# Búsqueda en Línea

Alan Reyes-Figueroa

Métodos Numéricos II

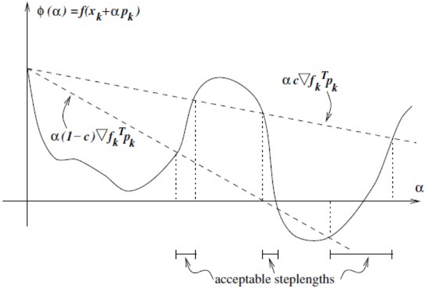
(Aula 27b) 23.octubre.2025

**Condiciones de Goldstein**:

Similares a las condiciones de Wolfe, las condiciones de Goldstein aseguran que el tamaño de paso $\alpha$ alcance un descenso suficiente para $f$. Se definen por

$$f(\mathbf{x}_k) + (1-c)\,\alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \leq f(\mathbf{x}_k + \alpha_k\,\mathbf{d}_k)^T \leq f(\mathbf{x}_k) + c\,\alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k, \tag{1}$$

para algún $c \in (0, \frac{1}{2})$.



La segunda desigualdad en (1) asegura un descenso suficiente, mientras que la primera controla que $\alpha_k$ esté lejos de 0.

## Teorema (Existencia de $\alpha$ con Condiciones de Goldstein)

*Suponga que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable, $\mathbf{d}_k$ es una dirección de descenso en $\mathbf{x}_k$, $f$ es limitada inferiormente sobre el rayo $\{\mathbf{x}_k + \alpha\,\mathbf{d}_k : \alpha > 0\}$. Si $0 < c < \frac{1}{2}$, entonces existe un intervalo para $\alpha$ donde se satisfacen las condiciones de Goldstein.*

Prueba: Sea $\ell(\alpha) = (\mathbf{x}_k) + \frac{1}{2}\,\alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$ la recta con pendiente relativa $\frac{1}{2}\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$. Como $f$ es limitada inferiormente para $\alpha > 0$, existe al menos un valor $\alpha' > 0$ donde $\ell(\alpha') = \varphi(\alpha')$, esto es

$$f(\mathbf{x}_k + \alpha'\,\mathbf{d}_k) = f(\mathbf{x}_k) + \tfrac{1}{2}\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k.$$

Como $0 < c < \frac{1}{2}$, entonces $1 - c > \frac{1}{2} > c > 0$ y como $\nabla f(_k)^T\mathbf{d}_k < 0$, tenemos

$$(1 - c)\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < \tfrac{1}{2}\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < c\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k.$$

# Búsqueda en Línea

Sumando $f(\mathbf{x}_k)$ a esta doble desigualdad, resulta

$$f(\mathbf{x}_k) + (1-c)\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < f(\mathbf{x}_k) + \tfrac{1}{2}\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < f(\mathbf{x}_k) + c\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k,$$

o equivalentemente

$$f(\mathbf{x}_k) + (1-c)\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < f(\mathbf{x}_k + \alpha'\,\mathbf{d}_k) < f(\mathbf{x}_k) + c\,\alpha'\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k.$$

Así, $\alpha'$ satisface las condiciones de Goldstein. Por la continuidad de $f$ y de $\nabla f$ existe una vecindad $I$ de $\alpha''$ donde se satisfacen ambas desigualdades. $\square$

**Obs!** Las condiciones de Wolfe y las condiciones de Goldstein poseen propiedades teóricas similares. La desventaja práctica de las condiciones de Goldstein respecto de las de Wolfe es que las primeras pueden excluir algunos (o todos) los mínimos locales de $\varphi$.

Usualmente las condiciones de Goldstein se usan en los métodos de tipo Newton, y pueden funcionar mal para métodos quasi-Newton o con hessiana modificada.

# Backtracking

La condición de Armijo (**??**) por si sola no es suficiente para asegurar que el descenso sea razonable en la dirección de búsqueda. Si el algoritmo de búsqueda en línea elige $\alpha$ de forma apropiada, podemos dispensar la condición de curvatura (**??**) y basta con usar la primer condición como criterio de paro. Este es el algoritmo llamado *Backtracking*.

**Algoritmo:** (*Backtracking*)
*Inputs*: $f : \mathbb{R}^n \to \mathbb{R}$ de clase $C^1$, $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{d}_k$ dirección de descenso, $\bar{\alpha} > 0$, $c \in (0, 1)$, $\rho \in (0, 1)$ tasa de decaimiento.
*Outputs*: $\alpha$ tamaño de paso satisfaciendo condiciones de Wolfe.
`Set` $\alpha = \bar{\alpha}$.
`While not` $f(\mathbf{x}_k + \alpha\,\mathbf{d}_k) \leq f(\mathbf{x}_k) + c\,\alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$:
    `redefine` $\alpha = \rho\alpha$.
`Return` $\alpha_k = \alpha$.

- El tamaño de paso inicial es $\bar{\alpha} = 1$ en los métodos de Newton y quasi-Newton.
- El factor de contracción $\rho$ puede variarse en cada iteración, por ejemplo, eligiendo $\rho_k \in [\rho_{min}, \rho_{max}]$, para $0 < \rho_{min} < \rho_{max} < 1$ fijos.

## Teorema (Finitud del Algoritmo Backtracking)

*Suponga que $f:\mathbb{R}^n\rightarrow\mathbb{R}$ es diferenciable, $\nabla f$ es Lipschitz, $c\in(0,1)$, $\rho\in(0,1)$, $\mathbf{d}_k$ es una dirección de descenso en $\mathbf{x}_k$. Entonces, la condición de Armijo se satisface para todo $\alpha_k\in(0,\omega_k)$, donde*

$$\omega_k = \frac{2(c-1)\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k}{\gamma\,||\mathbf{d}_k||^2},$$

*para $\gamma$ la constante de Lipschitz de $\nabla f$: $|\nabla f(\mathbf{x})-\nabla f(\mathbf{y})|\leq\gamma||\mathbf{x}-\mathbf{y}||$.*

<u>Prueba</u>: Del Teorema de Taylor,
$f(\mathbf{x}_k+\alpha\,\mathbf{d}_k) = f(\mathbf{x}_k)+\alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k+\frac{1}{2}\alpha^2\,\mathbf{d}_k^T\,D^2f(\mathbf{x}_k+t\alpha\,\mathbf{d}_k)\,\mathbf{d}_k$, con $t\in(0,1)$.

De la condición de Lispchitz sobre $\nabla f$, sabemos que

$$\begin{aligned}
|\mathbf{d}_k^T\,D^2f(\mathbf{x}_k+t\,\alpha\,\mathbf{d}_k)\,\mathbf{d}_k| &= \left|\lim_{h\rightarrow 0}\frac{\nabla f(\mathbf{x}_k+t\,\alpha\,\mathbf{d}_k+h\,\mathbf{d}_k)^T\mathbf{d}_k-\nabla f(\mathbf{x}_k+t\,\alpha\,\mathbf{d}_k)^T\mathbf{d}_k}{h}\right| \\
&= \lim_{h\rightarrow 0}\frac{||\nabla f(\mathbf{x}_k+t\,\alpha\,\mathbf{d}_k+h\,\mathbf{d}_k)^T\mathbf{d}_k-\nabla f(\mathbf{x}_k+t\,\alpha\,\mathbf{d}_k)^T\mathbf{d}_k||}{h}
\end{aligned}$$

$$\Longrightarrow \quad |\mathbf{d}_k^T D^2 f(\mathbf{x}_k + t\,\alpha\,\mathbf{d}_k)\,\mathbf{d}_k| \quad \leq \quad \lim_{h\to 0} \frac{\gamma\,||\mathbf{d}_k||\cdot||h\mathbf{d}_k||}{h} \quad = \gamma\,||\mathbf{d}_k||^2.$$

De ahí que $f(\mathbf{x}_k + \alpha\,\mathbf{d}_k) = f(\mathbf{x}_k) + \alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k + E$, con $|E| \leq \frac{1}{2}\alpha^2\,\gamma\,||\mathbf{d}_k||^2$.

Si $\alpha_k \leq \omega_k = \frac{2(c-1)\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k}{\gamma\,||\mathbf{d}_k||^2}$, entonces $\alpha_k\,\gamma\,||\mathbf{d}_k||^2 \leq 2(c-1)\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$, y

$$\begin{aligned} f(\mathbf{x}_k + \alpha_k\,\mathbf{d}_k) &\leq f(\mathbf{x}_k) + \alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k + \tfrac{1}{2}\alpha_k^2\,\gamma\,||\mathbf{d}_k||^2 \\ &\leq f(\mathbf{x}_k) + \alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k + \tfrac{1}{2}\alpha_k\,(c-1)\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \\ &\leq f(\mathbf{x}_k) + \alpha_k\,c\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k. \end{aligned}$$

Esto muestra que $\alpha_k$ satisface la condición de Armijo. $\square$

## Corolario

*Bajo las hipótesis del teorema anterior, el tamaño de paso en el algoritmo de Backtracking termina con* $\alpha_k \geq \min\{\bar{\alpha}, \rho\omega_k\}$.

Prueba: El algoritmos de *Backtracking* acaba si

i) $\bar{\alpha}$ inicial satisface la condición de Armijo,

ii) si $\alpha^{(l-1)} \geq \omega_k$, pero $\alpha_k = \alpha^{(l)} = \rho\alpha^{(l-1)} \leq \omega_k$.

Combinando estos resultados, tenemos que $\alpha_k \geq \min\{\bar{\alpha}, \rho\omega_k\}$. $\square$

Para obtener convergencia global, añadimos algunos requerimientos sobre los criterios de aceptación del tamaño de paso.

## Teorema (Convergencia Global del Algoritmo Backtracking)

*Suponga que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable con $\nabla f$ Lipschitz de constante $\gamma$, $c \in (0, 1)$, $\mathbf{d}_k$ es una dirección de descenso para $\mathbf{x}_k$. Entonces en el algoritmo de Backtracking ocurre:*

1. $\nabla f(\mathbf{x}_k) = \mathbf{0}$, *para algún* $k \geq 0$, *ó*
2. $\lim_{k\to\infty} f(\mathbf{x}_k) = -\infty$, *ó*
3. $\lim_{k\to\infty} |\nabla f(\mathbf{x}_k)^T\mathbf{d}_k| \cdot \min\{1, ||\mathbf{d}_k||^{-1}\} = 0$.

# Búsqueda en Línea

(1.) Significa que alcanzamos un punto estacionario en un número finito de pasos; (2.) significa que $f$ no es limitada inferiormente $\Rightarrow$ $f$ no tiene mínimo global; (3.) no implica convergencia, pero si $\nabla f(\mathbf{x}_k)$ y $\mathbf{d}_k$ no son ortogonales, y $||\mathbf{d}_k|| \not\to 0$, entonces $||\nabla f(\mathbf{x}_k)|| \to 0$.

<u>Prueba</u>: Suponga que (1) y (2) no se satisfacen. Mostramos (3). Considere

$$f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + c\,\alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \leq f(\mathbf{x}_0) + \sum_{j=0}^{k} c\,\alpha_j\,\nabla f(\mathbf{x}_j)^T\mathbf{d}_j.$$

Como cada $\mathbf{d}_j$ es dirección de descenso, y $f$ es limitada inferiormente,

$$\sum_{j=0}^{\infty} \alpha_j\,|\nabla f(\mathbf{x}_j)^T\mathbf{d}_j| = -\tfrac{1}{c}\lim_{k\to\infty}\sum_{j=0}^{k} c\,\alpha_j\,|\nabla f(\mathbf{x}_j)^T\mathbf{d}_j| = \tfrac{1}{c}\lim_{k\to\infty}\big(f(\mathbf{x}_0) - f(\mathbf{x}_k)\big) = \tfrac{1}{c}\big(f(\mathbf{x}_0) - f(\mathbf{x}^*)\big).$$

De ahí que $\lim\limits_{k\to\infty} \alpha_k\,|\nabla f(\mathbf{x}_k)^T\mathbf{d}_k| = 0$.

Como $\alpha_k \geq \min\{\bar{\alpha}, \rho\omega_k\}$, con ${}_k = 2(c-1)\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k/\gamma\,||\mathbf{d}_k||^2$, consideramos los conjuntos

# Búsqueda en Línea

$$K_1 = \{k \in \mathbb{N} : \ \alpha_k = \bar{\alpha}\}, \qquad K_2 = \{k \in \mathbb{N} : \ \alpha_k < \bar{\alpha}\}, \qquad \text{con } K_1 \cup K_2 = \mathbb{N}.$$

De $\lim\limits_{k\to\infty} \alpha_k \, |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| = 0$, resulta $\lim\limits_{k\to\infty,\, k\in K_1} \alpha_k \, |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| = 0$ y $\lim\limits_{k\to\infty,\, k\in K_2} \alpha_k \, |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| = 0$.

- Para $k \in K_1$: $\alpha_k = \bar{\alpha} > 0$, luego $\bar{\alpha} \lim\limits_{k\to\infty} |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| = 0$, lo que implica $\lim\limits_{k\to\infty} |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| = 0$.

- Para $k \in K_2$: $\rho\omega_k \leq \alpha_k \leq \omega_k$. Entonces

$$\alpha_k \, |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| \geq \rho\omega_k \, |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| \geq 2\rho(1-c) \, \frac{(\nabla f(\mathbf{x}_k)^T \mathbf{d}_k)^2}{\gamma \, ||\mathbf{d}_k||^2},$$

y esto implica que $\lim\limits_{k\to\infty} |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| \, \min\{1, ||\mathbf{d}_k||^{-1}\} = 0$. $\square$

# Búsqueda en Línea

Tenemos una versión más completa del algoritmo de descenso gradiente.

**Algoritmo:** (Descenso gradiente con Backtracking)
*Inputs*: $f : \mathbb{R}^n \to \mathbb{R}$ función de clase $C^1$, $\mathbf{x}_0 \in \mathbb{R}^n$, $\bar{\alpha} > 0$ tamaño de paso; $c \in (0, 1)$ constante en la condición de Armijo; $\rho \in (0, 1)$ parámetro de decaimiento.
*Outputs*: $\mathbf{x}$ punto crítico de $f$.

```
For k = 0,1,2,... hasta que se cumpla un criterio de paro:
    Compute d_k a descent direction
    (for example, any d_k such that ∠(−∇f(x_k), d_k) < |π/2|).
    Compute step-size α_k along d_k, using Backtracking
    (Backtraking es descrito en el algoritmo en página 14).
    Set x_{k+1} = x_k + α_k d_k.
Return x_{k+1}.
```

For k = 0,1,2,... hasta que se cumpla un criterio de paro:
  Compute $\mathbf{d}_k$ a descent direction
  (for example, any $\mathbf{d}_k$ such that $\angle(-\nabla f(\mathbf{x}_k), \mathbf{d}_k) < |\frac{\pi}{2}|$).
  Compute step-size $\alpha_k$ along $\mathbf{d}_k$, using Backtracking
  (Backtraking es descrito en el algoritmo en página 14).
  Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$.
Return $\mathbf{x}_{k+1}$.

- Aplica para cualquier dirección de descenso $\mathbf{d}_k$ en $\mathbf{x}_k$.
- La convergencia está garantizada por el Teorema de Convergencia Global, siempre que $f$ sea limitada inferiormente, y que no alcance un punto estacionario.