Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Zhenzhang Ye
Winter Semester 2019/20

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 1

Room: 02.09.023

Wednesday, 30.10.2019, 12:15-14:00

Submission deadline: Monday, 28.10.2019, 16:15, Room 02.09.023

Theory: Convex Sets (12+8 Points)

Exercise 1 (4 Points)**.** Let $\mathcal{C}$ be a family of convex sets in $\mathbb{R}^n$, $C_1, C_2 \in \mathcal{C}$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\lambda \in \mathbb{R}$. Prove convexity of the following sets:

- $\bigcap_{C\in\mathcal{C}} C$
- $P := \{x \in \mathbb{R}^n : Ax \leq b\}$
- $C_1 + C_2 := \{x + y : x \in C_1, y \in C_2\}$ (the Minkowski sum of C_1 and C_2)
- $\lambda C_1 := \{\lambda x : x \in C_1\}$ (the λ-dilatation of C_1).

Solution.

- Let $x_1, x_2 \in \bigcap_{C\in\mathcal{C}} C$. Then $x_1, x_2 \in C$ for all $C \in \mathcal{C}$. Since any C is convex, $\mu x_1 + (1-\mu)x_2 \in C$ for all $\mu \in [0,1]$ and $C \in \mathcal{C}$ and therefore $\mu x_1 + (1-\mu)x_2 \in \bigcap_{C\in\mathcal{C}} C$.
- Let $x_1, x_2 \in P$, which means that $Ax_1 \leq b$ and $Ax_2 \leq b$. Let $\mu \in [0,1]$. Then, $A(\mu x_1 + (1-\mu)x_2) = \mu A x_1 + (1-\mu)Ax_2 \leq \mu b + (1-\mu)b = b$. Therefore $\mu x_1 + (1-\mu)x_2 \in P$.
- Let $x, y \in C_1 + C_2$. Then there exist $x_1, y_1 \in C_1, x_2, y_2 \in C_2$ so that $x = x_1 + x_2$ and $y = y_1 + y_2$. Let $\mu \in [0,1]$. Then, since C_1, C_2 convex $\mu x + (1-\mu)y = \mu x_1 + \mu x_2 + (1-\mu)y_1 + (1-\mu)y_2 = \underbrace{\mu x_1 + (1-\mu)y_1}_{\in C_1} + \underbrace{\mu x_2 + (1-\mu)y_2}_{\in C_2} \in C_1 + C_2$.
- Let $x, y \in C_1$ and $\mu \in [0,1]$. Then, since C_1 convex, $\mu\lambda x + (1-\mu)\lambda y = \lambda\underbrace{(\mu x + (1-\mu)y)}_{\in C_1} \in \lambda C_1$.

Exercise 2 (4 Points)**.** Prove that if the set $C \subset \mathbb{R}^n$ is convex, then $\sum_{i=1}^N \lambda_i x_i \in C$ with $x_1, x_2, \ldots, x_N \in C$ and $0 \leq \lambda_1, \lambda_2, \ldots, \lambda_N \in \mathbb{R}$, $\sum_{i=1}^N \lambda_i = 1$.

Hint: Use induction to prove.

Solution. When N=2, it directly follows the definition of convex set.
Assume it holds for N. Now consider N+1 case:

$$\sum_{i=1}^{N+1} \lambda_i x_i = \sum_{i=1}^{N} \lambda_i x_i + \lambda_{N+1} x_{N+1}$$

If there exists a certain i such that $\lambda_i = 0$, it will be N case which is assumed to hold. Therefore, all $\lambda_i > 0$ and above equation turns into:

$$(1 - \lambda_{N+1}) \sum_{i=1}^{N} \frac{\lambda_i}{1 - \lambda_{N+1}} x_i + \lambda_{N+1} x_{N+1}$$

Using our assumption, $\sum_{i=1}^{N} \frac{\lambda_i}{1-\lambda_{N+1}} x_i$ is an element in C. Therefore, the convexity is proved.

Exercise 3 (4 Points)**.** Let $\emptyset \neq X \subset \mathbb{R}^n$. Prove the equivalence of the following statements:

- X is closed.
- Every convergent sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ attains its limit in X.

Solution. Let X be closed. By definition this means that the complement of X given as $X_C := \mathbb{R}^n \setminus X$ is open meaning that for all $x \in X_C$ there exists $\epsilon > 0$ s.t. the ball $B_\epsilon(x)$ is entirely contained in X_C:

$$B_\epsilon(x) \cap X = \emptyset.$$

Suppose that there exists a convergent sequence $X \supset \{x_n\}_{n\in\mathbb{N}} \to x$ with $x \notin X$. However, by definition of convergence for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ s.t.

$$X \ni x_n \in B_\epsilon(x)$$

for all $n \geq N$, which contradicts the assumption. Let conversely X not be closed (not the same as open). That means there exists $x \notin X$ s.t. for all $\epsilon > 0$ it holds that $B_\epsilon(x) \cap X \neq \emptyset$. This means that for all $\epsilon_n := \frac{1}{n} > 0$ there exists $x_n \in B_\epsilon(x) \cap X$. By construction we have a sequence $\{x_n\}_{n\in\mathbb{N}}$ converging to $x \notin X$ but with elements in X.

Exercise 4 (8 Points)**.** Some basic problems on calculus and linear algebra.

- Let $u \in \mathbb{R}^n$, compute the gradient of following function on u: $J(u) = \sqrt{u^\top A u}$, where $A \in \mathbb{R}^{n\times n}$ is full rank and $u \neq 0$.
- What happens if A is not full rank?

- Let $z \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, $f \in \mathbb{R}$ and ϵ is a positive real number, compute the gradient of following function on z:

$$R(z) = \frac{z}{f^2}\sqrt{f^2 \left\|Az\right\|^2 + \left\|-z - \mathbb{1}\langle x, Az\rangle\right\|^2 + \epsilon}$$

and $\mathbb{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$.

Solution.

- Using chain rule, we can get:

$$\nabla J(u) = \frac{(A + A^\top)u}{2\sqrt{u^\top A u}}$$

- We need to ensure that the denominator is not 0. Therefore, when A is not full rank, we need to ensure that u is not in null(A).

- Since the function is $\mathbb{R}^n \to \mathbb{R}^n$, we need to compute the Jacobian matrix. Applying chain rule, we can get:

$$J_{R(z)} = \text{diag}\left(\frac{1}{f^2}\sqrt{f^2 \left\|Az\right\|^2 + \left\|-z - \mathbb{1}\langle x, Az\rangle\right\|^2 + \epsilon}\right) + \frac{z}{f^2}\left(\frac{f^2 A^\top A z + (I + A^\top x \mathbb{1}^\top)^\top (z + \mathbb{1} x^\top A z)}{\sqrt{f^2 \left\|Az\right\|^2 + \left\|-z - \mathbb{1}\langle x, Az\rangle\right\|^2 + \epsilon}}\right)^\top$$

where I is the identity matrix.