Homework 4 - Continuous Optimization

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

December 25, 2022

Faudra pas oublier de mettre nos noms "comme il faut". J'ai juste voulu éviter de renforcer ma tite tête avec ton nom *officiel* à chaque fois que je lis cette ligne.

-Sam

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)

We want to solve the following problem:

$$\min_{x,y\in\mathbb{R}^n} \quad \text{subject to } h(x,y) = 0, \tag{P}$$

where f and h are specified on the homework sheet.

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

By definition, LICQ holds at $x \in S$ if and only if $\nabla h_1(x), \ldots \nabla h_p(x)$, and $\nabla g_i(x)$ for $i \in \mathcal{I}(x)$ are linearly independant.

Here, we do not have any constraint function g_i but we have three functions h_i:

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

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

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

If we compute their gradients, we get:
TODO : revoir gradient : les calculs sont justes et après la matrice est fausse, jsp si ça change les résultats pour le lin indep

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

    then $\nabla h_1(x,y) = \begin{bmatrix} -2\vec{x} \\ \vec{0} \end{bmatrix}$ where $\vec{0}$ is the vector $\vec{0} \in \mathbb{R}^n$.

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

    then $\nabla h_2(x,y) = \begin{bmatrix} \vec{0} \\ -2\vec{y} \end{bmatrix}$

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

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

They are linearly independent:
$\lambda_1 \nabla h_1(x,y) + \lambda_2 \nabla h_2(x,y) + \lambda_3 \nabla h_3(x,y) = 0 \iff -2\lambda_1 x + \lambda_3 y = 0$ and $\lambda_3 x - 2\lambda_2 y = 0$. We have that this is true without having all the $\lambda_i = 0$, if and only if $x = \lambda y$ (for some lambda that can be deduced from the previous equations), 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

Our two first constraints $h_1(x,y) = 1 - x^\top x = 0$ and $h_2(x,y) = 1 - y^\top y = 0$ can be rephrased $||x|| = ||y|| = 1$. We notice that there is no constraint on x and y at the same time, i.e. a constraint that tells us what x must be related to y and vice-versa. Similarly, we can notice that our function doesn't have a part where x and y are mixed.

It means that if we optimized x and y separately, the optimum found will be the optimum for our relaxed problem as well. So we rewrite our relaxed problem as:

$$\min_{x \in \mathbb{R}^n : ||x||=1} \frac{1}{2} x^\top A x + \min_{y \in \mathbb{R}^n : ||y||=1} \frac{1}{2} y^\top B y$$

The two problems are solved identically as the only thing changing is the matrix. To solve them, we first recall example 2.14 from the notes:
If A is a symmetric linear map with eigenvalues $\lambda_1, ..., \lambda_n$, then $\forall u \in \mathcal{E}$, we have

$$\lambda_{\min} ||u||^2 \leq \langle u, A(u) \rangle \leq \lambda_{\max} ||u||^2$$

and by rewriting the scalar product we get

$$\lambda_{\min} ||u||^2 \leq u^\top A u \leq \lambda_{\max} ||u||^2$$

Applied to our problems (let's look at the first as they are similar), knowing that we optimize on vector of norm 1, we get that for every feasible x, we have

$$\frac{1}{2} \lambda_{\min} \leq \frac{1}{2} x^\top A x \leq \frac{1}{2} \lambda_{\max}$$

It tells us that our optimal value can't be lower that $\frac{1}{2} \lambda_{\min}$. If we find an x that attains this bound, we will know that it is the optimal value. So let's find it:
As $\lambda_{\min}$ is an eigenvalue, it means that $\exists v_{\min} \in \mathbb{R}^n, v_{\min} \neq 0$ such that $A v_{\min} = \lambda_{\min} v_{\min}$. We take $x = \frac{v_{\min}}{||v_{\min}||}$, and then we have

$$x^\top A x = \frac{1}{||v_{\min}||^2} v_{\min}^\top (A v_{\min}) = \frac{1}{||v_{\min}||^2} v_{\min}^\top (\lambda_{\min} v_{\min}) = \frac{\lambda_{\min}}{||v_{\min}||^2} v_{\min}^\top v_{\min} = \frac{\lambda_{\min}}{||v_{\min}||^2} \cdot ||v_{\min}||^2 = \lambda_{\min}$$

So, this x, plotted in our function, would give us $\frac{1}{2} \lambda_{\min}$, which means our bound is attained, which means it is the optimal value.
We denote $\lambda_{\min}(A)$ and $\lambda_{\min}(B)$ our minimal eigenvalues for the matrices A and B. From our previous reasoning, we have that $\frac{\lambda_{\min}(A)}{2}$ and $\frac{\lambda_{\min}(B)}{2}$ are our optimal values, and so the optimal value for our relaxed problem is $\frac{\lambda_{\min}(A) + \lambda_{\min}(B)}{2}$.

Question 3

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

$$L(x, y, \mu) = f(x, y) + \mu^\top h(x, y)$$

$$= \frac{1}{2} x^\top A x + \frac{1}{2} y^\top B y + \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix} \begin{bmatrix} 1 - x^\top x \\ 1 - y^\top y \\ x^\top y \end{bmatrix}$$

$$= \frac{1}{2} x^\top A x + \frac{1}{2} y^\top B y + \mu_1 (1 - x^\top x) + \mu_2 (1 - y^\top y) + \mu_3 x^\top y$$

$$= \frac{1}{2} \left(x^\top A x - 2\mu_1 x^\top x + y^\top B y - 2\mu_2 y^\top y + 2\mu_3 x^\top y \right) + \mu_1 + \mu_2$$

$$= \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} \begin{bmatrix} A - 2\mu_1 I_n & \mu_3 I_n \\ \mu_3 I_n & B - 2\mu_2 I_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \mu_1 + \mu_2$$

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. $A - 2\mu_1 I_n \succeq 0$ and $B - 2\mu_2 I_n \succeq 0$. The two latter conditions are equivalent to $2\mu_1 \le \lambda_{\min}(A)$ and $2\mu_2 \le \lambda_{\min}(B)$, 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 $A - 2\mu_1 I_n \succeq 0$ and $B - 2\mu_2 I_n \succeq 0$, the optimal value for (D) is $\frac{\lambda_{\min}(A)+\lambda_{\min}(B)}{2}$.

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

Question 6

To use the strong duality theorem, we need to satisfy the two following assumptions:

(A1) the primal problem (P) admits a KKT point $(x^*, y^*) \in S$ with valid Lagrange multipliers $\mu^* \in \mathbb{R}^3$

(A2) the function $(x,y) \mapsto L(x,y,\mu^*)$ is convex

Consider $\mu_1^* = \lambda_{\min}(A)/2$, $\mu_2^* = \lambda_{\min}(B)/2$ and $\mu_3^* = 0$. We will try to find x^* and y^* a little later.

First, consider the condition (A2). Let $L_{\mu^*}(x,y) = L(x,y,\mu^*)$. By an analogous reasoning as for the previous point, the matrix $M_{\mu^*} = \nabla^2(L_{\mu^*}(x,y))$ is positive semi-definite. Hence by a theorem of the course, L_{μ^*} is convex. So (A2) holds for this choice of μ^*.

Now, let's consider (A1). We will try to find (x^*, y^*) that satisfy this assumption with Lagrange multipliers μ^* explicited above.

For $(x^*, y^*) \in S$ to be KKT with Lagrange multipliers μ^*, we need:

$$
h(x^*, y^*) = 0 \text{ and } -\nabla f(x^*, y^*) = \sum_{i=1}^3 \mu_i \nabla h_i(x^*, y^*).
$$

The gradient of h_1, h_2 and h_3 were already computed in question 1. We can compute easily:

$$\nabla f(x, y) = \begin{bmatrix} Ax \\ By \end{bmatrix}$$

Hence (x^*, y^*) and μ^* need to satisfy

$$-\begin{bmatrix} Ax^* \\ By^* \end{bmatrix} = \mu_1 \begin{bmatrix} -2x^* \\ \vec{0} \end{bmatrix} + \mu_2 \begin{bmatrix} \vec{0} \\ -2y^* \end{bmatrix} + \mu_3 \begin{bmatrix} y^* \\ x^* \end{bmatrix}$$

This gives the system $\begin{cases} -Ax^* = -2\mu_1 x^* + \mu_3 y^* \\ -By^* = -2\mu_2 y^* + \mu_3 x^* \end{cases}$.

Injecting $\mu^* = \left(\frac{\lambda_{\min}(A)}{2}, \frac{\lambda_{\min}(B)}{2}, 0 \right)^\top$, this gives:

$$\begin{cases} Ax^* = \lambda_{\min}(A)x^* \\ By^* = \lambda_{\min}(B)y^* \end{cases} .$$

So we may take $(x^*, y^*) = (v_{\min}(A), v_{\min}(B))$, where $v_{\min}(A)$ and $v_{\min}(B)$ are the unit-normed eigenvectors of $\lambda_{\min}(A)$ and $\lambda_{\max}(B)$ respectively.
At this point, we need to recall that (x^*, y^*) must belong to the set S, or equivalently we must have $h(x^*, y^*) = 0$. By taking unit-normed eigenvectors, we ensure that $h_1(x^*, y^*) = 0 = h_2(x^*, y^*)$. The last condition is $h_3(x^*, y^*) = 0$. By definition of h_3, this means $(x^*)^\top y^* = 0$, i.e. x^* and y^* need to be orthogonal in $\mathbb{R}^n$. This is the only assumption we do not know, but observe that as soon as we know it the previous reasoning implies that the assumptions of the strong duality theorem are satisfied.

Hence, we can always find $\mu^* \in \mathbb{R}^3$ such that (A2) is satisfied, but we cannot always ensure that (A1) is satisfied. One condition that can ensure it however, is that the eigenvectors of A and B corresponding to their smallest eigenvalues are orthogonal. We can not expect to have this before solving the problem if we don't have specific informations about A and B other that them being symmetric.

Question 7

We will first compute the gradient with respect to x. ∇_y will be done similarly.
First of all, we know, by linearity, that

$$\nabla_x L_\beta(x, y, \mu) = \nabla_x f(x, y) + \nabla_x \mu^\top h(x, y) + \frac{\beta}{2} \nabla_x ||h(x)||^2$$

We will compute each of these expressions:

$$\frac{\partial}{\partial x_i} f(x, y) = \frac{1}{2} \frac{\partial}{\partial x_i} x^\top Ax = \frac{1}{2} \frac{\partial}{\partial x_i} \sum_{j,k=1}^n x_j A_{ik} x_k = \frac{1}{2} \frac{\partial}{\partial x_i} \left(\sum_{j,k=1:j,k\neq 1}^n x_j A_{ik} x_k + A_{ii} x_i^2 + x_i \sum_{j=1:j\neq i}^n (A_{ij} + A_{ji}) x_j \right)$$

$$= \frac{1}{2} \left(2A_{ii} x_i + \sum_{j=1:j\neq i}^n (A_{ij} + A_{ji}) x_j \right) = \frac{1}{2} \left(2A_{ii} x_i + \sum_{j=1:j\neq i}^n 2A_{ij} x_j \right) = \sum_{j=1}^n A_{ij} x_j = row_i(A) \cdot x$$

As $\frac{\partial}{\partial x_i} f(x, y)$ is the i-th element of $\nabla_x f$, we get that $\nabla_x f = Ax$

Now for the second one, we can develop $\mu^\top h(x, y) = \mu_1(1 - x\top x) + \mu_2(1 - y^\top y) + \mu_3 x^\top y$, so we have:

$$\nabla_x \mu^\top h(x, y) = \nabla_x(\mu_1(1 - x\top x)) + \nabla_x(\mu_2(1 - y^\top y)) + \nabla_x(\mu_3 x^\top y)$$
$$= \mu_1 \nabla_x(1 - x\top x) + \mu_3 \nabla_x(x^\top y)$$
$$= \mu_1(2x) + \mu_3 y$$
$$= 2\mu_1 x + \mu_3 y$$

As $\frac{\partial}{\partial x_i}(x^\top x) = \frac{\partial}{\partial x_i}\sum_{i=1}^n x_i^2 = 2x_i$ and $\frac{\partial}{\partial x_i}(x^\top y) = \frac{\partial}{\partial x_i}\sum_{i=1}^n x_i y_i = y_i$

Now for the third part we will use the fact that $\nabla\|x\|^2 = 2x$ as $x^\top x = \|x\|^2$ and we've just computed it before. Plus we know the chain rule: $\frac{\partial}{\partial x_i} f \circ g(x) = \nabla f(g(x)) \cdot \frac{\partial}{\partial x_i} g(x)$. We can use it to compute our ∇ :

$$\frac{\partial}{\partial x_i} \frac{\beta}{2}\|h(x, y)\|^2 = \beta h(x, y) \cdot \frac{\partial}{\partial x_i}(h(x, y))$$

which gives us the formula

$$\nabla_x \frac{\beta}{2}\|h(x, y)\|^2 = \beta \nabla_x h(x, y) \cdot h(x, y) \text{ where } \nabla_x h(x, y) \text{ is a n} \times 3 \text{ matrix}$$
$$= \beta \begin{bmatrix} -2x & \vec{0} & y \end{bmatrix} \cdot \begin{bmatrix} 1 - x^\top x \\ 1 - y^\top y \\ x^\top y \end{bmatrix}$$
$$= \beta \begin{bmatrix} 2(x^\top x - 1)x_1 + (x^\top y)y_1 \\ \cdots \\ 2(x^\top x - 1)x_n + (x^\top y)y_n \end{bmatrix}$$
$$= \beta \left(2(x^\top x - 1)\vec{x} + (x^\top y)\vec{y} \right)$$

Putting together our three calculations, we get

$$\nabla_x L_\beta(x, y, \mu) = Ax + 2\mu_1 x + \mu_3 y + \beta \left(2(x^\top x - 1)\vec{x} + (x^\top y)\vec{y} \right)$$
$$= (A + 2\beta(x^\top x - 1) + 2\mu_1)x + (x^\top y + \mu_3)y$$

Similarly, we get

$$\nabla_y L_\beta(x, y, \mu) = By + 2\mu_2 y + \mu_3 x + \beta \left(2(y^\top y - 1)\vec{y} + (x^\top y)\vec{x} \right)$$
$$= (x^\top y + \mu_3)x + (B + 2\beta(y^\top y - 1) + 2\mu_2)y$$

Question 8