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

Problem Set 1 – Solutions (Convexity, Python Setup)

Convexity

Exercise 1. *Prove Jensen's inequality (Lemma 1.5)!*

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 2. *Prove that a convex function (with $\mathbf{dom}(f)$ open) is continuous (Lemma 1.6)!*

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: For a cube $C \subseteq \mathbf{dom}(f)$, let $V = \{v_1, \ldots, v_{2^d}\}$ be the set consisting of the 2^d vertices of C, and let $x \in C$. Because C is convex, we know that there exist $\lambda_1, \ldots, \lambda_{2^d} \geq 0$ with $\sum_{i=1}^{2^d} \lambda_i = 1$ such that

$$x = \sum_{i=1}^{2^d} \lambda_i v_i.$$

Thus, by Jensen's inequality we get that

$$f(x) \leq \sum_{i=1}^{2^d} \lambda_i f(v_i) \leq \sum_{i=1}^{2^d} \lambda_i \max_{i=1}^{2^d}\{f(v_i)\} \leq \max_{i=1}^{2^d}\{f(v_i)\}.$$

That is, the maximum value of the function is attained at some corner of C. Therefore, we can assume from now on that f is bounded in any cube $C \subseteq \mathbf{dom}(f)$.

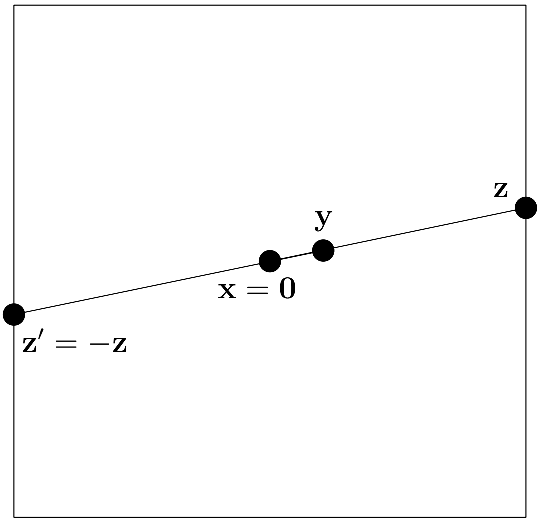
Now we can move on to prove the continuity of f. Let $\mathbf{x} \in \mathbf{dom}(f)$ be fixed, and let $C \subseteq \mathbf{dom}(f)$ be some axis-parallel cube of dimension d with $\mathbf{x}$ as its center (such a cube exists as $\mathbf{dom}(f)$ was assumed to be open). By changing the coordinates and shifting the value of f, we can assume without loss of generality that $\mathbf{x}$ coincides with the origin ($\mathbf{x} = \mathbf{0}$), and that $f(\mathbf{x}) = 0$. Moreover, since f is bounded in C, we can also assume without loss of generality that $\max\{f(\mathbf{z}) : \mathbf{z} \in C\} = 1$. Let $\mathbf{y} \in C$ such that $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y}\| \leq \delta$, for some $\delta > 0$ to be defined later. We want to prove that for sufficiently small δ, $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$. We now define

$$\lambda = \max\{\lambda' \in \mathbb{R}^+ : \lambda' \mathbf{y} \in C\} > 1,$$

and its corresponding point on the boundary of C

$$\mathbf{z} = \lambda\mathbf{y} = \lambda\mathbf{y} + (1-\lambda)\mathbf{x}.$$

Let also $\mathbf{z}' = -\mathbf{z}$.



Thus, we can express $\mathbf{y}$ as a convex combination of $\mathbf{z}$ and $\mathbf{x}$, and we can also express $\mathbf{x} = \mathbf{0}$ as a convex combination of $\mathbf{y}$ and $\mathbf{z}'$.

$$\begin{aligned}
\mathbf{y} &= \frac{1}{\lambda}\mathbf{z} + \frac{\lambda-1}{\lambda}\mathbf{x}, \\
\mathbf{x} = \mathbf{0} &= \mathbf{y} + \frac{1}{\lambda+1}(\mathbf{z}' - \mathbf{y}) = \frac{\lambda}{\lambda+1}\mathbf{y} + \frac{1}{\lambda+1}\mathbf{z}'.
\end{aligned}$$

Hence, by convexity, and the facts that $f(\mathbf{w}) \leq 1$ for all $\mathbf{w} \in C$ and $f(\mathbf{x}) = 0$, we get that

$$\begin{aligned}
f(\mathbf{y}) &\leq \frac{1}{\lambda}f(\mathbf{z}) + \frac{\lambda-1}{\lambda}f(\mathbf{x}) \leq \frac{1}{\lambda}, \\
0 &\leq \frac{\lambda}{\lambda+1}f(\mathbf{y}) + \frac{1}{\lambda+1}f(\mathbf{z}') \leq \frac{\lambda}{\lambda+1}f(\mathbf{y}) + \frac{1}{\lambda+1}.
\end{aligned}$$

Rearranging the latter inequality, we get that:

$$f(\mathbf{y}) \geq -\frac{1}{\lambda}.$$

Thus, to get $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$, we need to ensure that $\lambda > 1/\varepsilon$. Let r denote the distance of $\mathbf{x}$ to the boundary of C. Because $\mathbf{z}$ lies on the boundary of C, we get that

$$r \leq \|\mathbf{z}\| = \|\lambda\mathbf{y}\| = \lambda\|\mathbf{y}\|,$$

Hence, if $\delta < r\varepsilon$, then $\|\mathbf{y}\| < r\varepsilon$ and hence, $\lambda > 1/\varepsilon$ and $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$, so continuity holds.

Exercise 3. *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 1.19.)*

Solution: By Lemma 1.19, 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 4.

Solution:

(i) Let $\mathbf{x}, \mathbf{y} \in \mathbf{dom}(f)$ and $\lambda \in [0,1]$ be arbitrary. We simply 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 7. *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 1.13. 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 8. *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

Follow the Python setup tutorial `python_setup_tutorial.md` provided on our `github` repository here:

github.com/epfml/OptML_course/tree/master/labs/ex01/

After you are set up, clone the repository.

To get familiar with vector and matrix operations using NumPy arrays, you can go through the `numpy_primer.ipynb` notebook in the folder `/labs/ex01`. For computational efficiency, explicit `for`-loops should be avoided 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 e.g. Matlab) closer to lower-level languages like C.