Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Virginia Estellers
Exercises: Emanuel Laude
Winter Semester 2017/18

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 2

Room: 02.09.023
Friday, 10.11.2017, 09:15-11:00
Submission deadline: Monday, 06.11.2017, 10:15, Room 02.09.023

Theory: Convex Sets and Functions (24+8 Points)

Exercise 1 (8 Points)**.** Let $n \in \mathbb{N}$. Show that the following two statements are equivalent:

- $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex,
- $f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i)$, for $x_i \in \mathbb{E}$, $\alpha_i \in [0,1]$, $\sum_{i=1}^n \alpha_i = 1$, $n \geq 1$.

Solution. "$\Leftarrow$": For $n = 2$ it is precisely the definition of convexity.
"$\Rightarrow$": We prove this statement using induction. The cases $n = 1$ and $n = 2$ are trivial. Now assume the inequality holds for some $n \geq 1$. Without loss of generality we can assume $\alpha_{n+1} \neq 0$, since the case $\alpha_{n+1} = 0$ follows directly from the assumption.

$$
\begin{aligned}
f\left(\sum_{i=1}^{n+1} \alpha_i x_i\right) &= f\left(\sum_{i=1}^{n} \alpha_i x_i + \alpha_{n+1} x_{n+1}\right) \\
&= f\left((1-\alpha_{n+1}) \sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} x_i + \alpha_{n+1} x_{n+1}\right) \\
&\leq (1-\alpha_{n+1}) f\left(\sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} x_i\right) + \alpha_{n+1} f(x_{n+1}) \qquad (1) \\
&\leq (1-\alpha_{n+1}) \sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} f(x_i) + \alpha_{n+1} f(x_{n+1}) \\
&= \sum_{i=1}^{n} \alpha_i f(x_i) + \alpha_{n+1} f(x_{n+1}) = \sum_{i=1}^{n+1} \alpha_i f(x_i).
\end{aligned}
$$

Hence it also holds for $n+1$ and by the principle of induction we are finished.

Exercise 2 (8 Points)**.** Compute the subdifferential of the following functions:

- $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \|x\|_2$.

- $f : \mathbb{R}^n \to \overline{\mathbb{R}}, f(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise,} \end{cases}$
  for a closed convex set $C \subset \mathbb{R}^n$.

Solution.

- For $x \neq 0$ f is differentiable and we have $\partial f(x) = \left\{\frac{x}{\|x\|_2}\right\}$. For $p \in \mathbb{R}^n$ with $\|p\|_2 \leq 1$ we have $f(y) - f(x) = \|y\|_2 \geq \|y\|_2 \cdot \|p\|_2 \geq \langle y, p\rangle$. Therefore $p \in \partial f(0)$. For $\|p\|_2 > 1$ and $y = p$ we have

  $$f(p) - f(0) = \|p\|_2 < \|p\|_2^2 = \langle p, p\rangle.$$

  Together this yields

  $$\partial \|x\|_2 = \begin{cases} \frac{x}{\|x\|_2} & \text{if } x \neq 0 \\ B_1(0) & \text{if } x = 0. \end{cases}$$

- For $f(X) := \|X\|_{2,1} = \sum_{i=1}^m \|x^i\|_2$ we can again apply the sum rule of the subdifferential. Together with part 2 of the exercise we get

  $$\partial f(X) := \{P \in \mathbb{R}^{n\times m} : p^i \in \partial \|x^i\|_2\}.$$

- Take a point $x \in \operatorname{dom} f$. Then the subgradients $g \in \partial f(x)$ fulfill

  $$\langle g, y - x\rangle \leq 0, \forall y \in C \;\Leftrightarrow\; g \in N_C(x).$$

  Hence $\partial f(x) = N_c(x)$.

Definition (Convex Hull)**.** The convex hull $\operatorname{conv}(S)$ of a finite set of points $S \subset \mathbb{R}^n$ is defined as

$$\operatorname{conv}(S) := \left\{ \sum_{i=1}^{|S|} a_i x_i : x_i \in S, \sum_{i=1}^{|S|} a_i = 1, a_i \geq 0 \right\}$$

Exercise 3 (8 Points)**.** Prove the following statement: Let $n \in \mathbb{N}$ and let $A \subset \mathbb{R}^n$ contain $n + 2$ elements: $|A| = n + 2$. Then there exists a partition of A into two disjoint sets A_1, A_2

$$A = A_1 \dot\cup A_2,$$

(meaning that $A_1 \cap A_2 = \emptyset$) so that the convex hulls of A_1 and A_2 intersect:

$$\operatorname{conv}(A_1) \cap \operatorname{conv}(A_2) \neq \emptyset.$$

You may use the following hint. Don't forget to prove the hint!

Hint: Let $x_1, \ldots, x_{n+2} \in \mathbb{R}^n$. Then the set $\{x_1 - x_{n+2}, \ldots, x_{n+1} - x_{n+2}\}$ is linearly dependent and there exist multipliers $a_1, \ldots, a_{n+2}$, not all of which are zero, so that

$$\sum_{i=1}^{n+2} a_i x_i = 0, \quad \sum_{i=1}^{n+2} a_i = 0.$$

The desired partition is formed via all points corresponding with $a_i \geq 0$ and all points with $a_i < 0$.

Solution. Let $A := \{x_1, x_2, \dots, x_{n+2}\} \subset \mathbb{R}^n$. Since $n+1$ vectors in $\mathbb{R}^n$ are always linearly dependent there exist scalars $a_1, \dots, a_{n+1}$, not all of which are zero so that

$$\sum_{i=1}^{n+1} a_i(x_i - x_{n+2}) = \sum_{i=1}^{n+1} a_i x_i + \underbrace{\left(-\sum_{i=1}^{n+1} a_i\right)}_{=:a_{n+2}} x_{n+2} = 0.$$

Then, by construction $\sum_{i=1}^{n+2} a_i = 0$. Define $A_1 := \{x_i : a_i > 0\}$ and $A_2 := \{x_j : a_j \leq 0\}$. Clearly, $A = A_1 \dot\cup A_2$ forms a partition and A_1, A_2 are both nonempty. Suppose A_2 was empty. Then $a_i > 0$ for all $1 \leq i \leq n+2$. But $a_{n+2} := -\sum_{i=1}^{n+1} a_i < 0$ contradicts this assumption (The same holds for A_1). We have that

$$0 = \sum_{\{i:a_i<0\}} a_i x_i + \sum_{\{j:a_j\geq 0\}} a_j x_j \iff \sum_{\{i:a_i<0\}} \underbrace{-a_i}_{\geq 0} x_i = \sum_{\{j:a_j\geq 0\}} a_j x_j,$$

and on the other hand

$$0 = \sum_{\{i:a_i<0\}} a_i + \sum_{\{j:a_j\geq 0\}} a_j \iff \sum_{\{i:a_i<0\}} -a_i = \sum_{\{j:a_j\geq 0\}} a_j =: w > 0.$$

Altogether this yields

$$\underbrace{\sum_{\{i:a_i<0\}} \frac{-a_i}{w} x_i}_{\in \mathrm{conv}(A_1)} = \underbrace{\sum_{\{j:a_j\geq 0\}} \frac{a_j}{w} x_j}_{\in \mathrm{conv}(A_2)},$$

which completes the proof. The theorem is called Radon's Theorem.

Exercise 4 (8 Bonus points)**.** Prove the following statement using induction over m: Let $K_1, \dots, K_m \subset \mathbb{R}^n$, $m \geq n+1$, be convex, such that for all $\mathcal{I} \subset \{1, \dots, m\}$ with $|\mathcal{I}| = n+1$ it holds that $\bigcap_{i\in\mathcal{I}} K_i \neq \emptyset$. Then $\bigcap_{i=1}^m K_i \neq \emptyset$.

Hint: Use exercise 3 above.

Solution. <u>Base case:</u> for $m = n+1$ the statement clearly holds.

<u>Inductive step:</u> $m \to m+1$. For any $\mathcal{I} \subset \{1, \dots, m+1\}$ with $|\mathcal{I}| = n+1$ assume that $\bigcap_{i\in\mathcal{I}} K_i \neq \emptyset$. Fix $j \in \{1, 2, \dots, m+1\}$. The assumption implies that for all $\mathcal{I}' \subset \{1, \dots, m+1\} \setminus \{j\}$ with $|\mathcal{I}'| = n+1$ it holds that $\bigcap_{i\in\mathcal{I}'} K_i \neq \emptyset$. We may now apply the induction hypothesis to the sets $K_1, \dots, K_{m+1}$ excluding K_j and the sets $\mathcal{I}'$ and conclude that for any $\mathcal{J} \subset \{1, \dots, m+1\}$ with $\mathcal{J} \neq \emptyset$:

$$x_j \in \bigcap_{i=1, i\neq j}^{m+1} K_i \subset \begin{cases} \bigcap_{i\in\mathcal{J}} K_i & \text{if } j \notin \mathcal{J} \\ \bigcap_{i\notin\mathcal{J}} K_i & \text{if } j \in \mathcal{J}. \end{cases}$$

Now, consider the partitions $A_1 := \{x_j : j \notin \mathcal{J}\}$, $A_2 := \{x_j : j \in \mathcal{J}\}$ of the set $A := \{x_1, x_2, \cdots x_{m+1}\}$ determined via $\mathcal{J}$. Since $m+1 \geq n+2$ we know from exercise 4 of the last sheet that there exists an $\mathcal{J}' \subset \{1, \dots, m+1\}$ (the proof can easily be adapted to the more general case $m+1 \geq n+2$) so that $\mathrm{conv}(A_1) \cap \mathrm{conv}(A_2) \neq \emptyset$.

Since the K_i are convex and the intersection of convex sets is convex we have that $\operatorname{conv}(A_1) \subset \bigcap_{i \in \mathcal{J}'} K_i$ and $\operatorname{conv}(A_2) \subset \bigcap_{i \notin \mathcal{J}'} K_i$. Overall we have that

$$\emptyset \neq \operatorname{conv}(A_1) \cap \operatorname{conv}(A_2) \subset \bigcap_{i \in \mathcal{J}'} K_i \cap \bigcap_{i \notin \mathcal{J}'} K_i = \bigcap_{i=1}^{m+1} K_i.$$

The theorem is called Helly's Theorem.