Leonard David Vivas Dallos

TAREA # 1 PROGRAMACIÓN NO LINEAL

① Realizar los ejercicios 2.7 y 2.8 del texto guía.

2.7) Suppose that $f(x) = x^T Q x$, where Q is an $n \times n$ symmetric positive semidefinite matrix. Show using the definition (1.4) that $f(x)$ is convex on the domain $\mathbb{R}^n$

HINT: It may be convenient to prove the following equivalent inequality:

$$f(y + \alpha(x-y)) - \alpha f(x) - (1-\alpha) f(y) \leq 0$$

for all $\alpha \in [0,1]$ and all $x, y \in \mathbb{R}^n$

Dem: Supongamos que $f(x) = x^T Q x$, donde Q es una matriz semidefinida positiva simétrica y $n \times n$.

Veamos que, $f(x)$ es convexa en el dominio $\mathbb{R}^n$

De la hipótesis de que Q es semidefinida positiva, tenemos que

$$p^T Q p \geq 0 \quad \forall p \in \mathbb{R}^n$$

De donde particularmente,

$$f(x) = x^T Q x \geq 0$$
$$\therefore f(x) \geq 0$$

Ahora, por definición de función convexa, deseamos ver que $\forall x, y \in \mathbb{R}^n$, $\alpha \in [0,1]$

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

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

$$\iff f(y + \alpha(x-y)) - \alpha f(x) - (1-\alpha) f(y) \leq 0 \qquad (1)$$

Que es justamente la pista de demostración brindada.

Apliquemos la definición de $f(x)$ en (1), tenemos:

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

$$= [y+\alpha(x-y)]^T Q [y+\alpha(x-y)] - \alpha x^T Q x - (1-\alpha) y^T Q y$$

$$= (y^T + \alpha(x-y)^T) Q (y+\alpha(x-y)) - \alpha x^T Q x - (1-\alpha) y^T Q y$$

$$= (y^T + \alpha x^T - \alpha y^T) Q (y + \alpha x - \alpha y) - \alpha x^T Q x - (1-\alpha) y^T Q y$$

$$= (\alpha x^T + (1-\alpha) y^T) Q (\alpha x + (1-\alpha) y) - \alpha x^T Q x - (1-\alpha) y^T Q y$$

$$= \alpha^2 x^T Q x + \alpha(1-\alpha) x^T Q y + \alpha(1-\alpha) y^T Q x + (1-\alpha)^2 y^T Q y$$
$$- \alpha x^T Q x - (1-\alpha) y^T Q y$$

$$= (\alpha^2 - \alpha) x^T Q x + [(1-\alpha)^2 - (1-\alpha)] y^T Q y + \alpha(1-\alpha) x^T Q y$$
$$+ \alpha(1-\alpha) y^T Q x$$

$$= \alpha(\alpha-1) x^T Q x + [1 - 2\alpha + \alpha^2 - 1 + \alpha] y^T Q y$$
$$+ \alpha(1-\alpha) x^T Q y + \alpha(1-\alpha) y^T Q x$$

$$= \alpha(\alpha-1) x^T Q x + \alpha(\alpha-1) y^T Q y + \alpha(1-\alpha) x^T Q y$$
$$+ \alpha(1-\alpha) y^T Q x$$

$$= \alpha(\alpha-1)[x^T Q x + y^T Q y - x^T Q y - y^T Q x]$$

$$= \alpha(\alpha-1)(x-y)^T Q (x-y)$$

Ahora, como $\alpha \in [0,1]$, $(\alpha-1) \leq 0$, y como Q es semidefinida positiva, $(x-y)^T Q (x-y) \geq 0$. Así, tenemos

$$\alpha(\alpha-1)(x-y)^T Q (x-y) \leq 0$$

Así,

$$f(y+\alpha(x-y)) - \alpha f(x) - (1-\alpha) f(y) \leq 0$$

y por tanto, $f(x)$ es convexa en el dominio $\mathbb{R}^n$ //

2.8) Suppose that f is a convex function. Show that the set of global minimizers of f is a convex set.

Dem: Sea f una función convexa. Veamos que el conjunto de minimizadores globales de f es un conjunto convexo.

Sean x_1, x_2 minimizadores globales de f, queremos ver que para x_1, x_2 en el conjunto de minimizadores globales y $\alpha \in [0,1]$, tenemos

$$\alpha x_1 + (1-\alpha) x_2 \text{ es minimizador global}$$

Ahora, por hipótesis tenemos dos cosas;

(1) $f(x_1) = f(x_2) \leq f(x) \quad ; \quad \forall x \in \Omega \quad \Rightarrow x_1, x_2$ min glob

(2) $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \quad \forall x_1, x_2 \in \Omega, \ \alpha \in [0,1]$

(Pues f es convexa)

De (2) tenemos que,

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$$[\text{De } (1)] \Leftarrow \leq \alpha f(x) + (1-\alpha) f(x)$$

$$= \alpha f(x) + f(x) - \alpha f(x) = f(x)$$

Así, $f(\alpha x_1 + (1-\alpha)x_2) \leq f(x)$, $\alpha \in [0,1]$

De donde $\alpha x_1 + (1-\alpha) x_2$ es un minimizador global y por tanto, tenemos que el conjunto de minimizadores globales es convexo //

② Análisis de Ejemplos

- $f(x) = \cos(x + \pi/4)$ => Ej 1 - 4

Logramos ver con gran claridad la importancia de la elección de valores como ρ y α, valores importantes para la elección del paso

Primero, con $\rho = 0,75$ y $\alpha = 5,0$ (Ej 1), vemos un avance más discreto, con avance horizontal más grande pero menor avance al mínimo buscado

Con $\rho = 0,5$ y $\alpha = 5,0$ (Ej 2), vemos un avance mucho más precipitado, pues en una iteración se recorta mucho mas el valor de α en la multiplicación $\rho \cdot \alpha$, lo cual nos acerca mucho más rápido al mínimo

Para $\rho = 0,5$, $\alpha = 2,0$ (Ej 3) y $\rho = 0,75$, $\alpha = 2,0$ (Ej 4) obtenemos un resultado muy parecido, en algo que parece ser un "equilibrio" o ser valores que caen en la misma zona de condición cumplida, por lo que no hay muchas iteraciones, nuevamente se acerca bastante al punto mínimo.

- $f(x) = 100(x_1 - x_0^2)^2 + (1 - x_0^2)$ => Ej 5-8

Tal como en el anterior, vemos como la elección de las dos constantes ρ, α, juegan un papel fundamental en la decisión del tamaño de paso, siendo $\rho = 0.75$, $\alpha = 3,0$ (Ej 7) la elección que nos acerca con un paso mas grande en comparación a los valores testeados en los otros ejemplos

Es claro como con las elecciones de ciertas constantes restringimos los resultados y el # de iteraciones de un algoritmo que usa la evaluación y aprobación de dos condiciones (Wolfe y Armijo) con el fin de encontrar un tamaño de paso conveniente según la dirección y los valores de posición brindados.

③ IMPORTANCIA DE LA CONDICIÓN 3.14

$$\sum_{k\geq 0} \cos^2 \theta_k \|\nabla f_k\|^2 < \infty$$

Sabiendo que $\cos \theta_k$ relaciona el ángulo entre la dirección tomada y la dirección más decreciente, la condición 3.14 sin duda alguna nos da indicios de una convergencia del método, pues implica que al infinito esta relación se vuelve cero, es decir

$$\cos^2 \theta_k \|\nabla f_k\|^2 \to 0$$

Como es fácil notar, la condición 3.14 relaciona dos cosas, el ángulo entre la dirección hacia donde nos dirigimos y la dirección mas decreciente y el tamaño del vector gradiente, es decir, el tamaño del vector de la dirección más decreciente, y el hecho de que tienda a cero nos dice que nuestras elecciones de direcciones se van volviendo más y más precisas ($\cos \to 0$) o que llegará el punto en que independiente de la dirección, estamos muy cerca o llegamos a un minimizador local, por lo que no sería necesario más movimiento ($\nabla f_k \to 0$)

En conclusión, la condición 3.14 juega un papel importante, pues nos ayuda a derivar y/o deducir una convergencia global de los algoritmos de búsqueda lineal, haciendo comparaciones con el decrecimiento ideal.