MATH60005/70005: Optimisation (Autumn 22-23)

Weeks 6 and 7: Exercises

Dr Dante Kalise
Department of Mathematics
Imperial College London, United Kingdom
dkaliseb@imperial.ac.uk

Sara Bicego (GTA)
Department of Mathematics
Imperial College London, United Kingdom
s.bicego21@imperial.ac.uk

1. Show the convexity of the following functions:

   - the quad-over-lin function
   $$f(x_1, x_2) = \frac{x_1^2}{x_2}$$
   defined over $\mathbb{R} \times \mathbb{R}_{++} = \{(x_1, x_2) : x_2 > 0\}$.
   - the generalized quad-over-lin function
   $$g(\mathbf{x}) = \frac{\|\mathbf{A}\mathbf{x} + \mathbf{b}\|^2}{\mathbf{c}^\top \mathbf{x} + d} \qquad \left(\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n, d \in \mathbb{R}\right)$$
   is convex over $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} + d > 0\}$.
   - $f(x_1, x_2) = -\log(x_1 x_2)$, over $\mathbb{R}^2_{++}$.
   - $h(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$.

2. Show that $\sqrt{1 + \mathbf{x}^\top Q \mathbf{x}}$ is convex for Q positive definite.

3. Find the optimal solution of
$$\begin{aligned} \max_{\mathbf{x}\in\mathbb{R}^3} \quad & 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - 3x_2 + 4x_3 \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Solutions

1.1) For the quad-over-lin function, we proceed by computing the Hessian:

$$f(x_1, x_2) = \frac{x_1^2}{x_2^2}\,, \qquad \nabla f = \left(2\frac{x_1}{x_2}, -\frac{x_1^2}{x_2^2}\right), \qquad \nabla^2 f = 2\begin{pmatrix} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \\ -\frac{x_1}{x_2^2} & \frac{x_1^2}{x_2^3} \end{pmatrix}.$$

In order to determine the positiveness of $\nabla^2 f$, we study the sign of its trace and determinant:

$$Tr\left(\nabla^2 f\right) = 2\left(\frac{1}{x_2} + \frac{x_1^2}{x_2^3}\right) > 0\,, \qquad \text{since } x_2 \in \mathbb{R}_{++},$$

$$Det\left(\nabla^2 f\right) = 4\left[\frac{1}{x_2} \cdot \frac{x_1^2}{x_2^3} - \left(\frac{x_1}{x_2^2}\right)^2\right] = 0\,.$$

Thus, we have $\nabla^2 f \precsim 0$, and equivalently f is convex.

1.2) The generalized quad-over-lin function can be rewritten as follows

$$g(\mathbf{x}) := \frac{\|\mathbf{Ax} + \mathbf{b}\|^2}{\mathbf{c}^\top \mathbf{x} + d} = \frac{\|\mathbf{y}\|^2}{t} =: h(\mathbf{y}, t)$$

after the linear change of variables $(\mathbf{Ax} + \mathbf{b}, \mathbf{c}^\top \mathbf{x} + d) \mapsto (\mathbf{y}, t)$. Accordingly, the set $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} + d > 0\}$ becomes $D' = \{\mathbf{y} \in \mathbb{R}^m : t > 0\} = \mathbb{R}^m \times \mathbb{R}_{++}$. Moreover, the function h can be seen as a sum of quad-over-lin functions (ex 1.1) with $(x_1, x_2) = (y_i, t)$

$$h(\mathbf{y}, t) = \sum_{i=1}^{m} h_i(\mathbf{y}, t) = \sum_{i=1}^{m} \frac{y_i^2}{t}\,.$$

Being a sum of convex functions, h is convex, and so is $g(\mathbf{x})$, as it can be obtained as the composition of a convex function with a linear change of variables.

1.3) The function f can be written as

$$f(x_1, x_2) = -\log(x_1 x_2) = -\log(x_1) - \log(x_2)\,.$$

To prove convexity of f, it is enough to show that both the terms are convex. In order to do so, we start by noticing that both the addends are of the form $-\log(t)$ composed with a projection map

$$(x_1, x_2) \mapsto x_1 \qquad (x_1, x_2) \mapsto x_2$$

which is an affine transformation. We conclude by showing that the function $t \mapsto -\log(t)$ is convex over $\mathbb{R}_+$:

$$l = -\log(t)\,, \qquad l' = -\frac{1}{x}\,, \qquad l'' = \frac{1}{x^2} > 0.$$

Thus, f is convex as a sum of convex functions composed with affine transformations.

1.4) $h(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$ can be written as the following composition of functions

$$h(\mathbf{x}) = g(f(\mathbf{x})) \qquad \text{with} \;\; g(t) = e^t \;\; \text{and} \;\; f(\mathbf{x}) = \|\mathbf{x}\|^2.$$

Thanks to convexity of norms, we know that $f(\mathbf{x})$ is convex, and since

$$g' = e^t = g'' > 0\,,$$

we also have convexity of g. Furthermore, the outer function g is non-decreasing over any $I \subset \mathbb{R}$, hence we have convexity of $h(\mathbf{x})$.

Remark. Having only f, g convex is not enough, as shown with the following counterexample. Consider $s(x) = (x^2 - 4)^2$. We can write $s(x) = g(f(x))$ with $g(t) = t^2$ and $f(x) = (x^2 - 4)$. Even though both those functions are convex ($f'', g'' > 0$), we have $s'' < 0$ for $|x| < \sqrt{\frac{4}{3}}$.

2) We have

$$h(\mathbf{x}) := \sqrt{1 + \mathbf{x}^\top Q\, \mathbf{x}} = \sqrt{1 + \|\mathbf{x}\|_Q^2}$$

where $\|\mathbf{x}\|_Q = \sqrt{\mathbf{x}^\top Q\, \mathbf{x}}$. We can write $h(\mathbf{x}) = g(f(\mathbf{x}))$ as a composition of functions

$$f(\mathbf{x}) = \|\mathbf{x}\|_Q\,, \qquad g(r) = \sqrt{1 + r^2}\,.$$

In order to show convexity, we start by noting that, since $Q \succ 0$

$$\|\mathbf{x}\|_Q = \|\sqrt{\Lambda}\, U \mathbf{x}\|_2\,, \quad \text{for the diagonalization } Q = U^\top Q\, U$$

where Λ is a positive definite diagonal matrix. Since the diagonalization is a linear transformation and the 2−norm is convex, we have f convex.
On the other hand, we have

$$g(r) = \sqrt{1 + r^2}\,, \quad g'(r) = \frac{r}{\sqrt{1 + r^2}}\,, \quad g''(r) = \frac{1}{1 + r^2}\,,$$

and since $g''(r) > 0$ for every $r \in \mathbb{R}$ and $g'(r) > 0$ for every $r > 0$ (which is the case for $r = \mathbf{x}^\top Q\, \mathbf{x}$). Summing up, the function g is convex and non-decreasing, f is convex, hence h is in turn convex.

3) The optimization constraints prescribe $\mathbf{x} = \left(x_1, x_2, x_3\right)^\top \in \Delta_2$, that is the unit simplex (convex set) in $\mathbb{R}^3$ with extreme points $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$.
Furthermore, the objective function f can be written in quadratic form

$$f(\mathbf{x}) = \mathbf{x}^\top A\, \mathbf{x} + \mathbf{b}^\top \mathbf{x} = \mathbf{x}^\top \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 2 & -3 & 4 \end{bmatrix} \mathbf{x}$$

where $A \succ 0$, hence f is strictly convex. Maximizing a convex function over a convex set comes from evaluating the function at the extreme points . We conclude that the maximizer is $\mathbf{x}^* = (0, 0, 1)$, where the objective attains its maximum $f(\mathbf{x}^*) = 5$.

Extra exercises

3.1) We can generalize the above example to the matrix norm

$$\|A\|_{1,1} := \max\left\{\|A\mathbf{x}\|_1 : \|\mathbf{x}\|_1 \le 1\right\}, \qquad \|\mathbf{x}\|_1 := \sum_{i=1}^{n} |x_i|\,.$$

As before, we have the maximization of a convex function (norm composed with an affine transformation) over a convex set (closed ℓ_1−ball). Then, the optimizer lies at one of the extreme points of the set $\|\mathbf{x}\| \le 1$. To better understand how to characterize these extreme points, we consider the case $n = 2$:

$$\|\,\left[x_1,\, x_2\right]\,\|_1 \le 1 \quad\Longleftrightarrow\quad |x_1| + |x_2| \le 1 \quad\Longleftrightarrow\quad \begin{cases} x_1 + x_2 \le 1 \\ x_1 - x_2 \le 1 \\ -x_1 + x_2 \le 1 \\ -x_1 - x_2 \le 1 \end{cases}$$

Thus, the set of extreme points in the $n = 2$ case is given by $\{e_1, -e_1, e_2, -e_2\}$, where e_i is a vector with all zeros, except for the i−th entry, which has value 1. In the general case, the set becomes $\{e_1, -e_1, \ldots, e_n, -e_n\}$.

We now look at the value of the objective function at the extreme points. By definition of ℓ_1 norm, we have

$$\|A\,e_j\|_1 = \|A\,(-e_j)\|_1 = \sum_{i=1}^{m} |A_{i,j}|$$

and so

$$\|A\|_{1,1} = \max_{j=1,2,\ldots,n} \|Ae_j\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |A_{i,j}|\,.$$

4) Determine whether the set $C = \left\{\mathbf{x} \in \mathbb{R}^n : \min_i x_i \le 1\right\}$ is convex.

In the case $n = 1$, C is the left closed half-line originating from 1, hence C is convex. In the case $n = 2$, C is given by the union of the two half-planes defined by $C_1 = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 \le 1\right\}$ and $C_2 = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 \le 1\right\}$. We can show that C is a non-convex set by proving that a convex combination between two points in C is not in the set:

$$x^* = (2, 1),\;\; y^* = (1, 2) \in C \quad\not\Longrightarrow\quad \lambda\, x^* + (1 - \lambda)\, y^* := z^* \in C$$

as for $\lambda = \frac{1}{2}$ we have $z^* = (\frac{3}{2}, \frac{3}{2})$.

5) Determine whether the following function is convex:

$$g(\mathbf{x}) = \begin{cases} 0 & x \in K \\ \|\mathbf{x}\|_2 - 1 & \text{elsewhere} \end{cases} \qquad \text{where} \quad K = \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \le 1\right\}.$$

We start by noting that K is the unit ℓ_2–ball. Then, g is the distance function to K:

$$g(\mathbf{x}) = \min_{\mathbf{y} \in K} \|\mathbf{x} - \mathbf{y}\|_2 \,.$$

We conclude that g is convex because it is the distance function to a convex set.

6) Show that the following function is convex over $\mathbb{R}^n_{++}$:

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i \ln(x_i) - \left(\sum_{i=1}^{n} x_i\right) \ln\left(\sum_{i=1}^{n} x_i\right).$$

We first consider the case $n = 2$, in which $\mathbf{x} = (x_1, x_2)$ and

$$\begin{aligned} f(\mathbf{x}) &= x_1 \ln(x_1) + x_2 \ln(x_2) - (x_1 + x_2) \ln(x_1 + x_2) \\ &= x_1 \ln(x_1) + x_2 \ln(x_2) - x_1 \ln(x_1 + x_2) - x_2 \ln(x_1 + x_2) \\ &= x_1 \Big(\ln(x_1) - \ln(x_1 + x_2) \Big) + x_2 \Big(\ln(x_2) - \ln(x_1 + x_2) \Big) \\ &= x_1 \ln\left(\frac{x_1}{x_1 + x_2}\right) + x_2 \ln\left(\frac{x_2}{x_1 + x_2}\right). \end{aligned}$$

In the general case, we can rewrite f as

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{\sum_{k=1}^{n} x_k}\right) = \sum_{i=1}^{n} h_i(\mathbf{x}) \quad \text{for} \;\; h_i(\mathbf{x}) = x_i \ln\left(\frac{x_i}{\sum_{k=1}^{n} x_k}\right).$$

We now need to show that the functions $h_i(\mathbf{x})$ are convex. Consider the change of variables

$$\mathbf{x} \longmapsto (u, v) \qquad \text{where} \qquad u = x_i \,,\; v = \sum_{k=1}^{n} x_k \,.$$

Then, we can write h_i as

$$\varphi(u, v) = u \ln\left(\frac{u}{v}\right)$$

for which we can check convexity through the Hessian.

$$\frac{\partial \varphi}{\partial u} = \ln\left(\frac{u}{v}\right) + u \cdot \frac{v}{u} \cdot \frac{1}{v} = \ln(u) - \ln(v) \,, \qquad \frac{\partial \varphi}{\partial v} = -\frac{u}{v}$$

$$\frac{\partial^2 \varphi}{\partial u^2} = \frac{1}{u} \,, \qquad \frac{\partial^2 \varphi}{\partial v^2} = \frac{u}{v^2} \,, \qquad \frac{\partial^2 \varphi}{\partial u \partial v} = -\frac{1}{v}$$

$$\Longrightarrow \; \nabla^2 \varphi = \begin{bmatrix} \frac{1}{u} & -\frac{1}{v} \\ -\frac{1}{v} & \frac{u}{v^2} \end{bmatrix}, \quad Det(\nabla^2 \varphi) = \frac{1}{v^2} - \frac{1}{v^2} = 0 \,, \quad Tr(\nabla^2 \varphi) = \frac{1}{u} + \frac{u}{v^2} > 0 \,,$$

where the positiveness of the trace is given by $\mathbf{x} \in \mathbb{R}^n_{++}$.
To conclude, φ is convex, and so are the h_i's, as they are compositions of convex functions with linear transformation. Furthermore, f is convex as it is the sum of convex functions.