MA3701 Optimización: Programación No Lineal Irrestricta.

Héctor Ramírez C.

1. Principales nociones sobre las funciones convexas

Recordemos que una función $f : \mathbb{R}^n \to \mathbb{R}$ es **convexa** en un conjunto S si

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2), \quad \text{para todo } x_1,\, x_2 \in S,\ \lambda \in [0,1].$$

Si la desigualdad anterior es estricta para todo par de puntos distintos entonces la función f se dirá **estrictamente convexa**.[1]

Además, un conjunto S se dirá convexo si para todo $x_1, x_2 \in S$ y para todo escalar $\lambda \in (0,1)$ se tiene que $\lambda x_1 + (1-\lambda)x_2 \in S_\alpha(f)$.

Algunas propiedades interesante se establecen a continuación.

Teorema. *Sean $S \subset \mathbb{R}^n$ convexo no vacío y $f : \mathbb{R}^n \to \mathbb{R}$ una función convexa en S. Entonces todo mínimo local f en S es un de mínimo global de f en S.*

Demostración. *Ya demostrado. Directo de la definición de convexidad.*

Teorema. *Sean $S \subset \mathbb{R}^n$ convexo no vacío y $f : \mathbb{R}^n \to \mathbb{R}$ estrictamente convexa en S. Entonces f tiene a lo más un mínimo (global) en S.*

Demostración. *Sabemos que al ser f convexa todo mínimo local es mínimo global, por lo que ambos son equivalentes. Supongamos que existen entonces dos mínimos globales en S, denotados por x_1 y x_2, cuyo valor se denota (como es usual) por* val(f). *De la convexidad estricta de f se deduce que:*

$$f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) < \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = \mathrm{val(f)},$$

lo que contradice la optimalidad de x_1 o x_2.$\square$

Lema. *Sea $S \subset \mathbb{R}^n$ convexo no vacío y $f : \mathbb{R}^n \to \mathbb{R}$ convexa en S. Pruebe que el conjunto $S_\alpha(f) = \{x \in S : f(x) \leq \alpha\}$ es convexo.*

Demostración. *Sean $x_1, x_2 \in S_\alpha(f)$. Probemos que $\forall \lambda \in (0,1), \lambda x_1 + (1-\lambda)x_2 \in S_\alpha(f)$. Como $S_\alpha(f) \subseteq S$, entonces $\lambda x_1 + (1-\lambda)x_2 \in S$, pues S es convexo. Luego, de la convexidad de f se obtiene:*

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

Por lo tanto, $f(\lambda x_1 + (1-\lambda)x_2) \leq \alpha$ y se concluye el resultado.$\square$

Lema. *Sea $S \subset \mathbb{R}^n$ convexo no vacío y $f : \mathbb{R}^n \to \mathbb{R}$ (estrictamente) convexa en S. Sean $a \in S, d \in \mathbb{R}^n \setminus \{0\}$ para los cuales existe $\delta > 0$ tal que para $\lambda \in (0,\delta)$, $a + \lambda d \in S$. Pruebe que $\frac{f(a+\lambda d)-f(a)}{\lambda}$ es monotono creciente (estricta) en λ. En consecuencia, se tiene que:*

$$\lim_{\lambda \to 0^+} \frac{f(a+\lambda d) - f(a)}{\lambda} = \inf_{\lambda \in \mathbb{R}_+} \frac{f(a+\lambda d) - f(a)}{\lambda}$$

y por lo tanto el límite existe o es igual a $-\infty$.

[1] Por conveniencia, en esta sección diremos que f está definida en todo $\mathbb{R}^n$. Sin embargo, los resultados son también ciertos (y las demostraciones son las mismas) cuando f se restringe a un dominio dado (por ejemplo, $f(x) = -\ln(x)$ es convexa y tiene como domino $\mathbb{R}_{++}$). En este caso, S es siempre un subconjunto del dominio. Recordemos que el dominio de f se define como $Dom(f) = \{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$.

Demostración. *Sea $\eta > \lambda$. Así, de la convexidad de f se deduce que:*

$$f(a+\lambda d) = f\left(\frac{\lambda}{\eta}(a+\eta d) + (1-\frac{\lambda}{\eta})a\right) \leq \frac{\lambda}{\eta} f(a+\eta d) + \left(1-\frac{\lambda}{\eta}\right) f(a),$$

obteniendo

$$\frac{f(a+\lambda d) - f(a)}{\lambda} \leq \frac{f(a+\eta d) - f(a)}{\eta}.$$

La última desigualdad se obtiene tras dividir por λ y reordenar los términos. Claramente si f es estrictamente convexa la desigualdad anterior es también estricta.

La identidad $\lim_{\lambda\to 0^+} \frac{f(a+\lambda d)-f(a)}{\lambda} = \inf_{\lambda \in \mathbb{R}_+} \frac{f(a+\lambda d)-f(a)}{\lambda}$ es consecuencia directa de lo anterior. Finalmente, si el ínfimo existe, entonces existe el límite y es finito. Si no existe el ínfimo, entonces el límite es $-\infty$, pues la función es creciente.$\square$

Como consecuencia del resultado anterior se deduce el siguiente resultado.

Teorema. *Sea S convexo abierto no vacío en $\mathbb{R}^n$ y $f : \mathbb{R}^n \to \mathbb{R}$, diferenciable en S. Entonces f es convexa en S si y sólo si*

$$f(x) \quad \geq \quad f(\overline{x}) + \nabla f(\overline{x})^t (x - \overline{x}) \qquad \forall \bar{x}, x \in S.$$

Además, f es estrictamente convexa si y sólo si la desigualdad anterior es estricta.

Demostración. *Probaremos las dos implicancias:*

- $\Rightarrow$ *Supongamos que f es convexa. Basta evaluar, para $a = x$ y $d = (x - \bar{x})$ el término $\frac{f(a+\lambda d)-f(a)}{\lambda}$ en $\lambda \to 0^+$ y $\lambda = 1$. Así, la desigualdad pedida se deduce directamente del lema anterior.*

- $\Leftarrow$ *Sean x_1, $x_2 \in S$ y $\lambda \in [0,1]$. Usemos la desigualdad de la hipótesis para $\bar{x} = \lambda x_1 + (1-\lambda)x_2$ y $x = x_1$ y $x = x_2$. Esto nos da*

  $$\begin{aligned} f(x_1) &\geq f(\overline{x}) + \nabla f(\overline{x})^t (x_1 - \overline{x}) \\ f(x_2) &\geq f(\overline{x}) + \nabla f(\overline{x})^t (x_2 - \overline{x}). \end{aligned}$$

  Multiplicando estas ecuaciones por λ y $1-\lambda$, respectivamente, y sumandolas, se obtiene:

  $$\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\overline{x}) + \nabla f(\overline{x})^t (\lambda x_1 + (1-\lambda)x_2 - \overline{x}) = f(\bar{x}).$$

  $\square$

De lo anterior se concluye lo siguiente:

Corolarios. *Sea S convexo abierto no vacío en $\mathbb{R}^n$ y $f : \mathbb{R}^n \to \mathbb{R}$ dos veces continuamente diferenciable en S. Se concluye que:*

- $\nabla^2 f(x)$ *es semidefinida positiva para todo $x \in S$ si y sólo si f es convexa en S.*
- *Si $\nabla^2 f(x)$ es definida positiva para todo $x \in S$, entonces f es estrictamente convexa en S.*

2. Condiciones de Optimalidad para problemas no lineales irrestrictos

Teorema (Condición necesaria de primer orden, Regla de Fermat)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}$ diferenciable en $x \in int(Dom(f))$. Si x es mínimo local de f entonces $\nabla f(x) = 0$. Si además f es convexa, entonces esta condición es también suficiente y el mínimo es global.*

Teorema (Condición necesaria de segundo orden)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}$ dos veces continuamente diferenciable en $x \in int(Dom(f))$. Si x es mínimo local de f entonces $\nabla f(x) = 0$ y $\nabla^2 f(x)$es semidefinida positiva.*

Teorema (Condición suficiente de segundo orden)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}$ dos veces continuamente diferenciable en $x \in int(Dom(f))$. Si $\nabla f(x) = 0$ y $\nabla^2 f(x)$ es definida positiva, entonces x es mínimo local aislado de f.*

Problema (Desigualdad del promedio Aritmético-Geométrico (Cauchy, 1821))**.** *Pruebe que para $x_1, ..., x_n \in \mathbb{R} \setminus \{0\}$, se cumple*

$$(x_1 \cdot ... \cdot x_n)^{\frac{1}{n}} \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n} x_i$$

Solución. *Consideremos el siguiente cambio de variable $x_i = e^{y_i}$, con $y_i \in \mathbb{R}$. Con este cambio, la desigualdad a demostrar es:*

$$\begin{aligned}
\exp(\frac{y_1 + ... + y_n}{n}) &\leq \frac{e^{y_i} + ... + e^{y_n}}{n} \\
0 &\leq \frac{e^{y_i} + ... + e^{y_n}}{n} - \exp(\frac{y_1 + ... + y_n}{n}) \\
0 &\leq \Phi(y_1, ..., y_n)
\end{aligned}$$

con $\Phi : \mathbb{R}^n \to \mathbb{R}$. Si logramos probar que el mínimo de la función Φ sobre $\mathbb{R}^n$ es mayor o igual a cero, estamos listos.
Consideremos la restricción adicional $y_1 + ... + y_n = s$, con $s \in \mathbb{R}$. La función Φ queda de la forma

$$\Phi(y_1, ..., y_n, s) \quad = \quad \frac{e^{y_1} + ... + e^{y_n}}{n} - \exp(\frac{s}{n})$$

Con esa restricción el problema a resolver es:

$$\begin{aligned}
min \quad & \Phi(y_1, ..., y_n, s) \\
s.a \quad & y_1 + ... + y_n = s \\
& y_1, ..., y_n, s \in \mathbb{R}
\end{aligned}$$

Ahora, podemos considerar a la variable y_n de la forma $y_n = s - y_1 - ... - y_{n-1}$, con lo cual la función queda:

$$\Phi(y_1, ..., y_{n-1}, s) \quad = \quad \frac{e^{y_1} + ... + e^{y_{n-1}} + e^{s - y_1 - ... - y_{n-1}}}{n} - \exp(\frac{s}{n})$$

Y el problema a resolver es (se elimina la restricción antes incorporada):

$$\begin{aligned}
min \quad & \Phi(y_1, ..., y_{n-1}, s) \\
s.a \quad & y_1, ..., y_{n-1}, s \in \mathbb{R}
\end{aligned}$$

De la condición de primer orden, veamos las ecuaciones que debe satisfacen un óptimo local ($\nabla\Phi(y_1, ..., y_{n-1}, s) = 0$):

$$\begin{aligned}
e^{y_i} - e^{s - y_1 - ... - y_{n-1}} &= 0 \qquad \forall i = 1, ..., n-1 \\
e^{s - y_1 - ... - y_{n-1}} - \exp(\frac{s}{n}) &= 0
\end{aligned}$$

Es decir:

$$\begin{aligned}
y_i &= s - y_1 - ... - y_{n-1} \qquad \forall i = 1, ..., n-1 \\
s - y_1 - ... - y_{n-1} &= \frac{s}{n}
\end{aligned}$$

Luego, si y^* es óptimo local, debe ser de la forma $y_i = \frac{s}{n} \forall i = 1, ..., n$ y por lo tanto único.

Por otro lado, notando que $\Phi(y_1, ..., y_{n-1}, s)$ es continua y coerciva, del teorema de Weirstrass se concluye que Φ tiene un mínimo global, el cual debe coincidir con el mínimo anterior. Para este mínimo, el valor de $\Phi(y_1, ..., y_n)$ es

$$\Phi(\frac{s}{n}, ..., \frac{s}{n}) = \frac{1}{n}(ne^{\frac{s}{n}}) - e^{\frac{s}{n}} = e^{\frac{s}{n}} - e^{\frac{s}{n}} = 0,$$

de donde se concluye la desigualdad deseada.$\square$