MVA - Convex Optimization
DM2

Ariane ALIX

November 4th, 2019

Exercise 1 (LP Duality)

1. We can write the first constraint of the problem (P) as $Ax - b = 0$, and the second as $-x \preceq 0$ Therefore, the Lagrangian associated to the problem is :

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b)$$

The dual function g is defined as

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \quad L(x, \lambda, \nu) \\ &= \inf_x \quad c^T x - \lambda^T x + \nu^T (Ax - b) \\ &= \inf_x \quad -\nu^T b + (c^T - \lambda^T + \nu^T A)x \end{aligned}$$

We can see there that if $c^T - \lambda^T + \nu^T A$ is not equal to zero, then g tends to $-\infty$ when $x \to -\infty$ or $x \to +\infty$ depending on the sign. Therefore, we have :

$$g(\lambda, \nu) = \begin{cases} -\nu^T b, & \text{if } (c^T - \lambda^T + \nu^T A) = 0 \\ -\infty, & \text{else} \end{cases}$$

Since the dual problem as $\lambda \geq 0$ as a constraint, and that we need $c^T - \lambda^T + \nu^T A = 0$, the constraint can also be written as $c^T + \nu^T A \geq 0$. Finally, we get the Lagrange dual problem :

$$\begin{aligned} &\max_\nu -\nu^T b \\ \text{s.t} \quad & c + A^T \nu \succeq 0 \end{aligned}$$

2. Problem (D) can also be written as :

$$\min_{y} -b^T y$$
$$\text{s.t} \quad A^T y - c \preceq 0$$

We then have :

$$L(y, \lambda, \nu) = -b^T y + \lambda^T (A^T y - c)$$

And, with the same reasoning as for the problem (P), the dual function g is defined as :

$$g(\lambda, \nu) = \begin{cases} -\lambda^T c, & \text{if } (-b^T + \lambda^T A^T) = 0 \\ -\infty, & \text{else} \end{cases}$$

And the dual problem is :

$$\begin{aligned} &\max_{\lambda} -\lambda^T c \\ \text{s.t} \quad &\lambda \succeq 0 \\ &A\lambda = b \end{aligned}$$

3. The problem can also be written as:

$$\begin{aligned} &\min_{x,y} c^T x - b^T y \\ \text{s.t} \quad &Ax - b = 0 \\ &-x \preceq 0 \\ &A^T y - c \preceq 0 \end{aligned}$$

The corresponding Lagrangian is :

$$\begin{aligned} L((x,y), \lambda, \nu) &= c^T x - b^T y - \lambda_d^T x + \lambda_n^T (A^T y - c) + \nu^T (Ax - b) \\ &= (c^T - \lambda_d^T + \nu^T A)x + (-b^T + \lambda_n^T A^T)y - \lambda_n^T c - \nu^T b \\ &= (c - \lambda_d + A^T \nu)^T x + (A\lambda_n - b)^T y - \lambda_n^T c - \nu^T b \end{aligned}$$

For a clearer notation, we separated λ in two parts λ_d and λ_n of respective dimensions d and n, so that it matches the dimensions of x and y and we do not have to stack the vectors.

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x,y} \quad L((x,y), \lambda, \nu) \\ &= \begin{cases} -\lambda_n^T c - \nu^T b, & \text{if } (c - \lambda_d + A^T \nu) = 0 \text{ and } (A\lambda_n - b) = 0 \\ -\infty, & \text{else} \end{cases} \end{aligned}$$

The condition $(\lambda \geq 0)$ of the dual problem is equivalent to $(\lambda_d \geq 0$ and $\lambda_n \geq 0)$.

We can then consider this problem has a sum of the two problems (P) and (D) since it is indepednant in x and y. As a consequence, the dual problem can be written as :

$$
\begin{aligned}
&\max_{\lambda,\nu} -\nu^T b - \lambda^T c \\
\text{s.t} \quad & A\lambda = b \\
& \lambda \succeq 0 \\
& c + A^T \nu \succeq 0
\end{aligned}
$$

By passing to the min and replacing ν by $-\nu$, it is equivalent to :

$$
\begin{aligned}
&\min_{\lambda,\nu} -\nu^T b + \lambda^T c \\
\text{s.t} \quad & A\lambda = b \\
& \lambda \succeq 0 \\
& A^T \nu \preceq c
\end{aligned}
$$

which is the problem itself. Therefore it is self-dual.

4. We assume that the problem from **3.** is feasible and bounded and that $[x^*, y^*]$ is its optimal solution. Since the problem is convex (affine constraints and sum of affine functions to minimize), and feasible, the strong duality is guaranteed by the Slater's constraint.

- The third problem can be considered as the sum of problems (P) and (D) : its objective function is minimum when the objectives function of (P) and (D) are minimum. As a consequence, x^* is the solution of (P) and y^* is the solution of (D).
- For the problem (P), we have :
  - Primal and dual feasability.
  - Complementary slackness for $x = 0$ since $\lambda^T 0 = 0$.
  - The gradient of the Lagrangian vanishing for all x including $x = 0$ if we have $c^T - \lambda^T + \nu^T A = 0$ (condition in exercise **1.**).

  Therefore $x = 0$ and λ and ν such that $c^T - \lambda^T + \nu^T A = 0$ verify the Karush-Kuhn-Tucker conditions, and are optimal since the problem is convex. Therefore, the optimal value is 0.

  We notice that (D) is the dual problem of (P); it means by strong duality that (D) has the same optimal value : 0.

  As a consequence, the optimal value of the third problem is $0 + 0 = 0$.

Exercise 2 (Regularized Least-Square)

1. Let f be the function defined as $f(x) = \|x\|_1$. Its conjugate f^* is defined as :

$$\begin{aligned} f^*(y) &= \sup_{x \in \mathrm{dom} f} (y^T x - \|x\|_1) \\ &= \sup_{x \in \mathbb{R}^d} (y^T x - \sum_{i=1}^{d} |x_i|) \end{aligned}$$

- If there is a compononent y_i of y such as $y_i > 1$, then by looking at the x of the form $x = u\mathbf{e_i}$ with $u > 0$, we have : $f^*(y) = (y_i - 1)u$ that tends to $+\infty$ when $u \to +\infty$.
- If there is a compononent y_i of y such as $y_i < 1$, then by looking at the x of the form $x = u\mathbf{e_i}$ with $u < 0$, we have : $f^*(y) = (y_i + 1)u$ that tends to $+\infty$ when $u \to -\infty$.
- If $-1 \preceq y \preceq 1$, then $f^*(y) = \sup_{x \in \mathbb{R}^d}(\sum_{i=1}^{d} |y_i||x_i| - \sum_{i=1}^{d} |x_i|)$ (by choosing the signs of the x_i equal to the ones of y_i).

  From there we deduce that $f^*(y) = \sup_{x \in \mathbb{R}^d}(\sum_{i=1}^{d}(|y_i| - 1)|x_i|) = 0$

To conclude, we can define f^* as :

$$f^*(y) = \begin{cases} 0, & \text{if } -1 \preceq y \preceq 1 \\ +\infty, & \text{otherwise} \end{cases}$$

2. We introduce a variable y so that we can write the (RLS) as :

$$\begin{aligned} &\min_{x,y} \|y\|_2^2 + \|x\|_1 \\ \text{s.t} \quad & y = Ax - b \end{aligned}$$

As a consequence, the Lagrangian can be written as :

$$L((x, y), \lambda, \nu) = \|y\|_2^2 + \|x\|_1 + \nu^T(y + b - Ax)$$

And the dual function :

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x,y} \|y\|_2^2 + \|x\|_1 + \nu^T(y + b - Ax) \\ &= \inf_{x,y} \|y\|_2^2 + \nu^T(y + b) + \|x\|_1 - \nu^T Ax \end{aligned}$$

We notice that the two last terms correspond to the opposite of the conjugate of the norm L_1 :

$$\begin{aligned} \inf_x \|x\|_1 - \nu^T Ax &= -\sup_x -\|x\|_1 + \nu^T Ax \\ &= \begin{cases} 0, & \text{if } -1 \preceq A^T\nu \preceq 1 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

Therefore we can write the dual problem as :

$$\max_{\nu} \|y\|_2^2 + \nu^T(y+b)$$
$$\text{s.t} \quad A^T\nu \preceq 1$$

We could also simplify the dual function by looking at the terms in y. If we compute the gradient, we have $\nabla\|y\|_2^2 + \nu^T y = 0 \Leftrightarrow y = -\frac{\nu}{2}$; hence :

$$\inf_{y} \|y\|_2^2 + \nu^T y = -\frac{1}{4}\|\nu\|_2^2$$

Finally, the dual problem of (RLS) can be written as :

$$\max_{\nu} -\frac{1}{4}\|\nu\|_2^2 + \nu^T b$$
$$\text{s.t} \quad -1 \preceq A^T\nu \preceq 1$$

Exercise 3 (Data Separation)

1. From $(Sep.1)$ we can construct the equivalent problem (in terms of optimal value, and solution to a factor τ) :

$$\min_{\omega,z} \frac{1}{n\tau}\sum_{i=1}^{n} z_i + \frac{1}{2}\|\omega\|_2^2$$
$$\text{s.t} \quad z_i = \mathcal{L}(\omega, x_i, y_i) \quad \forall i = 1...n$$

by dividing the objectiving fucntion by τ and introducing the dummy variable z equal to the loss. The problem can also be written as :

$$\min_{\omega,z} \frac{1}{n\tau}\mathbf{1}^T z + \frac{1}{2}\|\omega\|_2^2$$
$$\text{s.t} \quad z_i = \max\{0; 1 - y_i(\omega^T x_i)\} \quad \forall i = 1...n$$

By definition of $\mathcal{L}$, $\forall i = 1...n$ we have $z_i \geq 0$ and $z_i \geq 1 - y_i(\omega^T x_i)$. Since the goal is to minimize the function based on the sum of z_i and that these inequalities provide the same lower bounds as the definition of $\mathcal{L}$, solving a problem with these 2 inequalities as constraints instead of the one on $z = \mathcal{L}$ would drive the same results.
Therefore $(Sep.2)$ solves $(Sep.1)$.

2. The Lagrangian of $(Sep.2)$ is :

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

And the dual function is :

$$g(\lambda, \pi) = \inf_{\omega} \frac{1}{n\tau}\mathbf{1}^T z + \frac{1}{2}\|\omega\|_2^2 + \sum_{i=1}^{n} \lambda_i(1 - y_i(\omega^T x_i) - z_i) - \pi^T z$$

Let g_ω be the part of g with terms in ω :

$$g_\omega(\lambda, \pi) = \inf_{\omega} \frac{1}{2}\|\omega\|_2^2 - \sum_{i=1}^{n} \lambda_i y_i(\omega^T x_i)$$

We then look at its gradient :

$$\nabla g_\omega(\lambda, \pi) = \omega - \sum_{i=1}^{n} \lambda_i y_i x_i$$

Since it is equal to 0 for $\omega = \sum_{i=1}^{n} \lambda_i y_i x_i$, we can deduce that :

$$g_\omega(\lambda, \pi) = -\sum_{i=1}^{n} \frac{1}{2}\lambda_i^2 y_i^2 \|x_i\|_2^2$$

Let g_z be the part of g with terms in z :

$$
\begin{aligned}
g_z(\lambda, \pi) &= \inf_z \frac{1}{n\tau}\mathbf{1}^T z - \sum_{i=1}^{n}(\lambda_i z_i) - \pi^T z \\
&= \inf_z (\frac{1}{n\tau}\mathbf{1} - \lambda - \pi)^T z \\
&= \begin{cases} 0, & \text{if } (\frac{1}{n\tau}\mathbf{1} - \lambda - \pi) \geq 0 \\ -\infty, & \text{otherwise } (z \text{ being positive}) \end{cases}
\end{aligned}
$$

Therefore, by plugging these 2 results in the g function, and simplifying the condition $(\frac{1}{n\tau}\mathbf{1} - \lambda - \pi) = 0$ to $(\frac{1}{n\tau}\mathbf{1} - \lambda) = 0$ since π does not appear in the rest and λ can play its role, we get the dual problem :

$$
\begin{aligned}
&\max_{\lambda, \pi} - \sum_{i=1}^{n} \frac{1}{2}\lambda_i^2 y_i^2 \|x_i\|_2^2 + \sum_{i=1}^{n} \lambda_i \\
\text{s.t} \quad & \frac{1}{n\tau}\mathbf{1} - \lambda \geq 0
\end{aligned}
$$