Módulo 5: Métodos de Optimización No Restringida

Departamento MACC
Matemáticas Aplicadas y Ciencias de la Computación
Universidad del Rosario

Primer Semestre de 2021

Agenda

Introducción

$x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$

$$\begin{aligned} &\min \ f(x) \\ &\text{s.a. } x \in \Omega \end{aligned}$$

Método de optimización:

- $\Omega = \mathbb{R}^n$ (problema *no restringido*)
- Dada una solución factible $x \in \mathbb{R}^n$, ...
- ... identificar una dirección de movimiento d
- ... y un tamaño de paso α

Direcciones de descenso

- Suponga una solución factible $x \in \mathbb{R}^n$
- d es una dirección de descenso en x si $\exists \delta > 0$:

$$f(x + \lambda d) < f(x), \; \lambda \in (0, \delta)$$

Si $\lim_{\lambda \to 0^+} \frac{f(x+\lambda d) - f(x)}{\lambda} < 0$

- $\Rightarrow d$ es una dirección de descenso de f en x

Si f es diferenciable

- $f'(x, d) = \lim_{\lambda \to 0^+} \frac{f(x+\lambda d) - f(x)}{\lambda} = \nabla f(x)' d$

Agenda

Método de descenso del gradiente

Lema

$f : \mathbb{R}^n \to \mathbb{R}$ diferenciable en x, con $\nabla f(x) \neq 0$. La solución óptima al problema $\min f'(x,d)$ s.a. $||d|| = 1$, es $d = -\frac{\nabla f(x)}{||\nabla f(x)||}$

En el punto x, el método de descenso de gradiente:

- Escoge dirección $d = -\frac{\nabla f(x)}{||\nabla f(x)||}$
- Búsqueda de línea en dirección d

Método de descenso del gradiente - Algoritmo

Paso 0: Determine $\epsilon > 0$, punto inicial de búsqueda x_1, $k = 1$

Paso 1:

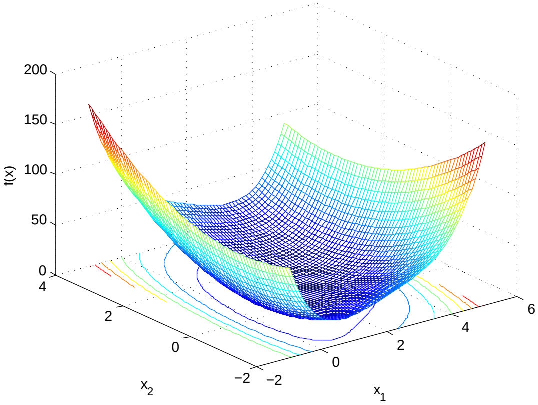
while $||\nabla f(x_k)|| > \epsilon$ **do**

$d_k = -\nabla f(x_k)$
$\lambda_k = \text{argmin}_{\lambda \geq 0} f(x_k + \lambda d_k)$
$x_{k+1} = x_k + \lambda_k d_k$
$k = k + 1$

end while

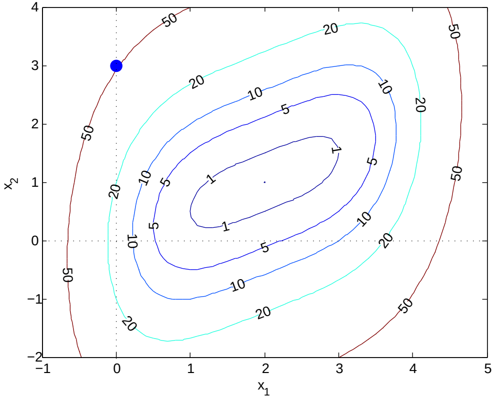
Método de descenso del gradiente - Ejemplo

$$\min(x_1 - 2)^4 + (x_1 - 2x_2)^2, \; x_A = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



Método de descenso del gradiente - Ejemplo (c.)

$$\min(x_1 - 2)^4 + (x_1 - 2x_2)^2,\ x_A = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



Método de descenso del gradiente - Ejemplo (c.)

$\text{mín}(x_1 - 2)^4 + (x_1 - 2x_2)^2$

- $x_A = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$
- $\nabla f(x) = \begin{bmatrix} 4(x_1 - 2)^3 + 2(x_1 - 2x_2) \\ -4(x_1 - 2x_2) \end{bmatrix}$
- $\nabla f(x_A) = \begin{bmatrix} -44 \\ 24 \end{bmatrix}$
- $d_A = -\nabla f(x_A) = \begin{bmatrix} 44 \\ -24 \end{bmatrix}$

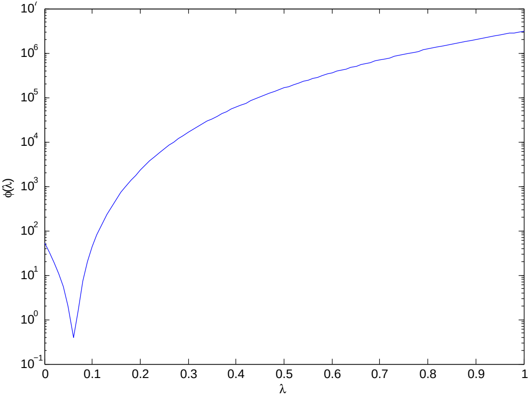
Método de descenso del gradiente - Ejemplo (c.)

$\lambda_A = \text{argmin}_{\lambda \geq 0} f(x_A + \lambda d_A)$

$$
\begin{aligned}
f(x_A + \lambda d_A) &= f\left(\begin{bmatrix} 0 \\ 3 \end{bmatrix} + \lambda \begin{bmatrix} 44 \\ -24 \end{bmatrix}\right) \\
&= f\left(\begin{bmatrix} 44\lambda \\ 3 - 24\lambda \end{bmatrix}\right) \\
&= (44\lambda - 2)^4 + (92\lambda - 6)^2
\end{aligned}
$$

Método de descenso del gradiente - Ejemplo (c.)

$$f(x_A + \lambda d_A) = (44\lambda - 2)^4 + (92\lambda - 6)^2$$



Método de descenso del gradiente - Ejemplo (c.)

Sección áurea:

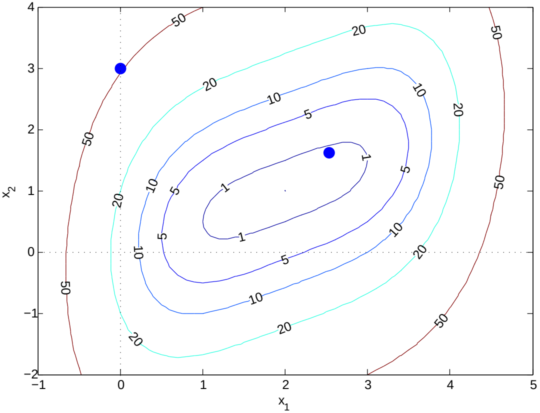
- $\phi(\lambda) = (44\lambda - 2)^4 + (92\lambda - 6)^2$
- $a_1 = 0,\ b_1 = 1,\ l = 10^{-3}$
- $\lambda_1 = 0{,}382,\ \mu_1 = 0{,}618$
- $\phi(\lambda_1) = 4{,}92 \times 10^4$
- $\phi(\mu_1) = 4{,}06 \times 10^5 \Rightarrow \phi(\mu_1) > \phi(\lambda_1)$
- $a_2 = a_1 = 0,\ b_2 = \mu_1 = 0{,}618\ (b_2 - a_2 = 0{,}618)$
- $\mu_2 = \lambda_1,\ \lambda_2 = a_2 + (1-\alpha)(b_2 - a_2) = 0{,}236$
- $\phi(\lambda_2) = 5{,}28 \times 10^3 \Rightarrow \phi(\mu_2) > \phi(\lambda_2)$

Método de descenso del gradiente - Ejemplo (c.)

- $a_3 = a_2 = 0,\ b_3 = \mu_2 = 0{,}382\ (b_3 - a_3 = 0{,}382)$
- $\mu_3 = \lambda_2,\ \lambda_3 = a_3 + (1-\alpha)(b_3 - a_3) = 0{,}146$
- $\phi(\lambda_3) = 4{,}61 \times 10^2 \Rightarrow \phi(\mu_3) > \phi(\lambda_3)$
- $\ldots$
- $\lambda^* = 0{,}0615$

Método de descenso del gradiente - Ejemplo (c.)

$$x_B = x_A + \lambda_A d_A = \begin{bmatrix} 2{,}71 \\ 1{,}52 \end{bmatrix}$$



Método de descenso del gradiente - Convergencia

- $\Omega = \{x^* : \nabla f(x^*) = 0\}$
- $D(x) = (x, -\nabla f(x))$: punto y dirección
- M: mapa de búsqueda de línea sobre intervalo L

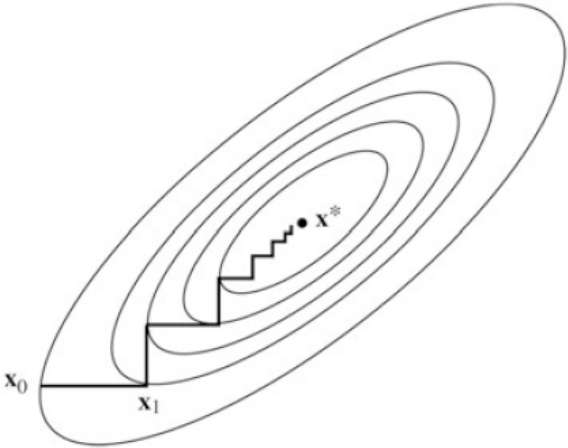
$$M(x, d) = \{y : y = x + \bar{\lambda} d, \bar{\lambda} \in L : f(y) \leq f(x + \lambda d), \forall \lambda \in L\}$$

- M es cerrado
- $A = MD$: mapa algorítmico del método de descenso del gradiente

Método de descenso del gradiente - Convergencia

- $\Omega = \{x^* : \nabla f(x^*) = 0\}$
- $A = MD$: mapa algorítmico del método de descenso del gradiente
- Si $f \in \mathcal{C}^1$, D es un mapa cerrado.
- Si D y M son cerrados, A es cerrado
- Función de descenso: f
- Suponiendo que $\{x_k\}$ permanece en un conjunto compacto, el método de descenso del gradiente definido por el mapa A converge a un punto en Ω

Método de descenso del gradiente - Convergencia



Método de descenso del gradiente - Tasa convergencia

- Función cuadrática: $f(x) = \frac{1}{2}x'Qx - b'x$
- Q: simétrica definida positiva
- $\nabla f(x) = Qx - b$, $H(x) = Q \succ 0$
- Descenso del gradiente:
  - En x_k, movimiento de tamaño λ_k en dirección $-\nabla f(x_k)$
- Función de error:

$$e(x) = \frac{1}{2}(x - x^*)'Q(x - x^*)$$

- $e(x) = 0 \Leftrightarrow x = x^*$, con $x^* \in \Omega$
- Si $x \neq x^* \Rightarrow e(x) > 0$

Método de descenso del gradiente - Tasa convergencia

Para descenso del gradiente y f cuadrática:

$$e(x_{k+1}) = (1 - \gamma_k)e(x_k)$$

con

$$\gamma_k = \frac{(\nabla f(x_k)'\nabla f(x_k))^2}{(\nabla f(x_k)'Q\nabla f(x_k))(\nabla f(x_k)'Q^{-1}\nabla f(x_k))}$$

Método de descenso del gradiente - Tasa convergencia

Teorema

Para cualquier $x_0 \in \mathbb{R}^n$, el método de descenso del gradiente converge al minimo x^* de f y

$$e(x_{k+1}) = \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 e(x_k)$$

con λ_n y λ_1 como los valores propios más grande y más pequeño de Q, respectivamente.

Con r la condición de la matriz,

$$\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 = \left(\frac{r-1}{r+1}\right)^2$$

Método de descenso del gradiente - Tasa convergencia

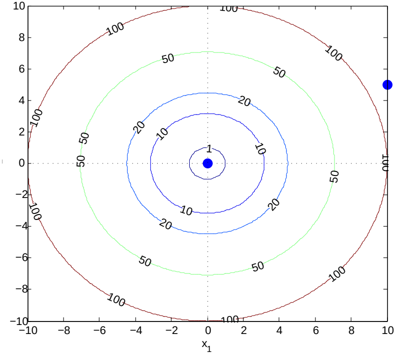
Entonces:

- Factor de reducción $\frac{e(x_{k+1})}{e(x_k)}$ acotado
- Si $r = 1$, convergencia en un paso (contornos circulares)
- Si $r >> 1$, convergencia más lenta (contornos excéntricos)

Método de descenso del gradiente - Tasa convergencia

Ejemplo: $f(x) = x_1^2 + x_2^2$

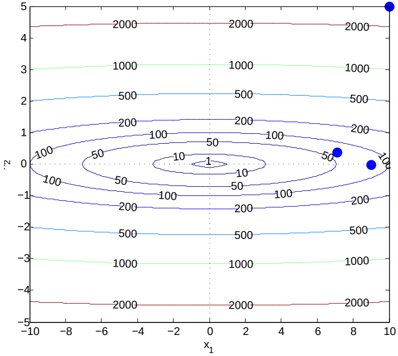
- $\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$, $H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
- $\alpha = 1$



Método de descenso del gradiente - Tasa convergencia

Ejemplo: $f(x) = x_1^2 + 100x_2^2$

- $\nabla f(x) = \begin{bmatrix} 2x_1 \\ 200x_2 \end{bmatrix}$, $H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 200 \end{bmatrix}$
- $\alpha = 100$



Agenda

Método de Newton

Usar aproximación cuadrática de la función en x_k:

$$q(x) = f(x_k) + \nabla f(x_k)'(x - x_k) + \frac{1}{2}(x - x_k)'H(x_k)(x - x_k)$$

Condición necesaria para un mínimo local de $q(x)$: $\nabla q(x) = 0$

$$\begin{aligned} \nabla q(x) &= \nabla f(x_k) + H(x_k)(x - x_k) = 0 \\ H(x_k)x &= H(x_k)x_k - \nabla f(x_k) \\ x &= x_k - H(x_k)^{-1}\nabla f(x_k) \end{aligned}$$

Método de Newton (cont.)

$$x_{k+1} = x_k - H(x_k)^{-1}\nabla f(x_k)$$

Suponiendo $\nabla f(\bar{x}) = 0$ y $H(\bar{x}) \succ 0$ en un mínimo local $\bar{x}$, y $f \in \mathcal{C}^2$:

- $H(x_k) \succ 0$ para todo punto x_k *cerca* de $\bar{x}$
- En cada iteración se resuelve un sistema lineal
- Sistema lineal: aproximación de primer orden
- Aproximación de primer orden de $\nabla f(x_k) = 0$:

$$\nabla f(x_k) + H(x_k)(x - x_k) = 0$$

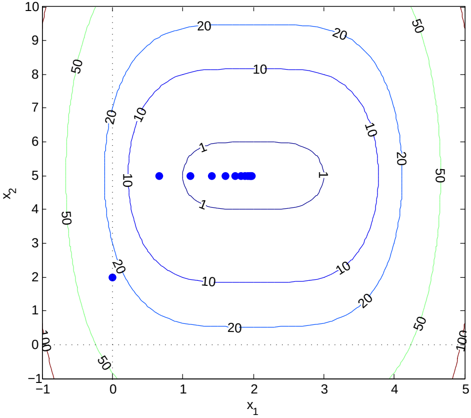
Método de Newton - Ejemplo

$f(x) = (x_1 - 2)^4 + (x_2 - 5)^2$

- $x_A = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$
- $\nabla f(x) = \begin{bmatrix} 4(x_1 - 2)^3 \\ 2(x_2 - 5) \end{bmatrix}$, $H(x) = \begin{bmatrix} 12(x_1 - 2)^2 & 0 \\ 0 & 2 \end{bmatrix}$
- $\nabla f(x_A) = -\begin{bmatrix} 32 \\ 6 \end{bmatrix}$, $H(x_A) = \begin{bmatrix} 48 & 0 \\ 0 & 2 \end{bmatrix}$
- $x_B = x_A - H(x_A)^{-1} \nabla f(x_A) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2/3 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 5 \end{bmatrix}$

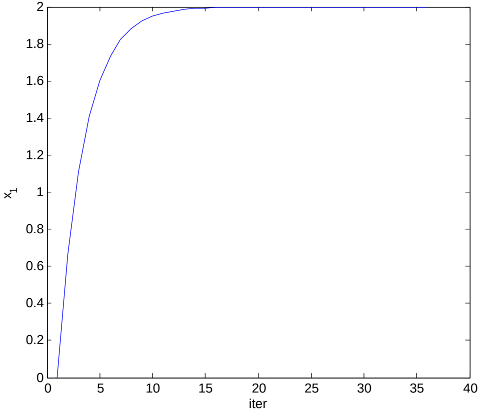
Método de Newton - Ejemplo (cont.)

$$f(x) = (x_1 - 2)^4 + (x_2 - 5)^2$$



Método de Newton - Ejemplo (cont.)

$$f(x) = (x_1 - 2)^4 + (x_2 - 5)^2$$



Método de Newton - Convergencia

- No necesariamente converge
- $H(x_k)$ puede ser singular
- $f(x_{k+1})$ puede ser mayor que $f(x_k)$
- Si x_k está suficientemente cerca a $\bar{x}$, método de Newton está bien definido y converge a $\bar{x}$

Agenda

Métodos de direcciones conjugadas

- Sea $H \in \mathbb{R}^{n \times n}$ una matriz simétrica
- $\{d_1, \ldots, d_n\}$ son vectores H-conjugados si son linealmente independientes y $d_i' H d_j = 0$, para $i \neq j$
- Objetivo: realizar búsqueda del mínimo local a través de direcciones H-conjugadas

Métodos de direcciones conjugadas - Caso cuadrático

- $f(x) = c'x + \frac{1}{2}x'Hx$ con H simétrica definida positiva
- $\{d_1, \ldots, d_n\}$: H-conjugadas
- $x = x_1 + \sum_{j=1}^{n} \lambda_j d_j$
- $\phi(\lambda) = f(x) = f(x_1 + \sum_{j=1}^{n} \lambda_j d_j)$

Métodos de direcciones conjugadas - Caso cuadrático

- $x = x_1 + \sum_{j=1}^{n} \lambda_j d_j$
- $\phi(\lambda) = f(x) =$
  $c' x_1 + \sum_{j=1}^{n} \lambda_j c' d_j + \frac{1}{2} \left(x_1 + \sum_{j=1}^{n} \lambda_j d_j \right)' H \left(x_1 + \sum_{j=1}^{n} \lambda_j d_j \right)$
- Por ser H-conjugadas:

$$\phi(\lambda) = c' x_1 + \frac{1}{2} x_1' H x_1 + \sum_{j=1}^{n} \left(\lambda_j c' d_j + \lambda_j d_j' H x_1 + \frac{1}{2} \lambda_j^2 d_j' H d_j \right)$$

Métodos de direcciones conjugadas - Caso cuadrático

- Valor óptimo de λ_j:

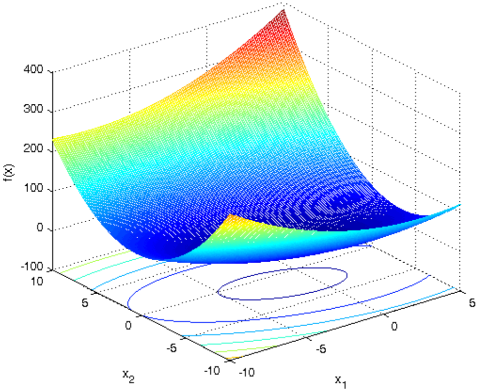
$$\frac{\partial \phi(\lambda)}{\partial \lambda_j} = c' d_j + d_j' H x_1 + \lambda_j d_j' H d_j = 0$$

$$\lambda_j^* = -\frac{c' d_j + d_j' H x_1}{d_j' H d_j}$$

- Tamaño de paso óptimo en cada dirección d_j

Métodos de direcciones conjugadas - Ejemplo

$$\min\ 5x_1 + 8x_2 + x_1x_2 + x_1^2 + 2x_2^2 = \begin{bmatrix} 5 & 8 \end{bmatrix} x + \frac{1}{2}x' \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} x$$



Métodos de direcciones conjugadas - Ejemplo (cont.)

$$\min\ 5x_1 + 8x_2 + x_1x_2 + x_1^2 + 2x_2^2 = \begin{bmatrix} 5 & 8 \end{bmatrix} x + \tfrac{1}{2}x' \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} x$$

- $\nabla f(x) = \begin{bmatrix} 5 + 2x_1 + x_2 \\ 8 + x_1 + 4x_2 \end{bmatrix}$
- $H = H(x) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \succ 0$
- $d_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $d_2 = \begin{bmatrix} a \\ b \end{bmatrix}$
- $d_1' H d_2 = 0$: $2a + b = 0 \Rightarrow b = -2a$
- $a = 1,\ b = -2,\ d_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

Métodos de direcciones conjugadas - Ejemplo (cont.)

$x_1 = 0$

- $d_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$,

$$\lambda_1 = -\frac{c' d_1 + d_1' H x_1}{d_1' H d_1} = -\frac{5}{2}$$

- $x_2 = x_1 + \lambda_1 d_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{5}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -5/2 \\ 0 \end{bmatrix}$

Métodos de direcciones conjugadas - Ejemplo (cont.)

$$x_2 = \begin{bmatrix} -5/2 \\ 0 \end{bmatrix}$$

- $d_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$,

$$\lambda_2 = -\frac{c' d_2 + d_2' H x_2}{d_2' H d_2} = \frac{11}{14}$$

- $x_3 = x_2 + \lambda_2 d_2 = \begin{bmatrix} -5/2 \\ 0 \end{bmatrix} + \frac{11}{14} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = - \begin{bmatrix} 12/7 \\ 11/7 \end{bmatrix} = x^*$

- $x^* = -H^{-1} c = -\frac{1}{7} \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \end{bmatrix} = -\frac{1}{7} \begin{bmatrix} 12 \\ 11 \end{bmatrix}$

Métodos de direcciones conjugadas - Ejemplo (cont.)

$$\min\ 5x_1 + 8x_2 + x_1x_2 + x_1^2 + 2x_2^2$$