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 7

Room: 02.09.023

Wednesday, 18.12.2019, 12:15-14:00

Submission deadline: Monday, 16.12.2019, 16:15, Room 02.09.023

Duality (12+6 Points)

Exercise 1 (6 Points)**.** Let $X, Y \in \mathbb{R}^{m\times n}$ be matrices and let $Y_i, X_i \in \mathbb{R}^m$ denote the i-th columns of X, Y. Then, the Frobenius scalar product is defined as follows:

$$\langle X, Y\rangle_F := \sum_{i=1}^{n} \langle X_i, Y_i\rangle, \tag{1}$$

where $\langle X_i, Y_i\rangle$ is the classical vector scalar product. For notational convenience we often omit the subscript F in $\langle \cdot, \cdot\rangle_F$. Compute the convex conjugates of the following functions:

1. $f_1 : \mathbb{R}^{m\times n} \to \mathbb{R} \cup \{\infty\}$ where $f_1(X) = \|X\|_{2,\infty} := \max_{1\leq i\leq n} \|X_i\|_2$.
2. $f_2 : \mathbb{R}^{m\times n} \to \mathbb{R} \cup \{\infty\}$ where

$$f_2(X) := \delta_{\|\cdot\|_{2,1}\leq 1}(X) = \begin{cases} 0 & \text{if } \|X\|_{2,1} := \sum_{i=1}^{n} \|X_i\|_2 \leq 1, \\ \infty & \text{otherwise.} \end{cases} \tag{2}$$

Solution. 1. Let $X \in \mathbb{R}^{m\times n}$, $\|X\|_{2,1} \leq 1$. We have any for $Y \in \mathbb{R}^{m\times n}$:

$$\begin{aligned} \langle X, Y\rangle_F &= \sum_{i=1}^{n} \langle X_i, Y_i\rangle \\ &\leq \sum_{i=1}^{n} |X_i| \cdot |Y_i| \\ &\leq \sum_{i=1}^{n} |X_i| \cdot \max_{1\leq j\leq n} |Y_j| \\ &= \|X\|_{2,1} \cdot \|Y\|_{2,\infty}. \end{aligned}$$

This implies that

$$\langle X, Y\rangle_F - \|Y\|_{2,\infty} \leq \|X\|_{2,1} \cdot \|Y\|_{2,\infty} - \|Y\|_{2,\infty} = (\|X\|_{2,1} - 1) \cdot \|Y\|_{2,\infty} \leq 0$$

Since $\langle X, 0\rangle_F - \|0\|_{2,\infty} = 0$ we get

$$f_1^*(X) = \sup_{Y\in\mathbb{R}^{m\times n}} \langle X, Y\rangle_F - \|Y\|_{2,\infty} = 0.$$

Now let $\|X\|_{2,1} > 1$. Define $Y \in \mathbb{R}^{m\times n}$ so that the i-th column Y_i of Y, $1 \le i \le n$ is given as $Y_i := \frac{X_i}{\|X_i\|_2}$, which implies $\|Y\|_{2,\infty} = 1$. We get

$$\langle X, Y\rangle_F = \sum_{i=1}^{n} \|X_i\|_2 = \|X\|_{2,1}.$$

For $\alpha > 0$ we get

$$\langle X, \alpha Y\rangle_F - \|\alpha Y\|_{2,\infty} = \alpha \underbrace{(\|X\|_{2,1} - 1)}_{>1}.$$

Therefore,

$$f_1^*(X) = \sup_{Y\in\mathbb{R}^{m\times n}} \langle X, Y\rangle_F - \|Y\|_{2,\infty} = \infty.$$

Altogether we obtain

$$f_1^*(X) = \delta_{\|\cdot\|_{2,1}\le 1}(X).$$

2. We have $f_2 = f_1^*$ and since f_1 is closed, proper and convex we have

$$f_2^* = f_1^{**} = f_1.$$

Exercise 2 (4 Points)**.** Assuming $J : \mathbb{R}^n \to \bar{\mathbb{R}}$, $\varepsilon > 0$, $c \in \mathbb{R}^n$, and J^* (i.e. the convex conjugate of J) are known, derive the expression of $(\langle c, \cdot\rangle + \varepsilon J(\cdot))^*$ in terms of J^*, ε, and c.

Solution.

$$\begin{aligned}\big(\langle c,\cdot\rangle + \varepsilon J(\cdot)\big)^*(p) &= \sup_{u\in\mathbb{R}^n} \langle u, p\rangle - \big(\langle c, u\rangle + \varepsilon J(u)\big)\\ &= \sup_{u\in\mathbb{R}^n} \langle u, p - c\rangle - \varepsilon J(u)\\ &= \varepsilon\big(\sup_{u\in\mathbb{R}^n} \left\langle u, \frac{p-c}{\varepsilon}\right\rangle - J(u)\big)\\ &= \varepsilon J^*\big(\frac{p-c}{\varepsilon}\big)\end{aligned}$$

Exercise 3 (8 Points)**.** Let $C \in \mathbb{R}^{m\times n}$, $\mu \in \mathbb{R}^m$, $\nu \in \mathbb{R}^n$, $\varepsilon > 0$ be given. Define $\mathbf{1}_m = (1, 1, ..., 1) \in \mathbb{R}^m$ and similarly for $\mathbf{1}_n \in \mathbb{R}^n$. Consider the "optimal mass transport" problem:

$$\min_X \; F(KX) + G(X),$$

where

$$F : (u, v) \in \mathbb{R}^m \times \mathbb{R}^n \mapsto \delta\{(u, v) = (\mu, \nu)\} \in \bar{\mathbb{R}},$$

$$G : X \in \mathbb{R}^{m\times n} \mapsto \sum_{i=1}^{m}\sum_{j=1}^{n} \Big(C_{ij}X_{ij} + \varepsilon X_{ij}(\log X_{ij} - 1) + \delta\{X_{ij} \ge 0\}\Big) \in \bar{\mathbb{R}},$$

$$K : X \in \mathbb{R}^{m\times n} \mapsto (X\mathbf{1}_n, X^\top\mathbf{1}_m) \in \mathbb{R}^m \times \mathbb{R}^n.$$

(1) Use the Fenchel-Rockafellar duality theorem to derive the dual formulation of the above problem. The formulae for the convex conjugates of F^* and G^* must be explicitly provided.
Hint: The adjoint of K (denoted by $K^\top$) can be derived as $K^\top : (u,v) \in \mathbb{R}^m \times \mathbb{R}^n \mapsto u\mathbf{1}_n^\top + \mathbf{1}_m v^\top \in \mathbb{R}^{m\times n}$;
(2) State the optimality conditions which involve both primal and dual variables. The formulae for all involved subdifferentials must be explicitly provided.

Solution. (1) We use $(p,q) \in \mathbb{R}^m \times \mathbb{R}^n$ to denote the dual variable. Firstly, we compute the convex conjugate of F.

$$\begin{aligned} F^*((p,q)) &= \sup\nolimits_{(u,v)} \langle p, u\rangle + \langle q, v\rangle - \delta\{(u,v) = (\mu,\nu)\} \\ &= \langle p, \mu\rangle + \langle q, \nu\rangle \end{aligned} \tag{3}$$

Denote $J(X) = \sum_{i=1}^m \sum_{j=1}^n X_{ij}(\log X_{ij} - 1) + \delta\{X_{ij} \geq 0\}$. We can write $G(X) = \langle C, X\rangle_{\mathrm{F}} + \epsilon J(X)$ with $\langle \cdot, \cdot\rangle_F$ is Frobenious product. By using the result from Q1, we get

$$G^*(Y) = \epsilon \sum_{i=1}^m \sum_{j=1}^n \exp(\frac{Y_{ij} - C_{ij}}{\epsilon}) \tag{4}$$

Thus, the dual formulation is

$$F^*((p,q)) + G^*(-K^\top(p,q)) = \langle p, \mu\rangle + \langle q, \nu\rangle + \epsilon \sum_{i=1}^m \sum_{j=1}^n \exp(\frac{-p_i - q_j - C_{ij}}{\epsilon}) \tag{5}$$

(2) The optimality condition for primal variable:

$$\begin{aligned} &KX \in \partial F^*((p,q)) \\ \Rightarrow &KX = (\mu,\nu) \\ \Rightarrow &\begin{cases} X\mathbf{1}_n = \mu \\ X^\top \mathbf{1}_m = \nu \end{cases} \end{aligned} \tag{6}$$

Denote a subset $Q := \{X \in \mathbb{R}^{m\times n} : X_{ij} \geq 0\}$ and $N_Q(X)$ as the normal cone of Q at X. The optimality condition for dual variable:

$$\begin{aligned} &-K^\top(p,q) \in \partial G(X) \\ \Rightarrow &-p\mathbf{1}_n^\top - \mathbf{1}_m q^\top \in C + \epsilon \log X_{ij} + N_Q(X) \\ \Rightarrow &-p\mathbf{1}_n^\top - \mathbf{1}_m q^\top - C - \epsilon \log X_{ij} \in N_Q(X) \\ \Rightarrow &\left\langle -p\mathbf{1}_n^\top - \mathbf{1}_m q^\top - C - \epsilon \log X_{ij}, Y - X \right\rangle \leq 0, \ \forall Y \in Q \\ \Rightarrow &p_i + q_j + C_{ij} + \epsilon \log X_{ij} = 0 \end{aligned} \tag{7}$$