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

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

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.

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.