Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Virginia Estellers
Exercises: Emanuel Laude
Winter Semester 2017/18

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 10

Room: 02.09.023

Friday, 02.02.2018, 09:15-11:00

Submission deadline: Monday, 29.01.2018, 10:15, Room 02.09.023

Theory: Recap (0+14 Points)

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 $\text{dom}(f_1^*), \text{dom}(f_2^*)$.

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 $x = t\text{sign}(y)$, $t \to +\infty$ 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 $\text{dom}(f^*) = [-1, 1]$.

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

Definition. A function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called 1-homogeneous if

$$g(\alpha x) = \alpha g(x),$$

for all $\alpha \geq 0$.

Exercise 2 (4 Points)**.** Let $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be convex, closed, proper and 1-homogeneous. Show that the proximity operator of the sum $\|\cdot\|_2 + g$ is the composition of the proximity operators of $\|\cdot\|_2$ and g, i.e.

$$\operatorname{prox}_{\|\cdot\|_2+g} = \operatorname{prox}_{\|\cdot\|_2} \circ \operatorname{prox}_g.$$

Solution. Let $y \in \operatorname{dom}(g)$. We have the following optimality conditions for $\operatorname{prox}_{\|\cdot\|_2+g}(y)$, $\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))$ and $\operatorname{prox}_g(y)$:

$$0 \in \operatorname{prox}_{\|\cdot\|_2+g}(y) - y + \partial(\|\cdot\|_2 + g)(\operatorname{prox}_{\|\cdot\|_2+g}(y)) \tag{3}$$

$$0 \in \operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y)) - \operatorname{prox}_g(y) + \partial(\|\cdot\|_2)(\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))) \tag{4}$$

$$0 \in \operatorname{prox}_g(y) - y + \partial g(\operatorname{prox}_g(y)) \tag{5}$$

Adding the last two inclusions yields:

$$0 \in \operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y)) - y + \partial g(\operatorname{prox}_g(y)) + \partial(\|\cdot\|_2)(\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))). \tag{6}$$

Assume that it holds for all $x \in \mathbb{R}^n$

$$\partial g(\operatorname{prox}_{\|\cdot\|_2}(x)) \supseteq \partial g(x). \tag{7}$$

Then for $x := \operatorname{prox}_g(y)$, and due to (**??**) and the sum rule of the subdifferential $\partial(\|\cdot\|_2 + g) \supseteq \partial\|\cdot\|_2 + \partial g$ we have that

$$\begin{aligned} 0 &\in \operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y)) - y + \partial g(\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))) + \partial(\|\cdot\|_2)(\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))) \\ &\subset \operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y)) - y + \partial(g + \|\cdot\|_2)(\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))). \end{aligned}$$

This shows that $\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y))$ satisfies (**??**) and therefore $\operatorname{prox}_{\|\cdot\|_2}(\operatorname{prox}_g(y)) = \operatorname{prox}_{\|\cdot\|_2+g}(y)$.

It remains to prove the sufficient condition (**??**). Clearly, for any $x, y \in \mathbb{R}^n$ with $x \perp y$ we have that $\|x + y\|_2 \geq \|y\|_2$, since $x \perp y$ implies $\langle x, y\rangle = 0$. Then we have that

$$\begin{aligned} \min_x \frac{1}{2}\|x - y\|_2^2 + \|x\|_2 &= \min_{\lambda, z \perp y} \frac{1}{2}\|z + \lambda y - y\|_2^2 + \|z + \lambda y\|_2 \\ &= \min_\lambda \frac{1}{2}\|\lambda y - y\|_2^2 + \|\lambda y\|_2 \\ &= \min_{\lambda \geq 0} \frac{1}{2}(\lambda - 1)^2\|y\|_2^2 + |\lambda|\|y\|_2. \end{aligned}$$

The constraint in the last equality can be seen as follows: Suppose $\lambda < 0$. Then increasing it to zero decreases both summands of the objective. Therefore, we have that $\operatorname{prox}_{\|\cdot\|_2}(y) = \lambda y$ for some $\lambda \geq 0$ and clearly $\operatorname{prox}_{\|\cdot\|_2}(y) = 0 \iff y = 0$. Since g is 1-homogeneous, its subdifferential is scaling invariant, meaning that $p \in \partial g(y) \implies p \in \partial g(\lambda y)$ for $\lambda > 0$, we have that (for $y \neq 0$) there exists $\lambda > 0$ so that,

$$\partial g(y) \subseteq \partial g(\lambda y) = \partial g(\operatorname{prox}_{\|\cdot\|_2}(y)).$$

It remains to prove the scaling invariance of the subdifferential for 1-homogeneous g. Let $\lambda > 0$: Via the substitution $z' = \frac{1}{\lambda} z$ we obtain that

$$\begin{aligned} p \in \partial g(y) &\implies \langle p, z - y \rangle + g(y) \leq g(z), \quad \forall z \in \mathrm{dom}(g) \\ &\implies \langle p, \lambda z - \lambda y \rangle + \lambda g(y) \leq \lambda g(z), \quad \forall z \in \mathrm{dom}(g) \\ &\implies \langle p, \lambda z - \lambda y \rangle + g(\lambda y) \leq g(\lambda z), \quad \forall z \in \mathrm{dom}(g) \\ &\implies \langle p, z' - \lambda y \rangle + g(\lambda y) \leq g(z'), \quad \forall z' \in \mathrm{dom}(g) \\ &\implies p \in \partial g(\lambda y). \end{aligned}$$

Exercise 3 (6 Points)**.** Let C be a nonempty, closed, convex subset of $\mathbb{R}^n$. For each $i \in \{1, ..., m\}$, let $\alpha_i \in (0,1)$, $\omega_i \in (0,1)$ and $\Phi_i : C \to \mathbb{R}^n$ be an α_i-averaged operator. Prove the following statements:

- Φ_i is α_i-averaged iff

  $$\|\Phi_i(u) - \Phi_i(v)\|_2^2 + \frac{1 - \alpha_i}{\alpha_i} \|(I - \Phi_i)(u) - (I - \Phi_i)(v)\|_2^2 \leq \|u - v\|_2^2,$$

  for all $u, v \in C$.

- If $\sum_{i=1}^m \omega_i = 1$ and $\alpha = \max_{1 \leq i \leq m} \alpha_i$, then

  $$\Phi = \sum_{i=1}^m \omega_i \Phi_i$$

  is α-averaged.

Solution. By the definition of the averaged operator, $\Phi_i = (1 - \alpha_i) I + \alpha_i \Psi_i$ for some nonexpansive operator $\Psi_i : C \to \mathbb{R}^n$, or $\Psi_i = (1 - \frac{1}{\alpha i}) I + \frac{1}{\alpha i} \Phi_i$. $\Leftrightarrow$ $\Psi_i = (1 - \frac{1}{\alpha i}) I + \frac{1}{\alpha i} \Phi_i$ is nonexpansive.

$$\begin{aligned} & \|\Psi_i(u) - \Psi_i(v)\| \leq \|u - v\| \\ \Leftrightarrow \quad & \alpha_i^2 \|u - v\|^2 \geq \|((\alpha_i - 1) I + \Phi_i)(u) - ((\alpha_i - 1) I + \Phi_i)(v)\|^2 \\ & = \|\Phi_i(u) - \Phi_i(v)\|^2 + (\alpha_i - 1)^2 \|u - v\|^2 \\ & \quad + 2(\alpha_i - 1) \langle u - v, \Phi_i(u) - \Phi_i(v) \rangle \end{aligned}$$

$$\begin{aligned} \Leftrightarrow \|\Phi_i(u) - \Phi_i(v)\|^2 + (1 - 2\alpha_i) \|u - v\|^2 \leq & (1 - \alpha_i) \|(I - \Phi_i)(u) - (I - \Phi_i)(v)\|^2 \\ & - (1 - \alpha_i) \|u - v\|^2 - (1 - \alpha_i) \|\Phi_i(u) - \Phi_i(v)\|^2 \end{aligned}$$

since $2\langle u - v, \Phi_i(u) - \Phi_i(v) \rangle = \|(I - \Phi_i)(u) - (I - \Phi_i)(v)\|^2 - \|u - v\|^2 - \|\Phi_i(u) - \Phi_i(v)\|^2$.

We have the estimate

$$
\begin{aligned}
&\|\Phi(u)-\Phi(v)\|_2^2+\frac{1-\alpha}{\alpha}\|(I-\Phi)(u)-(I-\Phi)(v)\|_2^2\\
&=\left\|\sum_{i=1}^m \omega_i(\Phi_i(u)-\Phi_i(v))\right\|_2^2+\frac{1-\alpha}{\alpha}\left\|\left(I-\sum_{i=1}^m\omega_i\Phi_i\right)(u)-\left(I-\sum_{i=1}^m\omega_i\Phi_i\right)(v)\right\|_2^2\\
&\le\sum_{i=1}^m\omega_i\|\Phi_i(u)-\Phi_i(v)\|_2^2+\frac{1-\alpha}{\alpha}\left\|\sum_{i=1}^m\omega_i((I-\Phi_i)(u)-(I-\Phi_i)(v))\right\|_2^2\\
&\le\sum_{i=1}^m\omega_i\|\Phi_i(u)-\Phi_i(v)\|_2^2+\frac{1-\alpha}{\alpha}\sum_{i=1}^m\omega_i\|(I-\Phi_i)(u)-(I-\Phi_i)(v)\|_2^2\\
&=\sum_{i=1}^m\omega_i\left(\|\Phi_i(u)-\Phi_i(v)\|_2^2+\frac{1-\alpha}{\alpha}\|(I-\Phi_i)(u)-(I-\Phi_i)(v)\|_2^2\right).
\end{aligned}
$$

Since $1>\alpha\ge\alpha_i>0$ for all i we have that $\frac{1}{\alpha}-1\le\frac{1}{\alpha_i}-1$. Then we can further bound:

$$
\begin{aligned}
\ldots&=\sum_{i=1}^m\omega_i\left(\|\Phi_i(u)-\Phi_i(v)\|_2^2+\frac{1-\alpha}{\alpha}\|(I-\Phi_i)(u)-(I-\Phi_i)(v)\|_2^2\right)\\
&\le\sum_{i=1}^m\omega_i\left(\|\Phi_i(u)-\Phi_i(v)\|_2^2+\frac{1-\alpha_i}{\alpha_i}\|(I-\Phi_i)(u)-(I-\Phi_i)(v)\|_2^2\right)\\
&\le\sum_{i=1}^m\omega_i\|u-v\|_2^2=\|u-v\|_2^2.
\end{aligned}
$$