PGM - Convex analysis - exercise solutions

Caio Corro

1 Convex combination

Question. Let U be a convex set and n a strictly positive integer. Prove that:

$$\forall \boldsymbol{u}^{(1)}...\boldsymbol{u}^{(n)}, \boldsymbol{\mu} \in \triangle(n): \quad \sum_{i=1}^{n} \mu_i \boldsymbol{u}^{(i)} \in U.$$

Answer. Proof by induction. Let $P(n)$ be the property that $\sum_{i=1}^{n} \mu_i \boldsymbol{u}^{(i)} \in U, \forall \boldsymbol{u}^{(1)}...\boldsymbol{u}^{(n)}, \boldsymbol{\mu} \in \triangle(n)$.

- Initialization: $P(2)$ is true by definition of a convex set.
- We assume $P(n)$, $n \geq 2$ is true, and prove that then $P(n+1)$ is true. We rewritte the convex combination as follows:

$$\sum_{i=1}^{n+1} \mu_i \boldsymbol{u}^{(i)} = \mu_{i+1} \boldsymbol{u}^{(i+1)} + \sum_{i=1}^{n} \mu_i \boldsymbol{u}^{(i)}$$

Wlog, we assume that $\mu_{i+1} < 1$.

$$= \mu_{i+1} \boldsymbol{u}^{(i+1)} + (1 - \mu_{i+1}) \sum_{i=1}^{n} \frac{\mu_i}{1 - \mu_{i+1}} \boldsymbol{u}^{(i)}$$

By definition of a convex set, we can see that $\sum_{i=1}^{n+1} \mu_i \boldsymbol{u}^{(i)} \mu_i \boldsymbol{u}^{(i)} \in U$ if $\sum_{i=1}^{n} \frac{\mu_i}{1-\mu_{i+1}} \in U$. We have:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\mu_i}{1 - \mu_{i+1}} \mu_i \boldsymbol{u}^{(i)} &= \sum_{i=1}^{n} \frac{\mu_i}{1 - (1 - \sum_{j=1}^{n} \mu_j)} \boldsymbol{u}^{(i)} \\
&= \sum_{i=1}^{n} \frac{\mu_i}{\sum_{j=1}^{n} \mu_j} \boldsymbol{u}^{(i)}
\end{aligned}$$

Note that the denominator ensure that the "weightning" term sums to one, therefore, as $P(n)$ is true:

$$\in U.$$

2 Sum of convex functions

Question. Let $f^{(1)}...f^{(n)}$ be convex functions and $\boldsymbol{w} \in \mathbb{R}_{+}^{n}$. Prove that the function f defined as follows is convex:

$$f(\boldsymbol{u}) = \sum_{i=1}^{n} w_i f^{(i)}(\boldsymbol{u}).$$

Answer. Let $\boldsymbol{u}, \boldsymbol{v} \in \operatorname{dom} f$ and $\epsilon \in [0, 1]$. Then:

$$f(\epsilon\boldsymbol{u} + (1-\epsilon)\boldsymbol{v}) = \sum_{i=1}^{n} w_i f^{(i)}(\epsilon\boldsymbol{u} + (1-\epsilon)\boldsymbol{v})$$

By convexity of each $f^{(i)}$ and the fact that $\boldsymbol{w} \geq \mathbf{0}$:

$$\begin{aligned}
&\leq \sum_{i=1}^{n} w_i \left(\epsilon f^{(i)}(\boldsymbol{u}) + (1-\epsilon) f^{(i)}(\boldsymbol{v})\right) \\
&= \epsilon \sum_{i=1}^{n} w_i f^{(i)}(\boldsymbol{u}) + (1-\epsilon) \sum_{i=1}^{n} w_i f^{(i)}(\boldsymbol{v}) \\
&= \epsilon f(\boldsymbol{u}) + (1-\epsilon) f(\boldsymbol{v})
\end{aligned}$$

Hence f is convex.

3 Biconjugate computation

Question. $f : \mathbb{R}^k \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \langle \boldsymbol{a}, \boldsymbol{u} \rangle + b$.

Answer.

$$\begin{aligned}
f^*(\boldsymbol{t}) &= \sup_{\boldsymbol{u} \in \operatorname{dom} f} \langle \boldsymbol{t}, \boldsymbol{u} \rangle - f(\boldsymbol{u}) \\
&= \sup_{\boldsymbol{u} \in \operatorname{dom} f} \langle \boldsymbol{t}, \boldsymbol{u} \rangle - \langle \boldsymbol{a}, \boldsymbol{u} \rangle - b \\
&= \sup_{\boldsymbol{u} \in \operatorname{dom} f} \langle \boldsymbol{t} - \boldsymbol{a}, \boldsymbol{u} \rangle - b \\
&= \begin{cases} -b & \text{if } \boldsymbol{t} = \boldsymbol{a} \\ \infty & \text{otherwise.} \end{cases} \\
f^{**}(\boldsymbol{u}) &= \sup_{\boldsymbol{t} \in \operatorname{dom} f^*} \langle \boldsymbol{t}, \boldsymbol{u} \rangle - f^*(\boldsymbol{u}) \\
&= \langle \boldsymbol{a}, \boldsymbol{u} \rangle + b
\end{aligned}$$

Question. $f : \mathbb{R} \to \mathbb{R}$ defined as $f(u) = \exp(u)$.

Answer.

$$\begin{aligned}
f^*(t) &= \sup_{u \in \operatorname{dom} f} ut - f(u) \\
&= \sup_{u \in \operatorname{dom} f} ut - \exp(u)
\end{aligned}$$

The objective is trivially concave and differentiable. By first order optimality conditions, $\hat{u}$ is an maximizer iff:

$$\begin{aligned}
\frac{\partial}{\partial \hat{u}}(\hat{u}t - \exp(\hat{u})) &= 0 \\
\exp(\hat{u}) &= t \\
\hat{u} &= \log t
\end{aligned}$$

Therefore, we have:

$$\begin{aligned} f^*(t) &= t\log t - t \\ f^{**}(u) &= \sup_{t\in \operatorname{dom} f^*} ut - f^*(t) \\ &= \sup_{t\in \operatorname{dom} f^*} ut - t\log t + t \end{aligned}$$

The objective is trivially concave and differentiable. By first order optimality conditions, $\hat{t}$ is an maximizer iff:

$$\begin{aligned} \frac{\partial}{\partial \hat{t}}\left(u\hat{t} - \hat{t}\log \hat{t} + \hat{t}\right) &= 0 \\ \log \hat{t} &= u \\ \hat{t} &= \exp(u) \end{aligned}$$

Therefore:

$$f^{**}(u) = u\exp(u) - u\exp(u) + \exp(u) = \exp(u).$$

Question. $f : \mathbb{R}^k \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \log \sum_i \exp u_i$.

Answer.

$$\begin{aligned} f^*(\boldsymbol{t}) &= \sup_{\boldsymbol{u}\in \operatorname{dom} f} \langle \boldsymbol{t}, \boldsymbol{u}\rangle - f(\boldsymbol{u}) \\ & \sup_{\boldsymbol{u}\in \operatorname{dom} f} \langle \boldsymbol{t}, \boldsymbol{u}\rangle - \log\sum_i \exp u_i \end{aligned}$$

The objective is trivially concave and differentiable. By first order optimality conditions, $\hat{\boldsymbol{u}}$ is an maximizer iff:

$$\begin{aligned} \nabla_{\hat{\boldsymbol{u}}}\left(\langle \boldsymbol{t}, \hat{\boldsymbol{u}}\rangle - \log\sum_i \exp \hat{u}_i\right) &= \boldsymbol{0} \\ \operatorname{softmax}(\hat{\boldsymbol{u}}) &= \boldsymbol{t} \end{aligned}$$

Importantly, this indicates us that $\operatorname{dom} f^* = \operatorname{int}(\triangle(k))$. It is easy to check that this is equivalent to:

$$\hat{\boldsymbol{u}} = \log \boldsymbol{t} + c\boldsymbol{1}$$

where $c \in \mathbb{R}$ is any constant. We have:

$$\begin{aligned} f^*(\boldsymbol{t}) &= \langle \boldsymbol{t}, \log \boldsymbol{t} + c\boldsymbol{1}\rangle - \log\sum_i \underbrace{\exp(c + \log t_i)}_{\exp(c)\times t_i} + \delta_{\triangle(k)}(\boldsymbol{t}) \\ &= \langle \boldsymbol{t}, \log \boldsymbol{t}\rangle + \underbrace{\sum_i t_i c}_{=c} \underbrace{- \log\left(\exp(c)\underbrace{\sum_i t_i}_{=1}\right)}_{=c} + \delta_{\triangle(k)}(\boldsymbol{t}) \\ &= \langle \boldsymbol{t}, \log \boldsymbol{t}\rangle + \delta_{\triangle(k)}(\boldsymbol{t}) \end{aligned}$$

which is the negative Shannon entropy function.

$$
\begin{aligned}
f^{**}(\boldsymbol{u}) &= \sup_{\boldsymbol{t} \in \operatorname{dom} f^*} \langle \boldsymbol{t}, \boldsymbol{u} \rangle - f^*(\boldsymbol{t}) \\
&= \sup_{\boldsymbol{t} \in \operatorname{dom} f^*} \langle \boldsymbol{t}, \boldsymbol{u} \rangle - \langle \boldsymbol{t}, \log \boldsymbol{t} \rangle - \delta_{\triangle(k)}(\boldsymbol{t})
\end{aligned}
$$

WARNING: $\operatorname{dom} f^* = \triangle(k)$, we have a constrained optimization problem! So we need to rely on KKT conditions!

$$
\begin{aligned}
\sup_{\boldsymbol{t} \in \mathbb{R}^k} \quad & \langle \boldsymbol{t}, \boldsymbol{u} \rangle - \langle \boldsymbol{t}, \log \boldsymbol{t} \rangle \\
\text{s.t.} \quad & \sum_i t_i = 1 \\
& t_i \geq 0, \forall i \in 1...k
\end{aligned}
$$

let $\lambda \in R$ and $\boldsymbol{\mu} \in \mathbb{R}^k_+$ be dual variables associates with equalities and inequalities, respectively. An optimal triple $\hat{\boldsymbol{t}}, \hat{\lambda}$ and $\hat{\boldsymbol{\mu}}$ must satisfy the following set of equations:

$$
\begin{aligned}
&\text{(stationarity)} && \frac{\partial}{\partial \hat{t}_i}\big(- \langle \hat{\boldsymbol{t}}, \boldsymbol{u} \rangle + \langle \hat{\boldsymbol{t}}, \log \hat{\boldsymbol{t}} \rangle + \lambda(1 - \sum_i \hat{t}_i) - \langle \boldsymbol{\mu}, \hat{\boldsymbol{t}} \rangle \big) = 0, \quad \forall i \in \{1...k\} \\
&\text{(primal feasibility)} && \sum_i \hat{t}_i = 1 \\
& && \hat{t}_i \geq 0, \quad \forall i \in 1...k \\
&\text{(primal feasibility)} && \hat{\boldsymbol{\mu}} \geq 0 \\
&\text{(complementary feasibility)} && -\langle \boldsymbol{\mu}, \hat{\boldsymbol{t}} \rangle = 0
\end{aligned}
$$

By stationarity, we have:

$$
\begin{aligned}
-\hat{u}_i + \log \hat{t}_i + 1 - \hat{\lambda} - \hat{\mu}_i = 0 \\
\log \hat{t}_i &= \hat{u}_i - 1 + \hat{\lambda} + \hat{\mu}_i \\
\hat{t}_i &= \exp(\hat{u}_i - 1 + \hat{\lambda} + \hat{\mu}_i) \\
\hat{t}_i &= \frac{\exp(\hat{u}_i + \hat{\mu}_i)}{\exp(1 - \hat{\lambda})}
\end{aligned}
$$

Note that this means that each element of $\hat{\boldsymbol{t}}$ is strictly positive, so $\hat{\boldsymbol{\mu}} = \boldsymbol{0}$, otherwise complementary slackness is not satisfied. Moreover, by primal feasilibilty we have:

$$
\begin{aligned}
\sum_i \hat{t}_i &= 1 \\
\sum_i \frac{\exp(\hat{u}_i - 1)}{\exp(1 - \hat{\lambda})} &= 1 \\
\exp(1 - \hat{\lambda}) &= \sum_i \exp(\hat{u}_i)
\end{aligned}
$$

So we have:

$$
\hat{t}_i = \frac{\exp(\hat{u}_i)}{\sum_j \exp(\hat{u}_j)}
$$

Replacing this in the biconjugate at optimality leads to the log-sum-exp function (which was expected!).