Labs
Optimization for Machine Learning
Spring 2023

EPFL
School of Computer and Communication Sciences
Martin Jaggi & Nicolas Flammarion
github.com/epfml/OptML_course

Problem Set 1 – Solutions (Convexity, Python Setup)

Convexity

Exercise 2. *Prove Jensen's inequality (Lemma 1.13)!*

Solution: For $m = 1$, there is nothing to prove, and for $m = 2$, the statement holds by convexity of f. For $m > 2$, we proceed by induction. If $\lambda_m = 1$ (and hence all other λ_i are zero), the statement is trivial. Otherwise, let $\mathbf{x} = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i$ and define

$$\mathbf{y} = \sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} \mathbf{x}_i.$$

Thus we have $\mathbf{x} = (1-\lambda_m)\mathbf{y} + \lambda_m \mathbf{x}_m$. Also observe that $\sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} = 1$. By convexity and Jensens's inequality that we inductively assume to hold for $m-1$ terms, we get

$$\begin{aligned}
f(\mathbf{x}) &= f((1-\lambda_m)\mathbf{y} + \lambda_m \mathbf{x}_m) \\
&\leq (1-\lambda_m) f\left(\sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} \mathbf{x}_i\right) + \lambda_m f(\mathbf{x}_m) \\
&\leq (1-\lambda_m)\left(\sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} f(\mathbf{x}_i)\right) + \lambda_m f(\mathbf{x}_m) = \sum_{i=1}^{m} \lambda_i f(\mathbf{x}_i).
\end{aligned}$$

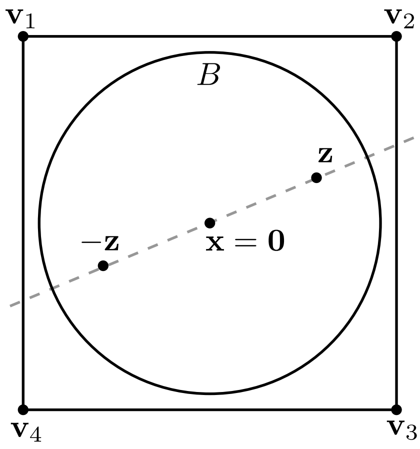
Exercise 3. *Prove that a convex function (with $\mathbf{dom}(f)$ open) is continuous (Lemma 1.14)!*

Hint: *First prove that a convex function f is bounded on any cube $C = [l_1, u_1] \times [l_2, u_2] \times \cdots \times [l_d, u_d] \subseteq \mathbf{dom}(f)$, with the maximum value occurring on some corner of the cube (a point $\mathbf{z}$ such that $z_i \in \{l_i, u_i\}$ for all i). Then use this fact to show that—given $\mathbf{x} \in \mathbf{dom}(f)$ and $\varepsilon > 0$—all $\mathbf{y}$ in a sufficiently small ball around $\mathbf{x}$ satisfy $|f(\mathbf{y}) - f(\mathbf{x})| < \varepsilon$.*

Solution: We will prove that, for any $\mathbf{x} \in \mathbf{dom}(f)$ the function f is continuous at point $\mathbf{x}$. For that we will prove:

1. There exists a ball $B \subset \mathbf{dom}(f)$ with center $\mathbf{x}$ with some radius $R > 0$ for which function difference is bounded, i.e. $|f(\mathbf{y}) - f(\mathbf{x})| \leq \gamma \ \forall \mathbf{y} \in B$ for some finite $\gamma \geq 0$.
2. If $\gamma > \varepsilon$, any point $\mathbf{y}$ in the smaller ball B' with center $\mathbf{x}$ with radius $\frac{R\varepsilon}{\gamma}$ satisfy $|f(\mathbf{y}) - f(\mathbf{x})| \leq \varepsilon$, so f is continuous at $\mathbf{x}$.

1. Existence of B



Assume without loss of generality that $\mathbf{x} = 0$ and $f(\mathbf{x}) = 0$. Now $f(\mathbf{y}) = f(\mathbf{y}) - f(\mathbf{x})$ and $\|y\| = \|y - x\|$.

Since the domain of f is open, there exists a cube with center $\mathbf{x} = \mathbf{0}$ that lies inside the domain. Because a cube is a convex set, any point $\mathbf{p}$ inside it can be written as a convex sum of the cube's 2^d vertices $\mathbf{v}_i$: $\mathbf{p} = \sum_{i=1}^{2^d} \lambda_i \mathbf{v}_i$, where $\lambda_i \geq 0 \; \forall i$ and $\sum_{i=1}^{2^d} \lambda_i = 1$. Due to convexity of f,

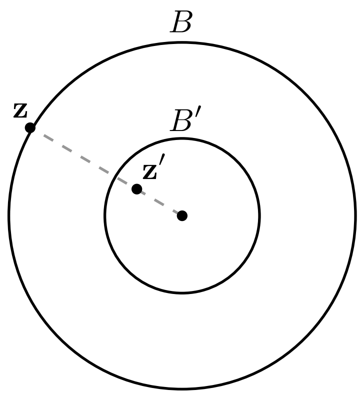
$$f(\mathbf{p}) \leq \sum_{i=1}^{2^d} \lambda_i f(\mathbf{v}_i) \leq \sum_{i=1}^{2^d} \lambda_i \max_i f(\mathbf{v}_i) = \max_i f(\mathbf{v}_i).$$

Because a cube has a finite number of vertices, this maximum exists, and the value of f inside the cube is bounded.

There exists a ball B with center $\mathbf{x}$ inside the cube with some radius R. Because the ball is a subset of the cube, f is bounded from above in the ball as well: $f(\mathbf{y}) \leq (\gamma := \max_i f(\mathbf{v}_i))$ for all $\mathbf{y} \in B$.

We will now show that f inside the ball is also bounded from below to finish this part of the proof. Consider any point $\mathbf{z} \in B$. By symmetry, $-\mathbf{z} \in B$ as well. Because the midpoint $\frac{1}{2}(\mathbf{z} + -\mathbf{z}) = \mathbf{0}$ is a convex combination of these two points, $0 = f(\mathbf{0}) \leq \frac{1}{2} f(\mathbf{z}) + \frac{1}{2} f(-\mathbf{z})$, or $f(\mathbf{z}) \geq -f(-\mathbf{z})$. This turns the upper bound $f(-\mathbf{z}) \leq \gamma$ into a lower bound $f(\mathbf{z}) \geq -\gamma$ for all $\mathbf{z} \in B$.

2. Shrinking of the ball



Again, assume without loss of generality that $\mathbf{x} = 0$ and $f(\mathbf{x}) = 0$. We use the first part of the proof to construct a ball B around the origin with radius R and $|f(\mathbf{y})| \leq \gamma$ for all $\mathbf{y} \in B$ and some $\gamma > 0$.

Consider the smaller ball B' around the origin with radius $r = \frac{R\varepsilon}{\gamma}$. We will use convexity to show that $|f(\mathbf{z}')| \leq \varepsilon$ for all $\mathbf{z}' \in B'$. Any point $\mathbf{z}' \in B'$ can be written as $\lambda \mathbf{z}$, where $\mathbf{z}$ is a point on the perimeter of the big ball B. The scale factor $\lambda \leq \frac{r}{R} = \frac{\varepsilon}{\gamma}$. Note that $0 \leq \lambda < 1$, so

$$f(\mathbf{z}') = f(\lambda \mathbf{z} + (1 - \lambda)\mathbf{0}) \leq \lambda f(\mathbf{z}) \leq \frac{\varepsilon}{\gamma} f(\mathbf{z}) \leq \varepsilon.$$

This is an upper bound $f(\mathbf{z}') \leq \varepsilon$ for $\mathbf{z}' \in B'$. To finish the proof, we just need to get a lower bound $f(\mathbf{z}') \geq -\epsilon$ as well. In part 1 of the proof, we turned an upper bound γ on the large ball B into a lower bound $-\gamma$. We can use the same argumentation here on the smaller ball B' with the previously derived upper bound ε to finish the proof.

Exercise 4. *Prove that the function $d_{\mathbf{y}} : \mathbb{R}^d \to \mathbb{R}$, $\mathbf{x} \mapsto \|\mathbf{x} - \mathbf{y}\|^2$ is strictly convex for any $\mathbf{y} \in \mathbb{R}^d$. (Use Lemma 1.25.)*

Solution: By Lemma 1.25, it suffices to show that $\nabla^2 d_{\mathbf{y}}(\mathbf{x})$ is positive definite for every $\mathbf{x} \in \mathbb{R}^d$ with $\mathbf{x} \neq \mathbf{0}$. We compute

$$d_{\mathbf{y}}(\mathbf{x}) = (\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y}), \quad \nabla d_{\mathbf{y}}(\mathbf{x}) = 2(\mathbf{x} - \mathbf{y}), \quad \nabla^2 d_{\mathbf{y}}(\mathbf{x}) = 2I \succ 0,$$

where I denotes the identity matrix. The claim follows.

Exercise 5. *Prove Lemma 1.19. Can (ii) be generalized to show that for two convex functions f, g, the function $f \circ g$ is convex as well?*

Solution:

(i) For $f = \max_{i=1}^m f_i$, we compute

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \max_{i=1}^m f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\leq \max_{i=1}^m (\lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})) \\ &= \lambda f_j(\mathbf{x}) + (1-\lambda) f_j(\mathbf{y}) \quad \text{(for some } j\text{)} \\ &\leq \lambda \max_{i=1}^m f_i(\mathbf{x}) + (1-\lambda) \max_{i=1}^m f_i(\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \end{aligned}$$

For $f = \sum_{i=1}^m \lambda_i f_i$, we compute

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \sum_{i=1}^m \lambda_i f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\leq \sum_{i=1}^m \lambda_i \left(\lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})\right) \\ &= \lambda \cdot \underbrace{\sum_{i=1}^m \lambda_i f_i(\mathbf{x})}_{f(\mathbf{x})} + (1-\lambda) \cdot \underbrace{\sum_{i=1}^m \lambda_i f_i(\mathbf{y})}_{f(\mathbf{y})}, \end{aligned}$$

where the inequality makes use of convexity of the individual f_i and of the fact that the λ_i are non-negative.

(ii) Let $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f \circ g)$ and $\lambda \in [0, 1]$ be arbitrary. We simply compute

$$\begin{aligned} (f \circ g)(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= f(A(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + \mathbf{b}) \\ &= f\big(\lambda \cdot (A\mathbf{x} + \mathbf{b}) + (1-\lambda) \cdot (A\mathbf{y} + \mathbf{b})\big) \\ &\leq \lambda \cdot \underbrace{f(A\mathbf{x} + \mathbf{b})}_{(f \circ g)(\mathbf{x})} + (1-\lambda) \cdot \underbrace{f(A\mathbf{y} + \mathbf{b})}_{(f \circ g)(\mathbf{y})}, \end{aligned}$$

where the inequality makes use of convexity of f and of the fact that both $g(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ and $g(\mathbf{y}) = A\mathbf{y} + \mathbf{b}$ are in the domain of f.

If two functions f and g are both convex, then their composition $f \circ g$ is not necessarily also convex. Consider for example convex functions $f(x) = x^2$ and $g(x) = x^2 - 1$. Then, the composition

$$(f \circ g)(x) = x^4 - 2x^2 + 1$$

satisfies $(f \circ g)(-1) = (f \circ g)(1) = 0$ and $(f \circ g)(0) = 1$, which is a clear violation of convexity.

Exercise 8. *Prove that the function* $f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^{d} |x_i|$ *(ℓ_1-norm) is convex!*

Solution: It suffices to prove that $f_i(\mathbf{x}) = |x_i|$ is convex and then use Lemma 1.19. Equivalently, that $f(x) = |x|$ is convex. For $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$, we compute

$$
\begin{aligned}
f(\lambda x + (1-\lambda)y) &= |\lambda x + (1-\lambda)y| \\
&\leq |\lambda x| + |(1-\lambda)y| \quad \text{(triangle inequality)} \\
&= |\lambda||x| + |(1-\lambda)||y| \\
&= \lambda|x| + (1-\lambda)|y| \\
&= \lambda f(x) + (1-\lambda)f(y).
\end{aligned}
$$

Exercise 10. *A seminorm is a function* $f : \mathbb{R}^d \to \mathbb{R}$ *satisfying the following two properties for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ *and all* $\lambda \in \mathbb{R}$.

(i) $f(\lambda \mathbf{x}) = |\lambda| f(\mathbf{x})$,

(ii) $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ *(triangle inequality).*

Prove that every seminorm is convex!

Solution: This just generalizes the previous exercise and shows what is actually going on. For $\lambda \in [0, 1]$ we get

$$
\begin{aligned}
f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) &\leq f(\lambda \mathbf{x}) + f((1-\lambda)\mathbf{y}) \quad \text{(triangle inequality)} \\
&= |\lambda| f(\mathbf{x}) + |(1-\lambda)| f(\mathbf{y}) \\
&= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).
\end{aligned}
$$

Getting Started with Python

Many exercises in this course use Python notebooks. We recommend running these notebooks in the cloud using Google Colab. This way, you do not have to install anything, and you can even get a free GPU. If you prefer to work locally, follow the `python_setup_tutorial.md` provided on our GitHub repository.

The first practical exercise is a primer on NumPy, a scientific computing library for Python. You can open the corresponding notebook in Colab with this link:

colab.research.google.com/github/epfml/OptML_course/blob/master/labs/ex01/npprimer.ipynb

For computational efficiency, avoid `for`-loops in favor of NumPy's built-in commands. These commands are vectorized and thoroughly optimized and bring the performance of numerical Python code (like for Matlab) closer to lower-level languages like C.