DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Introducción a la Investigación Operativa y Optimización

Primer cuatrimestre 2025

Optimización no lineal multivariable

Dada $f : \mathbb{R}^n \to \mathbb{R}$ y S $\subseteq \mathbb{R}^n$, queremos resolver el problema:

$$\min_{x \in S} f(x)$$

- Si $S = \mathbb{R}^n$, es el caso de dominio no restringido.
- Si $S \subset \mathbb{R}^n$, es un problema de optimización con restricciones.

Optimización multivariada sin restricciones.

Condiciones necesarias - $f : \mathbb{R}^n \to \mathbb{R}$ con $f \in C^2$

Condición de optimalidad de 1er orden

Sea $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$. Si x^* es un mínimo local de f, entonces $\nabla f(x^*) = 0$.

Demostración

Sea $d \in \mathbb{R}^n$.

- Definimos $\phi : \mathbb{R} \to \mathbb{R}$ como $\phi(\alpha) = f(x^* + \alpha d)$.
- Por regla de la cadena $\phi'(\alpha) = \nabla^T f(x^* + \alpha d)d$.
- Como x^* es un mínimo local de f, 0 es un mínimo local de ϕ.
- Por condición de optimalidad de 1er orden $\phi'(0) = 0$.
- $0 = \phi'(0) = \nabla^t f(x^*)d$, $\forall d \in \mathbb{R}^n$.
- Por lo tanto $\nabla f(x^*) = 0$.

Condiciones necesarias - $f : \mathbb{R}^n \to \mathbb{R}$ con $f \in C^2$

Condición de optimalidad de 2do orden

Sea $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$. Si x^* es un mínimo local de f, entonces $\nabla f(x^*) = 0$ y $Hf(x^*)$ semidefinida positiva.

Demostración

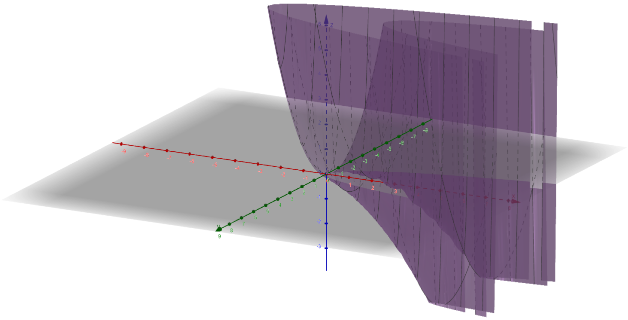
Sea $d \in \mathbb{R}^n$.

- Definimos $\phi : \mathbb{R} \to \mathbb{R}$ como $\phi(\alpha) = f(x^* + \alpha d)$.
- Por regla de la cadena $\phi''(\alpha) = d^t Hf(x^* + \alpha d)d$.
- Como x^* es un mínimo local de f, 0 es un mínimo local de ϕ.
- Por condición de optimalidad de 2do orden: $\phi'(0) = 0$ y $\phi''(0) \geq 0$.
- Entonces $\nabla^T f(x^*)d = 0$ y $\phi''(0) = d^t Hf(x^*)d \geq 0$, $\forall d \in \mathbb{R}^n$.
- Por lo tanto $\nabla f(x^*) = 0$ y $Hf(x^*)$ semidefinida positiva.

Condiciones necesarias - $f : \mathbb{R}^n \to \mathbb{R}$ con $f \in C^2$

No es condición suficiente

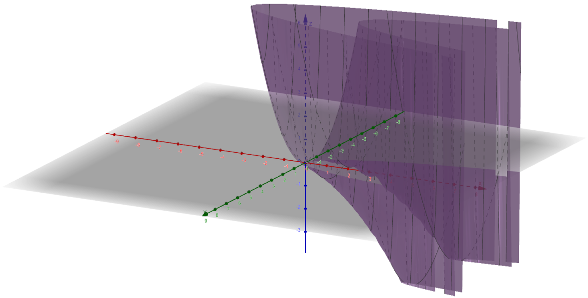
- $f(x_1, x_2) = (x_1 - x_2^2)(x_1 - 2x_2^2) = x_1^2 - 3x_1x_2^2 + 2x_2^4$
- $\nabla f(x_1, x_2) = (2x_1 - 3x_2^2, -6x_1x_2 + 8x_2^3)$
- $Hf(x_1, x_2) = \begin{pmatrix} 2 & -6x_2 \\ -6x_2 & -6x_1 + 24x_2^2 \end{pmatrix}$



Condiciones necesarias - $f : \mathbb{R}^n \to \mathbb{R}$ con $f \in C^2$

No es condición suficiente

- $\nabla f(0,0) = 0$
- $Hf(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ es semidefinida positiva.
- $(0,0)$ no es mínimo local. Tiene función objetivo $= 0$
- $x = (\epsilon, \sqrt{\frac{3\epsilon}{4}})$ tiene función objetivo $= -\frac{\epsilon^2}{8}$



Condición de optimalidad de 2do orden

Condición suficiente de 2do orden

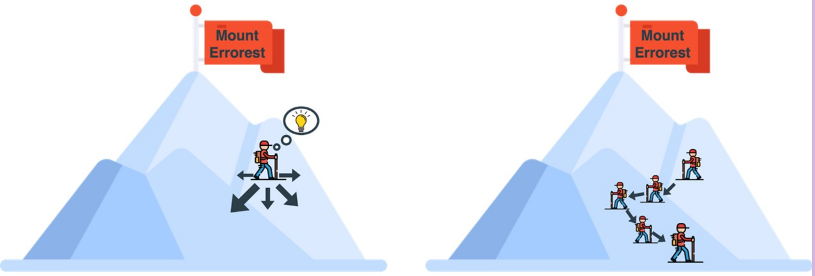
Sea $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$. Si x^* es un mínimo local de f, entoces $\nabla f(x^*) = 0$ y $Hf(x^*)$ definida positiva.

Demostración

Sea d una dirección factible.

- Definimos $g : \mathbb{R} \to \mathbb{R}$ como $g(\alpha) = f(x^* + \alpha d)$.
- Por regla de la cadena $g'(\alpha) = \nabla^T f(x^* + \alpha d)d$ y $g''(\alpha) = d^t Hf(x^* + \alpha d)d$.
- Como x^* es un mínimo local de f, 0 es un mínimo local de g. Entonces, por condición suficiente de optimalidad de 2do orden en una variable, $g'(0) = 0$ y $g''(0) > 0$.
- Como la región factible es todo $\mathbb{R}^n$, toda $d \in \mathbb{R}^n$ es dirección factible, y entonces $\nabla f(x^*) = 0$ y $g''(0) = d^t Hf(x^*)d > 0 \forall d \in \mathbb{R}^n$.
- Por lo tanto $\nabla f(x^*) = 0$ y $Hf(x^*)$ definida positiva.

Métodos de descenso



$f : \mathbb{R}^n \to \mathbb{R}$: Método de descenso coordenado

- Minimiza la función objetivo iterativamente a lo largo de una coordenada a la vez, mientras mantiene las demás coordenadas fijas.
- Resuelve de forma secuencial el problema

$$\min_{x_i \in \mathbb{R}} f(x_1, \ldots, x_n)$$

- Diferentes formas de seleccionar la siguiente coordenada sobre la cual optimizar:
  - aleatoria
  - cíclica
  - ida y vuelta: $e_1, e_2, \ldots, e_{n-1}, e_n, e_{n-1}, \ldots, e_2, e_1, \ldots$
  - método de Gauss-Southwell: elige la coordenada que maximiza el gradiente en valor absoluto. Esto puede acelerar la convergencia al enfocarse en la dirección de mayor descenso en cada paso

$f : \mathbb{R}^n \to \mathbb{R}$: Método de descenso coordenado

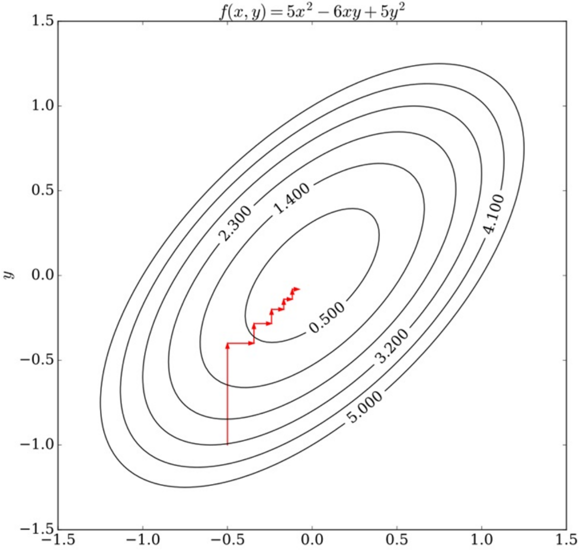
Algoritmo:

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$
2. Mientras no se cumpla el criterio de parada:
    1. Seleccionar una cordenada i según el criterio elegido
    2. Resolver el problema univariable:

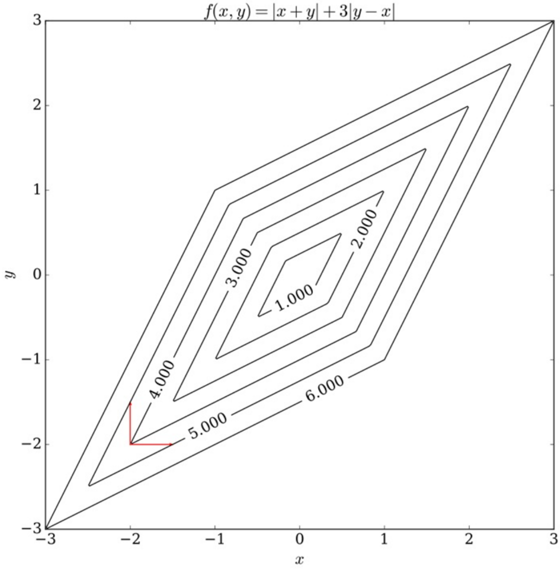
    $$\lambda^k = \arg\min_{\lambda \in R} f(x_1^k, \ldots, x_{i-1}^k, x_i^k + \lambda, x_{i+1}^k, \ldots, x_n^k)$$

    3. $x^{k+1} \longleftarrow x^k + \lambda^k e_i$
    4. $k \longleftarrow k + 1$

$f : \mathbb{R}^n \to \mathbb{R}$: Método de descenso coordenado - Convergencia



$f : \mathbb{R}^n \to \mathbb{R}$: Método de descenso coordenado - Divergencia



$f : \mathbb{R}^n \to \mathbb{R}$: Método de Hooke y Jeeves

- Combina dos tipos de direcciones de búsqueda iterativamente.
- Primero realiza una vuelta del método de descenso coordenado cíclico.
- Luego toma como dirección $x^{k+1} - x^k$ (patrón de búsqueda).

$f : \mathbb{R}^n \to \mathbb{R}$: Método de Hooke y Jeeves

Algoritmo:

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$, $y^0 \longleftarrow x^k$
2. Mientras no se cumpla el criterio de parada:
    1. Para $i = 1, \dots n$:
        1. Resolver el problema univariable:

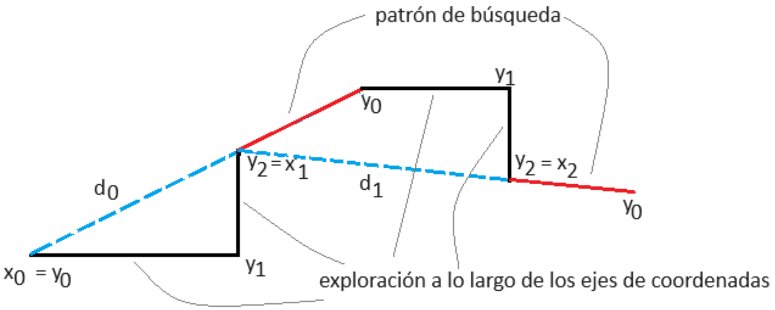
        $$\lambda^i = \arg\min_{\lambda \in R} f(y_1^{i-1}, \dots, y_{i-1}^{i-1}, y_i^{i-1} + \lambda, y_{i+1}^{i-1}, \dots, y_n^{i-1})$$

        2. $y^i \longleftarrow y^{i-1} + \lambda^i e_i$
    2. $x^{k+1} \longleftarrow y^n$
    3. $d^k \longleftarrow x^{k+1} - x^k$
    4. Resolver el problema univariable:

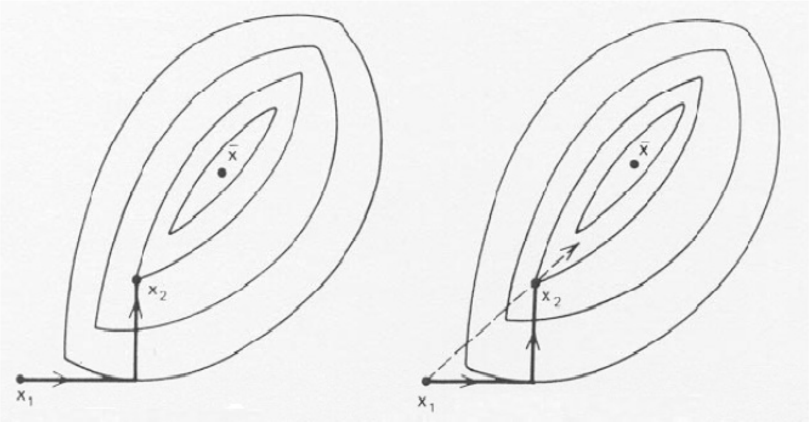
    $$\lambda^k = \arg\min_{\lambda \in R} f(y^n + \lambda d^k)$$

    5. $y^0 \longleftarrow y^n + \lambda^k d^k$
    6. $k \longleftarrow k + 1$
3. retornar y^0

$f : \mathbb{R}^n \to \mathbb{R}$: Método de Hooke y Jeeves



$f : \mathbb{R}^n \to \mathbb{R}$: Método de Hooke y Jeeves



Incluir una nueva direccion no coordenada, destraba el proceso.

$f : \mathbb{R}^n \to \mathbb{R}$: Método de descenso coordenado

- No es necesario el cálculo del gradiente (ni que exista).
- Convergencia aceptable si las variables están poco acopladas.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Dirección de descenso

Dado $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ es dirección de descenso en x si $f(x + \alpha d) < f(x)$ para α positivo suficientemente chico.

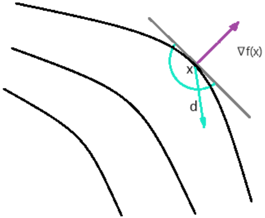
$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Dirección de descenso

Dado $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ es dirección de descenso en x si $f(x + \alpha d) < f(x)$ para α positivo suficientemente chico.

Dado $x \in \mathbb{R}^n$ y $d \in \mathbb{R}^n$. Si

$$\nabla f(x)^t d < 0$$

entonces d es dirección de descenso en x.



$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Dirección de descenso

Demostración:

- Definimos $\phi(\alpha) = f(x + \alpha d)$
- $\phi'(\alpha) = \nabla^T f(x + \alpha d)d$
- $\phi(0) = f(x)$ y $\phi'(0) = \nabla^T f(x)d$
- Como

  $$\phi'(0) = \lim_{t \to 0} \frac{\phi(\alpha) - \phi(0)}{\alpha}$$

  existe $\bar{t} > 0$ tal que $\forall t \in (0, \bar{t})$, los signos de $\phi'(0)$ y $\phi(\alpha) - \phi(0)$ son iguales.
- Como $\nabla^T f(x)d < 0$, entonces $\phi(\alpha) - \phi(0) < 0 \ \forall \alpha \in (0, \bar{\alpha})$.
- Esto implica que $f(x + \alpha d) < f(x) \ \forall \alpha \in (0, \bar{\alpha})$.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

Esquema básico

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$
2. Mientras no se cumpla el criterio de parada:
   1. Si $\nabla f(x_k) < \epsilon$, PARAR.
   2. Seleccionar d_k dirección de descenso: $\nabla f(x_k)^T d_k < 0$
   3. Resolver el problema univariable: Elegir paso α_k tal que $f(x_k + \alpha_k d_k) < f(x_k)$
   4. $x_{k+1} \longleftarrow x_k + \alpha_k d_k$
   5. $k \longleftarrow k + 1$

Tenemos que definir:

- cómo calcular la dirección d_k
- cómo calcular la longitud del paso α_k (búsqueda unidireccional o lineal)

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

Búsqueda unidireccional

- Trade-off entre calidad y tiempo.
- Reducción sustancial de f.
- No insumir mucho tiempo en su cálculo.
- Elección ideal:

$$\alpha_k = \min_{\alpha \in \mathbb{R}_{\geq 0}} \phi(\alpha) = \min_{\alpha \in \mathbb{R}} f(x_k + \alpha d_k)$$

- En general es muy caro identificar este valor.
- El cálculo de un mínimo local ya es caro.
- Puede requerir muchas evaluaciones de f y tal vez de ∇f.
- Puede no dar suficiente decrecimiento de f.
- Búsqueda lineal inexacta:
  - reducción adecuada de f,
  - a mínimo costo.

Condición simple y obvia:

$$f(x_k + \alpha_k d_k) < f(x_k)$$

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

Condición simple y obvia:

$$f(x_k + \alpha_k d_k) < f(x_k)$$

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$
2. Mientras no se cumpla el criterio de parada:
   1. Si $\nabla f(x_k) < \epsilon$, PARAR.
   2. Seleccionar d_k dirección de descenso: $\nabla f(x_k)^T d_k < 0$
   3. $\alpha_k \longleftarrow 1$
   4. Mientras $f(x_k + \alpha_k d_k) \geq f(x_k)$: $\alpha_k \longleftarrow \alpha_k / 2$
   5. $x_{k+1} \longleftarrow x_k + \alpha_k d_k$
   6. $k \longleftarrow k + 1$

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

- Si el algoritmo anterior para por $\nabla f(x_k) < \epsilon$, x_k será un candidato a minimizador.
- Si esta condición no se cumple, ¿siempre existe $\lim_{k\to\infty} x_k$?
- Si $\lim_{k\to\infty} x_k = x^*$, ¿$\nabla f(x^*) = 0$?

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

- $f(x) = x^2$
- $\nabla f(x) = 2x$
- Ejecutar el algoritmo con:
  - $x_0 = 2$
  - $d_k = -\frac{1}{(k+1)(k+2)}$
- $x_k = 1 + \frac{1}{k+1}$
- $x_k \longrightarrow 1$ cuando $k \longrightarrow \infty$
- $f'(1) = 2$, 1 no es punto estacionario.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

- $f(x) = x^2$
- $\nabla f(x) = 2x$
- Ejecutar el algoritmo con:
  - $x_0 = 2$
  - $d_k = -\frac{1}{(k+1)(k+2)}$
- $x_k = 1 + \frac{1}{k+1}$
- $x_k \longrightarrow 1$ cuando $k \longrightarrow \infty$
- $f'(1) = 2$, 1 no es punto estacionario.

- Lo que sucede es que la distancia entre x_k y x_{k+1} se aproxima a 0 muy rápido comparado al decrecimiento de la función.
- También puede pasar que aunque x_k y x_{k+1} estén a distancia, el decrecimiento de la función sea arbitrariamente chico.

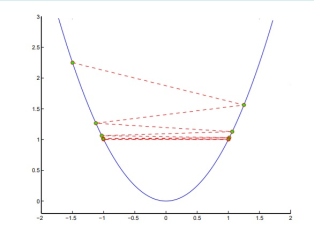
$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

- $f(x) = x^2$
- $\nabla f(x) = 2x$
- Ejecutar el algoritmo con:
  - $x_0 = 1$
  - $d_k = -\frac{1}{(k+1)(k+2)}$
- $x_k = \frac{1}{k+1}$
- $x_k \longrightarrow 0$ cuando $k \longrightarrow \infty$
- $f'(0) = 0$, 0 es punto estacionario.

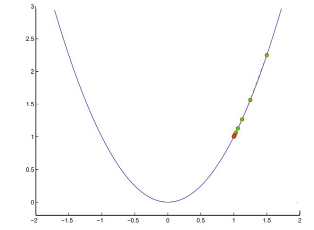
$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

¿Que ocurre con el paso α_k ?

Hay que evitar pasos **muy largos** y pasos **muy cortos**



Paso largo



Paso corto

$x_0 = 2$
$p_k = (-1)^{k+1}$
$\alpha_k = \frac{3}{2^{k+1}}$
$x_{k+1} = x_k - (-1)^{k+1}\frac{3}{2^{k+1}}$

$x_0 = 2$
$p_k = -1$
$\alpha_k = \frac{1}{2^{k+1}}$
$x_{k+1} = x_k - \frac{1}{2^{k+1}}$

Condición

$$||d_k|| \geq \sigma ||\nabla f(x_k)^T|| \quad \forall k$$

con $\sigma > 0$.

Condición

$$||d_k|| \geq \sigma ||\nabla f(x_k)^T|| \quad \forall k$$

con $\sigma > 0$.

- Impide que los pasos se aproximen muy rápido a 0.
- Si $||d_k|| \to 0$, entonces $||\nabla f(x_k)^T|| \to 0$.

Condición del ángulo

$$\nabla f^T(x_k) d_k \leq -\theta ||\nabla f^T(x_k)|| ||d_k||$$

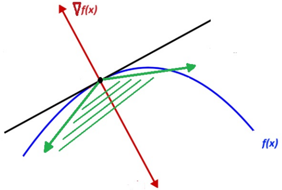
con $\theta \in (0, 1]$.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

Condición del ángulo

$$\nabla f^T(x_k)d_k \leq -\theta||\nabla f^T(x_k)||||d_k||$$

con $\theta \in (0, 1]$.



- Se busca que la dirección sea de descenso, forme una ángulo agudo *no cercano* a $\pi/2$ con $-\nabla f^T(x_k)$.
- Si $\nabla f^T(x_k)d_k \approx 0$ se darán pasos muy cortos.
- Si $\theta = 1$, $d_k = -\nabla f^T(x_k)$.

Criterio de Armijo

Decrecimiento suficiente si:

$$\phi(\alpha_k) = f(x_k + \alpha_k d_k) < f(x_k) + c_1 \alpha_k \nabla f^T(x_k) d_k$$

con $c_1 \in (0, 1)$.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Búsqueda lineal

Criterio de Armijo

Decrecimiento suficiente si:

$$\phi(\alpha_k) = f(x_k + \alpha_k d_k) < f(x_k) + c_1 \alpha_k \nabla f^T(x_k) d_k$$

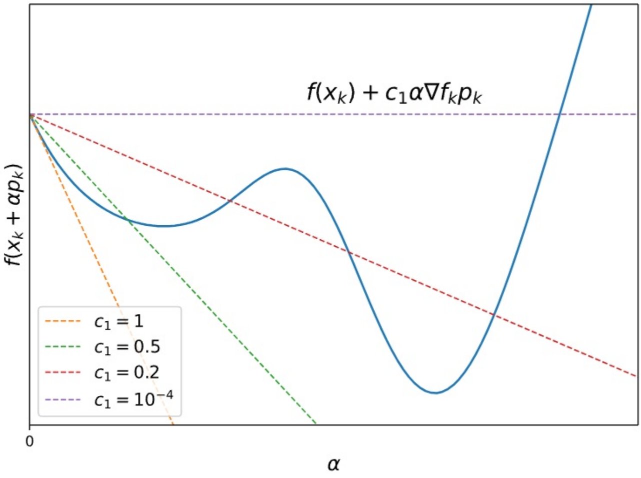
con $c_1 \in (0, 1)$.

- Asegurar que cada paso proporcione una reducción suficiente en la función objetivo.
- Obliga a que el decrecimiento sea proporcional al tamaño del paso.
- Evita pasos grandes con poco decrecimiento de f.

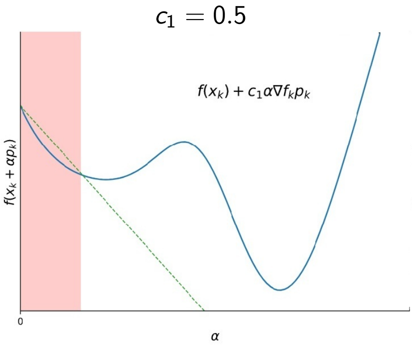
$f : \mathbb{R}^n \to \mathbb{R}$: Criterio de Armijo

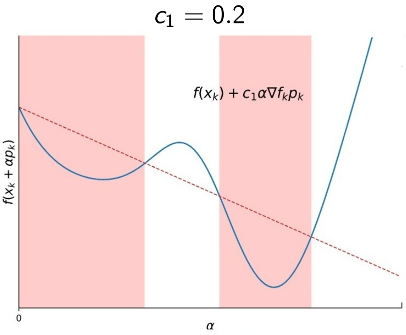
$$l(\alpha) = f(x_k) + c_1 \alpha \nabla f^T(x_k) d_k$$

- Pendiente de $l(\alpha) = c_1 \nabla f^T(x_k) d_k$ negativa.



$f:\mathbb{R}^n \to \mathbb{R}$: Criterio de Armijo





$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

Algoritmo básico

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$, $\sigma > 0$, $\theta \in (0, 1]$, $c_1 \in (0, 1)$
2. Mientras no se cumpla el criterio de parada:
   1. Si $\nabla f(x_k) < \epsilon$, PARAR.
   2. Seleccionar d_k dirección de descenso, tal que:
      - $||d_k|| \geq \sigma||\nabla f^T(x_k)||$
      - $\nabla f^T(x_k)d_k \leq -\theta||\nabla f^T(x_k)||||d_k||$
   3. $\alpha_k \longleftarrow 1$
   4. Mientras $f(x_k + \alpha_k d_k) \geq f(x_k) + c_1\alpha_k\nabla f^T(x_k)d_k$:
      - seleccionar $\alpha_k \in (0.1\alpha_k, 0.9\alpha_k)$
   5. $x_{k+1} \longleftarrow x_k + \alpha_k d_k$
   6. $k \longleftarrow k + 1$

El algoritmo básico:

- Termina por $\nabla f(x_k) < \epsilon$,
- o genera una secuencia infinita x_k tal que si x^* es punto de acumulación de la secuencia, entonces $\nabla f(x^*) = 0$.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Métodos de descenso

El algoritmo básico:

- Termina por $\nabla f(x_k) < \epsilon$,
- o genera una secuencia infinita x_k tal que si x^* es punto de acumulación de la secuencia, entonces $\nabla f(x^*) = 0$.

Mejora de criterios para el paso

- Criterio de Wolfe (condición de curvatura).
- Criterio de Goldstein.

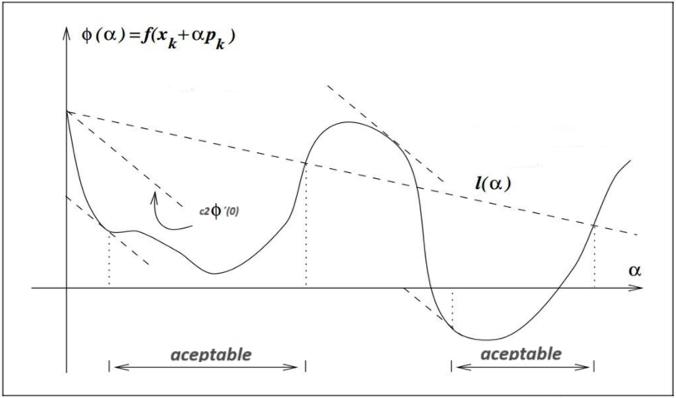
$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Búsqueda lineal

Criterio de Wolfe

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq c_2 \nabla f(x_k) d_k$$

para $c_2 \in (c_1, 1)$.

- Es decir,
$$\phi'(\alpha_k) \geq c_2 \phi'(0).$$
- Si la pendiente de $\phi(\alpha_k)$ es muy negativa, aumentando el valor de α se obtendrían reducciones significativas de f.
- En cambio, si la pendiente de $\phi(\alpha_k)$ es poco negativa o positiva, no es esperable obtener reducciones significativas de f en esa dirección.
- Evita pasos excesivamente pequeños.



$\phi(\alpha) = f(x_k + \alpha p_k)$
$c2\phi'(0)$
$l(\alpha)$
α
aceptable
aceptable

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Búsqueda lineal

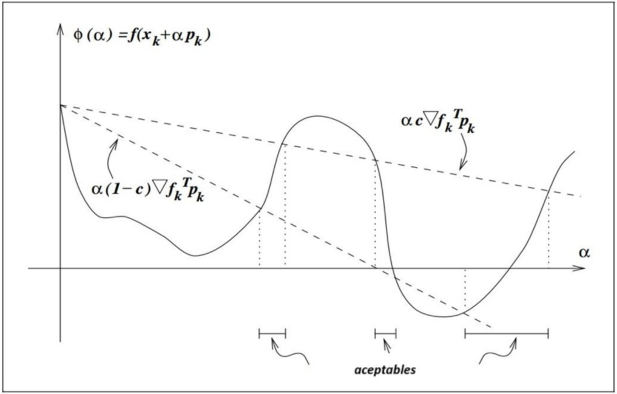
Criterio de Goldstein

$$f(x_k) + (1-c)\alpha_k \nabla f^T(x_k) d_k < \phi(\alpha_k) < f(x_k) + c\alpha_k \nabla f^T(x_k) d_k$$

con $c \in (0, 1/2)$.

- Evita pasos excesivamente pequeños.
- Puede excluir minimizadores de ϕ.
- Buenos resultados en métodos tipo Newton.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Condiciones de Goldstein



$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Direcciones de descenso

- Método del gradiente (mayor descenso)
- Método de direcciones conjugadas
- Método de Newton
- Método Quasi Newton

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método del gradiente

Se elige $d_k = -\nabla f(x_k)$

Algoritmo del gradiente

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$, $c_1 \in (0, 1)$
2. Mientras no se cumpla el criterio de parada:
   1. Si $\nabla f(x_k) < \epsilon$, PARAR.
   2. $\alpha_k \longleftarrow 1$
   3. Mientras $f(x_k - \alpha_k \nabla f(x_k)) \geq f(x_k) - c_1 \alpha_k \nabla^T f(x_k) \nabla f(x_k)$:
      - seleccionar $\alpha_k \in (0.1\alpha_k, 0.9\alpha_k)$
   4. $x_{k+1} \longleftarrow x_k - \alpha_k \nabla f(x_k)$
   5. $k \longleftarrow k + 1$

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método del gradiente

Caso cuadrático

- $f(x) = \frac{1}{2}x^T G x + b^T x$, con G definida positiva
- $\nabla f(x) = Gx + b$
- $\phi(\alpha) = f(x_k) - \alpha||\nabla f(x_k)||^2 + \alpha^2 \nabla f(x_k)^T G \nabla f(x_k)$
- $\phi'(\alpha) = -||\nabla f(x_k)||^2 + \alpha \nabla f^T(x_k) G \nabla f(x_k)$
- Haciendo $\phi'(\alpha) = 0$ se obtiene el mínimo global de $\phi(\alpha)$
- $\arg \min_{\alpha \in \mathbb{R}} \phi(\alpha) = \frac{||\nabla f(x_k)||^2}{\nabla^T f(x_k) G \nabla f(x_k)}$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$: Método del gradiente

Caso cuadrático

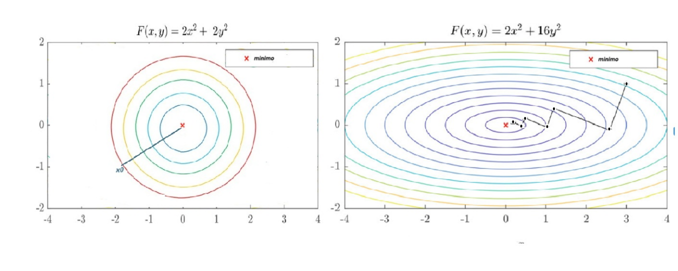
- $f(x) = \frac{1}{2}x^T Gx + b^T x$, con G definida positiva
- $\nabla f(x) = Gx + b$
- $\phi(\alpha) = f(x_k) - \alpha||\nabla f(x_k)||^2 + \alpha^2 \nabla f(x_k)^T G \nabla f(x_k)$
- $\phi'(\alpha) = -||\nabla f(x_k)||^2 + \alpha \nabla f^T(x_k) G \nabla f(x_k)$
- Haciendo $\phi'(\alpha) = 0$ se obtiene el mínimo global de $\phi(\alpha)$
- $\arg\text{mín}_{\alpha \in \mathbb{R}} \phi(\alpha) = \frac{||\nabla f(x_k)||^2}{\nabla^T f(x_k) G \nabla f(x_k)}$

Algoritmo caso cuadrático

1. Inicializar $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$
2. Mientras no se cumpla el criterio de parada:
    1. Si $\nabla f(x_k) < \epsilon$, PARAR.
    2. $x_{k+1} \longleftarrow x_k + \frac{||\nabla f(x_k)||^2}{\nabla^T f(x_k) G \nabla f(x_k)}(Gx_k + b)$
    3. $k \longleftarrow k + 1$

Caso cuadrático Método del gradiente

Algoritmo caso cuadrático



$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método del gradiente

Caso cuadrático

Se puede establecer una cota del descenso entre iteraciones.

$$f(x_{k+1}) - f(x^*) \leq (\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}})^2 (f(x_k) - f(x^*))$$

donde λ_{max} y λ_{min} son los autovalores máximo y mínimo de G

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de direcciones

Caso cuadrático

- Si G es diagonal, las curvas de nivel son elipses *alineadas* con los ejes cartesianos. Tomar como direcciones a los ejes coordenados, lleva al óptimo en un número finito de pasos.
- Si G no es diagonal, no hay una alineación con los ejes y la convergencia puede resultar muy lenta si nos movemos por ellos.

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de direcciones

Caso cuadrático

- Sean $p_0, p_1, \dots, p_{n-1}$ tales que $p_i G p_j = 0$.
- Se puede demostrar que estos vectores son l.i.
- Definimos S una matriz de $n \times n$ tal que $S = [p_0 p_1 \dots p_{n-1}]$.
- Sea $x = S\hat{x}$
- Realizando un cambio de variable, $\frac{1}{2}x^T Gx + b^T x = \frac{1}{2}\bar{x}^T S^T GS\bar{x} + S^T b^T \bar{x}$
- Como $p_i G p_j = 0$ resulta que $S^T GS = D$, matriz diagonal y por lo tanto servirán los ejes coordenados como direcciones.
- Viéndolo en las variables originales, corresponde a tomar como direcciones a los p_i.

¿Cómo encontramos $p_0, p_1, \dots, p_{n-1}$?

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de direcciones

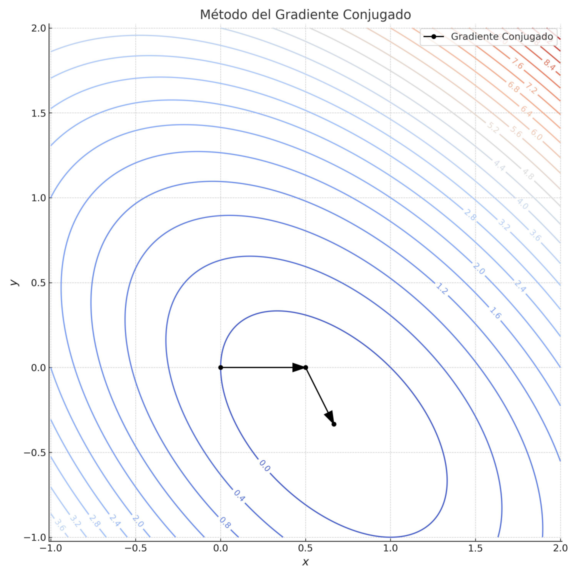
Caso cuadrático

- Por descomposición SVD, $A = UDU^T$ donde D tiene en la diagonal los autovalores de A y las columnas de U son una base ortonormal de autovectores de A^2.
- Método de gradiente conjugado: generarlos a demanda y la dirección es una combinación lineal del gradiente actual y la dirección anterior:

  $d_{k+1} = -\nabla f(x_{k+1}) + \beta d_k$, con $\beta = \frac{\nabla^T f(x_{k+1}) G d_k}{d_k^T G d_k}$
  y $d_0 = -\nabla f(x_0)$

Gradiente conjugado

Caso cuadrático



$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de Newton

Primera interpretación

- Se considera una *aproximación cuadrática* basada en la expresión de Taylor:
  $f(x_k + p) = f(x_k) + \nabla f(x_k)p + \frac{1}{2}p^T H(x_k)p$
  Donde $H(x_k)$ es la matriz hessiana de f en x_k. Asumimos $H(x_k)$ simétrica y definida positiva.

- Se busca el mínimo de esta función. Se alcanza en p_k con $H(x_k)p_k = -\nabla f(x_k)$

- Podemos considerar entonces a p_k como una dirección de descenso y definir

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

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de Newton

Segunda interpretación

- Se aproxima en orden 1 a la función gradiente

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

- Se busca x donde se anula

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

$f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2$: Método de Newton

Teorema

Si x_0 esta suficientemente cerca de x^*, f es C^2, la matriz hessiana es definida positiva y satisface condición de Lipschitz en una vecindad de x^* entonces la sucesión converge a x^* y el orden es cuadrático.

Modificaciones:

- Realizar una búsqueda en la dirección p_k, resulta entonces $x_{k+1} = x_k - \alpha_k p_k$
- Si la matriz hessiana no es definida positiva, remplazarla por $H(x_k) + \mu I$

Métodos Quasi Newton

- Utilizan modelos cuadráticos $M_k(p) = f(x_k) + \nabla f(x_k)p + \frac{1}{2}p^T B_k p$ con B_k definida positiva y se define

$$x_{k+1} = x_k + \alpha_k p_k \text{ con } p_k = -B_k^{-1}\nabla f(x_k)$$

- Se busca que el gradiente de $M_{k+1}(x)$ evaluado en x_{k+1} y x_k coincida con el valor del gradiente de f en esos puntos.
- De la condición anterior, surge que

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

.

ECUACIÓN SECANTE

Métodos Quasi Newton

- No está univocamente determinada. ¿Cómo hacerlo?

$$\min_{B \in \mathbb{R}^{nxn}} ||B - B_k||$$

  sujeto a B simétrica def positiva y satisface ecuación secante.
- Diferentes normas matriciales dieron origen a diferentes métodos.
- Los más conocidos son el método BFGS (Broyden-Fletcher-Goldfarb-Shanno) y el DFP (Davidon-Fletcher-Powell) con reglas de actualización de una iteración a otra.
- Se pierde la convergencia con orden cuadrático.