Proyecto Optimización

Giovanni Gamaliel López Padilla
Edgar Osvaldo López Zúñiga

10 de mayo de 2022

Introducción

Desarrollo

En 1988 Barzilai y Borwein propusieron dos tamaños de paso para mejorar el desempeño de métodos de descenso de gradiente, estos tamaños de paso son los siguientes:

$$\alpha_k^{BB1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \text{ y } \alpha_k^{BB2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}$$

donde $s_{k-1} = x_k - x_{k-1}$ y $y_{k-1} = g_k - g_{k-1}$. Ahora, si consideramos por la desigualdad de Cauchy-Schwarz que $\|s_{k-1}\|^2\|y_{k-1}\|^2 \geq (s_{k-1}^T y_{k-1})^2$ se puede observar que cuando $s_{k-1}^T y_{k-1} > 0$ se cumple que $\alpha_k^{BB1} \geq \alpha_k^{BB2}$. Por lo que se suele llamar al paso α_k^{BB1} el paso largo de Barzilai-Borwein y a α_k^{BB2} se le llama paso corto de Barzilai-Borwein. En el caso en el que la función es cuadrática, el paso α_k^{BB1} será el tamaño de paso de máximo descenso con un retardo de un paso y α_k^{BB2} será el paso del método de mínimo gradiente.

Se ha probado que el método de Barzilai-Borwein (BB) tiene convergencia R-superlineal para minimizar funciones cuadráticas bidimensionales fuertemente convexas y R-lineal para el caso general n-dimensional.

El método de BB cuenta con la propiedad de reducir los valores de la función objetivo de manera no monóntona, y esta propiedad es una característica intrínseca del mismo que causa su eficiencia. Sin embargo, es importante para los métodos de gradiente mantener la monotonicidad.

Debido a que en el caso de optimización general es difícil calcular un tamaño de paso α_k^{SD} y a que el algoritmo de BB ha sido exitoso a la hora de mejorar el desmpeño de los métodos de gradiente, el trabajo analizado se motiva en la búsqueda de una forma de aceleración para el método de Barzilai-Borwein incorporando pasos monótonos.

Considérese primero el problema de acelerar los métodos de descenso de gradiente que generan secuencias de iterados de la siguiente forma:

$$x_{k+1} = x_k - \alpha_k g_k$$

para resolver el problema

$$\min_{x \in \mathbb{R}^n} f(x)$$

donde $f : \mathbb{R}^n \to \mathbb{R}^n$ es continuamente diferenciable, $g_k = \nabla f(X_k)$ y $\alpha_k > 0$ es el tamaño de paso sobre la dirección del gradiente. Queremos optimizar una función cuadrática de la forma:

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

donde $b \in \mathbb{R}^n$ y $A \in \mathbb{R}^{n \times n}$ utilizando el siguiente tamaño de paso:

$$\alpha_k(\Psi(A)) = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T \Psi(A) A g_{k-1}}$$

En donde $\Psi(\cdot)$ es una función real analítica en $[\lambda_1, \lambda_n]$ que se puede expresar como:

$$\Psi(z) = \sum_{k=-\infty}^{\infty} c_k z^k, \; c_k \in \mathbb{R}^n$$

tal que $0 < \sum_{k=-\infty}^{\infty} c_k z^k < +\infty$ para todos los $z \in [\lambda_1, \lambda_n]$, donde λ_1 y λ_n son los eigenvalores más pequeño y más grande respectivamente. Se puede observar que los dos tamaños de paso de Barzilai-Borwein α_k^{BB1} y α_k^{BB2} se pueden obtener de este tamaño de paso más general haciendo $\Psi(A) = I$ y $\Psi(A) = A$ en la expresión para α_k respectivamente.

Más adelante se procederá a describir el tamaño de paso $\tilde{\alpha}_k$ que se deriva y se utiliza en el trabajo para minimizar funciones cuadráticas bidimensionales convexas en un máximo de 5 iteraciones. También se hablará de modificaciones y consideraciones que se hacen sobre este tamaño de paso para proponer métodos adaptativos no monótonos de gradiente que utilizan los pasos largo y corto del método de BB junto con algunos pasos montonos usando $\tilde{\alpha}_k(A)$.

A su vez, se discutirán dos variaciones del método desarrollado para minimizar de forma eficiente funciones más genertales utilizando tamaños de paso con retardo.

Resultados

Conclusiones