Subgradientes y subdiferenciales

Dada una función convexa $f: \mathbb{R}^n \to \mathbb{R}$, decimos que un vector $d \in \mathbb{R}^n$ es un subgradiente de f en un punto x si

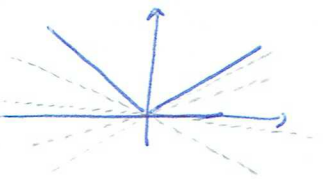
$$f(z) \geq f(x) + (z-x)^T d \qquad \forall z \in \mathbb{R}^n$$

Al conjunto de todos los subgradientes de f en x se le llama subdiferencial de f en x, y se denota $\partial f(x)$.

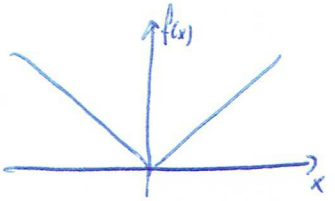
Ejemplos

$f(x) = |x|$ en $x = 0$

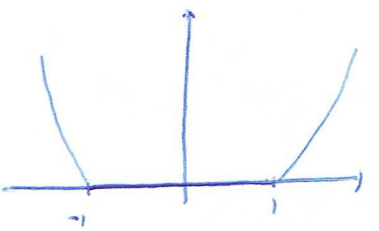
$\partial f(0) = [-1, 1]$

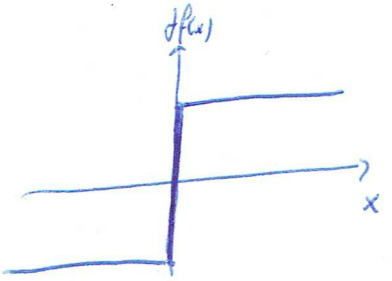


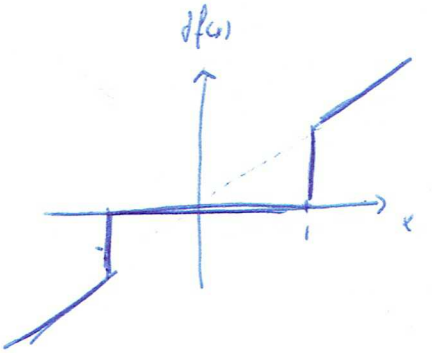
$f(x) = |x|$



$f(x) = \max\{0, \frac{1}{2}(x^2-1)\}$







Obs

Si f es diferenciable, el único subgradiente es el gradiente.

Prop

Sea $f: \mathbb{R}^n \to \mathbb{R}$ convexa y $x \in \mathbb{R}^n$. Entonces $\partial f(x)$ es un conjunto no vacío, convexo y compacto. Además, $\forall v \in \mathbb{R}^n$ se tiene

$$\frac{\partial f}{\partial v}(x) = \max_{g \in \partial f(x)} g^T v \qquad \left(\text{con } \frac{\partial f}{\partial v}(x) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha d) - f(x)}{\alpha} \right)$$

Dem (de una parte)

Por la definición de subgradiente, tenemos $\frac{f(x+\alpha d) - f(x)}{\alpha} \geq g^T d \quad \forall d \in \mathbb{R}^n, \alpha > 0$

Como el lado izquierdo converge monótonamente (con $\alpha \downarrow 0$) a $\frac{\partial f}{\partial d}(x)$, entonces

$$g \in \partial f(x) \iff \frac{\partial f}{\partial d}(x) \geq g^T d \quad \forall d \in \mathbb{R}^n$$

Por lo tanto $\partial f(x)$ se puede escribir como la intersección de los conjuntos $\{ g \,/\, \frac{\partial f}{\partial d}(x) \geq g^T d \}$ haciendo variar d en todos los vectores no nulos de $\mathbb{R}^n$.

Estos conjuntos son semi-espacios cerrados, y por lo tanto $\partial f(x)$ es cerrado y convexo. Además si fuera no acotado, podríamos hacer $g^T d$ no acotado, para algún d, lo que contradice $\frac{\partial f}{\partial d}(x) \geq g^T d$

Prop (Condición de Optimalidad)

Un punto $x \in \mathbb{R}^n$ minimiza una función convexa $f: \mathbb{R}^n \to \mathbb{R}$ sobre un conjunto convexo X sii existe un subgradiente $g \in \partial f(x)$ tal que $g^T(z-x) \geq 0 \quad \forall z \in X$

Obs:

- Esta condición generaliza la condición $\nabla f(x)^T (z-x) \geq 0 \quad \forall z \in X$
- Cuando $X = \mathbb{R}^n$, la condición necesaria y suficiente queda $0 \in \partial f(x)$

Interpretación geométrica

Dado un conjunto X y $x \in X$, llamamos cono normal a X en x al conjunto

$$N_X(x) = \{ g \mid g^T(z-x) \leq 0, \ \forall z \in X \}$$

Subgradiente

Fijamos $x \in \mathbb{R}^n$, y llamamos $c = f(x)$

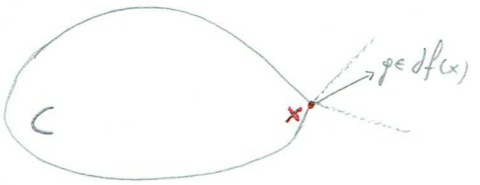
Consideramos el conjunto de nivel $C = \{ z \in \mathbb{R}^n \mid f(z) \leq c \}$

Si $g \in \partial f(x) \Rightarrow f(z) \geq f(x) + g^T(z-x) \quad \forall z \in \mathbb{R}^n$

En particular, si $z \in C \Rightarrow f(z) \leq f(x) \Rightarrow g^T(z-x) \leq 0$

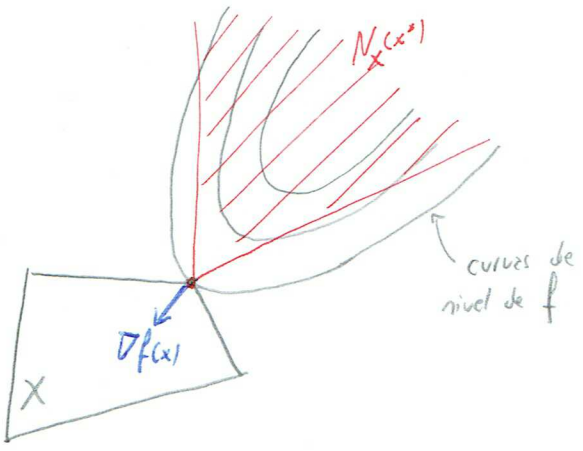
Esto quiere decir que $g \in N_C(x)$

(Esto generaliza el concepto que el gradiente es normal a la curva de nivel. Cuando f es diferenciable, el cono normal a C es una semi-recta)

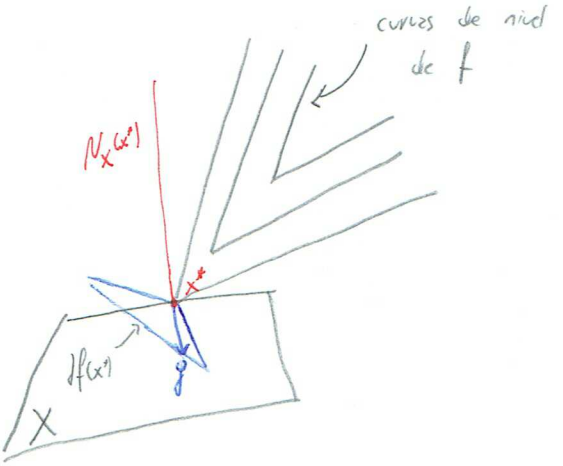


Condición de optimalidad

La condición de optimalidad $(\exists g \in \partial f(x) \mid g^T(z-x) \geq 0 \quad \forall z \in X)$ se traduce en que $\exists g \in \partial f(x) \mid -g \in N_X(x)$



f diferenciable $\Rightarrow$ la condición es $-\nabla f(x^*) \in N_X(x^*)$



f no diferenciable

Métodos de subgradiente

Veremos métodos de la forma:

$$x^{k+1} = P_X(x^k - \alpha^k g^k)$$

donde g^k es cualquier subgradiente $g^k \in \partial f(x^k)$ y $\alpha^k \geq 0$

Cuando $X = \mathbb{R}^n$ (sin restricciones)

$$x^{k+1} = x^k - \alpha^k g^k$$

Ejemplo

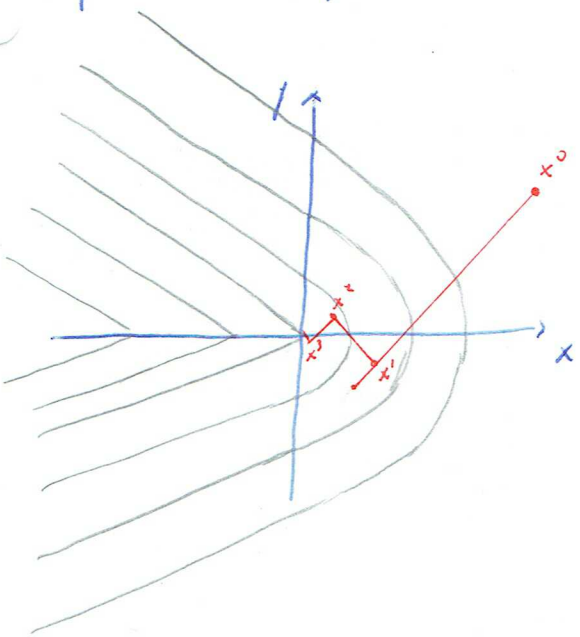
$$f(x,y) = \begin{cases} 5\sqrt{9x^2+16y^2} & \text{si } x > |y| \\ 9x + 16|y| & \text{si } x \leq |y| \end{cases}$$

Empezando en cualquier punto de $\{(x,y) / x > |y| > (\frac{9}{16})^2 x\}$

y usando el método de máximo descenso, con α^k determinado como

$$\alpha^k = \arg\min_{\alpha > 0} f(x^k - \alpha^k g^k)$$ (se puede encontrar este α^k analíticamente)

el método sigue una poligonal con ángulos rectos sucesivos, que converge al $(0,0)$ que no es un punto estacionario. (y de hecho $\lim_{x \to -\infty} f(x,0) = -\infty$)



A pesar de que en todos los puntos x^k la f es diferenciable, el algoritmo falla

Obs:

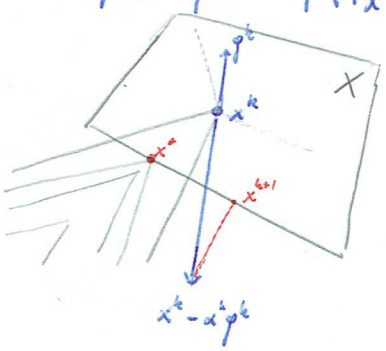
- Al proyectar, no nos alejamos del óptimo:

  De la no expansividad de la proyección: $\|P_X(x^k - \alpha^k g^k) - x\| \leq \|(x^k - \alpha^k g^k) - x\| \quad \forall x \in X$

  en particular para x^*

- No tenemos asegurado que en cada iteración bajemos el costo.

  Puede pasar que $f(P_X(x^k - \alpha g^k)) > f(x^k) \quad \forall \alpha > 0$



Sin embargo, si el paso es suficientemente chico, la distancia al óptimo se reduce:

Prop

Sea $\{x^k\}$ generado por el método de subgradiente. Entonces $\forall y \in X$ y $k \geq 0$:

(a) $\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha^k (f(x^k) - f(y)) + \alpha^{k^2} \|g^k\|^2$

(b) Si $f(y) < f(x^k)$, entonces $\|x^{k+1} - y\| < \|x^k - y\|$

para todo paso α^k / $0 < \alpha^k < \dfrac{2(f(x^k) - f(y))}{\|g^k\|^2}$

Dem

$$\|x^{k+1} - y\|^2 = \|P_X(x^k - \alpha^k g^k) - y\|^2 \leq \|x^k - \alpha^k g^k - y\|^2 = \|x^k - y\|^2 - 2\alpha^k g^{kT}(x^k - y) + \alpha^{k^2}\|g^k\|^2$$

$$\leq \|x^k - y\|^2 - 2\alpha^k (f(x^k) - f(y)) + \alpha^{k^2}\|g^k\|^2$$

(en la última desigualdad se usa la definición de subgradiente)

(b) es inmediato de (a)

Esto supone tomar un paso $\alpha^k = \dfrac{f(x^k) - f(x^*)}{\|p^k\|^2}$ (punto medio del intervalo de la prop. anterior)

Sin embargo, muy rara vez conocemos el valor óptimo.

Veamos algunos resultados para pasos "prácticos".

Paso constante:

Tomamos $\alpha^k = \alpha \ \forall k$, y sea $f_\infty = \lim_{k\to\infty} \inf f(x^k)$

Si $\|p^k\| \leq c \ \forall k$, entonces $f_\infty \leq f(x^*) + \dfrac{\alpha c^2}{2}$

Dem

Si no fuera cierto, entonces $\exists \varepsilon > 0$ / $f_\infty \geq f(x^*) + \dfrac{\alpha c^2}{2} + 2\varepsilon$

Sea $\bar{k}$ suficientemente grande tal que $f(x^k) \geq f_\infty - \varepsilon \quad \forall k \geq \bar{k}$

Sumando: $f(x^k) - f(x^*) \geq \dfrac{\alpha c^2}{2} + \varepsilon$

Usando la Prop anterior:

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha\left(f(x^k) - f(x^*)\right) + \alpha^2 c^2$$

$$\leq \|x^k - x^*\|^2 - 2\alpha\left(\frac{\alpha c^2}{2} + \varepsilon\right) + \alpha^2 c^2 = \|x^k - x^*\|^2 - 2\alpha\varepsilon$$

Entonces, $\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha\varepsilon$

$$\leq \|x^{k-1} - x^*\|^2 - 4\alpha\varepsilon$$

$$\vdots$$

$$\leq \|x^{\bar{k}} - x^*\|^2 - 2(k+1-\bar{k})\alpha\varepsilon$$

lo cual no puede ser cierto para k grande.

Es decir, la distancia al óptimo no se puede reducir una cantidad constante $(2\alpha\varepsilon)$ indefinidamente. Como la distancia al óptimo está relacionada con la diferencia de valores funcionales (por la Prop anterior), llegamos a lo deseado.

Paso decreciente

Suponemos que $\|g^k\| \leq c \ \forall k$.

Entonces si $\lim_{k\to\infty} \alpha^k = 0$ y $\sum \alpha^k = \infty \Rightarrow f_\infty = f^*$

Si además $\sum (\alpha^k)^2 < \infty \Rightarrow x^k$ converge al óptimo (o a un óptimo)

La demostración es igual que el caso anterior.