Lecturas de Optimización Numérica:

Optimización sin Restricciones

Prof. Dr. Fredy Vides
Scientific Computing Innovation Center, UNAH &
Centre for Analysis of Data-Driven Systems
E-mail: fredy.vides@unah.edu.hn

Índice

Objetivos

1. Estudiar los prinicipios fundamentales de optimización sin restricciones.
2. Aplicar principios de optimización al estudio de métodos de gradiente conjugado para la solución iterativa de sistemas de ecuaciones.

1. Optimización sin Restricciones

Definición 1.1. Dado $y \in \mathbb{R}^n$ y dado (un número real) $r > 0$, se denota por $B_r(y)$ la bola (abierta) de radio $r > 0$ centrada en y definida por la expresión:

$$B_r(y) = \{x \in \mathbb{R}^n : \|x - y\|_2 < r\}$$

Cuando se busca minimizar una función objetivo, es posible estar interesado en encontrar un mínimo o bien local, o bien global.

Definición 1.2. Un punto $x^* \in \mathbb{R}^n$ se denomina un **minimizador global** para $f \in \mathbb{C}^2(\mathbb{R}^n, \mathbb{R})$ si:

$$f(x^*) \leq f(x) \quad \forall x \in \mathbb{R}^n,$$

mientras que x^* es un **minimizador local** para f si existe un radio $0 < r < \infty$ tal que:

$$f(x^*) \leq f(x) \quad \forall x \in B_r(x^*) \subset \mathbb{R}^n.$$

Dada $f \in C^2(\mathbb{R}^n, \mathbb{R})$, denotamos por $\nabla f(x)$ el gradiente de f definido por la expresión:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

Dada una función $F \in C^2(\mathbb{R}^n, \mathbb{R}^n)$ escribimos $\mathbb{J}F(x)$ para denotar el Jacobiano de F definido por la expresión:

$$\mathbb{J}F(x) = \begin{bmatrix} (\nabla F_1(x))^\top \\ \vdots \\ (\nabla F_n(x))^\top \end{bmatrix} \in C(\mathbb{R}^n, \mathbb{R}^{n\times n})$$

La matriz Hessiana $\mathbb{H}f(x)$ puede definirse en términos de los operadores anteriores, utilizando la siguiente secuencia de operaciones:

$$\mathbb{H}f(x) = \mathbb{J}\nabla f(x)$$

Observación 1.3. Cuando $f \in C^2(\mathbb{R}^n, \mathbb{R})$, $\mathbb{H}f(x)$ es una matriz simétrica para cada $x \in \mathbb{R}^n$.

Definición 1.4. Un punto $x^* \in \mathbb{R}^n$ se denomina **estacionario** o **crítico** para $f \in \mathbb{C}^2(\mathbb{R}^n, \mathbb{R})$ si $\nabla f(x^*) = 0$, o un punto **regular** si $\nabla f(x^*) \neq 0$.

Definición 1.5. La función $f : \mathbb{R}^n \to \mathbb{R}$ se denomina convexa si $\forall x, y \in \mathbb{R}^n$ y $\forall \alpha \in [0, 1]$,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y);$$

y se denomina Lipschitz continua si existe una constante $L > 0$ tal que

$$|f(x) - f(y)| \leq L\|x - y\|_2, \ \ \forall x, y \in \mathbb{R}^n.$$

Propiedad 1.6. *(Condiciones de optimalidad.) Sea $x^* \in \mathbb{R}^n$. Si x^* es un minimizador para una función $f : \mathbb{R}^n \to \mathbb{R})$ (bien local o global) y si existe $r > 0$ tal que $f \in C^1(B_r(x^*))$, entonces $\nabla f(x) = 0$. Además, si $f \in C^2(B_r(x^*))$, $\mathbb{H}f(x^*)$ es SPSD. Viceversa, sea $r > 0$ tal que $f \in C^2(B_r(x^*))$. Si $\nabla f(x^*) = 0$ y $\mathbb{H}f(x)$ es SPSD para cada $x \in B_r(x^*)$, entonces x^* es un minimizador local para f. Finalemente, si $f \in C^1(\mathbb{R}^n, \mathbb{R})$, convexa en $\mathbb{R}^n$ y $\nabla f(x^*) = 0$, entonces x^* es un minimizador global para f.*

2. Aplicación: Prinicipios de Métodos de Gradiente Conjugado

Consideremos una matriz $A \in \mathbb{R}^{n\times n}$ arbitraria y un vector arbitrario $b \in \mathbb{R}^n$, y consideremos el funcional cuadrático:

$$\phi_{A,b}(x) = \frac{1}{2}x^\top Ax - x^\top b$$

Propiedad 2.1. *Si $A \in \mathbb{R}^{n\times n}$ es SPD, resolver el sistema $Ax = b$ equivale a resolver el siguiente problema de programación cuadrática.*

$$\min_{x\in\mathbb{R}^n} \phi_{A,b}(x) \tag{2.1}$$

La solución de los problemas cuadráticos de la forma (2.1) correspondientes a un sistema $Ax = b$ con matriz de coeficientes A SPD, pueden resolverse implementando métodos de gradiente conjugado que producen una secuencia $\{x_k\}_{k\geq 0} \subset \mathbb{R}^n$ tal que $x_k \to x = A^{-1}b$ y donde cada elemento x_k de la sucesión está dado por la expresión:

$$x_k = x_{k-1} + \alpha_{k-1}p_{k-1},\ k \geq 1$$

donde $\alpha_k \in \mathbb{R}$ y $p_k \in \mathbb{R}^n$ se calculan de acuerdo a criterios de optimzación que serán estudiados en detalle más adelante, y $x_0 \in \mathbb{R}^n$ es un elemento que se utiliza para inicializar el esquema iterativo.

Ejercicio para el lector 1. Demostrar que:

$$-\nabla\phi_{A,b}(x) = b - Ax$$

Ejercicio para el lector 2. Demostrar en detalle la Propiedad 2.1.

Observación 2.2. En un punto específico x_c, $\phi_{A,b}(x_c)$ decrece más rápidamente en la dirección del gradiente:

$$-\nabla\phi_{A,b}(x_c) = b - Ax_c$$

Sea r_c el resíduo $r_c = b - Ax_c$ de x_c. Si $r_c \neq 0$, entonces existe $\alpha > 0$ tal que $\phi_{A,b}(x_c + \alpha r_c) < \phi_{A,b}(x_c)$. En el método del descenso más empinado (con línea de búsqueda exacta) definimos:

$$\alpha = \frac{r_c^\top r_c}{r_c^\top A r_c}$$

con el fin de minimizar la expresión:

$$\phi_{A,b}(x_c + \alpha r_c) = \phi_{A,b}(x_c) - \alpha r_c^\top r_c + \frac{1}{2}\alpha^2 r_c^\top A r_c$$

Ejercicio para el lector 3. Verificar que $\alpha = \frac{r_c^\top r_c}{r_c^\top A r_c}$ minimiza la fórmula $\phi_{A,b}(x_c + \alpha r_c)$ definida por la expresión:

$$\phi_{A,b}(x_c + \alpha r_c) = \phi_{A,b}(x_c) - \alpha r_c^\top r_c + \frac{1}{2}\alpha^2 r_c^\top A r_c$$

Las considereciones anteriores permiten derivar el siguiente algoritmo prototípico para el método del descenso más empinado:

x_0: vector de inicialización
$r_0 \leftarrow b - Ax_0$
$k \leftarrow 0$
mientras $r_k \neq 0$
  $k \leftarrow k + 1$
  $\alpha_k \leftarrow (r_k^\top r_k)/(r_k^\top A r_k)$
  $x_k \leftarrow x_{k-1} + \alpha_k r_{k-1}$
  $r_k \leftarrow b - Ax_k$
fin

Ejercicio para el lector 4. Investigar sobre métodos computacionales de implementación de gradientes conjugados que permitan mejorar las características de convergencia de algoritmos de descenso más empinado como el anterior.

Ejercicio para el lector 5. Escribir un programa Octave que implemente tanto el algoritmo de descenso más empinado.

Ejercicio para el lector 6. Escribir un programa Octave que implemente tanto el algoritmo de gradiente conjugado resultante de su investigación.

REFERENCIAS

[1] R. L. Burden, D. J. Faires, A. M. Burden. (2017). Análisis Numérico. 10a Ed. Cengage Learning Editores.
[2] Golub, G. H., Van Loan C. F (1996). Matrix Computations (3aEd.). The Johns Hopkins University Press.
[3] Quarteroni A., Saleri F., Gervasio P. (2014). Scientific computing with MATLAB and Octave (Textbook).
[4] D. G. Luenberger, Y. Ye. (2016). Linear and Nonlinear Programming. 4a Ed. Springer International Publishing Switzerland.