Optimization
Homework 3

Erika Rivadeneira
erika.rivadeneira@cimat.mx

February 17, 2021

1. Is the set $S = \left\{a \in \mathbb{R}^k | p(0) = 1, |p(t)| \leq 1 \text{ for } t \in [\alpha, \beta]\right\}$, where

$$p(t) = a_1 + a_2 t + \cdots + a_k t^{k-1}$$

convex?

Solution
We have that $a \in \mathbb{R}^k$ and let's define

$$p_a(t) := a_1 + a_2 t + \cdots + a_k t^{k-1}.$$

Let's take x and y in S arbitrary such that

$$\begin{aligned} p_x(t) &:= x_1 + x_2 t + \cdots + x_k t^{k-1} \\ p_y(t) &:= y_1 + y_2 t + \cdots + y_k t^{k-1} \end{aligned}$$

then for $t \in [\alpha, \beta]$ we have that

$$\begin{array}{ccc} p_x(0) = x_1 = 1, & p_y(0) = y_1 = 1 & \\ |p_x(t)| \leq 1 & \text{and} & |p_y(t)| \leq 1 \end{array}$$

Let $\gamma : [0, 1] \to \mathbb{R}^{\urcorner}$ such that

$$\gamma(\lambda) = (1 - \lambda x + \lambda y)$$

Let's take $z \in \gamma(\hat{\lambda})$ with $\hat{\lambda} \in [0, 1]$, then

$$\gamma(\hat{\lambda}) = z = (1 - \hat{\lambda})x + \hat{\lambda}y.$$

Now, we have

$$\begin{aligned} p_z(t) &= z_1 + z_2 t + ... + z_k t^{k-1} \\ &= (1 - \hat{\lambda})x_1 + \hat{\lambda}y_1 + \left[(1 - \hat{\lambda})x_2 + \hat{\lambda}y_2\right] t + ... + \left[(1 - \hat{\lambda})x_k + \hat{\lambda}y_k\right] t^{k-1} \end{aligned}$$

Let's note that

$$p_z(0) = (1 - \hat{\lambda})x_1 + \hat{\lambda}y_1 = 1 - \hat{\lambda} + \hat{\lambda}.$$

On the other hand, since $|p_x(t)| \leq 1$ and $|p_y(t)| \leq 1$ we get

$$\begin{aligned} |p_z(t)| &= |(1 - \hat{\lambda})x_1 + \hat{\lambda}y_1 + \left[(1 - \hat{\lambda})x_2 + \hat{\lambda}y_2\right] t + ... + \left[(1 - \hat{\lambda})x_k + \hat{\lambda}y_k\right] t^{k-1}| \\ &= |(1 - \hat{\lambda})[x_1 + x_2 t + ... + x_k t^{k-1}] + \hat{\lambda}[y_1 + y_2 t + ... + y_k t^{k-1}]| \\ &\leq |(1 - \hat{\lambda})[x_1 + x_2 t + ... + x_k t^{k-1}]| + |\hat{\lambda}[y_1 + y_2 t + ... + y_k t^{k-1}]| \\ &= (1 - \hat{\lambda})|[x_1 + x_2 t + ... + x_k t^{k-1}]| + \hat{\lambda}|[y_1 + y_2 t + ... + y_k t^{k-1}]| \\ &= (1 - \hat{\lambda})|p_x(t)| + \hat{\lambda}|p_y(t)| \\ &\leq (1 - \hat{\lambda} + \hat{\lambda}) = 1 \end{aligned}$$

Therefore, S is convex.

2. Suppose f is convex, $\lambda_1 > 0$ and $\lambda_2 \leq 0$ with $\lambda_1 + \lambda_2 = 1$, and let $x_1, x_2 \in \operatorname{dom} f$. Show that the inequality

$$f\left(\lambda_1 x_1 + \lambda_2 x_2\right) \geq \lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right)$$

always holds.

Solution
Let's define $\eta_1, \eta_2 \geq 0$ such that

$$\eta_1 = \frac{1}{\lambda_1} \quad and \quad \eta_2 = -\frac{\lambda_2}{\lambda_1},$$

we can see that

$$\eta_1 + \eta_2 = \frac{1}{\lambda_1} - \frac{\lambda_2}{\lambda_1} = \frac{1-\lambda_2}{1} = 1,$$

and let's rewrite x_1 in terms of η_1 and η_2

$$x_1 = \eta_1(\lambda_1 x_1 + \lambda_2 x_2) + \eta_2 x_2,$$

then by Jensen's inequality we get

$$\begin{aligned} f(x_1) &= f(\eta_1(\lambda_1 x_1 + \lambda_2 x_2) + \eta_2 x_2) \\ &\leq \eta_1 f(x_1\lambda_1 + \lambda_2 x_2) + \eta_2 f(x_2) \\ &= \frac{1}{\lambda_1} f(x_1\lambda_1 + \lambda_2 x_2) - \frac{\lambda_2}{\lambda_1} f(x_2) \\ \Rightarrow \lambda_1 f(x_1) &\leq f(x_1\lambda_1 + \lambda_2 x_2) - \lambda_2 f(x_2). \end{aligned}$$

Therefore,

$$f(x_1\lambda_1 + \lambda_2 x_2) \geq \lambda_1 f(x_2) + \lambda_2 f(x_2).$$

3. Show that the following function $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

$$f(x) = -\exp(-g(x))$$

where $g : \mathbb{R}^n \to \mathbb{R}$ has a convex domain and satisfies

$$\begin{bmatrix} \nabla^2 g(x) & \nabla g(x) \\ \nabla^T g(x) & 1 \end{bmatrix} \succeq 0$$

for $x \in \operatorname{dom} g$.

Solution
Let's compute the gradient and the Hessian of $f(x)$

$$\begin{aligned} \frac{\partial}{\partial} f(x_i) &= \exp(-g(x)) \frac{\partial}{\partial x_i} g(x), \\ \frac{\partial^2}{\partial xj \partial x_i} &= -\exp(-g(x)) \left[\frac{\partial}{\partial x_j} g(x)\right] \left[\frac{\partial}{\partial x_i} g(x)\right] + \exp(-g(x)) \frac{\partial^2}{\partial x_j \partial x_i} g(x) \\ &= \exp(-g(x)) \left[\frac{\partial^2}{\partial x_j \partial x_i} g(x) - \left[\frac{\partial}{\partial x_j} g(x)\right] \left[\frac{\partial}{\partial x_i} g(x)\right]\right], \end{aligned}$$

thus,

$$\begin{aligned} \nabla^2 f(x) &= \frac{\partial^2}{\partial x_j \partial x_i} \\ &= exp(-g(x)) \left[\nabla^2 g(x) - \nabla g(x) \nabla g(x)^T\right]. \end{aligned}$$

Let's note that by hypothesis $\left[\nabla^2 g(x) - \nabla g(x) \nabla g(x)^T\right]$ is the determinant of

$$\begin{bmatrix} \nabla^2 g(x) & \nabla g(x) \\ \nabla^T g(x) & 1 \end{bmatrix} \succeq 0,$$

then it is positive semidefinite. On the other hand $exp(-g(x))$ is always positive. Therefore, f is convex since $\nabla^2 f(x)$ is positive semidefinite.

4. Show that $f(x,y) = x^2/y, y > 0$ is convex.

   Solution
   We have to compute the Hessian in orden to see if it is positive semidefinite. So,

$$\nabla f(x,y) = \left[\frac{2x}{y}, -\frac{x^2}{y^2}\right]$$
$$\nabla^2 f(x,y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix}.$$

   Let's note that $|\nabla^2 f(x,y)| = \frac{4x^2}{y^2} - \frac{4x^2}{y^2} = 0$ and the principal minors are $\frac{2}{y} > 0$ and $\frac{2x^2}{y^3} > 0$. Therefore, $f(x,y)$ is convex since its Hessian matrix is positive semidefinite.

5. Consider the function $f(x_1, x_2) = \left(x_1 + x_2^2\right)^2$. At the point $x^T = [1,0]$ we consider the search direction $p^T = [-1,1]$. Show that p is a descent direction and find all minimizers of the function.

   Solution
   In order to show that p is a descent direction we have to check that $\nabla f(1,0) \cdot p < 0$. So,

$$\nabla f(x_1,x_2) = (2(x_1 + x_2^2), 4(x_1 + x_2^2)x_2)$$
$$\Rightarrow \nabla f(1,0) = (2,0).$$

   So that

$$\nabla f(1,0) \cdot p = (2,0) \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$
$$= -2 < 0.$$

   Then, p is a descent direction.
   Now, in order to find the minimizers let's compute the gradient and set it to zero:

$$\nabla f(x_1,x_2) = (2(x_1 + x_2^2), 4(x_1 + x_2^2)x_2) = 0.$$

   Then, we have the following system

$$2(x_1 + x_2^2) = 0 \Rightarrow x_1 = -x_2^2$$
$$4(x_1 + x_2^2)x_2 = 0 \Rightarrow x_1 = -x_2^2,$$

   the Hessian matrix of $f(x_1,x_2)$ is

$$\nabla^2 f(x_1,x_2) = \begin{pmatrix} 2 & 4x_2 \\ 4x_2 & 4x_1 + 12x_2^2 \end{pmatrix}$$

   Now, evaluating the Hessian on $x_1 = -x_2^2$

$$\nabla^2 f(-x_2^2, x_2) = \begin{pmatrix} 2 & 4x_2 \\ 4x_2 & -4x_2^2 + 12x_2^2 \end{pmatrix}$$
$$= \begin{pmatrix} 2 & 4x_2 \\ 4x_2 & 8x_2^2 \end{pmatrix}$$

   Since the principal minors are no negative the Hessian matrix is definite positive on $(-x_2^2, x_2)$. Therefore, $(-x_2^2, x_2)$ is a minimizer of $f(x_1,x_2)$.

6. Find all the values of the parameter a such that $[1,0]^T$ is the minimizer or maximizer of the function $f(x_1,x_2) = a^3 x_1 e^{x_2} + 2a^2 \log(x_1 + x_2) - (a+2)x_1 + 8ax_2 + 16x_1x_2$

   Solution
   Let's compute the gradient of $f(x_1,x_2)$

$$\nabla f(x_1,x_2) = \left(a^3 e^{x_2} + 2a^2 \cdot \frac{1}{x_1 + x_2} - (a+2) + 16x_2, a^3 x_1 e^{x_2} + \frac{2a^2}{x_1 + x_2} + 8a + 16x_1\right),$$

evaluating the gradient of $f(x_1, x_2)$ on $[0, 1]$ we get the following system

$$a^3 + 2a^2 - a - 2 = 0 \tag{1}$$

$$a^3 + 2a^2 + 8a + 16 = 0, \tag{2}$$

subtracting (1) from (2) we get that

$$-9a = 18 \Rightarrow a = -2.$$

So that, for $a = -2$ we get

$$f(x_1, x_2) = -8x_1e^{x^2} + 8log(x_1 + x_2) - 16x_2 + 16x_1x_2,$$

then

$$\nabla f(x_1, x_2) = \left(-8e^{x^2} + \frac{8}{x_1 + x_2} + 16x_2, -8x_1e^{x^2} + \frac{8}{x_1 + x_2} - 16 + 16x_1\right)$$

and

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{-8}{(x_1+x_2)^2} & -8e^{x^2} - \frac{8}{(x_1+x_2)^2} + 16 \\ -8e^{x^2} - \frac{8}{(x_1+x_2)^2} + 16 & -8x_1e^{x^2} - \frac{8}{(x_1+x_2)^2} \end{pmatrix},$$

evaluating on $[1, 0]^T$ we get

$$\nabla^2 f(1, 0) = \begin{pmatrix} -8 & 0 \\ 0 & -16 \end{pmatrix}$$

We can see that the Hessian of $f(x_1, x_2)$ is definite negative therefore $[1, 0]^T$ is a maximizer of the function.

7. Consider the sequence $x_k = 1 + 1/k!, k = 0, 1, \cdots$. Does this sequence converge linearly to 1? Justify your response.

   Solution
   Let's find the value of μ for $p = 1$ and $x^* = 1$

$$\begin{aligned} \mu = \lim_{k\to\infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|\,p} &= \lim_{k\to\infty} \frac{\left|\left(1 + \frac{1}{(k+1)!}\right) - 1\right|}{\left|\left(1 + \frac{1}{k!}\right) - 1\right|} \\ &= \lim_{k\to\infty} \frac{\left|\frac{1}{(k+1)!}\right|}{\left|\frac{1}{k!}\right|} \\ &= \lim_{k\to\infty} \frac{k!}{(k+1)!} \\ &= \lim_{k\to\infty} \frac{1}{(k+1)} \\ &= 0 \end{aligned}.$$

   Therefore, the sequence x_k converges superlinear to 1 since $\mu = 0$.

8. Show that $f(\boldsymbol{x}) = \log \sum_{i=1}^{n} \exp(x_i)$ is convex.

   Solution
   We have to check that

$$f\left[\alpha\boldsymbol{x}_1 + (1 - \alpha)\boldsymbol{x}_2\right] \leq \alpha f\left(\boldsymbol{x}_1\right) + (1 - \alpha) f\left(\boldsymbol{x}_2\right).$$

with $\alpha \in (0,1)$. Then, by Hölder's inequality and logarithm's properties we get

$$\begin{aligned}
f(\alpha \mathbf{x_1} + (1-\alpha)\mathbf{x_2}) &= \log\left(\sum_{i=1}^{n} e^{\alpha x_{1i} + (1-\alpha)x_{2i}}\right) \\
&= \log\left(\sum_{i=1}^{n} \left(e^{x_{1i}}\right)^{\alpha} \left(e^{x_{2i}}\right)^{(1-\alpha)}\right) \\
&\leq \log\left[\left(\sum_{i=1}^{n} \left(e^{x_{1i}}\right)\right)^{\alpha} \left(\sum_{i=1}^{n} \left(e^{x_{2i}}\right)\right)^{(1-\alpha)}\right] \\
&= \alpha \log\left(\sum_{i=1}^{n} \left(e^{x_{1i}}\right)\right) + (1-\alpha)\log\left(\sum_{i=1}^{n} \left(e^{x_{2i}}\right)\right) \\
\Rightarrow f\left[\alpha x_1 + (1-\alpha)x_2\right] &\leq \alpha f\left(x_1\right) + (1-\alpha) f\left(x_2\right)
\end{aligned}$$

9. Show that $f(x) = \log \sum_{i=1}^{n} \exp\left(g_i(x)\right) : \mathbb{R} \to \mathbb{R}$ is convex if $g_i : \mathbb{R} \to \mathbb{R}$ are convex

   Solution
   We have to check that

   $$f\left[\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\right] \leq \alpha f\left(\boldsymbol{x}_1\right) + (1-\alpha) f\left(\boldsymbol{x}_2\right).$$

   with $\alpha \in (0,1)$. Then, by Jensen's inequality, Hölder's inequality and logarithm's properties we have that

   $$\begin{aligned}
f\left[\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\right] &= \log \sum_{i=1}^{n} exp(g_i(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2)) \\
&\leq \log \sum_{i=1}^{n} exp(\alpha g_i(\boldsymbol{x}_1) + (1-\alpha) g_i(\boldsymbol{x}_2)) \\
&= \log\left[\sum_{i=1}^{n} exp(\alpha g_i(\boldsymbol{x}_1)) \cdot exp\left((1-\alpha) g_i(x_2)\right)\right] \\
&\leq \log\left[\left(\sum_{i=1}^{n} exp(g_i(x_1))\right)^{\alpha} \cdot \left(\sum_{i=1}^{n} exp(g_i(x_2))\right)^{1-\alpha}\right] \\
&= \alpha \log\left(\sum_{i=1}^{n} exp(g_i(x_1))\right) + (1-\alpha)\log\left(\sum_{i=1}^{n} exp(g_i(x_2))\right) \\
&= \alpha f(x_1) + (1-\alpha) f(x_2)
\end{aligned}$$

10. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Show that f is convex over a nonempty convex set C if and only if

    $$(\nabla f(x) - \nabla f(y))^T (x-y) \geq 0, \forall x, y \in C$$

    Note: the proof we have is only for the case $(\Rightarrow)$
    Solution
    $(\Rightarrow)$ We know that f is convex if

    $$f(y) \geq f(x) + \nabla f(x)^T (y-x).$$

    Since x and y are arbitrary the following inequality also holds

    $$f(x) \geq f(y) + \nabla f(y)^T (x-y).$$

    Adding both inequalities we get

    $$\begin{aligned}
f(y) + f(x) &\geq f(x) + \nabla f(x)^T (y-x) + f(y) + \nabla f(y)^T (x-y) \\
0 &\geq \nabla f(x)^T (y-x) + \nabla f(y)^T (x-y) \\
0 &\leq \nabla f(x)^T (x-y) - \nabla f(y)^T (x-y) \\
0 &\leq (\nabla f(x) - \nabla f(y))^T (x-y)
\end{aligned}$$

    $(\Leftarrow)$ Let's assume that

    $$(\nabla f(x) - \nabla f(y))^T (x-y) \geq 0, \forall x, y \in C,$$

then

$$\nabla f(x)^T(x-y) \geq \nabla f(y)^T(x-y).$$

By the mean value theorem we know that

$$\nabla f(cx+(1-c)z) = \frac{f(x)-f(z)}{(x-z)},$$

let $y = cx+(1-c)z$ then

$$\begin{aligned} x-y &= (1-c)x-(1-c)z \\ &= (1-c)(x-z) \end{aligned}$$

So that,

$$\begin{aligned} (1-c)\nabla f(x)^T \cdot (x-z) &\geq \nabla f(y)^T \cdot (x-z)(1-c) \\ \Rightarrow \nabla f(x)^T \cdot (x-z) &\geq f(x)-f(z) \\ \Rightarrow -f(x)-\nabla f(x)^T \cdot (z-x) &\geq -f(z) \\ \Rightarrow f(z) &\geq f(x)+\nabla f(x)^T \cdot (z-x) \end{aligned}$$