Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Yuesong Shen, Zhenzhang Ye
Winter Semester 2018/19

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 3

Room: 01.09.014

Wednesday, 14.11.2018, 12:15-14:00

Submission deadline: Monday, 12.11.2018, 16:15, Room 01.09.014

Subdifferential (14+6 Points)

Exercise 1 (4 Points)**.** Let the convex function $J : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be differentiable at $u \in \text{int}(\text{dom}(J))$. Show that

$$\partial J(u) = \{\nabla J(u)\}.$$

Hint: Use the definition of the subdifferential and the directional derivative. For J being differentiable at the interior of its domain, some direction $v \in \mathbb{R}^n$ and some point $u \in \text{int}(\text{dom}(J))$ the directional derivative $\partial_v J$ of J is given as

$$\partial_v J(u) := \lim_{\epsilon \to 0} \frac{J(u + \epsilon v) - J(u)}{\epsilon} = \lim_{\epsilon \to 0} \frac{J(u) - J(u - \epsilon v)}{\epsilon} = \langle \nabla J(u), v \rangle.$$

Solution. Recall that the subdifferential $\partial J(u)$ of some convex J at $u \in \text{dom}(J)$ is given as

$$\{p \in \mathbb{R}^n : J(v) \geq J(u) + \langle p, v - u \rangle, \forall\, v \in \text{dom}(J)\}.$$

Since $u \in \text{int}(\text{dom}(J))$, we find that for all $v \in \mathbb{R}^n$, $u + \epsilon v \in \text{dom}(J)$ for ϵ small enough since the interior of a set is open. By the definition of the subdifferential, we have that if $p \in \partial J(u)$ then

$$J(u + \epsilon v) \geq J(u) + \epsilon \langle p, v \rangle, \quad J(u - \epsilon v) \geq J(u) - \epsilon \langle p, v \rangle,$$

for all $v \in \mathbb{R}^n$ and ϵ small enough. This implies that

$$\lim_{\epsilon \to 0} \frac{J(u + \epsilon v) - J(u)}{\epsilon} \geq \langle p, v \rangle, \quad \lim_{\epsilon \to 0} \frac{J(u) - J(u - \epsilon v)}{\epsilon} \leq \langle p, v \rangle,$$

which means (using the hint)

$$\langle \nabla J(u), v \rangle \geq \langle p, v \rangle, \quad \langle \nabla J(u), v \rangle \leq \langle p, v \rangle$$

or

$$\langle \nabla J(u) - p, v \rangle \geq 0, \quad \langle \nabla J(u) - p, v \rangle \leq 0$$

for all $v \in \mathbb{R}^n$. For the particular choice of $v := \nabla J(u) - p$ we have that

$$\langle \nabla J(u) - p, \nabla J(u) - p \rangle = \|\nabla J(u) - p\|_2^2 = 0$$

which means $p = \nabla J(u)$. Clearly, $\partial J(u)$ is non-empty (and bounded) since $u \in \text{int}(\text{dom}(x))$ implies $u \in \text{ri}(\text{dom}(x))$ (see Thm. Subdifferentiability). Together this concludes the proof.

Exercise 2 (6 Points)**.** Compute the subdifferential of norms in Euclidean space:

- Let $\|\cdot\|$ be a norm on an Euclidean space $\mathbb{E}$, and $\|\cdot\|_*$ its dual norm defined as

  $$\|p\|_* = \sup_{\|x\| \le 1} \langle p, x \rangle,$$

  prove that

  $$\partial \|\cdot\| (x) = \{p \in \mathbb{E} : \langle p, x \rangle = \|x\|, \|p\|_* \le 1\}. \tag{1}$$

  Hint: For $x \neq 0$, we have a generalized Cauchy-Schwarz inequality:

  $$\langle x, y \rangle = \|x\| \langle \frac{x}{\|x\|}, y \rangle \le \|x\| \cdot \sup_{\|z\| \le 1} \langle z, y \rangle = \|x\| \|y\|_*, \ \forall x, y \in \mathbb{E}. \tag{2}$$

- Using the result above, compute the subdifferential of the following functions:
  - $J : \mathbb{R}^n \to \mathbb{R}, J(u) = \|u\|_1$.
  - $J : \mathbb{R}^n \to \mathbb{R}, J(u) = \|u\|_2$.
  - $J : \mathbb{R}^n \to \mathbb{R}, J(u) = \|u\|_\infty$.

Solution.

- If $x = 0$ and assume p is one of the subdifferential, we have

  $$\begin{aligned} p \in \partial \|\cdot\| (0) &\Leftrightarrow \langle p, y \rangle \le \|y\|, \forall y \in \mathbb{E} \\ &\Leftrightarrow \frac{\langle p, y \rangle}{\|y\|} \le 1, \forall y \neq 0 \\ &\Leftrightarrow \sup_{y \neq 0} \frac{\langle p, y \rangle}{\|y\|} \le 1 \\ &\Leftrightarrow \sup_{\|y\| = 1} \langle p, y \rangle \le 1 \Leftrightarrow \|p\|_* \le 1 \end{aligned}$$

  For $x \neq 0$, let $p \in \mathbb{E}$ with $\langle p, x \rangle = \|x\|$, $\|p\|_* \le 1$. Then, using the generalized Cauchy-Schwarz inequality (Eq. 2), we have

  $$\langle p, y - x \rangle + \|x\| = \langle p, y \rangle - \langle p, x \rangle + \|x\| = \langle p, y \rangle \le \|y\| \|p\|_* \le \|y\|, \forall y \in \mathbb{E}.$$

  Hence $p \in \partial \|\cdot\| (x)$. Conversely take $p \in \partial \|\cdot\| (x)$. Then we have

  $$\begin{aligned} &\langle p, y - x \rangle + \|x\| \le \|y\|, \forall y \in \mathbb{E} \\ \Leftrightarrow \ &\|x\| - \langle p, x \rangle + \sup_y \ \langle p, y \rangle - \|y\| \le 0 \end{aligned} \tag{3}$$

The supremum evaluates as

$$\sup_y \langle p, y\rangle - \|y\| = \begin{cases} 0, & \|p\|_* \leq 1 \\ \infty, & \text{otherwise.} \end{cases}.$$

We show this as the following. Assume $\|p\|_* > 1$. Hence there is some vector $z \in \mathbb{E}$, $\|z\| \leq 1$ and $\langle p, z\rangle > 1$. It can be seen that the above supremum is unbounded, i.e. take some $y = tz$, $t(\langle p, z\rangle - \|z\|) \to \infty$ for $t \to \infty$. Now take $\|p\|_* \leq 1$, then we have $\langle p, y\rangle - \|y\| \leq \|y\| (\|p\|_* - 1) \leq 0$, where equality holds for $y = 0$.

Furthermore, we have

$$0 \geq -\langle p, x\rangle + \|x\| \geq - \|x\| \|p\|_* + \|x\| = \|x\| (1 - \|p\|_*) \geq 0$$

Hence $-\langle p, x\rangle + \|x\| = 0$ which implies $\|x\| = \langle p, x\rangle$.

- The dual norm of $\|\cdot\|_1$ is clearly $\|\cdot\|_\infty$ and vice versa. Hence,

$$\begin{aligned} \partial \|\cdot\|_1 (x) &= \{p \in \mathbb{R}^n : \|p\|_\infty \leq 1, \langle p, x\rangle = \|x\|_1\}, \\ &= \left\{ p \in \mathbb{R}^n : \begin{cases} p_i \in [-1, 1], & \text{if } \; x_i = 0 \\ p_i = \operatorname{sign}(x_i), & \text{otherwise.} \end{cases} \right\}. \end{aligned} \tag{4}$$

$$\partial \|\cdot\|_\infty (x) = \{p \in \mathbb{R}^n : \|p\|_1 \leq 1, \langle p, x\rangle = \|x\|_\infty\}. \tag{5}$$

Also, it is easy to show that the dual norm of $\|\cdot\|_2$ is $\|\cdot\|_2$ itself. Thus

$$\begin{aligned} \partial \|\cdot\|_2 (x) &= \{p \in \mathbb{R}^n : \|p\|_2 \leq 1, \langle p, x\rangle = \|x\|_2\} \\ &= \left\{ p \in \mathbb{R}^n : \begin{cases} p_i \in \mathcal{B}(0, 1), & \text{if } \; x = 0 \\ p = \frac{x}{\|x\|_2}, & \text{otherwise.} \end{cases} \right\}, \end{aligned} \tag{6}$$

where $\mathcal{B}(0, 1) = \{p \in \mathbb{R}^n : \|p\|_2 \leq 1\}$ denotes the unit ball around the origin according to the Euclidean metric.

Exercise 3 (4 points)**.** Given $b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$ with linearly independent rows, show that the normal cone N_C of the linear-inequality constraints

$$C = \{u \in \mathbb{R}^n : Au \leq b, \} \tag{7}$$

is

$$N_C(u) = \{A^\top \lambda : \lambda \geq 0, \lambda_i = 0 \text{ if } (Au - b)_i < 0\}. \tag{8}$$

Solution. Recall that by definition $N_C(u) = \{p \in \mathbb{R}^n : \langle p, v - u\rangle \leq 0 \text{ for all } v \in C\}$ for given $u \in C$.

Denote $S = \{A^\top \lambda : \lambda \geq 0, \lambda_i = 0 \text{ if } (Au - b)_i < 0\}$.

- First we prove that $S \subseteq N_C(u)$:

  Let $p_1 = A^\top \lambda_1 \in S$, then

$$\forall v \in C, \langle p_1, v-u \rangle = \langle \lambda_1, A(v-u) \rangle \leq 0, \tag{9}$$

  because component-wise it holds for $(\lambda_1)_i = 0$, and when $(\lambda_1)_i > 0$ we have $(b - Au)_i = 0$, thus $(A(v-u))_i \leq (b - Au)_i = 0$.

  Hence we have $p_1 \in N_C(u)$ and $S \subseteq N_C(u)$.

- Then we prove that $N_C(u) \subseteq S$:

  Let $p_2 \in N_C(u)$, since rows of A are independent, there exists a unique pair of $(\lambda_2, \mu_2) \in \mathbb{R}^m \times \mathbb{R}^n$ such that p_2 can be decomposed as follows:

$$p_2 = A^\top \lambda_2 + \mu_2, \text{ with } A\mu_2 = 0 \tag{10}$$

  where $A^\top \lambda_2$ is the projection of p_2 onto the subspace of $\mathbb{R}^n$ spanned by the rows of A and μ_2 is the component in the null space of A.

  Since $u + \mu_2 \in C$, we have that

$$\langle p_2, (u+\mu_2) - u \rangle = \langle \lambda_2, A\mu_2 \rangle + \langle \mu_2, \mu_2 \rangle = 0 + \|\mu_2\|^2 \leq 0, \tag{11}$$

  Thus the component in null space $\mu_2 = 0$.

  Let $a \in \mathbb{R}$ and $e_i \in \mathbb{R}^m$ be the i-th vector in the canonical basis. Since rows of A are independent, we can find $v_0 \in C$ such that $Av_0 = Au + ae_i$ so long as $Au + ae_i \leq b$, in which case we have

$$\langle p_2, v_0 - u \rangle = \langle \lambda_2, A(v_0 - u) \rangle = \langle \lambda_2, ae_i \rangle = a(\lambda_2)_i \leq 0. \tag{12}$$

  Since for $a \leq 0$, $Au + ae_i \leq b$ always holds, we must have $(\lambda_2)_i \geq 0$ for Eq. 12 to hold.

  Furthermore, when $(Au)_i < b_i$, a can also take any positive value up to $b_i - (Au)_i$. In this case, for Eq. 12 to hold, we must have $(\lambda_2)_i = 0$.

  The above reasoning shows that $p_2 \in S$, which proves that $N_C(u) \subseteq S$.

Thus we have proved that $N_C(u) = S = \{A^\top \lambda : \lambda \geq 0, \lambda_i = 0 \text{ if } (Au - b)_i < 0\}$.

Exercise 4 (6 points)**.** Compute the subdifferential of nuclear norm:

$$X \in \mathbb{R}^{n \times n} \mapsto \|X\|_{nuclear} = \sum_i \sigma_i(X),$$

i.e., sum of singular values.
Hint: Show that the subdifferential at point $X \in \mathbb{R}^{n\times n}$ with $s \geq 0$ zero singular values is given as

$$\partial \|\cdot\|_{\text{nuc}}(X) = \left\{ U_1 V_1^\top + U_2 M V_2^\top : M \in \mathbb{R}^{s\times s}, \|M\|_{\text{spec}} \leq 1 \right\}, \tag{13}$$

where $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ are given by the singular value decomposition of $X = U\Sigma V^\top$, with U_1 and V_1 having $n - s$ columns. Furthermore $\|\cdot\|_{\text{spec}}$ denotes the spectral norm, i.e., the largest singular value.

Solution. Denote by $\langle X, Y\rangle = \text{tr}(X^T Y)$. First we show that the dual norm of the nuclear norm is the spectral norm, i.e.,

$$\sup_{\sum_i \sigma_i(Y)\leq 1} \langle X, Y\rangle = \sigma_1(X).$$

Clearly, $\sup_{\sum_i \sigma_i(Y)\leq 1} \langle X, Y\rangle \geq \sigma_1(X)$ since the supremum is bigger than the function at the feasible candidate $Y = u_1 v_1^T$ (for $X = U\Sigma V^T$) for which the supremum evaluates to $\langle u_1 v_1^T, U\Sigma V^T\rangle = \sigma_1(X)$. The other inequality (again with $X = U\Sigma V^T$) follows from von Neumann's trace inequality $\text{tr}(AB) \leq \sum_{i=1}^n \sigma_i(A)\sigma_i(B)$.

$$\sup_{\sum_i \sigma_i(Y)\leq 1} \langle Y, X\rangle = \sup_{\sum_i \sigma_i(Y)\leq 1} \text{tr}(Y^T X) \leq \sup_{\sum_i \sigma_i(Y)\leq 1} \sum_{i=1}^n \sigma_i(X)\sigma_i(Y) = \sigma_1(X). \tag{14}$$

Hence, from the previous solution, it then follows that

$$\partial \|X\|_{\text{nuc}} = \{Y \in \mathbb{R}^{n\times n} \ : \ \langle X, Y\rangle = \|X\|_{\text{nuc}}, \|Y\|_{\text{spec}} \leq 1\}. \tag{15}$$

We finish the proof by showing that (13) and (15) are the same. Denote by $X = U_1 \Sigma V_1^T$ denote the compact SVD of X.

First we take some Y that satisfies (15), i.e., $\langle X, Y\rangle = \|X\|_{\text{nuc}}$ and $\|Y\|_{\text{spec}} \leq 1$ and show it is in (13). For that, consider the subspace $S = \{U_1 W V_1^T : W \in \mathbb{R}^{r\times r}\}$ where $r = n - s$ and its orthogonal complement $S^\perp = \{U_2 M V_2^T : M \in \mathbb{R}^{s\times s}\}$. Then we can write $Y = \Pi_S(Y) + \Pi_{S^\perp}(Y) = U_1 W V_1^T + U_2 M V_2^T$ for some W and M.

Since we have

$$\begin{aligned} \langle Y, X\rangle &= \langle U_1 W V_1^T + U_2 M V_2^T, U_1 \Sigma V_1^T\rangle = \langle U_1 W V_1^T, U_1 \Sigma V_1^T\rangle \\ &= \text{tr}(V_1^T W^T U_1^T U \Sigma V_1) = \text{tr}(W^T \Sigma) \overset{\text{assumption}}{=} \text{tr}(\Sigma) \end{aligned} \tag{16}$$

we can conclude that $W = I$ and hence $Y = U_1 V_1^T + U_2 M V_2^T$. Since projections always have Lipschitz constant less or equal one we have that

$$\|M\|_{\text{spec}} = \left\|U_2 M V_2^T\right\|_{\text{spec}} = \|\Pi_{S^\perp}(Y)\|_{\text{spec}} \leq \|Y\|_{\text{spec}} \overset{\text{assumption}}{\leq} 1,$$

where we used the unitary invariance of the spectral norm in the first equality.

Conversely take some $U_1 V_1^T + U_2 M V_2^T$ from (13) with $\|M\|_{\text{spec}} \leq 1$ and $X = U_1 \Sigma V_1^T$. We show that it satisfies (15):

$$\langle U_1 V_1^T + U_2 M V_2^T, U_1 \Sigma V_1^T\rangle = \text{tr}(V_1 U_1^T U \Sigma V_1^T) = \text{tr}(\Sigma) = \|X\|_{\text{nuc}}.$$

For the spectral norm we use the fact that if $\|Ax\|^2 \leq \|x\|^2$, then $\|A\|_{\text{spec}} \leq 1$.

$$\begin{aligned}
\left\|(U_1V_1^T + U_2MV_2^T)x\right\|^2 &= \langle U_1V_1^Tx + U_2MV_2^Tx, U_1V_1^Tx + U_2MV_2^Tx\rangle \\
&= \langle x, (U_1V_1^T + U_2MV_2^T)^T(U_1V_1^T + U_2MV_2^T)x\rangle \\
&= \langle x, (V_1U_1^TU_1V_1^Tx\rangle + \langle x, V_2M^TU_2^TU_2MV_2^Tx\rangle \\
&\underbrace{+\langle x, V_1U_1^TU_2MV_2^Tx\rangle + \langle x, V_2M^TU_2^TU_1V_1^Tx\rangle}_{=0} \\
&= \langle V_1^Tx, V_1^Tx\rangle + \langle MV_2^Tx, MV_2^Tx\rangle \\
&= \left\|V_1^Tx_1\right\|^2 + \left\|MV_2^Tx_2\right\| \\
&\overset{\text{assumption}}{\leq} \|x_1\|^2 + \|x_2\|^2 = \|x\|^2,
\end{aligned} \tag{17}$$

where we decomposed $x = x_1 + x_2$ onto the subspace spanned by V_2^T and its orthogonal complement in the second to last step.