

Aceleración del método de Barzilai-Borwein

Edgar Osvaldo López Zúñiga

Giovanni Gamaliel López Padilla

31 de mayo de 2022

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light beige shape occupies the bottom-left corner. The rest of the slide is white. The word 'Introducción' is centered in the white area.

Introducción

Introducción

Método Barzilai-Borwein

El estudio de la solución de la optimización de una función cuadrática de la forma

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

donde $b \in \mathbb{R}^n$ y A es una matriz definida positiva en $\mathbb{R}^{n \times n}$ resulta ser una base importante a la hora de resolver problemas más complejos. El resolver este problema equivalente a encontrar la solución del sistema de ecuaciones

$$Ax = b$$

La elección de los tamaños de paso propuestos están basados en métodos Cuasi-Newton. Donde α_k es remplazada por la matriz D_k tal que

$$D_k = \alpha_k I \quad (2)$$

Introducción

Método Barzilai-Borwein

El tamaño de paso es calculado a partir de la optimización de D_k^{-1} (BB1) y D_k (BB2) tal que satisfagan la ecuación de la secante desde un punto de vista de mínimos cuadrados (ecuación 3).

$$\min_{D=\alpha I} \|D^{-1}s_{k-1} - y_{k-1}\| \qquad \min_{D=\alpha I} \|s_{k-1} - Dy_{k-1}\| \qquad (3)$$

donde

$$s_{k-1} = x_k - x_{k-1} \qquad y_{k-1} = g_k - g_{k-1}$$

Introducción

Método Barzilai-Borwein

Las soluciones del problema son las descritas en la ecuación 4.

$$\alpha_k^{BB1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \quad \alpha_k^{BB2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \quad (4)$$

Al considerar la desigualdad de Cauchy-Schwarz observa que cuando $s_{k-1}^T y_{k-1}$ es mayor a cero, se cumple que $\alpha_k^{BB1} \geq \alpha_k^{BB2}$. Debido a esto se suele llamar a α_k^{BB1} tamaño de paso largo de Barzilai-Borwein y paso corto a α_k^{BB2} .



Desarrollo

Desarrollo

Método Barzilai-Borwein

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

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

para resolver el problema

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

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es continuamente diferenciable, $g_k = \nabla f(x_k)$ y $\alpha_k > 0$ es el tamaño de paso sobre la dirección del gradiente.

Desarrollo

Método Barzilai-Borwein

En particular se considerará una función cuadrática de la forma 1 utilizando el tamaño de paso mostrado en la ecuación 5.

$$\alpha_k(\Psi(A)) = \frac{g_{k-1}^T \Psi(A) g_{k-1}}{g_{k-1}^T \Psi(A) A g_{k-1}} \quad (5)$$

Donde $\Psi(\cdot)$ es una función real analítica en $[\lambda_1, \lambda_n]$ que se puede expresar como una serie de Laurent (ecuación 6).

$$\Psi(z) = \sum_{k=-\infty}^{\infty} c_k z^k, \quad c_k \in \mathbb{R}^n \quad \text{tal que} \quad 0 < \sum_{k=-\infty}^{\infty} c_k z^k < +\infty \quad \forall z \in [\lambda_1, \lambda_n] \quad (6)$$

Desarrollo

Tamaño de paso

Se explotaron algunas propiedades ortogonales en este subespacio bidimensional para acelerar los métodos.

Es por ello que si se quiere acelerar la convergencia de los métodos con tamaño de paso α_k (ecuación 5) en un subespacio de menor dimensionalidad se deben mantener propiedades ortogonales. Suponiendo que para un $k > 0$, existe un q_k que satisface

$$(I - \alpha_{k-1}A)q_k = g_{k-1}$$

Desarrollo

Tamaño de paso

Con esto, se tiene el lema 1 que presenta una propiedad para la derivación del tamaño de paso.

Lema

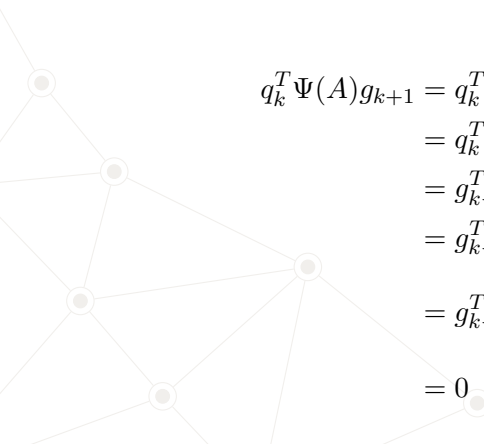
Suponiendo que la secuencia $\{g_k\}$ se obtiene aplicando un método de gradiente con pasos como el paso general (α_k) para minimizar una función cuadrática y q_k satisface que $(I - \alpha_{k-1}A)q_k = g_{k-1}$. Entonces

$$q_k^T \Psi(A) g_{k+1} = 0.$$

Desarrollo

Tamaño de paso

Para probar el lema 1 se tiene que:


$$\begin{aligned}q_k^T \Psi(A) g_{k+1} &= q_k^T \Psi(A) (I - \alpha_k A) (I - \alpha_{k-1} A) g_{k-1} \\&= q_k^T (I - \alpha_k A) \Psi(A) (I - \alpha_{k-1} A) g_{k-1} \\&= g_{k-1}^T \Psi(A) (I - \alpha_k A) g_{k-1} \\&= g_{k-1}^T \Psi(A) g_{k-1} - \alpha_k g_{k-1}^T \Psi(A) A g_{k-1} \\&= g_{k-1}^T \Psi(A) g_{k-1} \left[\frac{g_{k-1}^T \Psi(A) g_{k-1}}{g_{k-1}^T \Psi(A) A g_{k-1}} - \alpha_k \right] \\&= 0\end{aligned}$$

Desarrollo

Tamaño de paso

Considerando el problema de minimizar una función f en un subespacio bidimensional generado por

$$u = \frac{\Psi^r(A)q_{k-1}}{|\Psi^r(A)q_{k-1}|} \quad v = \frac{\Psi^r(A)g_k}{|\Psi^r(A)g_k|} \quad (7)$$

donde u y v forman una base ortogonal para \mathbb{R}^2 . Definiendo una función ϕ como

$$\phi(t, l) := f \left(x_k + t \frac{\Psi^r(A)q_{k-1}}{\|\Psi^r(A)q_{k-1}\|} + l \frac{\Psi^r(A)g_k}{\|\Psi^r(A)g_k\|} \right) \quad (8)$$

Desarrollo

Tamaño de paso

al expandir en una serie de Taylor obtenemos que

$$\phi(t, l) = f(x_k) + \nabla^T f(x_k) [tu + lv] + \frac{1}{2} [tu + lv]^T \nabla^2 f(x_k) [tu + lv] \quad (9)$$

Tomando a una matriz B_k como en la ecuación 10.

$$B_k = (u, v)^T \quad (10)$$

Con las ecuaciones 7 y 10 podemos escribir la ecuación 9 como en la ecuación 11.

$$\phi(t, l) = f(x_k) + g_k^T B_k^T \begin{pmatrix} t \\ l \end{pmatrix} + \frac{1}{2} \begin{pmatrix} t \\ l \end{pmatrix}^T B_k A B_k^T \begin{pmatrix} t \\ l \end{pmatrix} \quad (11)$$

Desarrollo

Tamaño de paso

Se define el vector ϑ como

$$\vartheta_k = B_k g_k = \begin{pmatrix} \frac{g_k^T \Psi^r(A) q_{k-1}}{|\Psi^r(A) q_{k-1}|} \\ \frac{g_k^T \Psi^{1-r}(A) g_k}{|\Psi^{1-r}(A) g_k|} \end{pmatrix} \quad (12)$$

y a la matriz H_k como

$$H_k = B_k A B_k^T = \begin{pmatrix} \frac{q_{k-1}^T \Psi^{2r}(A) A q_{k-1}}{\|\Psi^r(A) q_{k-1}\|^2} & \frac{q_{k-1}^T \Psi^r(A) A g_k}{\|\Psi^r(A) q_{k-1}\| \|\Psi^{1-r}(A) g_k\|} \\ \frac{q_{k-1}^T \Psi^r(A) A g_k}{\|\Psi^r(A) q_{k-1}\| \|\Psi^{1-r}(A) g_k\|} & \frac{g_k^T \Psi^{2(1-r)}(A) A g_k}{\|\Psi^r(A) g_k\|^2} \end{pmatrix} \quad (13)$$

Desarrollo

Tamaño de paso

Es posible expresar la función $\phi(t, l)$ como en la ecuación 14.

$$\phi(t, l) = f(x_k) + \vartheta_k^T \begin{pmatrix} t \\ l \end{pmatrix} + \frac{1}{2} \begin{pmatrix} t \\ l \end{pmatrix}^T H_k \begin{pmatrix} t \\ l \end{pmatrix} \quad (14)$$

Denotando las componentes de H_k por H_k^{ij} , $i, j = 1, 2$ y notando que $B_k B_k^T = I$ se tiene el teorema 1.

Desarrollo

Tamaño de paso

Teorema (Terminación finita)

Suponga que un método de gradiente se aplica para minimizar una función cuadrática bidimensional con α_k dado por el paso general mencionado anteriormente para todas las $k \neq k_0$ y usa el tamaño de paso

$$\tilde{\alpha}_{k_0} = \frac{2}{\left(H_{k_0}^{(11)} + H_{k_0}^{(22)}\right) + \sqrt{\left(H_{k_0}^{(11)} - H_{k_0}^{(22)}\right)^2 + 4\left(H_{k_0}^{(12)}\right)^2}}$$

en la iteración k_0 -ésima donde $k_0 \geq 2$. Entonces, el método encontrará el minimizador en máximo $k_0 + 3$ iteraciones.

Desarrollo

Tamaño de paso

Para demostrar el teorema 1 se supone que x_k no es un minimizador para toda $k = 1, \dots, k_0 + 2$. Para simplificar durante la prueba, se utilizará como notación k para referirse a k_0 .

Es necesario observar que $\tilde{\alpha}_k$ satisface la ecuación 15.

$$\tilde{\alpha}_k^2 \Delta - \tilde{\alpha}_k \left(H_k^{(11)} + H_k^{(22)} \right) + 1 = 0 \quad (15)$$

en donde $\Delta = \det(H_k) = \det(A) > 0$.

Desarrollo

Tamaño de paso

Sea

$$\Theta = \left(H_k^{(12)} \vartheta_k^{(1)} + H_k^{(22)} \vartheta_k^{(2)} \right) \vartheta_k^{(1)} - \left(H_k^{(11)} \vartheta_k^{(1)} + H_k^{(12)} \vartheta_k^{(2)} \right) \vartheta_k^{(2)}$$

en donde $\vartheta_k^{(i)}$ son las componentes de ϑ_k . Multiplicando Θ a la ecuación 15 se tiene lo siguiente

$$\tilde{\alpha}_k^2 \Delta \Theta - \tilde{\alpha}_k \left(H_k^{(11)} + H_k^{(12)} \right) \Theta + \Theta = 0,$$

que exactamente es

$$ab = cd$$

donde

$$a = H_k^{(22)} v_k^1 - H_k^{(12)} v_k^{(2)} - \tilde{\alpha}_k \Delta \vartheta_k^{(1)}$$

$$c = H_k^{(11)} v_k^2 - H_k^{(12)} v_k^{(1)} - \tilde{\alpha}_k \Delta \vartheta_k^{(1)}$$

$$b = \vartheta_k^{(2)} - \tilde{\alpha}_k (H_k^{(12)} \vartheta_k^{(1)} + H_k^{(22)} \vartheta_k^{(2)})$$

$$d = \vartheta_k^{(1)} - \tilde{\alpha}_k (H_k^{(11)} \vartheta_k^{(1)} + H_k^{(12)} \vartheta_k^{(2)}).$$

Desarrollo

Tamaño de paso

Consideremos dos vectores $\mathbf{A} = (a_1, a_2)^T$ y $\mathbf{B} = (b_1, b_2)^T$, tales que se cumple $a_1 b_2 = a_2 b_1$. Sustituyendo b_2 en \mathbf{B} , tenemos lo siguiente

$$\mathbf{B} = \begin{pmatrix} b_1 \\ \frac{b_1 a_2}{a_1} \end{pmatrix} = \frac{b_1}{a_1} \mathbf{A}$$

por lo que podemos decir que \mathbf{B} es paralelo a \mathbf{A} .

Desarrollo

Tamaño de paso

Haciendo uso del resultado anterior podemos decir que el vector

$$\begin{pmatrix} H_k^{(22)}\vartheta_k^{(1)} - H_k^{(12)}\vartheta_k^{(2)} - \tilde{\alpha}_k\Delta\vartheta_k^{(1)} \\ H_k^{(11)}\vartheta_k^{(2)} - H_k^{(12)}\vartheta_k^{(1)} - \tilde{\alpha}_k\Delta\vartheta_k^{(2)} \end{pmatrix}$$

es paralelo a

$$\begin{pmatrix} v_k^{(1)} - \tilde{\alpha}_k(H_k^{(11)}\vartheta_k^{(1)} + H_k^{(12)}\vartheta_k^{(2)}) \\ v_k^{(2)} - \tilde{\alpha}_k(H_k^{(12)}\vartheta_k^{(1)} + H_k^{(22)}\vartheta_k^{(2)}) \end{pmatrix}$$

Desarrollo

Tamaño de paso

Por ende el vector $\vartheta_k + H_k(-\tilde{\alpha}_k \vartheta_k)$ es paralelo a $H_k^{-1} \vartheta_j - \tilde{\alpha}_k \vartheta_k$. Es decir

$$\vartheta_k + H_k(-\tilde{\alpha}_k \vartheta_k) = \gamma(H_k^{-1} \vartheta_j - \tilde{\alpha}_k \vartheta_k) \quad (16)$$

donde $\gamma \neq 0 \in \mathbb{R}$. Si multiplicamos por la derecha a la ecuación 16 por B_k^T se tiene que

$$B_k^T[\vartheta_k + H_k(-\tilde{\alpha}_k \vartheta_k)] = \gamma B_k^T(H_k^{-1} \vartheta_j - \tilde{\alpha}_k \vartheta_k).$$

Desarrollo

Tamaño de paso

Se sabe que $B_k^T B_k = I$, $\vartheta_k = B_k g_k$ y que $H_k = B_k A B_k^T$. Además $g_{k+1} = g_k + \tilde{\alpha}_k A g_k$.
Haciendo uso de este conocimiento, se obtiene que

$$g_k + A(-\tilde{\alpha}_k B_k^T \vartheta_k) = \gamma(B_k^T H_k^{-1} \vartheta_k + \tilde{\alpha}_k g_k)$$

$$g_k + A(-\tilde{\alpha}_k g_k) = \gamma(B_k^T H_k^{-1} \vartheta_k + \tilde{\alpha}_k g_k)$$

$$g_{k+1} = \gamma(B_k^T H_k^{-1} \vartheta_k + \tilde{\alpha}_k g_k)$$

$$g_{k+1} = \gamma(B_k^T B_k A^{-1} B_k^T \vartheta_k + \tilde{\alpha}_k g_k)$$

$$g_{k+1} = \gamma(A^{-1} g_k + \tilde{\alpha}_k g_k)$$

Desarrollo

Tamaño de paso

Factorizando A^{-1} se tiene

$$g_{k+1} = \gamma A^{-1}(g_k + \tilde{\alpha}_k A g_k)$$

Considerando que $g_{k+1} = g_k + \tilde{\alpha}_k A g_k$ llegamos a

$$g_{k+1} = \gamma A^{-1} g_{k+1}$$

Es decir g_{k+1} es un eigenvector de la matriz A .

Desarrollo

Tamaño de paso

Por hipótesis, sabemos que x_{k+2} no es un minimizador, así que $g_{k+2} \neq 0$ y el algoritmo no se detendrá en la $k + 2$ -ésima iteración. Entonces, calculando α_{k+2} como

$$\alpha_{k+2} = \frac{g_{k+1}^T \Psi(A) g_{k+1}}{g_{k+1}^T \Psi(A) A g_{k+1}} = 1/\lambda$$

tenemos que

$$\begin{aligned} g_{k+3} &= (I - \alpha_{k+2} A) g_{k+2} \\ &= (1 - \alpha_{k+2} \lambda) g_{k+2} = 0 \end{aligned}$$

lo que implica que x_{k+3} debe ser el minimizador. Si tomamos $k_0 = 2$ en el teorema, el tamaño de paso encontrará al minimizador exacto en máximo 5 iteraciones cuando se tiene una función cuadrática bidimensional fuertemente convexa.

Desarrollo

Tamaño de paso

Corollary

Suponga que un método de gradiente se aplica a una función cuadrática bidimensional con α_{k_0+m} para $k_0 \geq 2$, algún entero positivo m y con $\alpha_{k_0+m} = \tilde{\alpha}_{k_0}$ para toda $k \neq k_0 + m$. Entonces, el método se detendrá en máximo $k + m + 3$ iteraciones.

Desarrollo

Tamaño de paso

Si hacemos $\Psi(A) = I$, $\Psi(A) = A$ y $r = 1/2$ en la ecuación 5 y tomando $k_0 = k$ se puede derivar los siguientes tamaños de paso:

$$\tilde{\alpha}_k^{BB1} = \frac{2}{\frac{q_{k-1}^T A q_{k-1}}{\|q_{k-1}\|^2} + \frac{1}{\alpha_k^{SD}} + \sqrt{\left(\frac{q_{k-1}^T A q_{k-1}}{\|q_{k-1}\|^2} - \frac{1}{\alpha_k^{SD}}\right)^2 + \frac{4(q_k^T A g_k)^2}{\|q_{k-1}\|^2 \|g_k\|^2}}} \quad (17)$$

$$\tilde{\alpha}_k^{BB2} = \frac{2}{\frac{1}{\hat{\alpha}_{k-1}} + \frac{1}{\alpha_k^{MG}} + \sqrt{\left(\frac{1}{\hat{\alpha}_{k-1}} + -\frac{1}{\alpha_k^{MG}}\right)^2 + \Gamma_k}} \quad (18)$$

Desarrollo

Tamaño de paso

en donde

$$\hat{\alpha}_k = \frac{q_k^T A q_k}{q_k^T A^2 q_k} \text{ y } \Gamma_k = \frac{4 (q_{k-1}^T A^2 g_k)^2}{q_{k-1}^T A q_{k-1} g_k^T A g_k} \quad (19)$$

Ahora, de 17 y 18 se tiene

$$\tilde{\alpha}_k^{BB1} \leq \min \left\{ \alpha_k^{SD}, \frac{\|q_{k-1}\|^2}{q_{k-1}^T A q_{k-1}} \right\} \text{ y } \tilde{\alpha}_k^{BB2} \leq \min \{ \alpha_k^{MG}, \hat{\alpha}_{k-1} \} \quad (20)$$

Desarrollo

Tamaño de paso

El método de gradiente no-monótono adaptativo (ANGM) toma el paso largo de BB α_k^{BB1} cuando $\alpha_k^{BB2} / \alpha_k^{BB1} \geq \tau_1$ para algún $\tau_1 \in (0, 1)$. De otra manera, toma un tamaño de paso α_k^{BB2} o $\tilde{\alpha}_k^{BB2}$ dependiendo del cociente $\|g_{k-1}\| / \|g_k\|$. El paso α_k^{BB2} minimiza al gradiente tal que

$$\alpha_k^{BB2} = \alpha_{k-1}^{MG} = \arg \min_{\alpha \in \mathbb{R}} \|g_{k-1} - \alpha A g_{k-1}\|.$$

Desarrollo

ANGM

El ANGM utiliza el tamaño de paso monótono $\tilde{\alpha}_k^{BB2}$ cuando $\|g_{k-1}\| > \tau_2 \|g_k\|$, de otra manera se deberían tomar ciertos pasos BB2. El ANGM aplica las siguientes estrategias adaptativas para elegir el tamaño de paso:

$$\alpha_k = \begin{cases} \min\{\alpha_k^{BB2}, \alpha_{k-1}^{BB2}\}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| < \tau_2 \|g_k\| \\ \tilde{\alpha}_k^{BB2}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| \geq \|g_k\| \\ \alpha_k^{BB1}, & \text{de otro modo} \end{cases} \quad (21)$$

Desarrollo

ANGR1

Se ha encontrado que los métodos de gradiente que utilizan pasos retardados pueden llevar a mejores desempeños, así reemplazando $\tilde{\alpha}_k^{BB2}$ en 21 por $\tilde{\alpha}_{k-1}^{BB2}$ se obtiene una variante llamada ANGR1:

$$\alpha_k = \begin{cases} \min\{\alpha_k^{BB2}, \alpha_{k-1}^{BB2}\}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| < \tau_2 \|g_k\| \\ \tilde{\alpha}_{k-1}^{BB2}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| \geq \|g_k\| \\ \alpha_k^{BB1}, & \text{de otro modo} \end{cases} \quad (22)$$

Desarrollo

ANGR2

Ahora, para simplificar el ANGR1 se puede remplazar $\tilde{\alpha}_{k-1}^{BB2}$ por su cota superior, es decir, se puede hacer uso de:

$$\alpha_{k-1}^{BB2} \leq \min \{ \alpha_k^{BB2}, \hat{\alpha}_{k-2} \}$$

Como resultado de esto, se obtiene otra variante de ANGM a la que se llama *ANGR2*:

$$\alpha_k \begin{cases} \min \{ \alpha_k^{BB2}, \alpha_{k-1}^{BB2} \}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| < \tau_2 \|g_k\| \\ \min \{ \alpha_k^{BB2}, \hat{\alpha}_{k-2} \}, & \text{si } \alpha_k^{BB2} < \tau_1 \alpha_k^{BB1} \text{ y } \|g_{k-1}\| \geq \|g_k\| \\ \alpha_k^{BB1}, & \text{de otro modo} \end{cases} \quad (23)$$

Desarrollo

ANGM, ANGR1 y ANGR2

Es necesario notar que para la implementación de los métodos ANGM, ANGR1 y ANGR2 no es necesario calcular productos de matriz-vector.

$$\hat{\alpha}_k = \frac{q_k^T A q_k}{q_k^T A^2 q_k} = \frac{\alpha_{k-1} q_k^T (q_k - g_{k-1})}{(q_k - g_{k-1})^T (q_k - g_{k-1})} \quad (24)$$

Desarrollo

ANGM, ANGR1 y ANGR2

Debido a que el cálculo de Ag_k se necesita para el término g_{k+1} y Γ_k en $\tilde{\alpha}_k^{BB2}$. Para $\tilde{\alpha}_{k-1}^{BB2}$, se necesita el calculo del término

$$g_{k-1}^T A^2 q_{k-2} = \frac{1}{\alpha_{k-3}} (q_{k-2} - g_{k-3})^T Ag_{k-1} \quad Ag_{k-1} = \frac{1}{\alpha_{k-1}} (g_{k-1} - g_k)$$

por ende

$$\Gamma_{k-1} = \frac{4((q_{k-2} - g_{k-3})^T (g_{k-1} - g_k))^2}{\alpha_{k-3} \alpha_{k-1} ((q_{k-2} - g_{k-3})^T q_{k-2}) g_{k-1}^T (g_{k-1} - g_k)}$$

por lo que no se requiere calcular productir de matriz-vector adicionales para Γ_{k-1} en $\tilde{\alpha}_{k1}^{BB2}$.

Desarrollo

ANGM, ANGR1 y ANGR2

Un problema a tener en cuenta es que el cálculo de q_k de forma exacta como se define en 6 puede llegar a ser tan difícil como minimizar la función cuadrática. Es por esto que a la hora de implementar, se puede utilizar una aproximación de q_k notando que satisface la ecuación

$$q_k^T g_k = \|g_{k-1}\|^2 \quad (25)$$

Desarrollo

ANGM, ANGR1 y ANGR2

Una forma eficiente de encontrar la q_k que satisface la ecuación de secante 25 es tratando a la matriz Hessiana A como una matriz diagonal, y así derivar q_k de

$$g_{k+1}^{(j)} = (1 - \alpha_k \lambda_j) g_k^{(j)} \text{ cuando } g_k^{(j)} \neq 0,$$

$$q_k^{(i)} = \frac{g_{k-1}^{(i)}}{1 - \alpha_{k-1} \lambda_i} = \frac{(g_{k-1}^{(i)})^2}{g_k^{(i)}}, \quad i = 1, \dots, n \quad (26)$$

y en el caso de que $g_k^{(i)} = 0$ se hace $q_k^{(i)} = 0$.

The background of the slide is composed of two large, overlapping geometric shapes. On the left, there is a teal-colored shape that tapers towards the top right. On the right, there is a light beige shape that tapers towards the bottom left. These two shapes meet at a diagonal line that runs from the top left towards the bottom right, creating a clean, modern aesthetic.

Resultados

Resultados

Función de Rosembrock

La función de Rosembrock se define en la ecuación 27.

$$f(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i) \quad (27)$$

Resultados

Función Wood

La función de Wood se define en la ecuación 28.

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3 - x_4)^2 + 10,1((x_2 - 1)^2 + (x_4 - 1)^2) + 19,8(x_2 - 1)(x_4 - 1) \quad (28)$$

Resultados

Función Lambda

En el artículo de Yakui Huang se utiliza una función definida de la siguiente manera:

$$f(x) = \frac{1}{2}x^T Ax \quad A = \begin{cases} 0 & \text{para } i \neq j \\ 10^{\frac{ncond}{n-1}(n-j)} & \text{para } i = j \end{cases}$$

donde $ncond = \log \kappa$ con $\kappa = 10^3$ y $n = 10$.

Resultados

Función cuadrática

En el mismo artículo⁴ se propone una función cuadrática donde la matriz esta definida de la siguiente manera:

$$A = \text{diag}\{1, \lambda\}$$

En nuestro caso tomamos $\lambda = 10$.

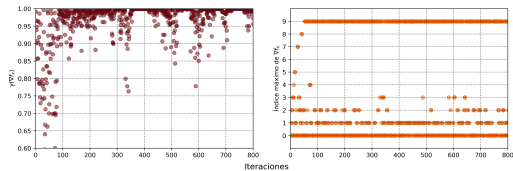
Resultados

Porcentaje de contribución

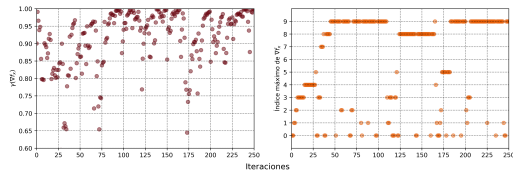
Se definió una función γ para medir el porcentaje de las dos componentes más grandes del gradiente en cada iteración de la optimización. La función γ tiene la siguiente forma

$$\gamma(g_k) = \frac{|g_k^{(1)}| + |g_k^{(2)}|}{\sum_i |g_k^{(i)}|}$$

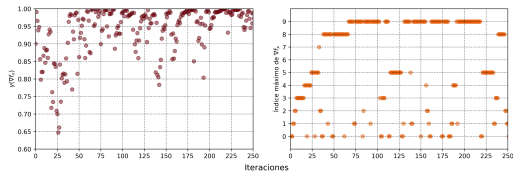
Realizando el calculo de la función γ con la función lambda en el punto inicial $x = (10, 10, \dots, 10)^T$ se obtuvieron los resultados mostrados en la figura 1 para los diferentes métodos.



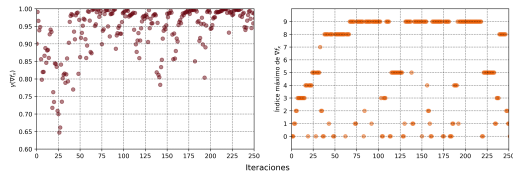
(a) Método BB.



(b) Método ANGM.



(c) Método ANGR1



(d) Método ANGR2.

Figura 1: Función γ para diferentes métodos (izquierda) y la posición en el vector de la componente más grande (derecha).

Resultados

Porcentaje de contribución

Método	$\gamma(g_k) > 0,8$	Total
SD	8057	8104
BB	815	851
ANGM	293	316
ANGR1	240	253
ANGR2	200	245

Tabla 1: Número de iteraciones donde el valor de la función γ tuvo un valor mayor a 0.8 para cada método de optimización.

Resultados

Estabilidad de resultados

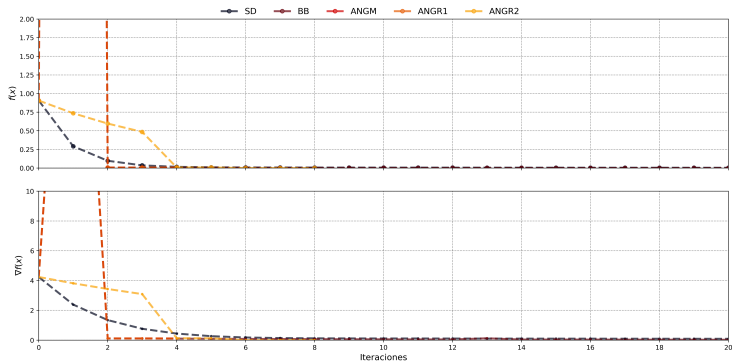


Figura 2: Valor de la función y norma del gradiente en cada iteración de la mejor ejecución de cada método de optimización de la función cuadrática con matriz diagonal.

Resultados

Estabilidad de resultados



Figura 3: Valor de la función y norma del gradiente en cada iteración de la mejor ejecución de cada método de optimización de la función de Rosembrock con matriz diagonal.

Función	Valor	SD	BB	ANGM	ANGR1	ANGR2
Lambda	Función	0.00	0.00	0.00	0.00	0.00
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	7764.81	2770.34	252.43	289.49	255.83
Cuadrática	Función	0.00	0.00	0.00	0.00	0.00
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	713.00	1531.23	5.24	7.36	9.70
Rosembrock	Función	4.16	42.25	0.00	0.00	0.00
	Gradiente	0.57	45.45	0.00	0.00	0.00
	Iteraciones	10000.00	1603.21	68.84	70.17	67.15
Wood	Función	0.00	3.16	5.07	4.56	8.43
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	9428.69	1806.43	360.36	292.43	199.21

Tabla 2: Media de las 100 ejecucciones partiendo de puntos aleatorios.

Función	Valor	SD	BB	ANGM	ANGR1	ANGR2
Lambda	Función	0.00	0.00	0.00	0.00	0.00
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	612.38	209.34	32.55	28.34	25.04
Cuadrática	Función	0.00	0.00	0.00	0.00	0.00
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	43.81	96.11	1.19	1.85	2.88
Rosembrock	Función	14.24	393.58	0.00	0.00	0.00
	Gradiente	1.36	396.05	0.00	0.00	0.00
	Iteraciones	0.00	1867.61	17.32	20.59	14.92
Wood	Función	0.00	10.09	12.22	11.86	14.40
	Gradiente	0.00	0.00	0.00	0.00	0.00
	Iteraciones	465.60	633.53	190.56	204.76	119.76

Tabla 3: Desviaciones estandar de las 100 ejecucciones partiendo de puntos aleatorios.

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light gray shape occupies the bottom-left corner. The rest of the slide is white. The word "Conclusiones" is centered in the white area.

Conclusiones

Code



[https://github.com/Edgar-LZ/
Barzilai-Borwein-monotone](https://github.com/Edgar-LZ/Barzilai-Borwein-monotone)

References I

J. Barzilai and J. M. Borwein.

Two-Point Step Size Gradient Methods.

IMA Journal of Numerical Analysis, 8(1):141–148, 01 1988.

Y.-H. Dai and R. Fletcher.

Projected barzilai-borwein methods for large-scale box-constrained quadratic programming.

Numerische Mathematik, 100(1):21–47, 2005.

R. Fletcher.

On the barzilai-borwein method.

In *Optimization and control with applications*, pages 235–256. Springer, 2005.

References II

Y. Huang, Y.-H. Dai, X.-W. Liu, and H. Zhang.

On the acceleration of the barzilai–borwein method.

Computational Optimization and Applications, pages 1–24, 2022.

Y. Huang, Y.-H. Dai, X.-W. Liu, and H. Zhang.

On the asymptotic convergence and acceleration of gradient methods.

Journal of Scientific Computing, 90(1):1–29, 2022.

B. Oleg.

Stabilized barzilai–borwein method.

JCM, 37(6):916–936, jun 2019.

References III

M. Raydan.

On the barzilai and borwein choice of steplength for the gradient method.

IMA Journal of Numerical Analysis, 13(3):321–326, 1993.

M. Raydan.

The barzilai and borwein gradient method for the large scale unconstrained minimization problem.

SIAM Journal on Optimization, 7(1):26–33, 1997.