Convex optimization
Exercises

Luis Antonio Ortega Andrés
Antonio Coín Castro

March 22, 2021

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

Solution. Given that S is open, we got that $\forall x \in S\ \exists \delta > 0\ :\ B(x,\delta) \subset S$. By definition $S^c \subset \bar{S^c}$, where

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

Let $x \in \bar{S^c}$, then $\forall \delta > 0$,

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

Therefore, we have shown that $\bar{S^c} \subset S^c \implies \bar{S^c} = S^c$.

Exercise 2. *If S_1, s_2 are convex subsets, prove that $S_1 \cup S_2, S_1 + S_2$ and $S_1 - S_2$ are also convex sets.*

Solution. Let us begin with $S_1 \cup S_2$, let $x, x' \in S_1 \cup S_2$ and $\lambda \in (0,1)$. We got that

$$x, x' \in S_1 \cup S_2 \implies \lambda x + (1-\lambda)x' \in S_1 \text{ and } \lambda x + (1-\lambda)x' \in S_2$$

Therefore, $\lambda x + (1-\lambda)x' \in S_1 \cup S_2$, meaning the set is convex.

Consider now $x, x' \in S_1 + S_2$, we can decompose them as $x = x_1 + x_2$ and $x' = x'_1 + x'_2$, such that

$$\lambda x + (1-\lambda)x' = \lambda x_1 + (1 - \lambda x_2) + \lambda x'_1 + (1 - \lambda x'_2) \in S_1 + S_2.$$

Lastly, consider the set $S_1 - S_2$, using the same reasoning used in the previous set, it is convex.

Exercise 3. *If $f : S \to \mathbb{R}$ is a convex function on the convex set S, the set $\{x :\ x \text{ is a minimum of} f\}$ is a convex set.*

Solution. Let $A = \{x \in S :\ x \text{ is a minimum of f}\}$ and consider $x, x' \in A$, $\lambda \in (0,1)$, since f is convex, there is only one minimal value, ie, $f(x) = f(x') = c$. Using this, $x \in A \iff f(x) = c$ and $\nexists x \in S$ such that $f(x) < c$. Therefore

$$f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda)f(x') = f(x) \implies \lambda x + (1-\lambda)x' \in A.$$

As a result A is a convex subset of S.

Exercise 4. *Let $f : S \subset \mathbb{R}^d \to \mathbb{R}$ be a convex function on the convex set S and we extend it to an $\hat{f} : \mathbb{R}^d \to \mathbb{R}$ as*

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in S, \\ +\infty & \text{if } x \notin S \end{cases}$$

Show that $\hat{f}$ is a convex function on $\mathbb{R}^D$.

Solution: Let $x, x' \in \mathbb{R}^d$ and $\lambda \in (0,1)$,

- If $x, x' \in S$, $\hat{f}(\lambda x + (1-\lambda)x') \leq \lambda \hat{f}(x) + (1-\lambda)\hat{f}(x') = f(x)$.
- If $x \notin S$, $\lambda \hat{f}(x) + (1-\lambda)\hat{f}(x) = +\infty$ and the save happens for $x' \notin S$.

Exercise 5. *If Q is a symmetric, positive definite $d \times d$ matrix, show that $f(x) = x^T Q x,\ x \in \mathbb{R}^d$, is a convex function.*

Solution. Using Taylor's series, we know that $f(x+h) = f(x) + \nabla f(x) h + o(h)$, then

$$\begin{aligned} f(x+h) &= (x+h)^T Q (x+h) = x^T Q x + h^T Q x + x^T Q h + h^T Q h \\ &= f(x) + x^T (Q + Q^T) h + h^T Q h \end{aligned}$$

Where $\|h^T Q h\| \leq \|Q\| \|h\|^2 = o(h)$. Therefore $\nabla f(x) = x^T (Q + Q^T) = 2x^T Q$. Reusing the same argument,

$$\nabla f(x+h) = 2x^T 2 + 2h^T Q = f(x) + 2h^T Q \implies H(f)(x) = 2Q > 0.$$

We know use that f is convex given that its Hessian its semi-definite positive.

Exercise 6. *Given a quadratic form $q(w) = w^T Q w + bw + c$, with Q a symmetric $d \times d$ matrix, w, b $d \times 1$ vectors and c a real number, derive its gradient and Hessian.*

Solution. From the expanded form

$$q(w) = \sum_{i=1}^{d} \sum_{j=1}^{d} Q_{i,j} w_i w_j + \sum_{i=1}^{d} b_i w_i + c$$

we can take partial derivatives as

$$\frac{\partial q}{\partial w_i} = \sum_{j=1}^{d} 2 Q_{i,j} w_j + b_i \quad \text{and} \quad \frac{\partial^2 q}{\partial w_i w_j} = 2 Q_{i,j}$$

As a result

$$\nabla q(w) = \sum_i \frac{\partial q}{\partial w_i} = \sum_{i=1}^{d} \sum_{j=1}^{d} Q_{i,j} w_j + b_i = Qw + b,$$

and

$$H(q) = \begin{pmatrix} \frac{\partial^2 q}{\partial w_1 w_1} & \cdots & \frac{\partial^2 q}{\partial w_1 w_d} \\ \vdots & & \vdots \\ \frac{\partial^2 q}{\partial w_d w_1} & \cdots & \frac{\partial^2 q}{\partial w_d w_d} \end{pmatrix} = \begin{pmatrix} 2Q_{1,1} & \cdots & 2Q_{1,d} \\ \vdots & & \vdots \\ 2Q_{d,1} & \cdots & 2Q_{d,d} \end{pmatrix} = 2Q$$

Exercise 7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function and assume that $epi(f) \subset \mathbb{R}^d \times \mathbb{R}$ is convex. Prove that then f is convex.*

Solution. Knowing that

$$epi(f) = \{(x,t) \ : \ t \geq f(x)\}$$

is the graph above f. Let $a = (x, f(x))$ and $b = (x', f(x'))$ both in $epi(f)$. Then

$$\lambda a + (1-\lambda)b \in epi(f) \implies (\lambda x + (1-\lambda)x', \lambda f(x) + (1-\lambda)f(x')) \in epi(f)$$

Therefore, $f(\lambda x + (1-\lambda)x') \leq \lambda f(x) + (1-\lambda)f(x')) \implies f$ is convex.

Exercise 8. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Prove that $epi(f)$ is a closed set and that $(x, f(x)) \in \partial epi(f)$.*

Solution. Let us show that $epi(f)$ is closed by showing that $epi(f) = e\bar{pi}(f)$.

Let $(x,t) \in e\bar{pi}(f)$, then $epi(f) \cap B((x,t),\delta) \neq \emptyset \ \forall \delta > 0$. Consider that $(x,t) \notin epi(f)$, then $t < f(x)$ but given a fixed $\delta > 0$ $epi(f) \cap B((x,t),\delta) \neq \emptyset$. Consider the closed and bounded set $epi(f) \cap B((\bar{x,t}),\delta)$.

Given that is closed, bounded and non empty, there must exist a minimum distance α from (x,t) to that set, such that for $\delta < \alpha$, $epi(f) \cap B((\bar{x,t}),\delta) = \emptyset$, which is not possible. Therefore $(x,t) \in epi(f)$.

Exercise 9. *Prove that if f is strictly convex, it has a unique global minimum.*

Solution. Let x_1 and x_2 be two local minimum of f such that

$$f(x_1) \leq f(x_2), \quad x_1 \neq x_2.$$

Given that f is strictly convex,

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2), \quad \lambda \in (0,1).$$

Since $\lambda > 0$, $\lambda f(x_1) \leq \lambda f(x_2)$, which implies that

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

Therefore, we have proved that

$$f(\lambda x_1 + (1-\lambda)x_2) < f(x_2), \quad \forall \lambda \in (0,1)$$

However, if x_2 is a local minima, there must exists a neighborhood where every value is higher that $f(x_2)$ but adjusting the above λ we might get as close as we want to x_2. For this reason, the initial assumption $x_1 \neq x_2$ is false.

Exercise 10. *Let $f, g : S \subset \mathbb{R}^d \to \mathbb{R}$ be two convex functions on the convex set S . Prove that, as subsets, $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$.*

Solution.

$$\partial(f+g)(x) = \{c \in \mathbb{R}^d \ : \ (f' + g')(x) \geq f(x) + g(x) + c(x'-x) \ \forall x' \in S\}$$

Exercise 11. *Compute the proximal of $f(x) = 0$ and of $f(X) = \frac{1}{2}\|x\|^2$.*

Solution. Recall that $pox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\}$.

- $prox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\} = prox_f(x) = argmin_z\{\frac{1}{2}\|z - x\|^2\}$, but $\|z - x\|$ is minimized at $z = x$:
$$prox_f(x) = x$$

- $prox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\} = prox_f(x) = argmin_z\{\frac{1}{2}\|z\|^2 + \frac{1}{2}\|z - x\|^2\}$. But
$$\nabla_z \frac{1}{2}\|z\|^2 + \frac{1}{2}\|z - x\|^2 = z + (z - x) = 0 \implies z = \frac{1}{2}x.$$
Therefore $prox_f(x) = \frac{1}{2}x$.

Exercise 12. *Assume that f is convex. Prove that for any $\lambda > 0, \partial(\lambda f)(x) = \lambda \partial f(x)$ as subsets.*

Solution. By definition,

$$\begin{aligned}
\partial(\lambda f)(x) &= \{c \in \mathbb{R}^d \ : \ (\lambda f')(x) \geq \lambda f(x) + c(x' - x) \ \forall x' \in S\} \\
&= \{\lambda c \in \mathbb{R}^d \ : \ f(x) \geq f(x) + \lambda c(x' - x) \ \forall x' \in S\} \\
&= \lambda \partial f(x).
\end{aligned}$$

Exercise 13. *Compute the proximals of the hinge $f(x) = max\{0, -x\}$ and the ϵ-insensitive $max\{0, |x| - \epsilon\}$ loss function.*

Solution. Recall that $prox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\}$.

- $prox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\} = argmin_z\{\max\{0, -z\} + \frac{1}{2}\|z - x\|^2\}$.

  Using that
$$\begin{cases} argmin_z\{\frac{1}{2}\|z - x\|^2\} = x & \text{if } z \geq 0 \\ argmin_z\{-z + \frac{1}{2}\|z - x\|^2\} = x + 1 & \text{if } z < 0 \end{cases}$$
$$prox_f(x) = x$$

- $prox_f(x) = argmin_z\{f(z) + \frac{1}{2}\|z - x\|^2\} = argmin_z\{\max\{0, |z| - \epsilon\} + \frac{1}{2}\|z - x\|^2\}$

  Using that
$$\begin{cases} argmin_z\{\frac{1}{2}\|z - x\|^2\} = x & \text{if } |z| \leq \epsilon \\ argmin_z\{|z| - \epsilon + \frac{1}{2}\|z - x\|^2\} = & \text{if } ||z| > \epsilon \end{cases}$$
$$prox_f(x) = x$$

Exercise 14. *If $p_1, \ldots, p_K$ is a probability distribution, prove that its entropy $H(p_1, \ldots, p_K) = -\sum_{i=1}^K p_i \log p_i$ is a concave function. Show also that its maximum is $\log K$, attained when $p_i = \frac{1}{K} \forall i$.*

Solution. Given that $L(p_1, \ldots, p_k; \mu) = -\sum_{i=1}^k p_i \log p_i + \mu \left(\sum_{i=1}^k p_i - 1\right)$,

$$\frac{\partial L}{\partial p_i} = -\log p_i - 1 + \mu \ \forall i = 1, \ldots, k$$

Resolvemos $\begin{cases} \log p_i & = \mu - 1, \quad i = 1, \dots, k \\ \sum p_i & = 1 \end{cases} \implies p_i = e^{\mu-1} \; for all i$, that is, all p_i are equal

Exercise 15. *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 $y^p(wx^p + b) \geq 1$. *What are here the KKT conditions?*