Búsqueda en Línea

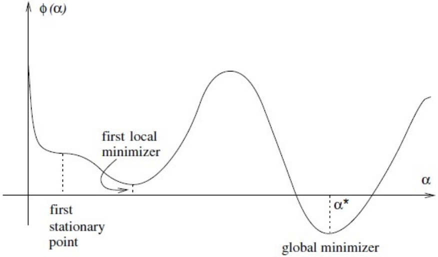
Alan Reyes-Figueroa
Métodos Numéricos II (Aula 23) 03.octubre.2023

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

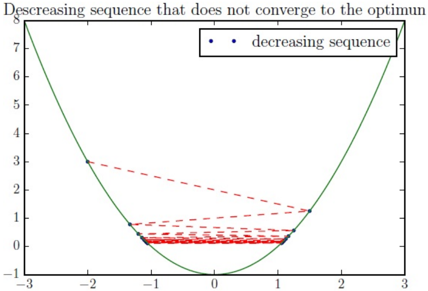
$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k).$$

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.



UNIVERSIDAD DEL VALLE DE GUATEMALA

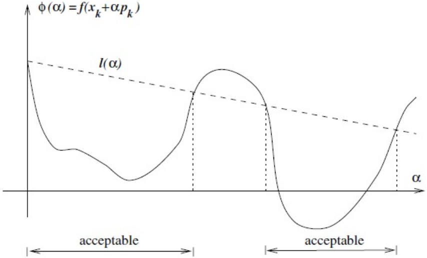
Búsqueda en Línea

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.

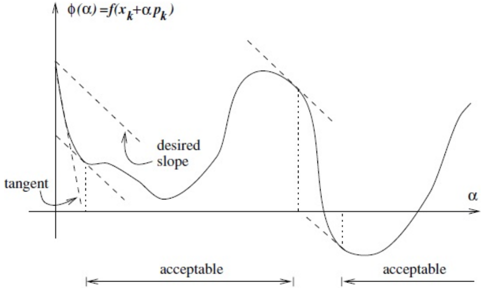
Para evitar esto, se introduce un segundo requisito, conocido como la **condición de curvatura**:

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

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.

Búsqueda en Línea

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**:

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

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**:

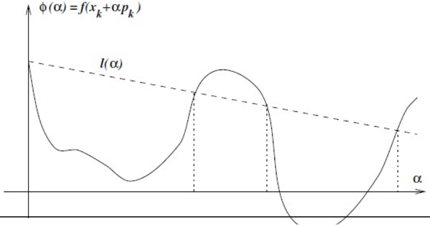
$$
\begin{aligned}
f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) &\leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k, & (5)\\
c_2 |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| &\geq |\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k|, & (6)
\end{aligned}
$$

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.

Prueba: 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 $\alpha > 0$.



(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.

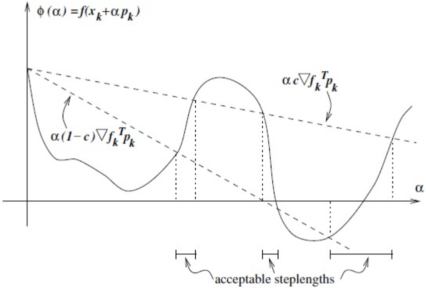
$\square$

Condiciones de Goldstein:

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

$$f(\mathbf{x}_k) + (1-c)\,\alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \leq f(\mathbf{x}_k + \alpha_k\,\mathbf{d}_k)^T \leq f(\mathbf{x}_k) + c\,\alpha_k\,\nabla f(\mathbf{x}_k)^T\mathbf{d}_k, \qquad (7)$$

para algún $c \in (0, \frac{1}{2})$.



Búsqueda en Línea

La segunda desigualdad en (7) asegura un descenso suficiente, mientras que la primera controla que α_k esté lejos de 0.

Teorema (Existencia de α 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 α donde se satisfacen las condiciones de Goldstein.

<u>Prueba</u>: 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

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

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

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

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, α' satisface las condiciones de Goldstein. Por la continuidad de f y de ∇f existe una vecindad I de α'' 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 φ.

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 (1) 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 α de forma apropiada, podemos dispensar la condición de curvatura (2) 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: α 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 ρ puede variarse en cada iteración, por ejemplo, eligiendo $\rho_k \in [\rho_{min}, \rho_{max}]$, para $0 < \rho_{min} < \rho_{max} < 1$ fijos.

Búsqueda en Línea

Teorema (Finitud del Algoritmo Backtracking)

Suponga que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable, ∇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 γ la constante de Lipschitz de ∇f: $|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})| \leq \gamma||\mathbf{x} - \mathbf{y}||$.

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

De la condición de Lispchitz sobre ∇f, sabemos que

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

Búsqueda en Línea

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

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 α_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 $\alpha_k \geq \min\{\bar{\alpha}, \rho\omega_k\}$.

<u>Prueba</u>: 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* ∇f *Lipschitz de constante* γ, $c \in (0,1)$, $\mathbf{d}_k$ *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. $\lim_{k\to\infty} |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k| \cdot \min\{1, ||\mathbf{d}_k||^{-1}\} = 0$.

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

UNIVERSIDAD DEL VALLE DE GUATEMALA

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` $\mathbf{d}_k$ `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` α_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.