Convex Optimization for Machine Learning and Computer Vision

Lecture: T. Wu
Exercises: E. Laude, T. Möllenhoff
Summer Semester 2017

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 4

Room: 02.09.023

Monday, 29.05.2017, 12:15-14:00

Submission deadline: Wednesday, 24.05.2017, Room 02.09.023

Convex Duality (4 Points + 8 Bonus)

Exercise 1 (4 Points)**.** Compute the convex conjugates of the following functions:

1. $f_1 : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ where $f_1(x) = \sqrt{1+x^2}$.
2. $f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ where $f_2(x) = \log\left(\sum_{i=1}^n e^{x_i}\right)$.

Don't forget to specify the domains $\mathrm{dom}(f_1^*), \mathrm{dom}(f_2^*)$.

Solution. 1. The conjugate is defined as

$$f_1^*(y) = \sup_x \; xy - \sqrt{1+x^2}.$$

For $|y| > 1$, the supremum is $+\infty$, since due to the subadditivity of $\sqrt{\cdot}$ we have

$$xy - \sqrt{1+x^2} \geq xy - \sqrt{1} - \sqrt{x^2} = xy - |x| - 1,$$

and the choice $x = t\mathrm{sign}(y)$, $t \to +\infty$ yields $t(|y|-1)-1$ with drives the lower bound to infinity.

Now take $|y| < 1$. Setting the derivative of the function inside the supremum to zero yields

$$y = \frac{x}{\sqrt{1+x^2}} \Rightarrow x = \pm \frac{y}{\sqrt{1-y^2}}.$$

Taking the choice $x = \frac{y}{\sqrt{1-y^2}}$ (it leads to a larger value inside the supremum), we have:

$$f^*(y) = y\frac{y}{\sqrt{1-y^2}} - \sqrt{1+\frac{y^2}{1-y^2}} = \frac{y^2}{\sqrt{1-y^2}} - \sqrt{\frac{1}{1-y^2}} = -\sqrt{1-y^2},$$

with $\mathrm{dom}(f^*) = [-1, 1]$.

2. The conjugate is defined as:

$$f_2^*(y) = \sup_{x\in\mathbb{R}^n} \sum_{i=1}^n x_i y_i - \log\left(\sum_{i=1}^n \exp(x_i)\right).$$

We start by computing $\mathrm{dom}(f_2^*)$. Take some $y_i < 0$, then set $x_i = -a$ and $x_j = 0$ for $i \neq j$. Then the conjugate simplifies to

$$\sup_{a\in\mathbb{R}} \; -ay_i - \log\left(n - 1 + \exp(-a)\right),$$

and one can see that for $a \to \infty$ this becomes $+\infty$.

Next, take $x_i = a \cdot \mathrm{sgn}(y_i)$, then we have for the conjugate:

$$\begin{aligned}
&\sup_{a\in\mathbb{R}} \; a\,\|y\|_1 - \log\left(\sum \exp(a\cdot \mathrm{sgn}(y_i))\right) \\
&= \sup_{a\in\mathbb{R}} \; a\,\|y\|_1 - a - \log\left(\sum \exp(\mathrm{sgn}(y_i))\right) \\
&= \sup_{a\in\mathbb{R}} \; a(\|y\|_1 - 1) - \log\left(\sum \exp(\mathrm{sgn}(y_i))\right),
\end{aligned} \tag{1}$$

which becomes infinite if $\|y\|_1 \neq 1$.

Setting the gradient of the function inside the supremum to zero yields

$$y_i = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \Leftrightarrow x_i = \log y_i \sum_j \exp(x_j)$$

First, we conclude that $\mathrm{dom}(f_2^*) = \{y \in \mathbb{R}^n : y_i \geq 0, \|y\| = 1\}$. Plugging in x_i into the function inside the supremum yields

$$\begin{aligned}
&\sum_{i=1}^n y_i \log y_i \sum_j \exp(x_j) - \log\left(\sum_{i=1}^n \exp(\log y_i \sum_j \exp(x_j))\right) \\
&= \sum_{i=1}^n y_i \log y_i + y_i \log \sum_j \exp(x_j) - \log\left(\sum_{i=1}^n y_i \sum_j \exp(x_j)\right) \\
&\overset{\|y\|_1=1}{=} \sum_{i=1}^n y_i \log y_i + \log \sum_j \exp(x_j) - \log \sum_j \exp(x_j) \\
&= \sum_i y_i \log y_i.
\end{aligned} \tag{2}$$

Hence, $f^*(y) = \sum_i y_i \log y_i + \delta_{\Delta^n}(y)$.

Exercise 2 (4 Points)**.** Compute the convex envelope f^{**} of the functions

1. $f : \mathbb{R} \to \overline{\mathbb{R}}$, $f(x) = \begin{cases} 0 & \text{if } \; x = 0, \\ \lambda & \text{if } \; x \neq 0, |x| \leq 1, \\ \infty & \text{otherwise.} \end{cases}$

2. $f : \mathbb{R}^{n\times n} \to \overline{\mathbb{R}}$, $f(X) = \operatorname{rank}(X) + \delta\{\|X\|_{\text{spec}} \le 1\}$.

by taking the convex conjugate twice.

Solution. 1. The conjugate is given as

$$f^*(y) = \sup_{|x|\le 1} xy - \begin{cases} 0, & \text{if } x = 0, \\ \lambda, & \text{otherwise.} \end{cases}$$

In the case $x = 0$ the supremum is 0, otherwise the supremum is attained at $|y| - \lambda$. Hence, we have

$$f^*(y) = \max\{0, |y| - \lambda\}.$$

Now, for the biconjugate we have

$$f^{**}(x) = \sup_y xy - \max\{0, |y| - \lambda\} \tag{3}$$

If $|x| > 1$ we can see that the supremum becomes $+\infty$. Now assume $|x| \le 1$. Assume $|y| \le \lambda$. Then we have $f^{**}(x) = \sup_{|y|\le\lambda} xy = \lambda|x|$. For $|y| > \lambda$ we have that

$$xy - |y| - \lambda \le |x||y| - |y| - \lambda = |y|\underbrace{(|x| - 1)}_{\le 0} - \lambda < \lambda|x|.$$

Hence, the supremum is always attained for $|y| \le \lambda$. To summarize, we have

$$f^{**}(x) = \lambda|x| + \delta\{|x| \le 1\}.$$

2. For the proof, we will again make use of von Neumann's trace inequality

$$\operatorname{tr}(X^TY) \le \sum_{i=1}^n \sigma_i(X)\sigma_i(Y).$$

We start by computing the conjugate:

$$\begin{aligned} f^*(Y) &= \sup_{\|X\|_{\text{spec}}\le 1} \operatorname{tr}(X^TY) - \operatorname{rank}(X) \\ &= \sup_{\|\Sigma_X\|_{\text{spec}}\le 1} \operatorname{tr}(V_X\Sigma_X U_X^T U_Y \Sigma_Y V_Y^T) - \operatorname{rank}(\Sigma_X) \\ &= \sup_{\|\Sigma_X\|_{\text{spec}}\le 1} \operatorname{tr}(\Sigma_X\Sigma_Y) - \operatorname{rank}(\Sigma_X) \\ &= \sup_{\|X\|_{\text{spec}}\le 1} \sum_{i=1}^n \sigma_i(X)\sigma_i(Y) - \operatorname{rank}(X). \end{aligned} \tag{4}$$

We had the freedom to choose $U_X = U_Y$ and $V_X = V_Y$, since the spectral norm and rank function are unitarily invariant and this choice is optimal due to von Neumann's trace inequality.

Now if $\operatorname{rank}(X) = r$, we have $f^*(Y) = \sum_{i=1}^r \sigma_i(Y) - r$, and hence the conjugate can be expressed as

$$f^*(Y) = \max\{0, \sigma_1(Y) - 1, \dots, \sum_{i=1}^r \sigma_i(Y) - r, \dots, \sum_{i=1}^n \sigma_i(Y) - n\} \tag{5}$$

The largest term in the above maximum is the whose sums only consist of positive terms, which yields:

$$f^*(Y) = \sum_{i=1}^n \max\{0, \sigma_i(Y) - 1\}.$$

We continue with the biconjugate as above:

$$\begin{aligned} f^{**}(X) &= \sup_Y \; \sum_{i=1}^n \sigma_i(Y)\sigma_i(X) - \sum_{i=1}^n \max\{0, \sigma_i(Y) - 1\} \\ &= \sup_Y \; \sum_{i=1}^n \sigma_i(Y)\sigma_i(X) - \sum_{i=1}^{k(Y)} \sigma_i(Y), \end{aligned} \tag{6}$$

where $k(Y)$ is the number of singular values bigger than 1. Now if $\|X\|_{\text{spec}} > 1$ we see that the coefficient $(\sigma_1(X) - 1)$ for $\sigma_1(Y)$ is positive, and hence the supremum is $+\infty$.

Next, let $\|X\|_{\text{spec}} \leq 1$. Then if $\|Y\| \leq 1$ we have that $f^*(Y) = 0$ and the supremum is achieved for $\sigma_i(Y) = 1$, which evaluates to $\sum_i \sigma_i(X) = \|X\|_{\text{nuc}}$.

Finally, we prove that the supremum is always achieved for some Y with $\|Y\| \leq 1$. Indeed for Y with $\|Y\| > 1$ we have the following bound:

$$\begin{aligned} &\sum_{i=1}^n \sigma_i(Y)\sigma_i(X) - \sum_{i=1}^{k(Y)} \sigma_i(Y) \\ &\quad = \sum_{i=1}^n \sigma_i(Y)\sigma_i(X) - \sum_{i=1}^{k(Y)} \sigma_i(Y) + \sum_{i=1}^n \sigma_i(X) - \sum_{i=1}^n \sigma_i(X) \\ &\quad = \underbrace{\sum_{i=1}^{k(Y)} (\sigma_i(Y) - 1)(\sigma_i(X) - 1)}_{\leq 0} + \underbrace{\sum_{i=k(Y)+1}^{n} (\sigma_i(Y) - 1)\sigma_i(X)}_{\leq 0} - k(Y) + \sum_{i=1}^n \sigma_i(X) \\ &\quad < \sum_{i=1}^n \sigma_i(X). \end{aligned} \tag{7}$$

To summarize the above, we have shown that

$$f^{**}(X) = \|X\|_{\text{nuc}} + \delta\{\|X\|_{\text{spec}} \leq 1\},$$

i.e., the convex envelope of the rank function on the spectral norm unit-ball is the nuclear norm.

Definition. The convex hull of an arbitrary set $C \subset \mathbb{R}^n$ is defined as

$$\operatorname{conv}(C) = \left\{ \sum_{i=1}^{p} \lambda_i x_i : x_i \in C, \lambda_i \geq 0, \sum_{i=1}^{p} \lambda_i = 1, p \geq 0 \right\}. \tag{8}$$

Exercise 3 (4 Points)**.** Show that for a set $C \neq \emptyset$ in $\mathbb{R}^n$, every point of $\operatorname{conv}(C)$ can be expressed as a convex combination of $n+1$ points of C (not necessarily different).

Solution. Suppose we can write a point $x \in \operatorname{conv}(C)$ as a convex combination

$$x = \sum_{i=1}^{p} \lambda_i x_i, \lambda_i \geq 0, \sum_{i=1}^{p} \lambda_i = 1, \tag{9}$$

for some $p > n+1$. We will prove that we can write the same point x as a different convex combination involving $p-1$ points. Inductively, this will yield the claim.

Since $p \geq n+2$, we can use exercise 1.4 (Radon's theorem) to partition our p points into 2 nonempty sets $\{x_1, \ldots, x_l\}$, $\{x_{l+1}, \ldots, x_p\}$ whose convex hulls share a common point, i.e. $\exists a \in \mathbb{R}^l, b \in \mathbb{R}^{p-l}$:

$$\sum_{i=1}^{l} a_i x_i = \sum_{i=l+1}^{p} b_i x_i,\ a_i \geq 0, \sum_{i=1}^{l} a_i = 1, b_i \geq 0, \sum_{i=l+1}^{p} b_i = 1. \tag{10}$$

For notational convenience, we index the $p-l$ entries b_i from $i = l+1, \ldots, p$. Now, take the multiplier $b_j \neq 0$ for which $\frac{\lambda_j}{b_j} \leq \frac{\lambda_i}{b_i}$. Solving (10) for x_j yields

$$x_j = \frac{\sum_{i=1}^{l} a_i x_i - \sum_{i=l+1, i\neq j}^{p} b_i x_i}{b_j}$$

To finish the proof, we substitute x_j into (9) and show that the result is a valid convex combination. The substitution yields

$$\begin{aligned} x &= \sum_{i\neq j} \lambda_i x_i + \lambda_j \frac{\sum_{i=1}^{l} a_i x_i - \sum_{i=l+1, i\neq j}^{p} b_i x_i}{b_j} \\ &= \sum_{i=1}^{l} \left(\lambda_i + \frac{\lambda_j a_i}{b_j} \right) x_i + \sum_{i=l+1, i\neq j}^{p} \left(\lambda_i - \frac{\lambda_j b_i}{b_j} \right) x_i. \end{aligned} \tag{11}$$

It remains to show that this is indeed a convex combination, i.e., all the $p-1$ multipliers are bigger than zero and sum up to one. The first l multipliers are positive by construction. The last $p-l-1$ multipliers are positive since

$$\lambda_i - \frac{\lambda_j b_i}{b_j} \geq \lambda_i - \frac{\lambda_i b_i}{b_i} = 0.$$

Finally, for the sum of all coefficients we have:

$$\begin{aligned} &\sum_{i\neq j}\lambda_i + \sum_{i=1}^{l}\frac{\lambda_j a_i}{b_j} - \sum_{i=l+1, i\neq j}^{p}\frac{\lambda_j b_i}{b_j} \\ =&\sum_{i\neq j}\lambda_i + \frac{\lambda_j}{b_j} - \frac{\lambda_j}{b_j}(1-b_j) = \sum_i \lambda_i = 1. \end{aligned} \tag{12}$$

Image Cartooning

Finish the programming exercise from the second exercise sheet.