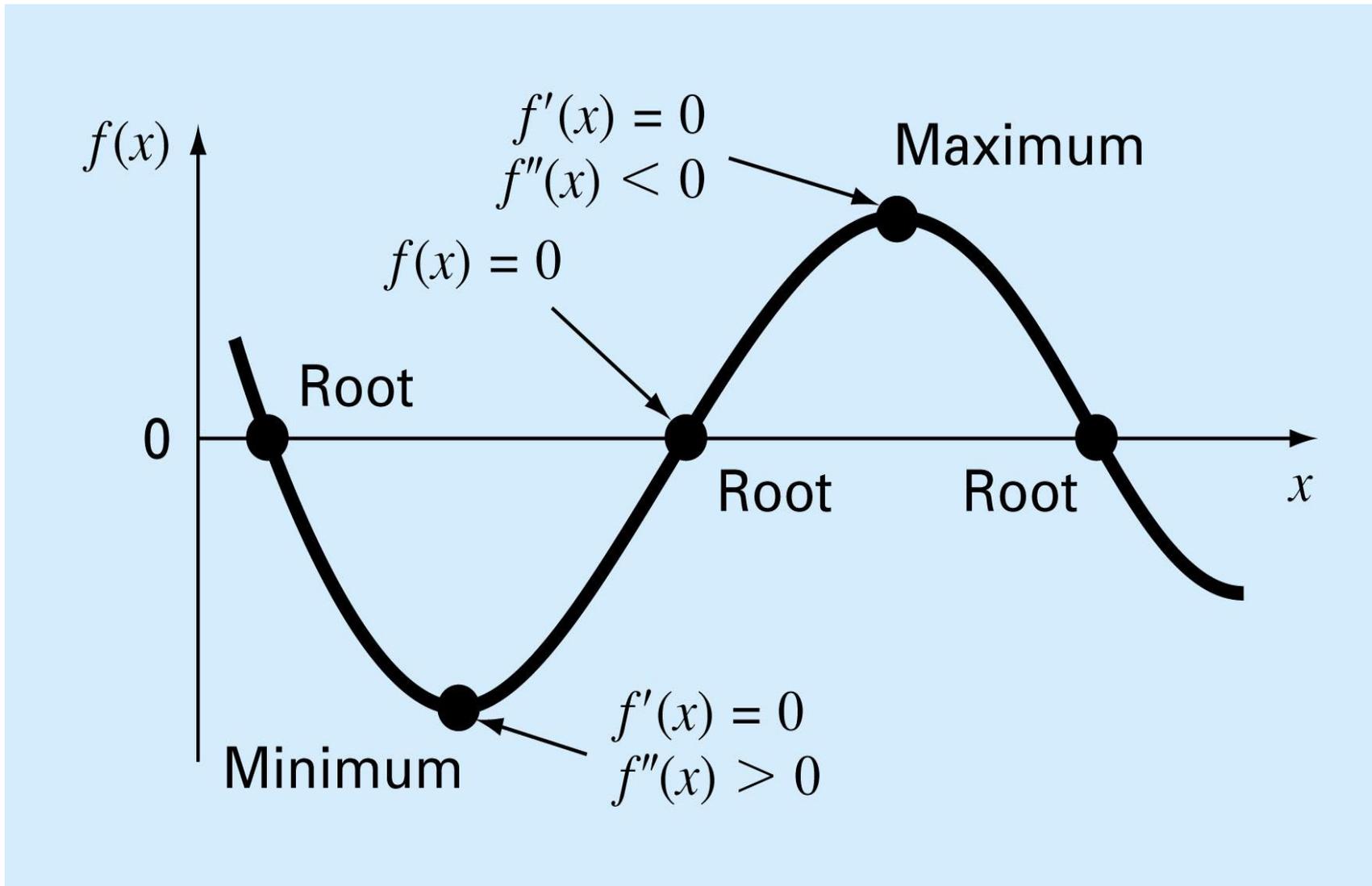


Optimización

Profesor: Dr. Alejandro S. M. Santa Cruz
JTP: Dr. Juan Ignacio Manassaldi
Auxiliar: Srta. Amalia Rueda

Esquema

- La determinación de raíces y la optimización están relacionadas, ambas involucran la estimación y la búsqueda de un punto del dominio de una función de características particulares.
- La diferencia fundamental es:
 - La determinación de raíces es la búsqueda de los ceros de una función o de varias funciones.
 - La optimización es la determinación del máximo o el mínimo de una función de una o varias variables.



Fundamentos Matemáticos

- Un problema de optimización o de programación matemática generalmente se enuncia como:

Encuentre x que minimiza o maximiza $f(x)$ sujeto a:

$$d_i(x) \leq a_i \quad i = 1, 2, \dots, m^*$$

$$e_i(x) = b_i \quad i = 1, 2, \dots, p^*$$

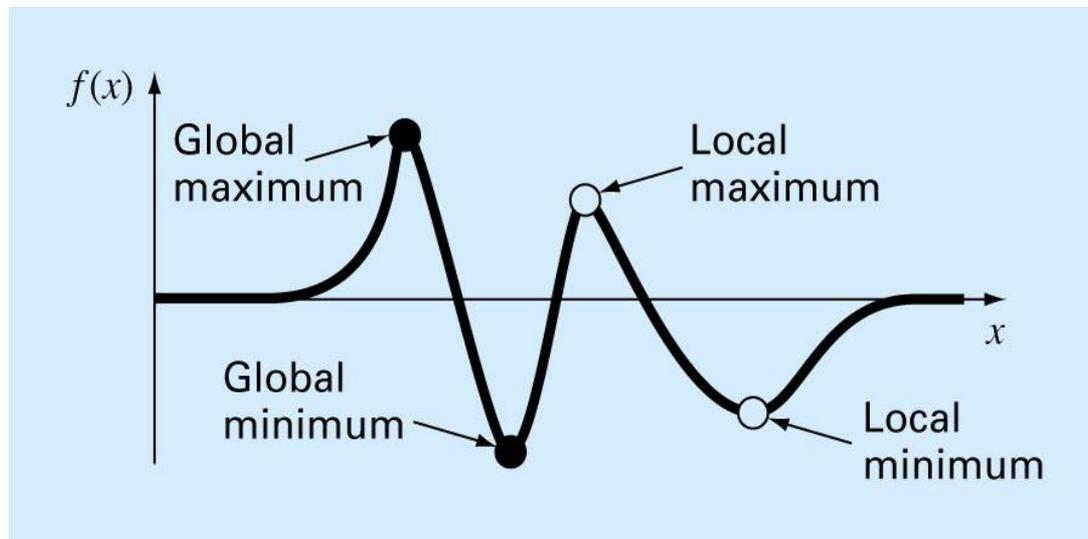
Donde x es un *vector de diseño n -dimensional*, $f(x)$ es la *función objetivo*, $d_i(x)$ son restricciones de desigualdad, $e_i(x)$ son restricciones de igualdad y a_i y b_i son constantes.

Fundamentos Matemáticos

- Los problemas de optimización pueden clasificarse según la forma de $f(x)$:
 - Si $f(x)$ y las restricciones son lineales, tenemos programación lineal.
 - Si $f(x)$ es cuadrática y las restricciones son lineales, tenemos programación cuadrática.
 - Si $f(x)$ no es lineal o cuadrática y/o las restricciones no son lineales, tenemos una programación no lineal.
- Cuando se incluyen las ecuaciones (*), tenemos un problema de optimización restringido; de lo contrario, es un problema de optimización sin restricciones.

Fundamentos Matemáticos

- En funciones multimodales, pueden ocurrir óptimos tanto locales como globales. En casi todos los casos, nos interesa encontrar el valor absoluto más alto o más bajo de una función.



¿Cómo distinguimos un óptimo global de uno local?

- Graficando para conocer mejor el comportamiento de la función.
- Usando estimaciones iniciales generadas aleatoriamente y eligiendo el más grande de los óptimos como el global.
- Perturbando la estimación inicial para ver si la rutina devuelve un mejor punto o el mismo extremo local (máximo o mínimo).

Ejemplo

- Supongamos que estamos diseñando latas para envasar cerveza. Queremos minimizar la cantidad de metal que necesitamos para envasar un determinado volumen de cerveza.
- En otras palabras, queremos minimizar la relación superficie/volumen.
- Obviamente una esfera satisfaría esta relación, pero una lata de estas características sería difícil de fabricar y más aún de abrir.
- Suponemos entonces que la forma está dada y es un cilindro.

Ejemplo

- Superficie metálica del cilindro: $A = 2\pi R^2 + 2\pi Rh$
Dos tapas *Superficie lateral*
- Volumen del cilindro: $V = \pi R^2 h$
- Queremos minimizar A sujeta a la restricción que el volumen encerrado sea $V_0 = 250 \text{ cm}^3$.
- Así planteado, este es un problema de optimización restringida 2-D (depende de 2 variables, R y h).
- Podemos transformar este problema a un problema de optimización 1-D no restringida reconociendo que:

$$h = \frac{V_0}{\pi R^2} \therefore A = 2\pi R^2 + \frac{2V_0}{R}$$

Ejemplo

- Planteamos la condición necesaria de existencia de extremo:

$$\frac{dA}{dR} = \frac{d\left(2\pi R^2 + \frac{2V_0}{R}\right)}{dR} = 4\pi R - \frac{2V_0}{R^2} = 0$$

- Despejando R obtenemos:

$$R = \left(\frac{V_0}{2\pi}\right)^{1/3} = \left(\frac{250}{2\pi}\right)^{1/3} = 3.41392031 \text{ cm}$$

- Finalmente resulta:

$$h = \frac{V_0}{\pi R^2} = \frac{250}{\pi(3.41392031)^2} = 6.82784065 \text{ cm}$$

Planteo General del Problema

- Sea $f(x)$ una función real continua y diferenciable, definida sobre un dominio S contenido en R^n ,

$$f : \underline{x} \in S \subseteq R^n \rightarrow R$$

- Planteamos el problema:

$$\min_{\underline{x}} f(\underline{x}) \equiv f(x_1, x_2, \dots, x_n) \text{ o } \max_{\underline{x}} f(\underline{x}) \equiv f(x_1, x_2, \dots, x_n)$$

sobre un dominio S en un espacio n -dimensional.

- Para el problema de la lata, el problema se escribiría así:

$$\min_{\underline{x}} f(x_1, x_2) \text{ con } f(x_1, x_2) = 2\pi x_1^2 + 2\pi x_1 x_2$$

$$\text{sujeto a } \underline{x} \in S = \left\{ (x_1, x_2) \mid \pi x_1^2 x_2 = V_0 \right\}$$

Planteo General del Problema

- Si $S = \mathbb{R}^n$ (todo el espacio), entonces el problema es no restringido.
- Los $\underline{x} \in S$ son las soluciones factibles .
- El problema de regresión o de ajuste de datos experimentales es un problema de *optimización no restringida*.
- 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

Planteo General del Problema

- 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:

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

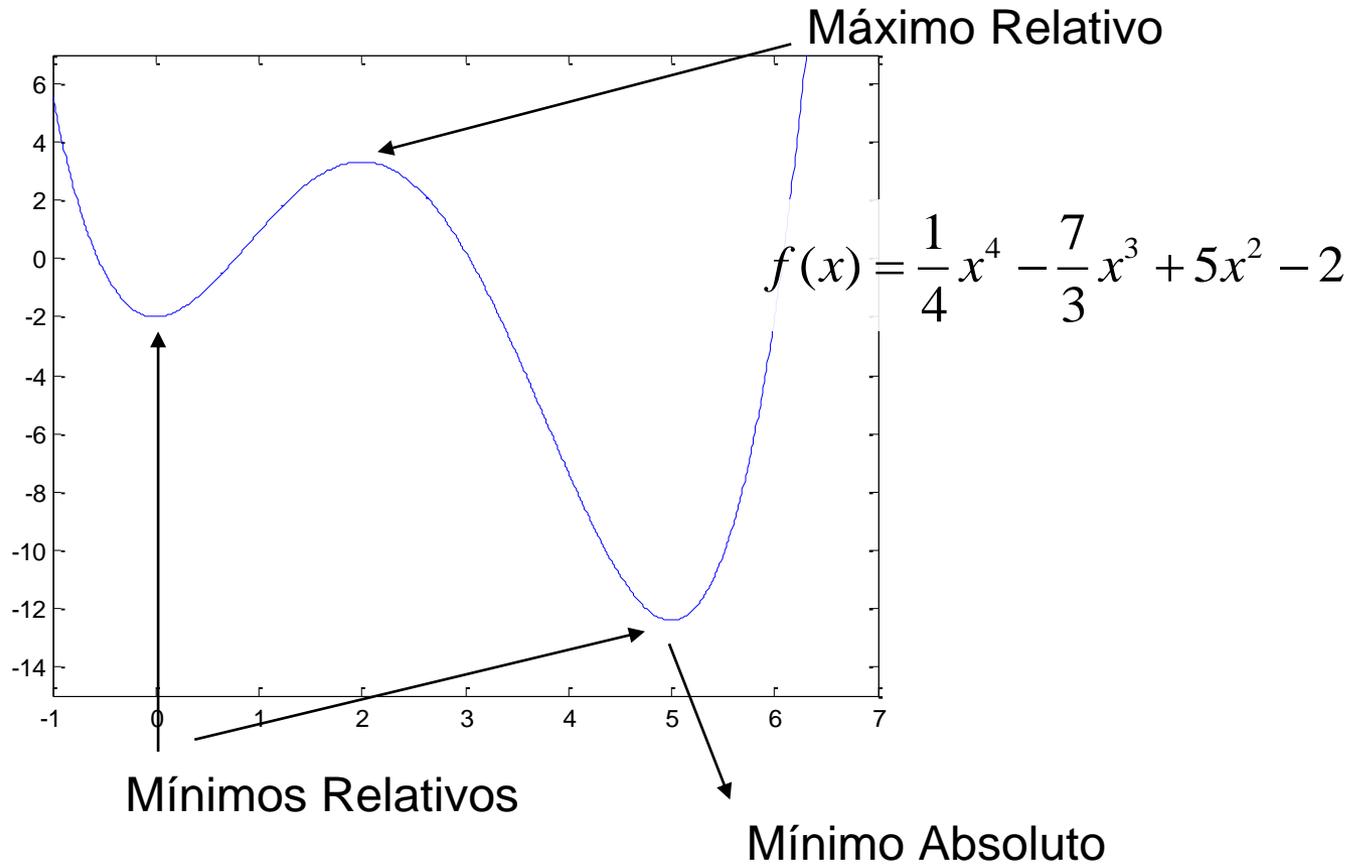
- Decimos que \underline{x}^* es un máximo local si:

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

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

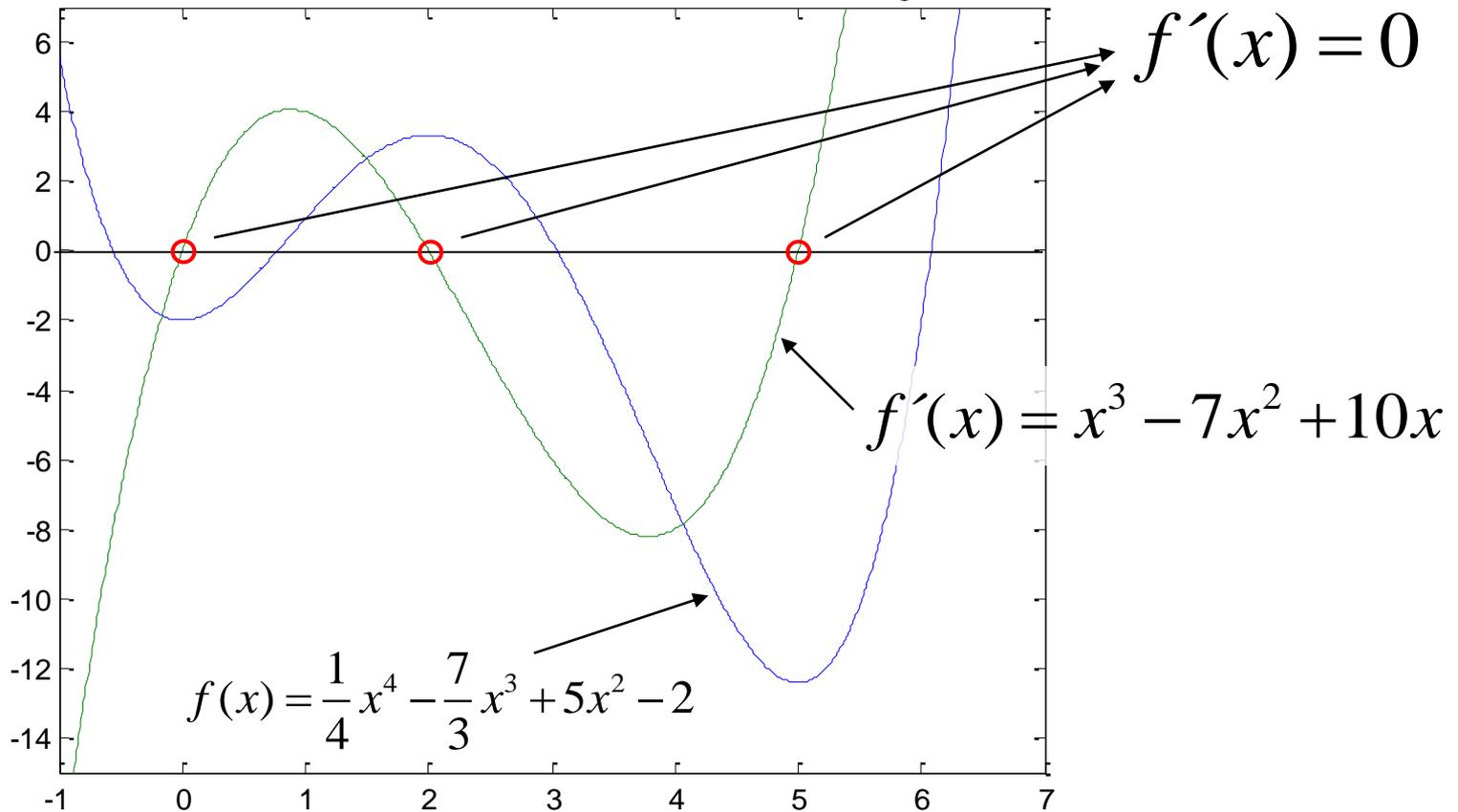
$$\underline{x}^* \in S \text{ tal que } f(\underline{x}^*) > f(\underline{x}) \forall \underline{x} \in S$$

Optimización 1-D No Restringida



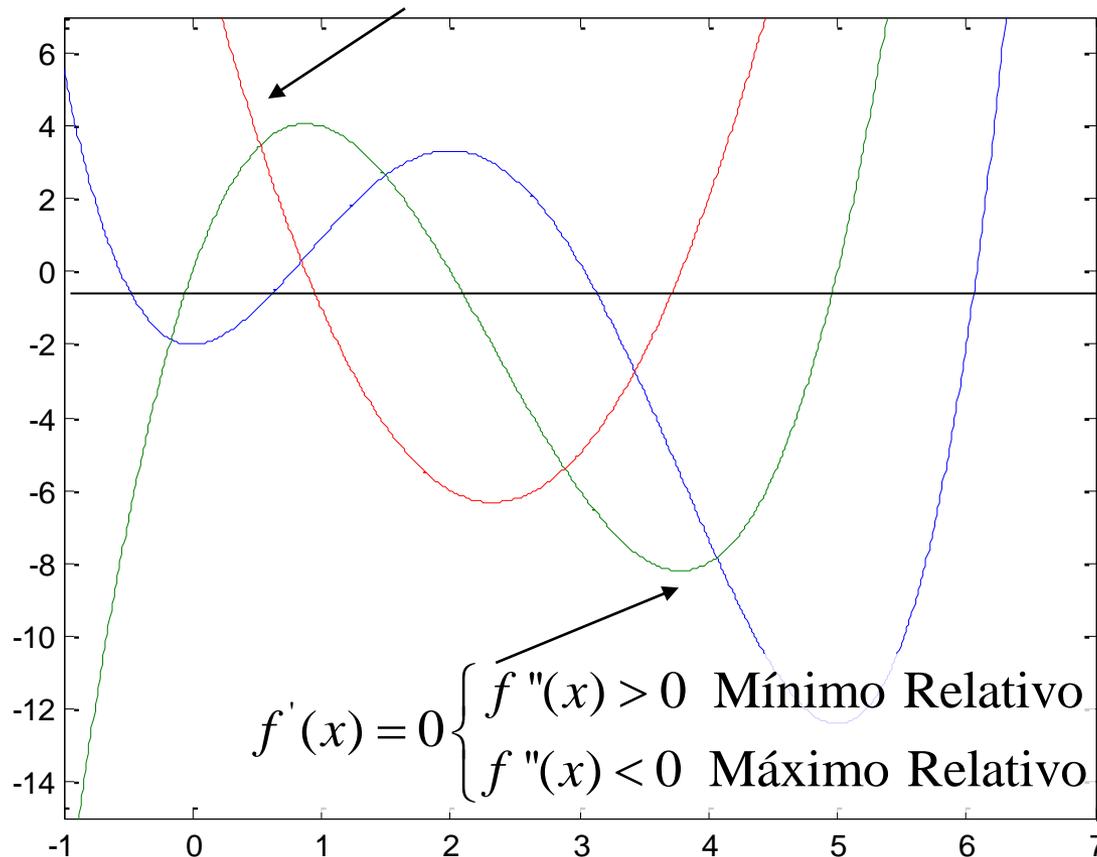
Optimización 1-D No Restringida

Condición necesaria de mínimo o máximo: $f'(x) = 0$



Optimización 1-D No Restringida

$$f''(x) = 3x^2 - 14x + 10$$



Optimización 1-D No Restringida

➤ Vamos a analizar tres enfoques básicos análogos a las tres técnicas de determinación de raíces.

Ellos son:

1. Método de Newton
2. Método de interpolación parabólica sucesiva
3. Método de la búsqueda dorada (Fibonacci)

Método de Newton

- Tenemos una función real de una variable.
- Podemos aproximarla con una parábola.
- Desarrollamos la función en serie de Taylor y nos quedamos con los tres primeros términos del desarrollo, esto es, incluimos al término cuadrático.

$$f(x) \approx f(x_k) + (x - x_k) f'(x_k) + \frac{1}{2} (x - x_k)^2 f''(x_k)$$

- Utilizamos la aproximación para obtener la siguiente estimación del punto crítico.
- Buscamos determinar x^* para el que $f'(x^*) = 0$. Luego,

$$f'(x) \approx f'(x_k) + (x - x_k) f''(x_k) = 0$$

Método de Newton

- Si x_{k+1} es una aproximación a la raíz de la derivada de la función, entonces resulta:

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

- Esto es virtualmente idéntico a la determinación de raíces de una función por el método de Newton, sólo que en este caso, en lugar de determinar la raíz de la función original, se determina la raíz de su derivada.
- Las consideraciones acerca de la convergencia del método son la mismas que se hicieron para el caso de las ecuaciones no-lineales.

En general, si buscamos $f(x) = 0$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \xrightarrow{\text{Para encontrar } f'(x) = 0} x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}$$

```

h=1e-5;
x=1;
i=1;
dx=1;
while abs(dx)>1e-4
    cbc=fo(x);
    cbp=fo(x+h);
    cbm=fo(x-h);
    fp=(cbp-cbm)/(2*h);
    fpp=(cbp+cbm-2*cbc)/h^2;
    dx=-fp/fpp;
    x=x+dx;
    i=i+1;
end
    
```

$$f(x) = \frac{1}{4}x^4 - \frac{7}{3}x^3 + 5x^2 - 2$$

```

function y=fo(x)
y=1/4*x.^4-7/3*x.^3+10/2*t.^2-2;
endfunction
    
```

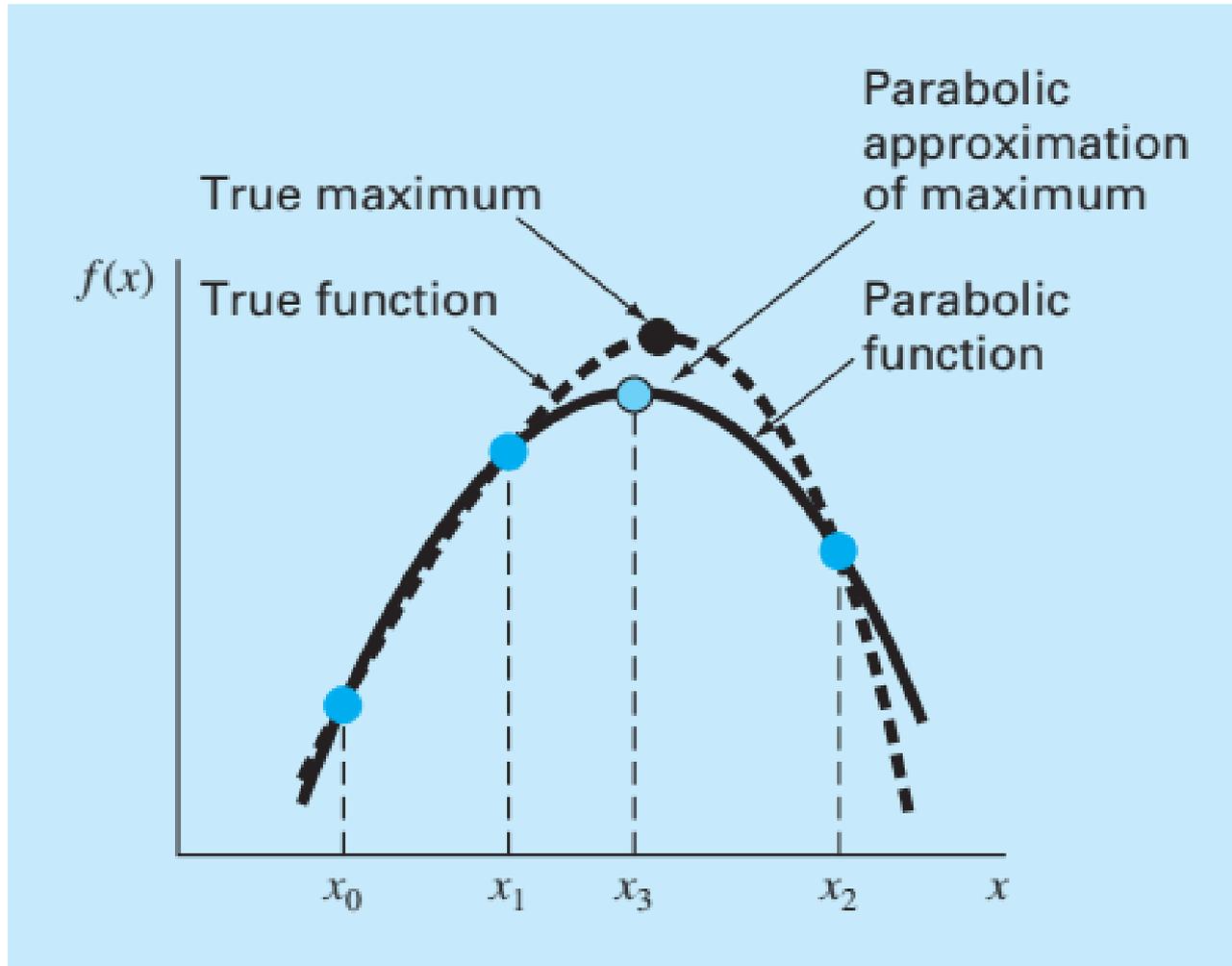
$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

Método de Interpolación Parabólica Sucesiva

- La interpolación parabólica aprovecha el hecho de que un polinomio de segundo orden a menudo proporciona una buena aproximación a la forma de $f(x)$ cerca de un óptimo.
- Así como sólo hay una línea recta que conecta dos puntos, sólo hay un polinomio cuadrático o parábola que conecta tres puntos.
- Por lo tanto, si tenemos tres puntos que encierran un óptimo, podemos ajustar una parábola a tales puntos.

Método de Interpolación Parabólica Sucesiva



Método de Interpolación Parabólica Sucesiva

- Ajustamos la parábola:

$$g(x) = ax^2 + bx + c$$

a los puntos:

$$\left[x_k, f(x_k) \right]; \left[x_{k-1}, f(x_{k-1}) \right]; \left[x_{k-2}, f(x_{k-2}) \right]$$

- En cada iteración resolvemos un SEAL, así:

$$f(x_{k-2}) = ax_{k-2}^2 + bx_{k-2} + c$$

$$f(x_{k-1}) = ax_{k-1}^2 + bx_{k-1} + c$$

$$f(x_k) = ax_k^2 + bx_k + c$$

Método de Interpolación Parabólica Sucesiva

- Una vez que determinamos a , b y c buscamos el punto crítico:

$$\frac{dg}{dx} = 2ax + b = 0$$

- Por consiguiente, el nuevo valor será:

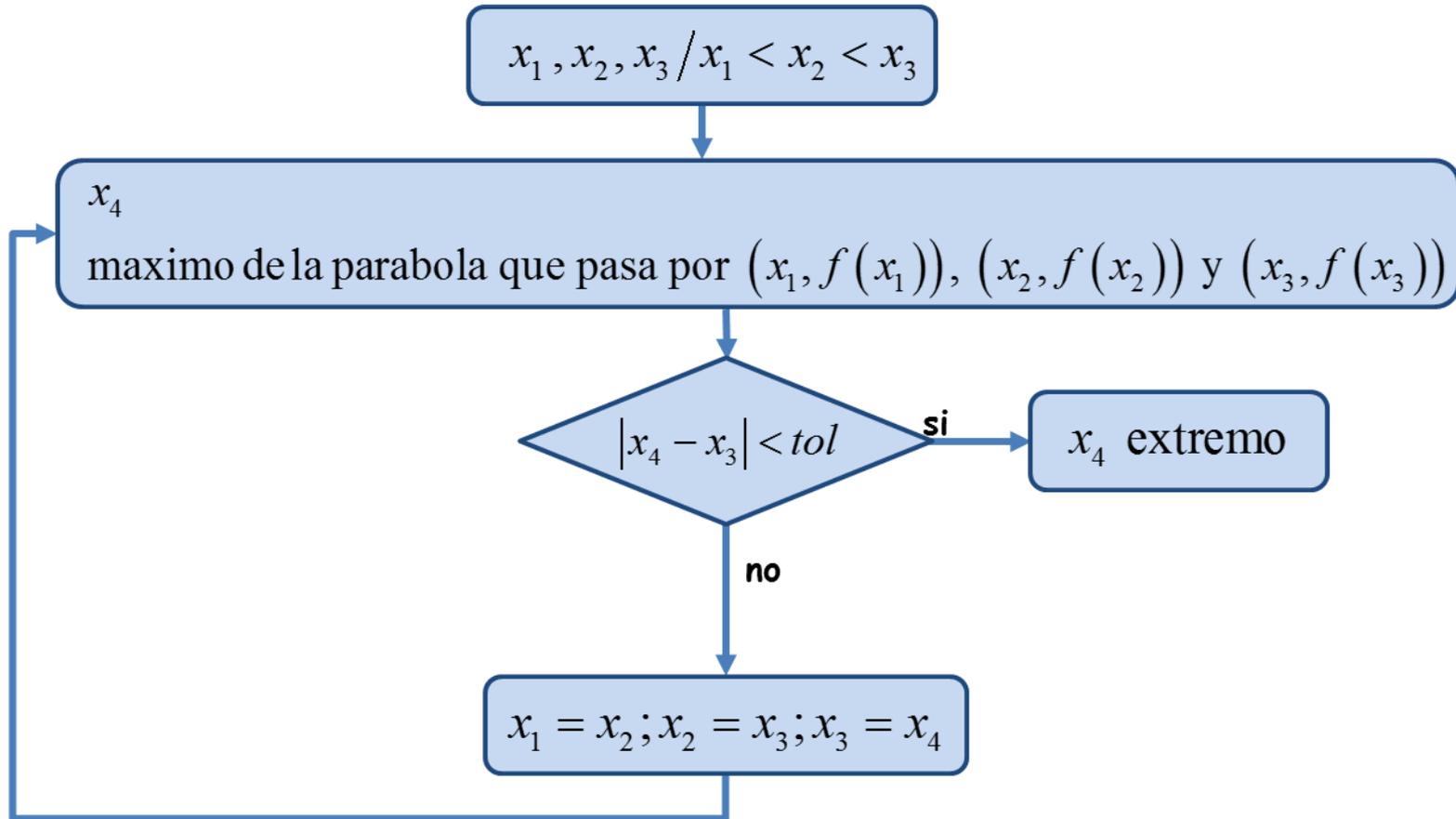
$$x_{k+1} = -\frac{b}{2a}$$

e iteramos hacia adelante manteniendo los últimos 3 puntos.

- Se puede demostrar, mediante algunas manipulaciones algebraicas, que el resultado anterior se puede obtener a través de la siguiente expresión:

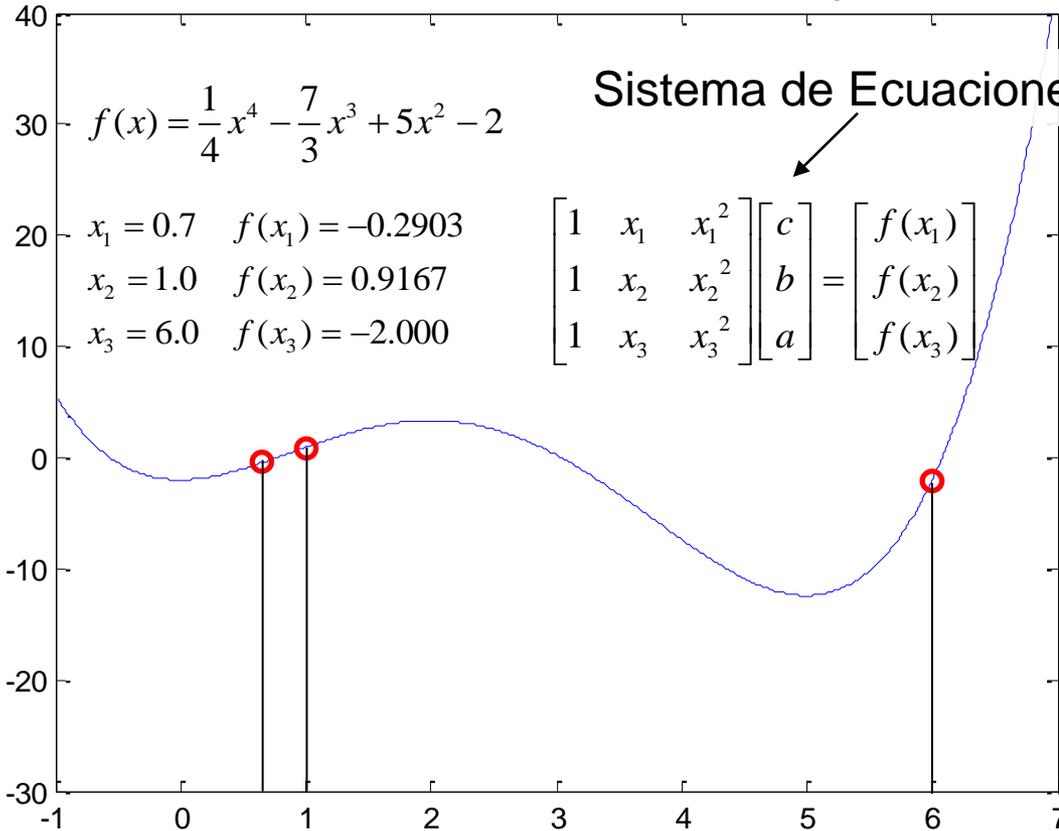
$$x_{k+1} = \frac{f(x_{k-2})(x_{k-1}^2 - x_k^2) + f(x_{k-1})(x_k^2 - x_{k-2}^2) + f(x_k)(x_{k-2}^2 - x_{k-1}^2)}{2f(x_{k-2})(x_{k-1} - x_k) + 2f(x_{k-1})(x_k - x_{k-2}) + 2f(x_k)(x_{k-2} - x_{k-1})}$$

Método de Interpolación Parabólica Sucesiva



MIPS: Ejemplo de Aplicación

$g(x) = ax^2 + bx + c \longrightarrow$ Ecuación de la parábola



$$f(x) = \frac{1}{4}x^4 - \frac{7}{3}x^3 + 5x^2 - 2$$

$$x_1 = 0.7 \quad f(x_1) = -0.2903$$

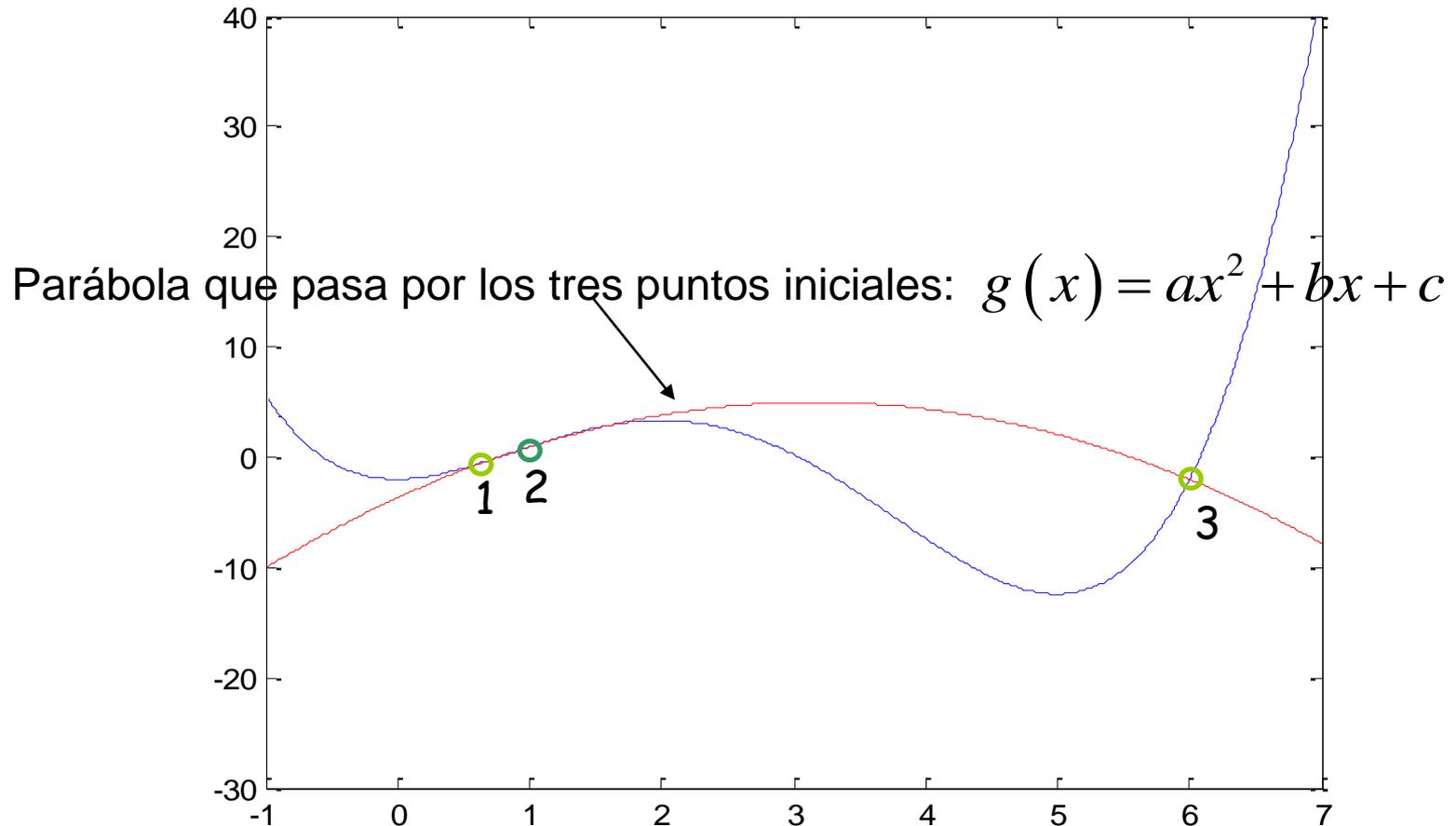
$$x_2 = 1.0 \quad f(x_2) = 0.9167$$

$$x_3 = 6.0 \quad f(x_3) = -2.000$$

Sistema de Ecuaciones Lineales (3x3)

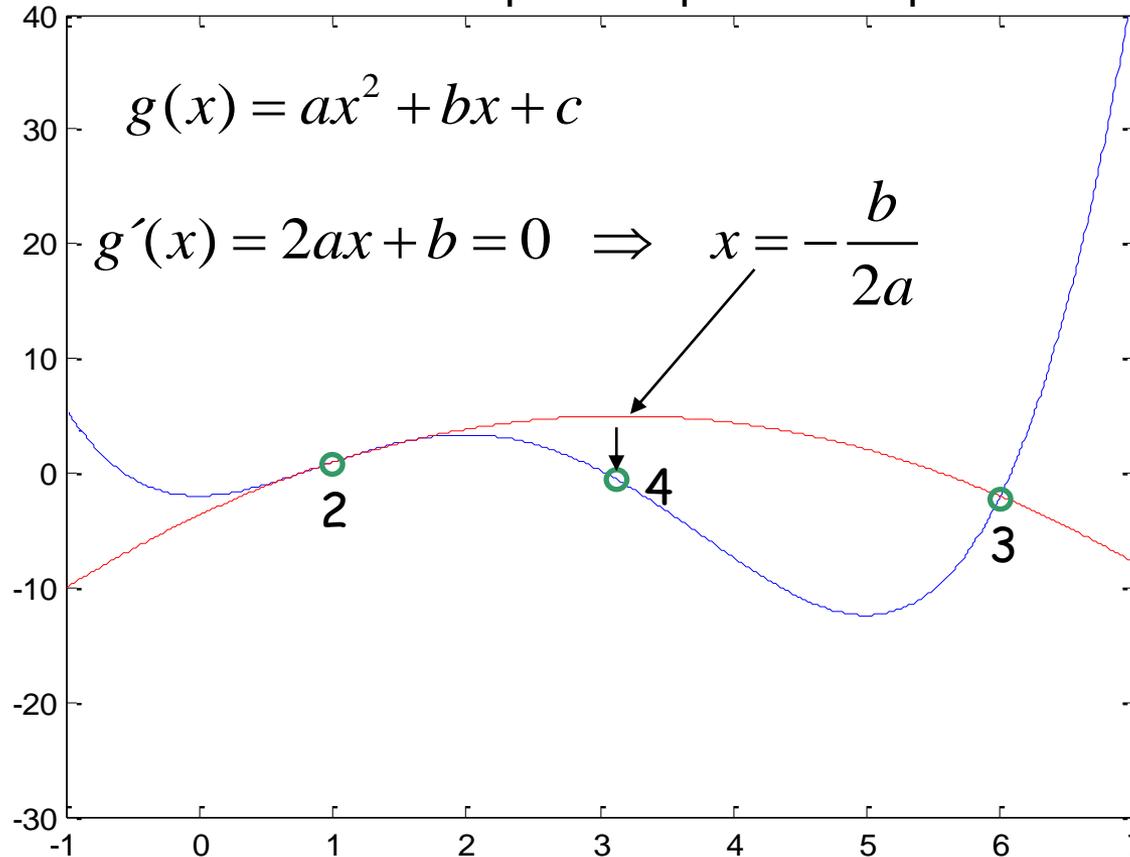
$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix}$$

MIPS: Ejemplo de Aplicación

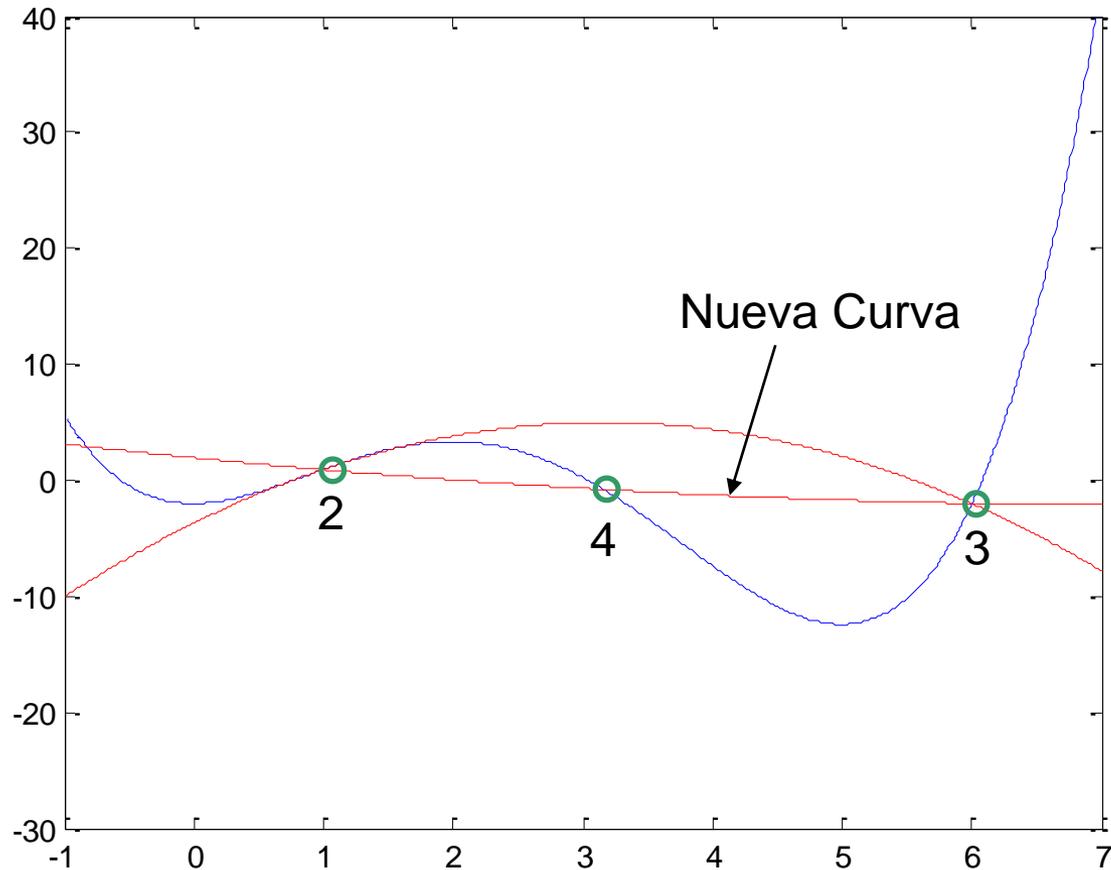


MIPS: Ejemplo de Aplicación

Determinación de un nuevo punto a partir del óptimo de la curva



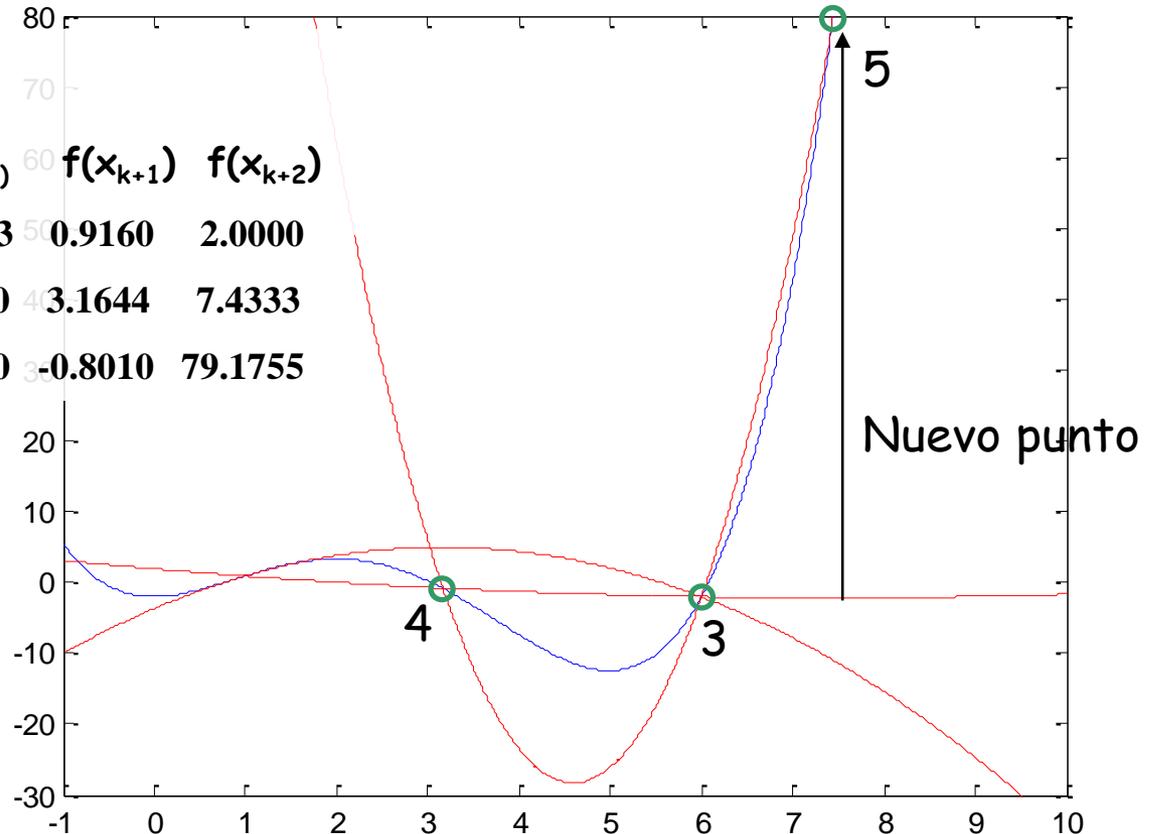
MIPS: Ejemplo de Aplicación



MIPS: Ejemplo de Aplicación

Secuencia de Puntos:

It.	x_k	x_{k+1}	x_{k+2}	$f(x_k)$	$f(x_{k+1})$	$f(x_{k+2})$
1-	0.7000	1.0000	6.0000	-0.2903	0.9160	2.0000
2-	1.0000	6.0000	3.1644	6.0000	3.1644	7.4333
3-	0.9166	-2.0000	-0.8010	-2.0000	-0.8010	79.1755

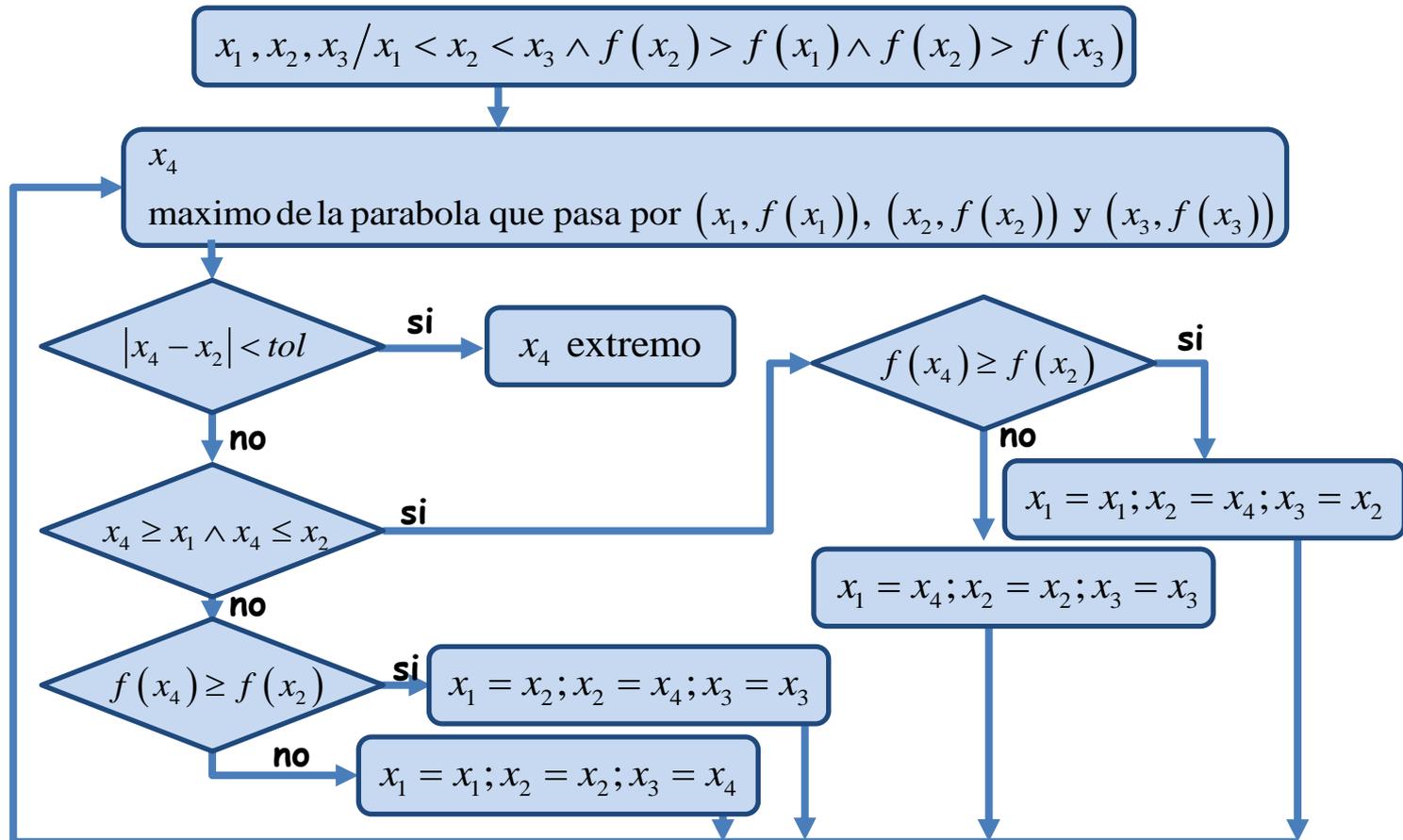


MIPS: Forma Alternativa de Implementar el Algoritmo

- En el ejemplo ocurre que los puntos de arranque elegidos están muy alejados entre sí, de modo que en el intervalo de búsqueda, la función presenta 2 extremos, un mínimo y un máximo.
- A menos que no se sepa si lo que se está buscando es un mínimo o un máximo, el algoritmo no garantiza a que extremo va a ir.
- Por este motivo, para usar el algoritmo bajo este procedimiento, resulta conveniente elegir los puntos iniciales próximos al óptimo (máximo o mínimo).
- Alternativamente, como se ilustra en el siguiente esquema, se puede emplear un enfoque similar a la bisección o a la búsqueda de la sección dorada.

MIPS: Forma Alternativa de Implementar el Algoritmo

Esta metodología converge a un máximo o a un mínimo, según lo que se esté buscando. Podemos decidir qué buscar. Para un máximo se tiene:



Método de la Búsqueda Dorada

- Es análogo al método de bisección para determinar raíces de ecuaciones no lineales.
- Pero, en este caso, en lugar de determinar un intervalo $[x_l, x_u]$ donde la función cambie de signo como en el método de bisección, se requiere que la función sea *unimodal* en el intervalo de búsqueda.
- Si una función es unimodal entonces tiene un único máximo o mínimo en el intervalo de búsqueda.
- Esto significa que (si estamos buscando un mínimo):

$$f'(x) \begin{cases} < 0 & \text{si } x < x^* \text{ y } x \in [x_l, x_u] \\ > 0 & \text{si } x > x^* \text{ y } x \in [x_l, x_u] \end{cases}$$

- Obsérvese que f'' puede cambiar de signo en este intervalo!
- Si $f'' = 0$ tanto el Método de Newton como el de Interpolación Parabólica Sucesiva fallarán, pero el Método de la Búsqueda Dorada no.

Método de la Búsqueda Dorada

- Supongamos que tenemos el intervalo $[x_l, x_u]$ y que la función es unimodal en él y contiene a un máximo.
- Evaluamos la función en 2 puntos interiores del intervalo. Por ejemplo:

$$x_1 = x_l + r(x_u - x_l)$$

$$x_2 = x_u - r(x_u - x_l)$$

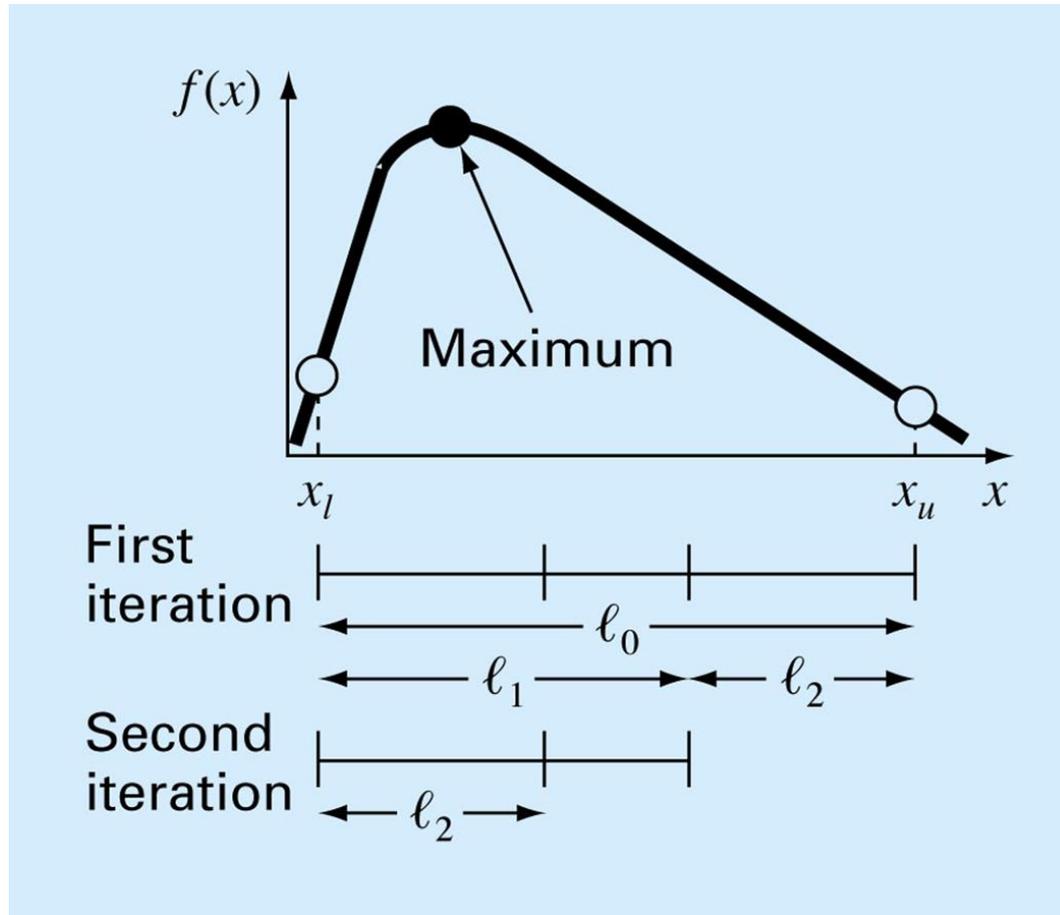
$$\text{con } 0.5 < r < 1$$

- Esto es, x_1 representa al punto interior derecho y x_2 al punto interior izquierdo en el intervalo $[x_l, x_u]$.
- Obsérvese que cualquiera sea el valor de r que satisfaga $0.5 < r < 1$ los dos puntos resultantes son simétricos respecto al punto medio del intervalo de búsqueda.
- Luego, evaluamos la función en estos dos puntos y nos quedamos con el que nos suministra el mayor valor (1era. estimación del máximo) y descartamos la parte del dominio donde no se encuentra el máximo.

Método de la Búsqueda Dorada

- En cada iteración debemos evaluar la función en 2 nuevos puntos y descartar. Esto tiene un costo computacional.
- Por consiguiente, se trata de determinar la relación de partición de modo que en cada iteración evaluemos sólo una vez la función utilizando uno de los valores calculados previamente.

Método de la Búsqueda Dorada



Método de la Búsqueda Dorada

$$l_0 = l_1 + l_2$$

$$\frac{l_1}{l_0} = \frac{l_2}{l_1}$$

- La primera condición especifica que la suma de las amplitudes, l_1 y l_2 , de los dos subintervalos debe ser igual a la longitud del intervalo original.
- La segunda condición nos dice que la razón de las longitudes debe ser igual.

$$\frac{l_1}{l_1 + l_2} = \frac{l_2}{l_1} \quad R = \frac{l_2}{l_1}$$

$$1 + R = \frac{1}{R} \quad R^2 + R - 1 = 0$$

$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0.61803$$

Relación Dorada

Método de la Búsqueda Dorada

El método comienza con dos estimaciones iniciales, x_l y x_u , que incluyen al extremo local de $f(x)$:

- Los siguientes dos puntos interiores x_1 y x_2 se eligen de acuerdo con la relación dorada:

$$d = \frac{\sqrt{5} - 1}{2} (x_u - x_l)$$

$$x_1 = x_l + d$$

$$x_2 = x_u - d$$

- La función se evalúa en estos dos puntos interiores.

Método de la Búsqueda Dorada

Supongamos que buscamos un máximo, pueden ocurrir dos cosas:

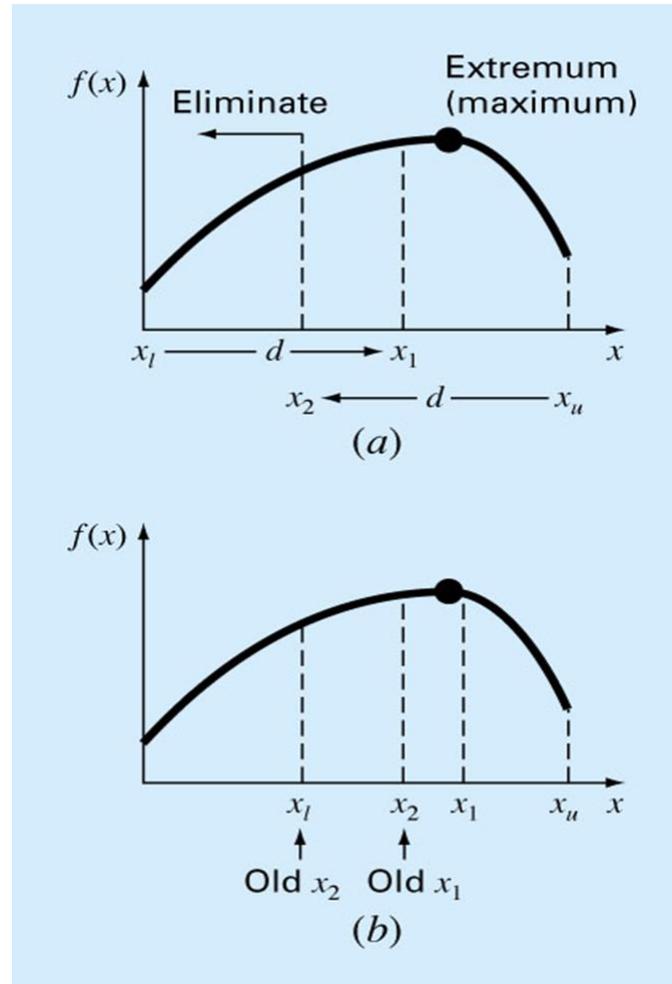
- 1) Si $f(x_1) > f(x_2)$ entonces el dominio de x a la izquierda de x_2 desde x_1 hasta x_2 , se puede eliminar porque no contiene el máximo. Entonces, x_2 se convierte en el nuevo x_1 para la siguiente iteración y x_1 en el punto interior izquierdo en el nuevo intervalo. El nuevo x_2 se determina como antes:

$$x_2 = x_u - \frac{\sqrt{5} - 1}{2} (x_u - x_l)$$

- 2) Si $f(x_2) > f(x_1)$, entonces el dominio de x a la derecha de x_1 desde x_1 hasta x_u , se puede eliminar. En este caso, x_1 se convierte en el nuevo x_u para la siguiente iteración y x_2 en el punto interior derecho en el nuevo intervalo. El nuevo x_1 se determina como antes:

$$x_1 = x_l + \frac{\sqrt{5} - 1}{2} (x_u - x_l)$$

Método de la Búsqueda Dorada



Método de la Búsqueda Dorada

El beneficio real del uso de la relación dorada es porque las x_1 y x_2 originales se eligieron usando la relación dorada, no es necesario volver a calcular todos los valores de la función para la siguiente iteración.

Error del Método de la Búsqueda Dorada

- Es posible determinar una cota superior exacta para el error en cada iteración.
- Una vez que se termina una iteración, el valor óptimo estará en el intervalo que contiene los 3 primeros puntos o en el intervalo que contiene los últimos 3.
- Supongamos un mínimo:
 - Si $f(x_2) < f(x_1)$ entonces el óptimo estará en el intervalo inferior que incluye a (x_1, x_2, x_1)
 - Si $f(x_1) < f(x_2)$, entonces el óptimo estará en el intervalo superior que incluye a (x_2, x_1, x_u)
- Dado que los puntos interiores son simétricos con respecto al centro del intervalo de búsqueda, se puede utilizar cualquiera de ellos para definir el error.

Error del Método de la Búsqueda Dorada

- Si $f(x_1) < f(x_2)$ y el mínimo estuviese entre x_2 y x_1 , la máxima distancia entre el valor estimado del mínimo y el valor verdadero sería:

$$\begin{aligned}\Delta x_a &= x_1 - x_2 = x_l + d - (x_u - d) \\ &= 2d - (x_u - x_l) = 2R(x_u - x_l) - (x_u - x_l) \\ &= (2R - 1)(x_u - x_l) = 0.362(x_u - x_l)\end{aligned}$$

- Por otra parte, si el mínimo estuviese entre x_1 y x_u la máxima distancia al mínimo sería:

$$\begin{aligned}\Delta x_b &= x_u - x_1 = x_u - (x_l + d) \\ &= (x_u - x_l) - d = (x_u - x_l) - R(x_u - x_l) \\ &= (1 - R)(x_u - x_l) = 0.382(x_u - x_l)\end{aligned}$$

Error del Método de la Búsqueda Dorada

- Por consiguiente, este último caso representa el error máximo.
- El resultado después se normaliza al valor óptimo (último punto interior que minimiza (maximiza) la función), x_{opt} , para dar:

$$\varepsilon_a = (1 - R) \left| \frac{x_u - x_l}{x_{opt}} \right| 100\%$$

- Esta estimación de la cota del error relativo proporciona una base para terminar las iteraciones del algoritmo.

Si se busca un mínimo

$[a_0, b_0]$ intervalo de búsqueda original

Punto interior izquierdo: $\lambda_0 = b_0 - \alpha(b_0 - a_0)$

Punto interior derecho: $\mu_0 = a_0 + \alpha(b_0 - a_0)$

$k = 0$

$|a_k - b_k| < tol$

si

extremo

no

$f(\lambda_k) < f(\mu_k)$

no

si

$$a_{k+1} = \lambda_k$$

$$b_{k+1} = b_k$$

$$\lambda_{k+1} = \mu_k$$

$$\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1})$$

$$k = k + 1$$

$$a_{k+1} = a_k$$

$$b_{k+1} = \mu_k$$

$$\mu_{k+1} = \lambda_k$$

$$\lambda_{k+1} = b_{k+1} - \alpha(b_{k+1} - a_{k+1})$$

$$k = k + 1$$

Optimización Multidimensional No Restringida

- El problema se vuelve mucho más complejo para un número de variables independientes $n > 1$.
- En general, se comienza con un punto dado en el dominio n -dimensional, se elige una dirección de búsqueda y se busca el óptimo (mínimo o máximo) en esa dirección, transformando al problema en un problema 1-D.

Método de la Pendiente Más Pronunciada

- Si buscamos un mínimo, efectuamos la búsqueda en la dirección de $-\underline{\nabla}f$; si buscamos un máximo en la dirección del $\underline{\nabla}f$
- Supongamos que buscamos un mínimo, entonces resolvemos el siguiente problema unidimensional:

$$\underset{\alpha}{\text{mín}} f \left[\underline{x}^{(k)} - \alpha \underline{\nabla} f \left(\underline{x}^{(k)} \right) \right]$$

Para obtener:

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

donde α es el óptimo deseado.

Método de la Pendiente Más Pronunciada

- Este método tiende a ser más que lento.
- Esto se debe a que el gradiente no apunta necesariamente en la dirección del óptimo (máximo o mínimo).
- Veamos un ejemplo:

$$\text{Sea } f(x) = x_1^2 + 10x_2^2 + 100x_3^2$$

- Esto tiene un mínimo global en: $\underline{x}^* = \underline{0}$; $(x_1^* = x_2^* = x_3^* = 0)$

- Ahora:

$$\underline{\nabla} f = [2x_1, 20x_2, 200x_3]$$

- En cada iteración buscamos resolver:

$$\min_{\alpha} f[\underline{x} - \alpha \underline{\nabla} f] = \min_{\alpha} \left\{ (x_1 - 2\alpha x_1)^2 + (x_2 - 20\alpha x_2)^2 + (x_3 - 200\alpha x_3)^2 \right\}$$

Método de la Pendiente Más Pronunciada

- Planteamos la condición necesaria de extremo para α :

$$\frac{d \left\{ (x_1 - 2\alpha x_1)^2 + (x_2 - 20\alpha x_2)^2 + (x_3 - 200\alpha x_3)^2 \right\}}{d\alpha} = 0$$

- Se obtiene una ecuación lineal en α :

$$4(x_1 - 2\alpha x_1)x_1 + 400(x_2 - 20\alpha x_2)x_2 + 40000(x_3 - 200\alpha x_3)x_3 = 0$$

- Despejando α resulta:

$$\alpha = \frac{x_1^2 + 10^2 x_2^2 + 10^4 x_3^2}{2(x_1^2 + 10^3 x_2^2 + 10^6 x_3^2)}$$

- Si adoptamos como punto de partida $\underline{x}^{(0)} = [1, 1, 1]$, después de 180 iteraciones obtenemos: $\underline{x}^{(180)} = [1.07 \times 10^{-4}, 0, 3.01 \times 10^{-6}]$.
- Esto se debe a que la función tiene diferentes curvaturas en diferentes direcciones.
- Uno podría imaginar al algoritmo como recorriendo hacia adelante y hacia atrás el estrecho valle de un río que fluye lentamente hasta desembocar en el mar.

Método de Newton

- Desarrollamos en serie de Taylor la función de varias variables alrededor de un punto \underline{x}_0 :

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

- La matriz $\underline{\nabla} \underline{\nabla} f$ recibe el nombre de matriz Hessiana. Es una matriz cuyos elementos son las derivadas segundas de $f(x)$, esto es:

$$\left[\underline{\underline{H}} \right]_{ij}(\underline{x}) \equiv \left[\underline{\nabla} \underline{\nabla} f(\underline{x}) \right]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Método de Newton

- La condición necesaria de existencia de puntos críticos para funciones de 2 o más variables es:

$$\underline{\nabla} f(\underline{x}) \cong \underline{\nabla} f(\underline{x}_0) + \underline{\nabla} \underline{\nabla} f(\underline{x}_0) \cdot (\underline{x} - \underline{x}_0) = \underline{0}$$

donde hemos truncado los términos de orden superior.

- Si $\underline{x}^{(k)}$ es una aproximación del punto crítico, entonces podemos obtener una nueva estimación despejando de la expresión anterior,

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \left[\underline{\nabla} \underline{\nabla} f(\underline{x}^{(k)}) \right]^{-1} \cdot \underline{\nabla} f(\underline{x}^{(k)}) = \underline{x}^{(k)} - \underline{\underline{H}}^{-1} \cdot \underline{\nabla} f(\underline{x}^{(k)})$$

- Este método converge cuadráticamente cerca del punto crítico, pero puede fallar en la convergencia.

Esquema General de Resolución

- Supongamos que queremos encontrar un mínimo.
- Queremos desarrollar un algoritmo que haga lo siguiente:

1) Que calcule $f(\underline{x}^{(k)})$ y $\underline{\nabla}f(\underline{x}^{(k)})$

2) Que calcule alguna dirección descendente \underline{p} tal que:

$$f(\underline{x}^{(k)} + \varepsilon \underline{p}) < f(\underline{x}^{(k)}) \text{ para pequeños } \varepsilon$$

Podemos hacer esto de dos formas:

i. Por el método de la pendiente más pronunciada: $\underline{p} = -\underline{\nabla}f(\underline{x}^{(k)})$

ii. Por el método de Newton: $\underline{p} = -[\underline{\nabla}\underline{\nabla}f]^{-1} \cdot \underline{\nabla}f = -\underline{H}^{-1} \cdot \underline{\nabla}f$ más rápido, pero no necesariamente señala un mínimo.

3) Que busque a lo largo del vector \underline{p} :

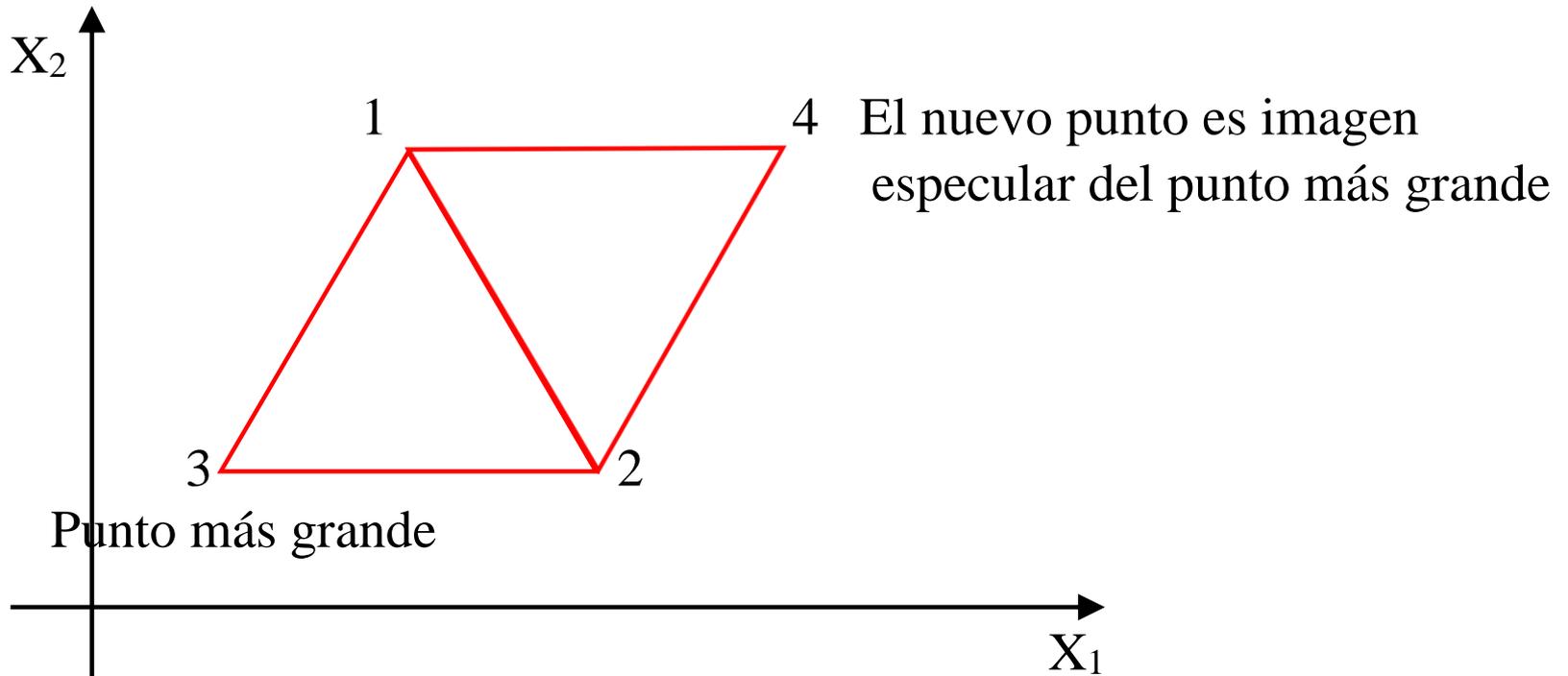
$$\min_{\alpha} f(\underline{x}^{(k)} + \alpha \underline{p}) \rightarrow \underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_{opt.} \underline{p} ; \varepsilon = \alpha_{opt.}$$

Volver al Paso 1.

Método Simplex

- Un algoritmo completamente diferente es el Método Simplex.
- Este algoritmo utiliza una búsqueda basada en triángulos equiláteros en un espacio 2-D y en pirámides multidimensionales en espacios n-D.
- Veamos el problema en 2-D
- Supongamos que estamos buscando un mínimo.
- Se elijen 3 puntos del dominio en forma de triángulo equilátero.
- Se evalúa la función en esos 3 puntos y seguidamente se elije el nuevo punto como imagen especular del punto donde la función es más grande.

Método Simplex



Método Simplex

- La técnica funciona bien si todos los elementos de $\underline{\nabla}f$ son comparables.
- Si ese no fuese el caso, se requiere reescalar las variables.
- El método se modifica cuando se va aproximando al mínimo (se reduce el tamaño del triángulo equilátero).
- El método Nelder-Mead también cambia la forma del triángulo, haciéndolo más largo en la dirección del mínimo.
- En Scilab la function *fminsearch* utiliza este enfoque.

Optimización Restringida 1-D

- Hasta aquí se ha tratado con métodos de optimización no restringida.
- La mayoría de las veces (excepto en problemas de regresión) se tienen restricciones sobre el espacio de las soluciones factibles. Como se trata esta cuestión?
- En el caso de optimización unidimensional (1-D), se pueden tener restricciones de desigualdad.
- En este caso, el óptimo se calcula en el interior del dominio y se lo compara con los valores que adopta la función sobre la frontera de la región de soluciones factibles.
- En particular, si se busca un mínimo y ese óptimo es menor que los valores que la función asume en la frontera, entonces ese valor es la respuesta al problema de optimización.

Optimización Restringida Multidimensional

- Si la optimización es multidimensional, el problema es mucho más complejo.
- En este caso podemos tener restricciones de igualdad y de desigualdad.
- Veamos primero las restricciones de igualdad.
- En general se busca resolver:

$$\underset{\underline{x}}{\text{mín}} f(\underline{x}) \text{ o } \underset{\underline{x}}{\text{máx}} f(\underline{x})$$

$$\text{sujeto a } g_i(\underline{x}) = 0 \text{ con } i = 1, 2, \dots, m$$

(m restricciones, con $m < n$, siendo n el número de variables)

- La mejor forma de tratar esta situación es utilizar las m restricciones $g_i(\underline{x}) = 0$ para eliminar m variables de $f(\underline{x})$.

Optimización Restringida Multidimensional

- Esto se había hecho con el ejemplo de la lata de cerveza.

$$\min_{\underline{x}} f(x_1, x_2) = 2\pi x_1^2 + 2\pi x_1 x_2$$

$$\text{sujeto a } g(x_1, x_2) = \pi x_1^2 x_2 - V_0 = 0$$

(m restricciones, con $m < n$, siendo n el número de variables)

- Utilizamos la restricción para eliminar la variable x_2 :

$$x_2 = \frac{V_0}{\pi x_1^2}$$

$$f(x_1) = 2\pi x_1^2 + \frac{2V_0}{\pi x_1^2}$$

- Por lo tanto, el problema se redujo a un problema de optimización 1-D no restringido!!

Optimización Restringida Multidimensional Método de los Multiplicadores de Lagrange

- Usualmente, no podemos ir más allá con el planteo anterior si se trata de optimizar una función de más de dos variables con varias restricciones.
- Un enfoque es utilizar los multiplicadores de Lagrange, Sea:

$$f^*(\underline{x}) = f(\underline{x}) + \underline{\lambda} \cdot \underline{g}(\underline{x})$$

donde $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)$ contiene m multiplicadores
para las m restricciones

$$\underline{g}(\underline{x}) = \begin{pmatrix} g_1(\underline{x}) \\ g_2(\underline{x}) \\ \vdots \\ g_m(\underline{x}) \end{pmatrix}$$

Optimización Restringida Multidimensional Método de los Multiplicadores de Lagrange

- El problema de optimización era determinar la solución de:

$$\underline{\nabla} f(\underline{x}) = \underline{0}$$

- Ahora tenemos el problema aumentado:

$$\underline{\nabla}^* f^*(\underline{x}, \underline{\lambda}) = \underline{0}$$

donde :

$$\underline{\nabla}^* \equiv \begin{pmatrix} \underline{\nabla}_x \\ \underline{\nabla}_\lambda \end{pmatrix}$$

- Efectivamente,

$$\underline{\nabla}^* f^*(\underline{x}, \underline{\lambda}) = \underline{\nabla}_x f^*(\underline{x}, \underline{\lambda}) + \underline{\nabla}_\lambda f^*(\underline{x}, \underline{\lambda}) = \underline{\nabla}_x f^*(\underline{x}, \underline{\lambda}) + \underline{g}(\underline{x}) = \underline{0}$$

Optimización Restringida Multidimensional Método de los Multiplicadores de Lagrange

- Ahora, si tenemos restricciones de desigualdad,

$$\underline{g}(\underline{x}) \leq \underline{0}$$

- Podemos convertir a las restricciones de desigualdad en restricciones de igualdad adicionando *variables de holgura*, \underline{S}

- Sea:

$$\underline{g}(\underline{x}) + \underline{S} = \underline{0}$$

donde:

$$S_i = \underbrace{x_{n+i}^2}_{\text{Elección conveniente asegurando que } S_i \geq 0} \geq 0 ; i = 1, 2, \dots, m$$

- Luego, tratamos el problema utilizando los multiplicadores de Lagrange.

Optimización Restringida Multidimensional Funciones de Penalización

- Supongamos que deseamos transformar un problema de optimización multidimensional restringido en uno no restringido.
- Podemos hacer esto de una manera artificial. Sea entonces:

$$\min_{\underline{x}} f(\underline{x}) = \underline{0} \text{ sujeto a } g(\underline{x}) = \underline{0}$$

- Podemos definir:

$$f^*(\underline{x}) = f(\underline{x}) + p \left\| g(\underline{x}) \right\|_2^2$$

- Procedemos a determinar el $\underline{x}_{\text{opt}}$ para valores moderados de p y lo incrementamos levemente.
- $p \rightarrow \infty$ corresponde a $g(\underline{x}) = \underline{0}$
- Esto también puede utilizarse con restricciones de desigualdad utilizando variables de holgura.