Homework 4 - Continuous Optimization

~~Estelle~~ Ezra Baup, Samuel Bélisle, Cassandre Renaud

December 14, 2022

holà les gus
ce devoir, Sam'enchante pas, vous?
(un peu cass-é comme blague je l'avoue, mais gardez ez-poir en moi please)

Question 1

NOTE/TODO je suis pas du tout sure que cette notation est très belle ducoup dites si vous voulez changer. Aussi c'est très moche les produit scalaires "horizontaux" mais plus lisible? jsp

The feasible set is $S = \{x, y \in \mathbb{R}^n | h(x,y) = 0\} = \{x, y \in \mathbb{R}^n | 1 - x^\top x = 0, 1 - y^\top y = 0, x^\top y = 0\}$.
It is not convex. Indeed, we will give two points z_1 and $z_2 \in S$, but such that $z = \lambda z_1 + (1-\lambda) z_2 \notin S$ for a given λ. We will work with these $z_i \in \mathbb{R}^2 \times \mathbb{R}^2$ i.e. $n = 2$.

We will take $z_1 = (x_1, y_1) = ((1,0),(0,1))$. First we check that $z_1 \in S$.

- $1 - x_1^\top x_1 = 1 - \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = 1 - 1 = 0$
- $1 - y_1^\top y_1 = 1 - \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 - 1 = 0$
- $x_1 \top y_1 = \langle (1,0),(0,1) \rangle = 0$

And we will take $z_2 = (x_2, y_2) = ((0,1),(1,0))$. we also check that $z_2 \in S$.

- $1 - x_2^\top x_2 = 1 - \langle (0,1),(0,1) \rangle = 1 - 1 = 0$
- $1 - y_2^\top y_2 = 1 - \langle (1,0),(1,0) \rangle = 1 - 1 = 0$
- $x_2 \top y_2 = \langle (0,1),(1,0) \rangle = 0$

Lastly, we will take $\lambda = \frac{1}{2}$. Now we can compute our $z = \lambda z_1 + (1-\lambda) z_2$

$$z = \lambda z_1 + (1-\lambda) z_2 = \frac{1}{2}((1,0),(0,1)) + \frac{1}{2}((0,1),(1,0)) = ((\frac{1}{2}, \frac{1}{2}),(\frac{1}{2}, \frac{1}{2})) = (x, y)$$

But if we compute $x^\top y = \left\langle (\frac{1}{2}, \frac{1}{2}),(\frac{1}{2}, \frac{1}{2}) \right\rangle = (\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{2} \neq 0$, so the third condition of our function h does not hold on this point, hence our set is not convex.

By definition, LICQ holds if and only if the ∇h_i and ∇g_i for g_i that are active holds.
Here we do not have any g_i but we have three function h:

- $h_1(x,y) = 1 - x^\top x$
- $h_2(x,y) = 1 - y^\top y$

- $h_3(x, y) = x^\top y$

If we compute their gradients, we get:

- $\nabla_x h_1(x, y) = \nabla_x \left(1 - x^\top x\right) = \nabla_x \left(1 - \sum_{i=1}^n x_i^2\right) = -2x$ and $\nabla_y h_1(x, y) = 0$

$$\text{then } \nabla h_1(x, y) = \begin{bmatrix} \vec{x} \\ \vec{0} \end{bmatrix} \text{ where } \vec{0} \text{ is the vector } \vec{0} \in \mathbb{R}^n.$$

- $\nabla_x h_2(x, y) = 0$ and $\nabla_y h_2(x, y) = \nabla_y \left(1 - y^\top y\right) = \nabla_y \left(1 - \sum_{i=1}^n y_i^2\right) = -2y$

$$\text{then } \nabla h_2(x, y) = \begin{bmatrix} \vec{0} \\ \vec{y} \end{bmatrix}$$

- $\nabla_x h_3(x, y) = \nabla_x \sum_{i=1}^n x_i \cdot y_i = \vec{y}$ and $\nabla_y h_3(x, y) = \nabla_y \sum_{i=1}^n x_i \cdot y_i = \vec{x}$

$$\text{then } \nabla h_3(x, y) = \begin{bmatrix} \vec{y} \\ \vec{x} \end{bmatrix}$$

They are linearly independent:
$\lambda_1 \nabla h_1(x, y) + \lambda_2 \nabla h_2(x, y) + \lambda_3 \nabla h_3(x, y) = 0 \iff \lambda_1 x + \lambda_3 y = 0$ and $\lambda_3 x + \lambda_2 y = 0$. We have that this is true without having all the $\lambda_i = 0$, if and only if $x = \lambda y$, or $y = 0$. But if $y = 0$, we have $h_2(x, y) = 1 \neq 0$ so our point is not feasible. Same with h_1 if $x = 0$. Lastly, if $x = \lambda y$ for $\lambda, x, y \neq 0$, we get that $h_3(x, y) = x^\top y = \lambda y^\top y = \lambda ||y|| \neq 0$, so our point is not feasible either.
To conclude, we get that for all of our feasible points, i.e. the points in the set S, $\nabla h_1(x, y), \nabla h_2(x, y), \nabla h_3(x, y)$ are linearly independent, which means by definition that LICQ holds.

Question 2

Question 3

Let's find an expression for the Lagrangian function $L(x, y, \mu)$. We denote I_n for the identity matrix in $\mathbb{R}^{n \times n}$.

$$\begin{aligned}
L(x, y, \mu) &= f(x, y) + \mu^\top h(x, y) \\
&= \frac{1}{2} x^\top A x + \frac{1}{2} y^\top B y + \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix} \begin{bmatrix} 1 - x^\top x \\ 1 - y^\top y \\ x^\top y \end{bmatrix} \\
&= \frac{1}{2} x^\top A x + \frac{1}{2} y^\top B y + \mu_1 (1 - x^\top x) + \mu_2 (1 - y^\top y) + \mu_3 x^\top y \\
&= \frac{1}{2} \left(x^\top A x - 2\mu_1 x^\top x + y^\top B y - 2\mu_2 y^\top y + 2\mu_3 x^\top y\right) + \mu_1 + \mu_2 \\
&= \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} A - 2\mu_1 I_n & \mu_3 I_n \\ \mu_3 I_n & B - 2\mu_2 I_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \mu_1 + \mu_2
\end{aligned}$$

As A and B are symmetric by assumptions, we see that the above matrix is symmetric.

Question 4

By definition, $L_D(\mu) = \inf_{x,y\in\mathbb{R}^n} L(x, y, \mu)$. Using the previous question, and denoting $M_\mu := \begin{bmatrix} A - 2\mu_1 I_n & \mu_3 I_n \\ \mu_3 I_n & B - 2\mu_2 I_n \end{bmatrix}$, we find:

$$
\begin{aligned}
L_D(\mu) &= \inf_{x,y\in\mathbb{R}^n} L(x, y, \mu) \\
&= \inf_{x,y\in\mathbb{R}^n} \left\{ \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} M_\mu \begin{bmatrix} x \\ y \end{bmatrix} + \mu_1 + \mu_2 \right\} \\
&= \inf_{x,y\in\mathbb{R}^n} \left\{ \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} M_\mu \begin{bmatrix} x \\ y \end{bmatrix} \right\} + \mu_1 + \mu_2 \\
&= \begin{cases} \mu_1 + \mu_2 & \text{if } M_\mu \succeq 0 \\ -\infty & \text{else} \end{cases}
\end{aligned}
$$

Hence we can write the dual as:

$$
\max_{\mu\in\mathbb{R}^3} \mu_1 + \mu_2 \text{ subject to } \begin{bmatrix} A - 2\mu_1 I_n & \mu_3 I_n \\ \mu_3 I_n & B - 2\mu_2 I_n \end{bmatrix} \succeq 0 \tag{D}
$$

Question 5

We know that M_μ is symmetric for all $\mu \in \mathbb{R}^3$. If μ is a solution of the dual problem, then the matrix M_μ associated with the values of μ should be positive semidefinite. This implies in particular that the diagonal blocks would be positive semidefinite too, i.e. $A - \mu_1 I_n \succeq 0$ and $B - \mu_2 I_n \succeq 0$. The two latter conditions are equivalent to $\mu_1 \leq \lambda_{\min}(A)$ and $\mu_2 \leq \lambda_{\min}(B)$, where $\lambda_{\min}(A)$ and $\lambda_{\min}(B)$ are the smallest eigenvalues of A and B respectively.

Hence, if we focus only on satisfying the conditions $A - \mu_1 I_n \succeq 0$ and $B - \mu_2 I_n \succeq 0$, the optimal value for (D) is $\lambda_{\min}(A) + \lambda_{\min}(B)$.

In fact, this is the optimal value even for the whole condition $M_\mu \succeq 0$. Indeed, by question 2,

Question 6

Question 7

Question 8