Exercises on Convex and Constrained Optimization

Álvaro José Álvarez Arranz

May 8, 2024

1. (*) Show that if S is an open set, its complement S^c is closed, and viceversa.

Solución:

Supongamos que S es un conjunto abierto. Vamos a probar que S^c es cerrado mostrando que coincide con su cierre $cl(S^c)$. Por un lado, tenemos por definición que $S^c \subset cl(S^c)$ como:

$$cl(S^c) = \{x : \forall\delta \; B(x, \delta) \cap S^c \neq \emptyset\}$$

Por otra parte, $x \in cl(S^c)$ y $\delta > 0$. Entonces, tenemos:

$$B(x, \delta) \subset S^c \neq \emptyset \implies B(x, \delta) \not\subset S$$

Pero como S es abierto, se cumple que $\forall x \in S$, $\exists\delta > 0$ tal que $B(x, \delta) \subset S$. Por lo tanto, como δ era arbitraria en la última ecuación, necesariamente $x \notin S$, entonces $x \in S^c$. Como x también era arbitraria, hemos demostrado que $cl(S^c) \subset S^c$, y por tanto que el complemento de un conjunto abierto es cerrado, como se deseaba. La inversa es obvia tomando los complementos.

2. (*) If S_1, S_2 are convex subsets, prove the following are also convex sets:

   (a) $S_1 \cap S_2 = \{x : x \in S_1 \text{ and } x \in S_2\}$

   (b) $S_1 + S_2 = \{x + x' : x \in S_1, x' \in S_2\}$

   (c) $S_1 - S_2 = \{x - x' : x \in S_1, x' \in S_2\}$

Solución:

- Sea $S_1 \cap S_2 = \{x : x \in S_1 \text{and } x \in S_2\}$. Sea $x, x' \in S_1 \cap S_2$ y consideremos el segmento
$$z \equiv \lambda x + (1-\lambda)x', \lambda \in [0,1]$$
Como $x \in S_1$, que es un conjunto convexo, $z \in S_1$ para cualquier $\lambda \in [0,1]$. Del mismo modo, $z \in S_2$, y por lo tanto en la intersección $S_1 \cap S_2$. Dado que el segmento está contenido en la intersección, $S_1 \cap S_2$ es un subconjunto convexo.

- Sea $S_1 + S_2 = \{x + z : x \in S_1, z \in S_2\}$. También, que $x + z \in S_1 + S_2$ y $x' + z' \in S_1 + S_2$. Además, considere el segmento
$$\begin{aligned}\lambda(x+z) + (1-\lambda)(x'+z') &= \lambda x + \lambda z + (1-\lambda)x' + (1-\lambda)z' \\ &= \underbrace{\lambda x + (1-\lambda)x'}_{\in S_1} + \underbrace{\lambda z + (1-\lambda)z'}_{\in S_2} \in S_1 + S_2\end{aligned}$$

- Usamos el mismo proceso realizado en el conjunto anterior. Sea $S_1 - S_2 = \{x - z : x \in S_1, z \in S_2\}$. También, que $x - z \in S_1 - S_2$ y $x' - z' \in S_1 - S_2$. Además, considere el segmento
$$\begin{aligned}\lambda(x-z) + (1-\lambda)(x'-z') &= \lambda x - \lambda z + (1-\lambda)x' - (1-\lambda z') \\ &= \underbrace{\lambda x + (1-\lambda)x'}_{\in S_1} - \underbrace{(\lambda z + (1-\lambda)z')}_{\in S_2} \in S_1 - S_2\end{aligned}$$

3. If $f : S \rightarrow \mathbf{R}$ is a convex set S, prove that the set $\{x : x$ is a minimum of $f\}$ is a convex set.

Solución:

Omitimos el caso en que $S_{min} = \emptyset$, ya que el conjunto vacío es convexo. Ahora, sea y el mínimo de $f(x) : y = \min\limits_{x} f(x)$. Entonces, $S_{min} = \{x \in S : f(x) = y\}$. Tenemos que demostrar que $\forall x, x' \S_{min}$ y $\forall \lambda \in [0,1]$:

$$z \equiv \lambda x + (1-\lambda) x' \in S_{min} \iff f(z) = y$$

Como S es convexo, $z \in S$, y como f también es convexo:

$$\begin{aligned} f(z) &= f(\lambda x + (1-\lambda) x') \\ &\leq \lambda f(x) + (1-\lambda) f(x') \\ &= \lambda y + (1-\lambda) y \\ &= y \end{aligned}$$

donde hemos empleado $x, x' \in S_{min}$. Pero como y es el mínimo de f, se cumple la igualdad $f(z) = y$. Esto significa que $z \in S_{min}$, por lo tanto S_{min} es un conjunto convexo.

4. Given a quadratic form $q\left(w\right) = \frac{1}{2}w^tQw + bw + c$, with Q a symmetric $d \times d$ matrix, w,b $d \times 1$ vectors and c a real number, derive its gradient and Hessian

$$\nabla q\left(w\right) = Qw + b,\ \ Hq\left(w\right) = Q$$

Hint: expand $q(w) = \sum_{i=1}^{d}\sum_{j=1}^{d} Q_{ij}w_iw_j + \sum_{i=1}^{d} b_iw_i + c$, *and take the partials* $\frac{\partial q}{\partial w_i}$ *y* $\frac{\partial^2 q}{\partial w_i \partial w_j}$.

Solución:

Partiendo de la forma expandida

$$q\left(w\right) = 0.5\sum_{i=1}^{d}\sum_{j=1}^{d} Q_{ij}w_iw_j + \sum_{i=1}^{d} b_iw_i + c$$

y teniendo en cuenta que $Q_{ij} = Q_{ji}$, podemos calcular fácilmente las derivadas parciales

$$\begin{aligned}\frac{\partial q\left(w\right)}{\partial w_i} &= Q_{ii}w_w + \sum_{\substack{j=1\\ j\neq i}}^{d} w_j\left(Q_{ij} + Q_{ji}\right) + b_i\\ &= \sum_{j=1}^{d} Q_{ij}w_j + b_i, \quad i = 1, ..., d,\end{aligned}$$

y

$$\frac{\partial^2 q\left(w\right)}{\partial w_i \partial w_j} = Q_{ij}, \quad i, j = 1, ..., d$$

Como resultado, tenemos

$$\begin{aligned}\nabla q\left(w\right) &= \left(\frac{\partial q\left(w\right)}{\partial w_i}, ..., \frac{\partial q\left(w\right)}{\partial w_d}\right)^T\\ &= \left(\sum_{j=1}^{d} Q_{1j}w_j, ..., \sum_{j=1}^{d} Q_{dj}w_j\right)^T + \left(b_1, ..., b_d\right)^T\\ &= 2Qw + b\end{aligned}$$

y

$$
\begin{aligned}
Hq\,(w) &= \begin{pmatrix} \frac{\partial^2 q(w)}{\partial w_1 \partial w_1} & \cdots & \frac{\partial^2 q(w)}{\partial w_1 \partial w_d} \\ \vdots & & \vdots \\ \frac{\partial^2 q(w)}{\partial w_d \partial w_1} & \cdots & \frac{\partial^2 q(w)}{\partial w_d \partial w_d} \end{pmatrix} \\
&= \begin{pmatrix} Q_{11} & \cdots & Q_{1d} \\ \vdots & & \vdots \\ Q_{d1} & \cdots & Q_{dd} \end{pmatrix}
\end{aligned}
$$

5. (*) If $(p_1, ..., p_K)$ is a probability distribution, prove that its entropy $H(p_1, ..., p_K) = -\sum_{i=1}^{K} p_i \text{log} p_i$ is a concave function. Show also that its maximum is lg K, attained when $p_i = \frac{1}{K}\ \forall i$.

Solución:

En este problema, al tratarse de probabilidades, aparecen dos nuevas restricciones:

$$\sum_{i=1}^{K} p_i = 1 p_i \geq 0\ \forall i = 1, ..., n$$

Se utilizarán más adelante.

Calculemos el gradiente y el hessiano de H para ver que es cóncavo. En primer lugar, tenemos que

$$\frac{\partial H}{\partial p_i} = -\log(p_i) - 1,\ \forall i = 1, ..., K$$

y, por lo tanto,

$$\frac{\partial^2 H}{\partial p_i \partial p_j} = -\frac{\delta_{ij}}{p_i},$$

donde δ_{ij} es el delta de Kroneker. Por último, dado que $p_i \geq 0$, tenemos que resolver el siguiente problema de optimización:

$$\max_{(p_1,...,p_K)} H\,(p_1, ..., p_K)\,\text{s.t.} \sum_{i=1}^{K} p_i - 1 = 0 p_i \geq 0\ \forall i = 1, ..., K$$

Consideremos el lagrangiano de este problema:

$$L\left(\{p_i\}_{i=1}^{K}, \lambda\right) = -\sum_{i=1}^{K} p_i \log(p_i) + \lambda\left(\sum_{i=1}^{K} p_i - 1\right).$$

Podemos obtener su gradiente diferenciando con respecto a cada variable

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\log p_i - 1 + \lambda, \ \frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{K} p_i - 1.$$

Usando estas derivadas, tenemos que igualarlas a cero, lo que es resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \log p_i = \lambda - 1, \ i = 1, ..., K \\ \sum_{i=1}^{K} p_i = 1 \end{cases}$$

Observando la primera ecuación, las p_i no tienen interdependencias, son constantes y tienen el mismo valor. Además, como tienen que sumar 1, la única solución posible es que cada $p_i = \frac{1}{K}$. Por último, podemos calcular el valor máximo de la entropía:

$$\begin{aligned} H\left(\{p_i\}_{i=1}^{K}\right) &= -\sum_{i=1}^{K} \frac{1}{K} \log\left(\frac{1}{K}\right) \\ &= -\frac{1}{K}\left(\sum_{i=1}^{K} \log 1 - \log K\right) \\ &= \frac{1}{K} \cdot K \log K = \log K \end{aligned}$$

como queríamos demostrar.

6. (*) We want to solve the following constrained restriction problem:

$$\begin{aligned} \min\ & x^2 + 2y^2 + 4xy \\ \text{s.t.}\ & x + y = 1 \\ & x \geq 0, y \geq 0 \end{aligned}$$

(a) Write its Lagrangian with α, β the multipliers of the inequality constrains.

(b) Write the KKT conditions.

(c) Use them to solve the problem. For this consider separately the $(\alpha = \beta = 0), (\alpha > 0, \beta = 0), (\alpha = 0, \beta > 0), (\alpha > 0, \beta > 0)$ cases.

Solución:

- Escribir la **Lagrangiana** en términos de $\alpha,\ \beta, \lambda$ es bastante sencillo:

$$\mathcal{L}(x, y, \alpha, \beta, \lambda) = x^2 + 2y^2 + 4xy + \lambda(x + y - 1) + \alpha x + \beta y$$

- **Condiciones KKT**. Resumiendo mucho, las condiciones KKT son el gradiente de la Lagrangiana igual a 0 y las restricciones de desigualdad, multiplicadas por su correspondiente constante, también igual a cero. Así, las **condiciones de KKT** son:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} = 2x + 4y + \lambda + \alpha = 0 \\ \frac{\partial \mathcal{L}}{\partial y} = 4y + 4x + \lambda + \beta = 0 \\ \alpha x = 0 \\ \beta y = 0 \end{aligned}$$

- Ahora, vamos a resolver este sistema de ecuaciones mediante los siguientes casos:

  (a) Caso $(\alpha = \beta = 0)$. El sistema es:

  $$\begin{aligned} 2x + 4y + \lambda = 0 \\ 4y + 4x + \lambda = 0 \end{aligned}$$

  Sustituyendo $4y$ de la primera ecuación en la segunda ecuación, obtenemos

  $$-2x - \lambda + 4x + \lambda = 0 \implies 2x = 0 \implies x = 0$$

y, como $x + y = 1$, obtenemos que out *KKT point* es$(0, 1)$. Cualquier valor no negativo de α y β es válido para satisfacer tanto las condiciones KKT como las restricciones de nuestro problema.

(b) Caso $(\alpha > 0, \beta = 0)$. Observando nuestras condiciones KKT, dado que $\alpha > 0$, tenemos que $x = 0$, resultando $y = 1$ y un punto KKT $(0, 1)$, que es el mismo que obtuvimos en el primer caso.

(c) Caso $(\alpha = 0, \beta > 0)$. Obtenemos $(1, 0)$ como un nuevo *KKT punto.*

(d) Caso $(\alpha > 0, \beta > 0)$. En este caso, obtenemos de las condiciones KKT que $x = y = 0$, lo que no coincide con nuestras condiciones iniciales $x + y = 1$, por lo que no se obtienen puntos *KKT*.

Hasta ahora, tenemos dos candidatos a ser el óptimo: $\{(0, 1), (1, 0)\}$. Ahora hacemos uso del siguiente teorema:

Teorema 1: *Si en un problema de minimización con restricciones* $g_i(x), h_j(x) \in C^1$*, si suponemos que* f *es convexo y* h_j *es afín, entonces un punto KKT* $x*$ *es un óptimo de este problema. (Diapositiva 18)*

Por lo tanto, podemos evaluar la función en nuestros puntos KKT para encontrar el mínimo. Obtenemos que $f(1, 0) = 1$, $f(0, 1) = 2$, por lo que el mínimo se alcanza en $(1, 0)$ con valor óptimo 1.

7. We have worked out the dual problem for the soft SVC problem. Do the same for the simpler **hard** SVC problem

$$\min_{w,b} \frac{1}{2}||w||^2$$

subject to $y^p(w \cdot x^p + b) \geq 1$. What are the KKT conditions?

Solución:

En primer lugar, consideremos el Lagrangiano para este problema

$$\begin{aligned}
\mathcal{L}(w, b; \alpha) &= \frac{1}{2}||w||^2 - \sum_p \alpha_p \left[y^p \left(w \cdot x^p + b \right) - 1 \right] \\
&= \frac{1}{2} w \cdot w - w \sum_p \alpha_p y^p x^p - b \sum_p \alpha_p y^p + \sum_p \alpha_p \\
&= w \left(\frac{1}{2} w - \sum_p \alpha_p y^p x^p \right) - b \sum_p \alpha_p y^p + \sum_p \alpha_p
\end{aligned}$$

A continuación, tenemos que calcular el gradiente de la Lagrangiana

$$\nabla_w \mathcal{L}(w, b; \alpha) = w - \sum_p \alpha_p y^p x^p = 0 \implies w = \sum_p \alpha_p y^p x^p$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_p \alpha_p y^p = 0$$

Por último, tenemos que utilizar estas igualdades en la expresión de la Lagrangiana:

$$\begin{aligned}
\mathcal{L}(w, b; \alpha) &= \sum \alpha_p - 0.5 \left(\sum_p \alpha_p y^p x^p \right) \left(\sum_q \alpha_q y^q x^q \right) \\
&= \sum \alpha_p - 0.5 \sum_{p,q} \alpha_p \alpha_q y^p y^q x^p x^q
\end{aligned}$$

Definiendo la matriz Q con valor $y^p y^q x^p x^q$ en la posición (p, q), nuestro problema de optimización dual se simplifica en

$$\begin{cases} \max_\alpha \sum_p \alpha^p - 0.5 \alpha^T Q \alpha \\ \text{s.t. } \alpha^p, \sum \alpha_p y^p = 0 \end{cases}$$

Llegados a este punto, podemos darnos cuenta de que hemos eliminado completamente la dependencia b de nuestro problema. Las condiciones KKT para este problema son:

$$\begin{cases} \nabla_w \mathcal{L} = w - \sum_p \alpha_p y^p x^p & = 0 \\ \frac{\partial \mathcal{L}}{\partial b} = -\sum_p \alpha_p y^p & = 0 \\ \alpha_p \left(1 - y^p \left(w \cdot x^p + b\right)\right) & = 0 \end{cases}$$

8. (*) A typical Linear Programming (LP) problem can be stated as the following constrained optimization problem:

$$\min_x c \cdot x \text{ s.t. } x \geq 0, Ax \leq b$$

with $x \in \mathbf{R}^d$, A an $m \times d$ matrix and $b \in \mathbf{R}^m$. A tool often used in LP is to study the so called **dual problem**, which in this case is

$$\min_z b \cdot z \text{ s.t. } z \geq 0, A^t z \leq -c$$

with now $z \in \mathbf{R}^m$. Apply our Lagrangian dual construction technique to show that is indeed the dual formulation of the initial LP problem.

Solución:

En primer lugar, tenemos el Langrangiano para este problema:

$$\begin{aligned} \mathcal{L}\left(x, \lambda, \mu\right) &= c \cdot x - \sum_{i=1}^{d} \lambda_i x_i + \sum_{j=1}^{m} \mu_j \left(a_j \cdot x - b_j\right) \\ &= xc - x\lambda + x\left(A^T \mu\right) - b\mu \end{aligned}$$

donde $\mu \geq 0$. Por lo tanto, el gradiente respecto a x de la lagrangiana es:

$$\nabla_x \mathcal{L} = c - \lambda + A^T \mu \implies c = \lambda - A^T \mu$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -x = 0 \implies x = 0$$

Por tanto, el problema dual es:

$$\begin{cases} \max_\mu -b\mu \\ \text{s.t. } c = \lambda - A^T \mu, \ \mu \geq 0 \end{cases}$$

Como λ no aparece en la función objetivo, podemos eliminarlo cambiando la restricción de $c = \lambda - A^T\mu$ a $-c \geq A^T\mu$. Además, podemos cambiar la función de optimización de $\max_m u - b\mu$ a $\min_\mu b\mu$, obteniendo:

$$\begin{cases} \min_\mu b\mu \\ \text{s.t. } -c \geq A^T\mu, \ \mu \geq 0 \end{cases}$$

como debía demostrarse.

9. We know that, theoretically, the minimum SVC primal f^* and the maximum SVC dual q^* are equal. Check this in case by writing q^* and f^* in terms of the α_p^* and checking that both expressions coincide.

Solución:

El problema SVC primario se define como

$$\min_w \frac{1}{2}||w||^2 + C\sum_p \xi_p \text{ s.t. } y^p\left(wx^p + b\right) \geq 1 - \xi_p, \ \xi_p \geq 0$$

Vamos a ver que el mínimo de esta función, digamos f^*, es igual al máximo de la función dual $\mathcal{D}$, digamos q^*.

Recordemos que, la relación entre el óptimo w^* del problema primal y el α^* del problema dual es:

$$w^* = \sum_p \alpha_p^* y^p x^p$$

y también que

$$\xi_p^* = 1 - y^p\left(w^* x^P + b^*\right) \text{ and } C = \alpha_p^* + \beta_p^*$$

Podemos utilizar estas expresiones en el problema SVC primario para obtener la siguiente cadena de igualdad:

$$\begin{aligned}
\frac{1}{2}||w||^2 + C\sum_p \xi_p &\overset{(1)}{=} \frac{1}{2}\sum_{p,q}\alpha_p^* y^p x^p y^q x^q \left(\alpha_q^*\right)' + \left(\alpha_p^* + \beta_p^*\right)\sum_p \left(1 - y^p\left(w^* x^p + b^*\right)\right) \\
&= \frac{1}{2}\left(\alpha^*\right)\mathcal{Q}\alpha^* + \alpha_p^*\sum_p\left(1 - y^p\left(w^* x^p + b^*\right)\right) + \beta_p^*\sum_p\left(1 - y^p\left(w^* x^p + b^*\right)\right) \\
&\overset{(2)}{=} \frac{1}{2}\left(\alpha^*\right)'\mathcal{Q}\alpha^* + \sum_p \alpha_p^* - \alpha_p^*\sum_p y^p\left(w^* x^p + b^*\right) + \sum \xi_p^*\beta_p^* \\
&= \frac{1}{2}\left(\alpha^*\right)'\mathcal{Q}\alpha^* + \sum_p \alpha_p^* - \sum_{p,q}\alpha_p^* y^p y^q x^p x^q \left(\alpha_p^*\right)' + b^*\sum_p \alpha_p^* y^p + \sum_p \xi_p^*\beta_p^* \\
&= \frac{1}{2}\left(\alpha^*\right)'\mathcal{Q}\alpha^* + \sum_p \alpha_p^* - \left(\alpha^*\right)'\mathcal{Q}\alpha^* + b^*\sum_p \alpha_p^* y^p + \sum_p \xi_p^*\beta_p^* \\
&= \sum_p \alpha_p^* - \frac{1}{2}\left(\alpha^*\right)'\mathcal{Q}\alpha^* + b^*\sum_p \alpha_p^* y^p + \sum_p \xi_p^*\beta_p^*
\end{aligned}$$

Donde, en (1) hemos utilizado las expresiones de w^* y C y en (2) hemos hecho lo mismo con ξ_p^*. Además, como los dos últimos sumatorios de la expresión resultante es igual a 0, se puede concluir que:

$$\frac{1}{2}||w||^2 + C\sum_p \xi_p = \sum_p \alpha_p^* - \frac{1}{2}\left(\alpha^*\right)'\mathcal{Q}\alpha^*$$

Observamos que esta expresión es menos la **óptima** del problema dual SVC, lo cual tiene sentido ya que en el problema dual estamos maximizando y aquí estamos minimizando, por lo que hemos demostrado que $f^* = q^*$.

10. (*) If Q is a symmetric, positive definite $d \times d$ matrix, show that $f(x) = x^t Q x, x \in \mathbf{R}^d$, is a convex function.

Solución:

Si una función f es dos veces diferenciable, entonces es convexa si y sólo si su matriz hessiana es definida positiva. En este caso, Hess $f = \mathcal{Q}$, que es simétrica y definida positiva por hipótesis, lo que demuestra que f es convexa.

> 11. Let $f : \mathbf{R}^d \to \mathbf{R}$ be a function and assume that $\text{epi}(f) \subset \mathbf{R}^d \times \mathbf{R}$ is convex. Prove that then f is convex.

Solución:

Consideremos

$$\text{epi}(f) = \left\{(x,t) \in \mathbf{R}^d \times \mathbf{R} : t \geq f(x)\right\}$$

Es evidente que, $\forall x \in \mathbf{R}^d, (x, f(x)) \in \text{epi}(f)$. Ahora, $\text{epi}(f)$ es convexa por hipótesis, por lo que si consideramos $x, x' \in \mathbf{R}^d$, tenemos:

$$\lambda(x, f(x)) + (1-\lambda)(x', f(x')) = (\lambda x + (1-\lambda)x', \lambda f(x) + (1-\lambda) f(x')) \in \text{epi}(f) \ \forall \lambda \in [0,1]$$

Puesto que $\forall \lambda \in [0,1]$ el punto obtenido está en $\text{epi}(f)$, tenemos

$$f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda) f(x'), \ \forall \lambda \in [0,1]$$

Entonces f es convexo.

> 12. Let $f : \mathbf{R}^d \to \mathbf{R}$ be a convex function. Prove that $\text{epi}(f)$ is a closed set and that $(x, f(x)) \in \partial\, \text{epi}(f)$.

Solución:

Sabemos que un conjunto S es cerrado si $cl(S) = S$. Vamos a demostrar esto cuando $S = \text{epi}(f)$, utilizando un argumento de doble inclusión.

En primer lugar, está claro que $\text{epi}(f) \subseteq cl(\text{epi}(f))$, por la definición de $cl(S)$.Para comprobar la segunda inclusión, veremos que si $(x,t) \in cl(\text{epi}(f))$, entonces $(x,t) \in \text{epi}(f)$.

Sea $(x,t) \in cl(\text{epi}(f))$. Entonces, $\forall \delta > 0$,

$$B((x,t)\,\delta) \cap \text{epi}(f) \neq \emptyset$$

Supongamos que $(x,t) \notin \text{epi}(f)$. Dado que f es convexa, $\text{epi}(f)$ también es convexa. Utilizando el *Teorema de la Proyección*, podemos encontrar una distancia mínima $d > 0$ de (x,t) a $\text{epi}(f)$. Por tanto,

$$\text{epi}(f) \cap B\left((x,t), \frac{d}{2}\right) = \emptyset,$$

que es una contradicción con la ecuación anterior, por lo que $(x,t) \in \text{epi}(f)$, y $cl(\text{epi}(f)) = \text{epi}(f)$.

Veamos ahora que $(x, f(x)) \in \partial\,\text{epi}(f)$. Podemos escribir $\text{epi}(f) = cl(\text{epi}(f)) = int(\text{epi}(f)) \cup \partial\,\text{epi}(f)$, so

$$\text{epi}(f) = cl(\text{epi}(f)) = int(\text{epi}(f)) \cup \partial\,\text{epi}(f) \implies \partial\,\text{epi}(f) = cl(\text{epi}(f)) \setminus int(\text{epi}(f))$$

Si $(x, f(x)) \in int(\text{epi}(f))$, $\exists \delta > 0$ tal que:

$$B((x, f(x)), \delta) \subset \text{epi}(f)$$

Si esto ocurriera, podríamos decir que mover el valor de $\left(x, f(x) \frac{\delta}{2}\right) \in \text{epi}(f)$. Sin embargo, esto no es cierto, ya que podría significar que $f(x)(x) - \frac{\delta}{2}$, lo que no puede ocurrir ya que $\delta > 0$. Esto demuestra que $(x, f(x)) \notin int(\text{epi}(f))$, pero $(x, f(x))$ está claramente en $\text{epi}(f)$, por lo que este punto tiene que estar en $\partial\,\text{epi}(f)$, como queríamos ver.

13. Let $f, g : S \subset \mathbf{R}^d \to \mathbf{R}$ be two convex functions on the convex set S. Prove that, as subsets, $\partial f(x) + \partial g(x) \subset \partial(f+g)(x)$.

Solución:

Sabemos que

$$\xi \in \partial f(x) \implies f(x') > f(x) + \xi(x - x') \ \forall x' \in S$$

Apliquemos esta definición para obtener el resultado.

Consideremos $\xi_1 \in \partial f(x)$ y $\xi_2 \in \partial g(x)$. Entonces, $\xi_1 + \xi_2 \in \partial f(x) + \partial g(x)$. Ahora, usando la definición para cada ξ_i anterior, obtenemos:

$$f(x') > f(x) + \xi_1(x - x'), \ g(x') > g(x) + \xi_2(x - x')$$

Y, si sumamos ambas desigualdades:

$$\begin{gathered} f(x') + g(x') > f(x) + g(x) + (\xi_1 + \xi_2)(x - x') \implies \\ (f+g)(x') > (f+g)(x) + (\xi_1 + \xi_2)(x - x') \end{gathered}$$

lo que significa que $\xi_1 + \xi_2 \in \partial(f+g)(x)$.

14. (*) Compute the proximals of $f(x) = 0$ and of $g(x) = \frac{1}{2}||x||^2$.

Solución:

Podemos calcular la proximal de f directamente:

$$\text{prox}_f(x) = \arg\min_z 0 + \frac{1}{2}||x - z||^2 = x$$

Ahora vamos con g. LSupongamos que g nace en $\mathcal{R}$. Su proximal es:

$$\text{prox}_g(x) = \arg\min_z \underbrace{0.5z^2 + 0.5||x - z||^2}_{\equiv h(z)}$$

Ahora, igualamos el gradiente de h a 0 para encontrar el mínimo:

$$0 = \nabla h(z) = z + z - x \implies z = 0.5x \implies \text{prox}_g(x) = 0.5x$$

15. (*) Show that the function $f(x) = x^3$ when $x \geq 0$ and $f(x) = -x^3$ when $x < 0$ is convex and compute its subgradient ∂f and its proximal $\text{prox}_f(x) = (I + \partial f)^{-1}(x)$.

Solución:

Para demostrar que la función

$$f(x) = \begin{cases} x^3 & x \geq 0 \ (1) \\ -x^3 & x < 0 \ (2) \end{cases}$$

Vamos a mirar la función f por las partes descritas:

- (1). Para comprobar la convexidad, podemos usar la **desigualdad de Jensen**. Para ello, elegimos dos puntos:

  $$x_1 = 0, x_2 = x \ \lambda_1 = 1 - \lambda, \quad \lambda_2 = \lambda \ \lambda \in [0, 1]$$

  Aplicamos la desigualdad de Jensen:

  $$\begin{aligned} f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \implies \\ (\lambda \cdot x)^3 \leq (1 - \lambda) \cdot 0^3 + \lambda \cdot x^3 \implies \\ \lambda^3 \cdot x^3 \leq \lambda \cdot x^3 \implies \lambda^2 \leq 1 \end{aligned}$$

  Dado que $\lambda \in [0, 1]$, se demuestra que (1) es convexa en el intervalo $x \geq 0$

- (2). Aplicaremos la **desigualdad de Jensen**. Elegimos dos puntos:

$$x_1 = 0, x_2 = x \quad \lambda_1 = 1 - \lambda, \lambda_2 = \lambda \quad \lambda \in [0,1]$$

  Aplicamos la desigualdad:

$$\begin{aligned} -\left(\lambda \cdot x\right) \geq (1-\lambda) \cdot 0^3 + \lambda \cdot \left(-x\right)^3 \implies \\ -\lambda^3 \cdot x^3 \geq -\lambda^3 \cdot x^3 \implies \lambda^2 \leq 1 \end{aligned}$$

La razón para usar $\leq$ en (1) y $\geq$ en (2) es porque (2) es una función cóncava en $\mathbf{R}$, pero se ha demostrado que en convexa en $x < 0$.

Como (1) y (2) se ha demostrado que con convexas, siguiendo la demostración vista en el ejercicio 2, la función f, que es la intersección de dos funciones, es convexa.