Búsqueda en Línea

Alan Reyes-Figueroa

Métodos Numéricos II

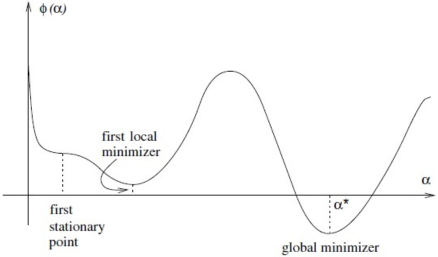
(Aula 27a) 23.octubre.2025

Búsqueda en Línea

Desarrollamos estrategias para calcular un tamaño de paso α_k en los métodos de descenso gradiente. Recordemos que estos métodos calculan una dirección de descenso $\mathbf{d}_k$ y luego se mueven a lo largo de esa dirección con la iteración

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k.$$

Al calcular α_k tenemos un *trade-off*: queremos que α_k haga una reducción sustancial del valor de f, pero al mismo tiempo no queremos invertir mucho costo en hacer la búsqueda. La elección ideal sería el mínimo de la función $\varphi(\alpha) = f(\mathbf{x}_k + \alpha\,\mathbf{d}_k), \quad \alpha > 0.$



Búsqueda en Línea

En general, identificar este valor global es costoso (requiere muchas evaluaciones de f y posiblemente ∇f.

Existen estrategias más prácticas producen un valor inexacto de α^*, pero que producen resultados a bajo costo. Típicamente, estos métodos producen una secuencia de candidatos $\{\alpha_i\}_{i\geq 0}$ y paran cuando alguno de éstos satisface ciertas condiciones. Este proceso se llama **búsqueda en línea**.

La búsqueda en línea se hace en dos etapas:

- Una primera fase halla intervalos o regiones con valores deseables para α.
- Un método de bisección o interpolación calcula α^* dentro de estos intervalos.

Una condición de paro simple para la búsqueda en línea sería

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k).$$

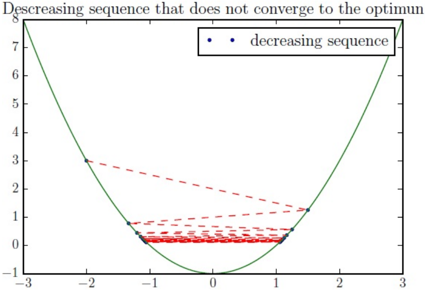
Esta condición no es suficiente para garantizar la convergencia al óptimo $\mathbf{x}^*$ de f.

Ejemplo: Consideramos la función $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$, $\mathbf{x}_k = 0$, $\mathbf{d}_k = 1$, y la secuencia $\{\alpha_i = (-1)^{i+1}\sqrt{1 + \frac{1}{i}}\}_{i \geq 1}$. Observe que $f(\mathbf{x}_k + \alpha_i \mathbf{d}_k) = 1 + \frac{1}{i}$.

Luego,

$$f(\mathbf{x}_k + \alpha_{i+1}\mathbf{d}_k) = 1 + \tfrac{1}{i+1} < 1 + \tfrac{1}{i} = f(\mathbf{x}_k + \alpha_i \mathbf{d}_k),$$

es una secuencia decreciente. Sin embargo, $\lim\limits_{i\to\infty} f(\mathbf{x}_k \alpha_i \mathbf{d}_k) = \lim\limits_{i\to}(1 + \frac{1}{i}) = 1$, no converge al valor del mínimo global $\mathbf{x}^*$ de f.



UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

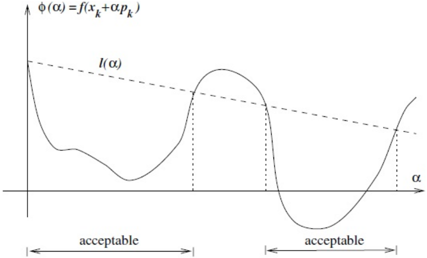
Búsqueda en Línea

Para evitar este descenso insuficientes que impide la convergencia, precisamos condiciones de descenso adecuadas.

Condiciones de Wolfe:
Uno de las condiciones más populares sobre α_k es requerir que satisfaga la **condición de Armijo**:

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + c_1\, \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k, \tag{1}$$

para alguna $c_1 \in (0, 1)$. La reducción es proporcional a α_k y a $\frac{\partial f}{\partial \mathbf{d}_k} = \nabla f(\mathbf{x}_k)^T \mathbf{d}_k$.



El parámetro c_1 controla la pendiente de la recta $\ell(\alpha_k) = f(\mathbf{x}_k) + c_1\,\alpha_k \nabla f(\mathbf{x}_k)^T\mathbf{d}_k$ en (1). En la práctica, se toman valores muy pequeños (*e.g.* $C_1 \approx 10^{-4}$).

Obs! La condición de Armijo por sí sola aún no es suficiente para garantizar la convergencia. Sin embargo, sabemos que existe un intervalo $(0, c_0)$ donde los valores de α en este intervalo satisfacen (1).

En efecto, si $\mathbf{d}_k$ es una dirección de descenso y $0 < c_1 < 1$, entonces $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < 0$. Por Taylor, sabemos que existe $\delta > 0$ tal que si $0 < \alpha_k < \delta$, entonces

$$f(\mathbf{x}_k) \leq f(\mathbf{x}_k) + \alpha_k \nabla f(\mathbf{x}_k)^T\mathbf{d}_k \; < \; f(\mathbf{x}_k) + c_1\,\alpha_k \nabla f(\mathbf{x}_k)^T\mathbf{d}_k.$$

De ahí que para $0 < \alpha_k < \delta$, se satisface (1), aunque no hagan reducción suficiente de f.

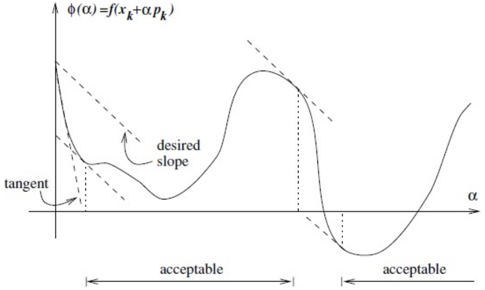
Para evitar esto, se introduce un segundo requisito, conocido como la **condición de curvatura**:

$$\nabla f(\mathbf{x}_k + \alpha_k\mathbf{d}_k)^T\mathbf{d}_k \geq c_2 \nabla f(\mathbf{x}_k)^T\mathbf{d}_k, \tag{2}$$

para algún $c_2 \in (0, 1)$, $c_1 < c_2 < 1$.

Búsqueda en Línea

El término en el lado izquierdo de (2) es justamente $\varphi'(\alpha_k)$, de modo que la condición de curvatura segura que la pendiente $\varphi'(\alpha_k) \geq c_2\,\varphi'(0)$ es al menos c_2 veces la pendiente en el punto inicial $\varphi'(0) = \nabla f(\mathbf{x}_k)^T\mathbf{d}_k$.



- Si $\varphi'(\alpha_k)$ es fuertemente negativa, esto es un indicador de que podemos reducir f significativamente moviéndose en esa dirección, con ese valor de α_k.
- Si $\varphi'(\alpha_k)$ es cercano a 0, no debemos esperar mucha reducción de f en esa dir.

Valores típicos para c_2 son $c_2 \approx 0.9$, cuando $\mathbf{d}_k$ se elige usando métodos de tipo Newton o quasi-Newton. Ó $c_2 \approx 0.1$ cuando $\mathbf{d}_k$ proviene de método de gradiente conjugado.

A las condiciones de Armijo y de curvatura juntas, se les conoce como las **condiciones de Wolfe**:

$$\begin{aligned} f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) &\leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k, \quad (3) \\ c_2 \nabla f(\mathbf{x}_k)^T \mathbf{d}_k &\leq \nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k, \quad (4) \end{aligned}$$

con $0 < c_1 < c_2 < 1$.

La idea es que un valor de α_k de tamaño de paso que satisface las condiciones de Wolfe, produce puntos $\mathbf{x}_k + \alpha_k \mathbf{d}_k$ que están suficientemente cerca de un mínimo local de φ.
La condición de curvatura (2) puede modificarse para forzar que α_k esté cerca de un mínimo local de f. Así, se definen las **condiciones de Wolfe fuertes**:

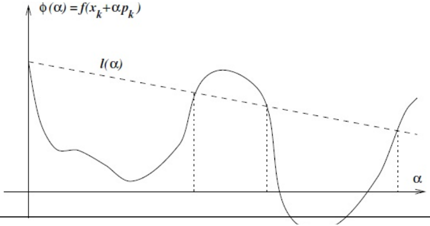
$$\begin{aligned} f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) &\leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k, \quad (5) \\ c_2 |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| &\geq |\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k|, \quad (6) \end{aligned}$$

con $0 < c_1 < c_2 < 1$.

Teorema (Existencia de α con Condiciones de Wolfe)

Sea $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^1, $\mathbf{d}_k$ una dirección de descenso en $\mathbf{x}_k$, y suponga que f es limitada inferiormente en el rayo $\{\mathbf{x}_k + \alpha\,\mathbf{d}_k : \alpha > 0\}$. Si $0 < c_1 < c_2 < 1$, entonces existe un intervalo para α donde se satisfacen las condiciones de Wolfe y las condiciones de fuertes de Wolfe.

<u>Prueba</u>: La función $\varphi(\alpha) = f(\mathbf{x}_k + \alpha\,\mathbf{d}_k)$ es limitada inferiormente, para $\alpha > 0$. Como $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < 0$, la recta $\ell(\alpha) = f(\mathbf{x}_k) + \alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$ interseca a φ en al menos un punto $\alpha > 0$.



Búsqueda en Línea

(Esto es porque $\ell(0) = \varphi(0)$, y si b es una cota inferior para φ, entonces ℓ interseca a $y = b$. Como $0 < c_1 < 1$, entonces ℓ interseca también a φ, en algún punto α, (ℓ decrece menos que φ)).

Sea $\alpha' > 0$ el menor valor donde ℓ corta a φ, $\alpha' = \inf\{\alpha > 0 : \ell(\alpha) = \varphi(\alpha)\}$. Luego, $(\mathbf{x}_k + \alpha'\mathbf{d}_k) = f(\mathbf{x}_k) + c_1\,\alpha'\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$. Como $0 < c_1 < 1$, por definición de α' se sabe que $(\mathbf{x}_k + \alpha\,\mathbf{d}_k) < f(\mathbf{x}_k) + c_1\,\alpha\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$, para todo $0 < \alpha < \alpha'$, y se satisface la condición de Armijo (1) en $(0, \alpha')$.

Por el Teorema del Valor Medio, como φ es diferenciable en $(0, \alpha')$, existe $\alpha'' \in (0, \alpha')$ tal que

$$f(\mathbf{x}_k + \alpha''\mathbf{d}_k) - f(\mathbf{x}_k) = \varphi(\alpha') - \varphi(0) = \alpha' \cdot \varphi'(\alpha'') = \alpha'\,\nabla f(\mathbf{x}_k + \alpha''\mathbf{d}_k)^T\mathbf{d}_k.$$

Luego, $\nabla f(\mathbf{x}_k + \alpha''\mathbf{d}_k) = c_1\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k > c_2\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k$, ya que $c_1 < c_2$ y $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < 0$. Así, $\alpha'' \in (0, \alpha')$ satisface la condición de curvatura (2), y portanto las condiciones de Wolfe.

Ahora, como $\nabla f(\mathbf{x}_k)$ es continua, entonces la condición de curvatura también se cumple en una vecindad U de α'', y las condiciones de Wolfe se satisfacen en el intervalo $I = U \cap (0, \alpha')$.

Por otro lado, como $\nabla f(\mathbf{x}_k + \alpha''\mathbf{d}_k)^T\mathbf{d}_k < 0$ y $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k < 0$, entonces

$$-\nabla f(\mathbf{x}_k + \alpha''\mathbf{d}_k)^T\mathbf{d}_k < -c_2\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k,$$

de modo que $|\nabla f(\mathbf{x}_k + \alpha''\mathbf{d}_k)^T\mathbf{d}_k| < c_2\,|\nabla f(\mathbf{x}_k)^T\mathbf{d}_k|$.
Esto muestra que α'' y el intervalo I también satisfacen las condiciones fuertes de Wolfe.

$\square$