Homework 2 Convex Optimization

Franck Laborde

October 2024

Exercice 1 (LP Duality)

1. Let us compute the Lagrangian of the problem (P). We have :

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (b - Ax) = (c - \lambda - A^T \nu)^T x + \nu^T b \tag{1}$$

with $\lambda \in \mathbb{R}^d$ and $\nu \in \mathbb{R}^n$.

To find $g(\lambda, \nu)$ the Lagrange dual function, we need to minimize the Lagrangian in respect to x. To do that, the coefficient of x need to vanish, it is to say : $c - \lambda - A^T \nu = 0 \Leftrightarrow \lambda = c - A^T \nu$.

In this case, $L(x, \lambda, \nu) = \nu^T b = b^T \nu$

Thus, the dual problem is :

$$\begin{aligned} \max_{\nu} \quad & b^T \nu \\ \text{s.t.} \quad & A^T \nu \leq c \end{aligned} \tag{2}$$

2. The Lagrangian of the problem (D) is :

$$L(y, \lambda) = -b^T y + \lambda^T (A^T y - c) = (-b + A\lambda)^T y - \lambda^T c \tag{3}$$

with $\lambda \in \mathbb{R}^d$.

As in the previous question, the minimum of the Lagrangian in respect to y happend when the coefficient of y vanishes, meaning : $A\lambda = b$.

In this case, $L(x, \lambda, \nu) = -\lambda^T c = -c^T \lambda$

Thus, the dual problem is given by :

$$\begin{aligned} \min_{\lambda} \quad & c^T \lambda \\ \text{s.t.} \quad & A\lambda = b, \\ & \lambda \geq 0 \end{aligned} \tag{4}$$

3. The Lagrangian of the self-dual problem is :

$$\begin{aligned} L(x, y, \lambda_1, \lambda_2, \nu) &= c^T x - b^T y - \lambda_1^T x + \lambda_2 (A^T y - c) + \nu (b - Ax) \\ &= (c - \lambda_1 - A^T \nu)^T x + (A\lambda_2 - b)^T y - \lambda_2^T c + \nu^T b \end{aligned} \tag{5}$$

It comes that the Lagrangian is minimal in respect to x and y when both coefficients on x and y vanish, which gives two conditions : $\lambda_1 = c - A^T\nu$ and $A\lambda_2 = b$.

The Lagrange dual function is then :

$$g(\lambda, \nu) = \nu^T b - \lambda^T c = b^T \nu - c^T \lambda \tag{6}$$

Thus, the dual problem is :

$$\begin{aligned} \max_{\lambda,\nu} \quad & b^T\nu - c^T\lambda \\ \text{s.t.} \quad & \lambda \geq 0 \\ & A\lambda = b \\ \lambda_1 \geq 0 \Leftrightarrow & A^T\nu \leq c \end{aligned} \tag{7}$$

If we change it for a minimization problem, we have the original problem. This shows that the problem (Self-Dual) is indeed self dual.

4. Assuming that the (Self-Dual) problem is feasible and bounded, let $[x^*, y^*]$ be its optimal solution. The problem (P) respects the Slater's constraint qualification as it is a linear thus convex problem. This implies that the strong duality holds for the problem (P).

   This means that both $[x^*, y^*]$ satisfies the constraints of the problem (P) for x^* and the constraints of the problem (D) for y^*. Moreover, x^* must be an optimal solution to (P) and y^* to (D) (consequence of the strong duality). This shows that the vector $[x^*, y^*]$ can also be obtained by solving the problems (P) and (D)

   To solve the optimal value, we need to solve

$$\min_{x,y} \quad c^T x^* - b^T y^* \tag{8}$$

   but we also have, from the strong duality of the problems (P) and (D) :

$$c^T x^* = b^T y^* \Leftrightarrow c^T x^* - b^T y^* = 0 \tag{9}$$

   So the optimal value is exactly 0.

Exercise 2 (Regularized Least-Square)

1. We need to compute the conjugate of $||x||_1$, meaning $f^*(y) = \sup_{x\in\mathbb{R}^d}(y^T x - ||x||_1)$. To improve readability, we will note the l_1-norm $||.||$ instead of $||.||_1$.

   To compute f^*, we will perform a case analysis. Let $y \in \mathbb{R}^d$:

   - **Case** $||y||_\infty \leq 1$: As the l_1-norm can be expressed as $\max_{||p||<1} x^T p > x^T y$.

     Then :

  * $x^T y - ||x|| \leq 0, \quad \forall x \in \mathbb{R}^d,$
  * $x^T y - ||x|| \leq 0, \quad \text{at} \quad x = 0$

  Therefore, in this case, $f^*(y) = 0$
- **Case** $||y||_\infty > 1$: We can rewrite $x = ||x|| \frac{x}{||x||} = ||x|| x_0$.
  Then, $y^T x - ||x|| = ||x||(y^T x_0 - 1)$ with $y^T x_0 - 1 > 0$ so the maximum is clearly $+\infty$ as it is maximizing $||x||$ over $\mathbb{R}^d$.

Thus, the conjugate of $||x||$ is :

$$f^*(y) = \begin{cases} 0 & \text{if } ||y||_\infty \leq 1, \\ +\infty & \text{otherwise} \end{cases} \tag{10}$$

2. Let us begin by noticing that, if we introduce the variable $z = Ax - b$, the optimization problem (RLS) can be rewrite as :

$$\begin{aligned} \min_x \quad & ||z||_2^2 + ||x||_1 \\ \text{s.t.} \quad & z = Ax - b \end{aligned} \tag{11}$$

In this case, the Lagrangian is :

$$L(x, z, \nu) = ||z||_2^2 + ||x||_1 + \nu^T (Ax - b - z) \tag{12}$$

To compute the dual problem, we need to solve two optimization problems : $\min_x(||x||_1 + \nu^T Ax)$ and $\min_z(||z||_2^2 - \nu^T z)$

- $\min_x(||x||_1 + \nu^T Ax)$
  That problem can be solved by using the conjugate of the l_1-norm, found in the previous question. In fact, we have the equivalent problem: $\max_x(-\nu^T Ax - ||x||_1)$.
  The solution is then equal to 0 when $||\nu^T A||_\infty \leq 1$.
- $\min_z(||z||_2^2 - \nu^T z)$
  To solve this problem, we can developp the expression of the l_2-norm and noticing that it can be expressed as a distance :

$$\begin{aligned} ||z||_2^2 - \nu^T z =& z^T z - \nu^T z \\ =& \left(z - \frac{\nu}{2}\right)^T \left(z - \frac{\nu}{2}\right) - \frac{1}{4}||\nu||_2^2 \end{aligned} \tag{13}$$

  As the first is a distance, it is positive, which means that the solution of the optimization problem on z is reached when $z = 0$ and its minimum is $-\frac{1}{4}||\nu||_2^2$

If we use the two results showed above, the Lagrange dual function is $g(\nu) = -\frac{1}{4}||\nu||_2^2 - \nu^T b$ and the dual problem of (RLS) is

$$\begin{aligned} \max_\nu \quad & -\frac{1}{4}||\nu||_2^2 - \nu^T b \\ \text{s.t.} \quad & ||\nu^T A||_\infty \leq 1 \end{aligned} \tag{14}$$

Exercise 3 (Data Separation)

1. First of all, multiplying the problem (Sep 2) by τ does not change it.

   The new variable introduced z handle the loss function of the problem (Sep 1), specifically since $z \geq 0$ and :

   $\forall i, z_i \geq 1 - y_i(\omega^T x_i)$

   We observe that

   - If $1 - y_i(\omega^T x_i) \geq 0$, $z_i \geq 1 - y_i(\omega^T x_i)$
   - At the contrary, if $1 - y_i(\omega^T x_i) \leq 0$, $z_i = 0$.

   Thus, z_i represents the loss term $\max(0; 1 - y_i(\omega^T x_i))$ for each data point.

   Moreover, as we are summing all the components of z, we are minimizing $\sum_{i=1}^n z_i$ over ω and z while satisfying the constraints, we are replicating the loss function in problem (Sep 1).

   Thus, the optimization problem (Sep 2) indeed solves the problem (Sep 1)

2. Let us compute the Lagrangian of the problem (Sep 2).

   Let $\lambda = (\lambda_1, \ldots, \lambda_n), \pi \in \mathbb{R}^n$.

$$\begin{aligned} L(z, \omega, \lambda, \pi) &= \frac{1}{n\tau}\mathbf{1}^T z + \frac{1}{2}||\omega||_2^2 + \sum_{i=1}^n \lambda_i (1 - y_i(\omega^T x_i) - z_i) - \pi^T z \\ &= (\frac{1}{n\tau}\mathbf{1} - \lambda - \pi)^T z + \frac{1}{2}||\omega||_2^2 + \sum_{i=1}^n \lambda_i - \sum_{i=1}^n \lambda_i y_i (\omega^T x_i) \end{aligned} \tag{15}$$

   To mininize L w.r.t z and ω, both the coefficient of z and the derivative of L w.r.t ω need to be null.

   - The coefficient of z null gives the condition : $\frac{1}{n\tau}\mathbf{1} - \lambda - \pi = 0$.
     With $\pi \geq 0$ for the dual problem, this gives us $\lambda_i \leq \frac{1}{n\tau}$ for all $i \in \{1, \ldots, n\}$.
   - Let us compute the derivative of L w.r.t ω:
     $\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \lambda_i y_i x_i = 0 \Leftrightarrow \omega = \sum_{i=1}^n \lambda_i y_i x_i$

   Thus, the Lagrange dual function is :

$$g(\lambda) = \frac{1}{2}\left|\left| \sum_{i=1}^n \lambda_i y_i x_i \right|\right|_2^2 + \sum_{i=1}^n \lambda_i - \sum_{i=1}^n \sum_{j=1}^n \lambda_i y_i \lambda_j y_j x_j^T x_i \tag{16}$$

   We can observe that :

   $\left|\left| \sum_{i=1}^n \lambda_i y_i x_i \right|\right|_2^2 = \left(\sum_{i=1}^n \lambda_i y_i x_i \right)^T \sum_{i=1}^n \lambda_i y_i x_i = \sum_{i=1}^n \sum_{j=1}^n \lambda_i y_i \lambda_j y_j x_j^T x_i$

Thus, $g(\lambda) = \sum_{i=1}^n \lambda_i - \frac{1}{2}\left|\left| \sum_{i=1}^n \lambda_i y_i x_i \right|\right|_2^2$

To conclude, the dual problem of (Sep 2) is :

$$\begin{aligned} \max_{\lambda} \quad & \sum_{i=1}^n \lambda_i - \frac{1}{2}\left|\left| \sum_{i=1}^n \lambda_i y_i x_i \right|\right|_2^2 \\ \text{s.t.} \quad & 0 \leq \lambda_i \leq \frac{1}{n\tau} \quad \forall i = 1, \dots, n \end{aligned} \tag{17}$$