Homework 2 Convex Optimization

For Monday the 4th of November

Exercise 1 (LP Duality)

For given $c \in \mathbb{R}^d$, $b \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n \times d}$, consider the two following linear optimization problems,

$$\begin{aligned} \text{(P)} \quad & \min_x c^T x \\ & \text{s.t.} \quad Ax = b, \\ & x \geq 0, \end{aligned}$$

and

$$\begin{aligned} \text{(D)} \quad & \max_y b^T y \\ & \text{s.t.} \quad A^T y \leq c. \end{aligned}$$

1.) *Compute the dual of problem (P) and simplify it if possible.*

Problem (P) is LP program which constraint inequalities are given by the functions $f_i : x \mapsto -x_i$ for $1 \leq i \leq d$ with x_i the ith component of x.

In order to form the Lagrangian of problem (P) we introduce the constraint Lagrange multipliers λ_i for the inequality constraints and the ν_j for the equality constraints. We can write $\nu^T = (\nu_1, \ldots, \nu_n)$. We have

$$L(x, \lambda, \nu) = c^T x - \Sigma_{i=1}^d \lambda_i x_i + \nu^T (Ax - b) = -b^T + \left(c + A^T \nu - \lambda\right)^T x$$

The dual function is

$$\begin{aligned} g(\lambda, \nu) &= \underset{x}{inf}\{-b^T \nu + \left(c + A^T \nu - \lambda\right)^T x\} && (1) \\ &= -b^T \nu + \underset{x}{inf}\{\left(c + A^T \nu - \lambda\right)^T x\} && (2) \end{aligned}$$

At this stage there are two options

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

Indeed if the factor $A^T \nu - \lambda + c$ is not equal to 0 it's clear that playing on the signs of the component of x and take those components really big all the expression converges to $-\infty$.
If it's equal to 0, then the expression is written $-b^T \nu$. In that case we can write the dual problem as being

$$\begin{aligned} & \max_\nu -b^T \nu \\ \text{subject to } & A^T \nu - \lambda + c = 0 \end{aligned}$$

we know that the solution of that problem will be less than the solution $p^\star$ of the original problem.
2.) *Compute the dual of problem (D).*
The problem (D) can be written as follows

$$\begin{aligned} \text{(D)} \quad & \min_y -b^T y \\ & \text{s.t.} \quad A^T y - c \leq 0 \end{aligned}$$

Similarly to **1.)** we introduce the constraint Lagrange multipliers λ_i for the inequality constraints. There's no equality constraints. Hence the Lagrangian of the problem (D) (re-written) can be written as follows

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

The dual function is

$$g(\lambda) = -\lambda^T c + \underset{x}{inf}\{(-b + A\lambda)^T y\}$$

So as similar to our resolution in **1.)** there are two options

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

In the first case we know that a lower bound of our problem (D) is $-\lambda^T c$

$$\max_{\lambda} -\lambda^T c$$
$$\text{subject to } A\lambda = b$$

3.) *A problem is called self-dual if its dual is the problem itself. Prove that the following problem is self-dual.*

$$\text{(Self-Dual)} \quad \min_{x,y} c^T x - b^T y$$
$$\text{s.t.} \quad Ax = b,$$
$$x \geq 0,$$
$$A^T y \leq c.$$

We can re write the problem as follows
We write

$$X = \begin{pmatrix} x \\ y \end{pmatrix}$$

and

$$K = \begin{pmatrix} c \\ -b \end{pmatrix}$$

Also we note $h_1, \dots, h_n$ the functions corresponding to the equality constraint $Ax = b$ so that for each $1 \leq i \leq n$

$$h_i : X \mapsto \sum_{j=1}^{d} a_{ij} X_j - b_j = \sum_{j=1}^{d} a_{ij} x_j - b_j$$

This corresponds to multiply X with a bigger matrix of dimension $n \times (d+n)$ taking from A:

$$\begin{pmatrix} A & 0 \end{pmatrix}$$

Also the first d constraint functions which we write f_i for $1 \leq i \leq d$ corresponding to $x \leq 0$ correspond to
$f_i : X \mapsto -X_i (= x_i)$
and the n constraint functions that follows which we write f_{d+i} for $1 \leq i \leq n$
$f_{d+i} : X \mapsto \sum_{j=1}^{n} a_{ji} X_{d+j} - c_j = \sum_{j=1}^{n} a_{ji} y_j - c_j$
This corresponds to multiply X with a bigger matrix using A^T

$$\begin{pmatrix} 0 & A^T \end{pmatrix}$$

We will write $C = \begin{pmatrix} 0 \\ \dots \\ 0 \\ c_1 \\ \dots \\ c_d \end{pmatrix}$ and $B = \begin{pmatrix} b_1 \\ \dots \\ b_n \\ 0 \\ \dots \\ 0 \end{pmatrix}$ and $\lambda = \begin{pmatrix} \lambda_1 & \dots & \lambda_d & 0 & \dots & 0 \end{pmatrix}$

The Lagrangian of this problem is

$$\begin{aligned} L(X, \lambda, \nu) &= K^T X + \nu^T \left(\begin{pmatrix} 0 & A^T \end{pmatrix} X - C \right) - \lambda X + \mu^T \left(\begin{pmatrix} A & 0 \end{pmatrix} X - B \right) \\ &= \left(K^T + \nu^T \begin{pmatrix} 0 & A^T \end{pmatrix} - \lambda + \mu^T \begin{pmatrix} A & 0 \end{pmatrix} \right) X - \nu^T C - \mu^T B \end{aligned}$$

With $\nu \geq 0$

So by expliciting the dual function g of that problem and assembling the Lagragian parameters for the inequality constraints as $\begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_d \\ \nu_1 \\ \dots \\ \nu_n \end{pmatrix}$, we got

$$g(\mu, \begin{pmatrix} \lambda \\ \nu \end{pmatrix}) = \begin{cases} -\nu^T C - \mu^T B & \text{if } K^T + \nu^T \begin{pmatrix} 0 & A^T \end{pmatrix} - \lambda^T + \mu^T \begin{pmatrix} A & 0 \end{pmatrix} = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

The first condition can be written differently consider the first d rows and the n rows that follow

$$\begin{cases} c^T - \begin{pmatrix} \lambda_1 & \dots & \lambda_d \end{pmatrix} + \begin{pmatrix} \mu_1, \dots, \mu_n \end{pmatrix}^T A & = 0 \\ -b^T + \begin{pmatrix} \nu_1 & \dots & \nu_d \end{pmatrix} A^T & = 0 \end{cases}$$

Which can be also written when we take the transpose:

$$\begin{cases} A^T \begin{pmatrix} \mu_1, \dots, \mu_n \end{pmatrix} & = c + \begin{pmatrix} \lambda_1 & \dots & \lambda_d \end{pmatrix}^T \\ A \begin{pmatrix} \nu_1 & \dots & \nu_d \end{pmatrix}^T & = b \end{cases}$$

As we took $\begin{pmatrix} \lambda_1 & \dots & \lambda_d \end{pmatrix} \geq 0$, and remembering that $\nu \geq 0$ we can now say that the self dual problem is similar to

in the non trivial case maximizing $-\nu^T C - \mu^T B$ (or transposing and minimizing the opposite $c^T \begin{pmatrix} \nu_1 \\ \dots \\ \nu_n \end{pmatrix} + b^T \begin{pmatrix} \mu_1 \\ \dots \\ \mu_d \end{pmatrix}$ [1]

(resp. in the trivial case: $g(\mu, \begin{pmatrix} \lambda \\ \nu \end{pmatrix}) = -\infty$ with the same following constraints)

$$\text{s.t} \begin{cases} A^T \begin{pmatrix} \mu_1, \dots, \mu_n \end{pmatrix} & \leq c \\ A \begin{pmatrix} \nu_1 & \dots & \nu_d \end{pmatrix}^T & = b \\ \begin{pmatrix} \nu_1 \\ \dots \\ \nu_d \end{pmatrix} \geq 0 \end{cases}$$

Which is exactly the original (Self Dual) problem itself !

[1] probably there's a way to have the exact signs of the self dual problem at the beginning

4.) *Assume the above problem is feasible and bounded, and let $[x^*, y^*]$ be its optimal solution. Using the strong duality property of linear programs, show that*

- *the vector $[x^*, y^*]$ can also be obtained by solving (P) and (D),*
- *the optimal value of (Self-Dual) is exactly 0.*

The strong duality theorem tells us that if there exist feasible primal (P) and dual (D) solutions then there exist feasible primal and dual solutions which have the same objective value. In our case that means that $c^T x^\star = b^T y^\star$ because they are optimal respectively for the problem (P) and (D).
And so $c^T x^\star - b^T y^\star = 0$. So this is an achievable value. We know by the weak duality theorem that $c^T x \geq b^T y$ for any feasible point (x, y) of the self-dual problem.

So knowing this, we see that $(x^\star, y^\star)$ are optimal for the self dual problem. Thus the objective value of the Self-Dual problem is 0

In the other way round, if $(x^\star, y^\star)$ achieves the objective value of 0 it follows that

1. $x^\star$ satisfies $Ax^\star = b$ and $x^\star \geq 0$ so $x^\star$ is feasible for (P)
2. $y^\star$ satisfies $A^T y^\star \leq c$ and so $y^\star$ is feasible for (D)

So we can obtain them by solving (P) and (D).

Exercice 2 (Regularized Least Square)

1.) *Compute the conjugate of* $||x||_1$ Following the definition of what a conjugate is

$$f^\star(y) = \underset{x}{sup}\left(y^T x - ||x||_1\right)$$

If $||y||_{+\infty}(=\max_{1\leq i\leq n}(|y_i|)\leq 1$, then for each i, $y_i x_i \leq |x_i|$ then when summing

$$y^T x \leq ||x||_1 = \sum_{i=1}^{n} |x_i|$$

hence

$$y^T x - ||x||_1 \leq 0$$

That bound is reached in 0
If $||y||_{+\infty} > 1$ then $f^\star = +\infty$ So

$$f^\star(y) = \begin{cases} 0, & \text{if } \|y\|_\infty \leq 1, \\ +\infty, & \text{if } \|y\|_\infty > 1. \end{cases}$$

2.) *Compute the dual of (RLS)*
Let's turn that problem into an optimization problem with constraints. (RLS) is actually the same thing that the following problem (*)

$$\begin{cases} \min_{x,y} ||y||_2^2 + ||x||_1 \\ Ax - b = y \end{cases}$$

So the Lagrangian of that problem is

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

So as the dual Lagrange function is $g\left(\lambda\right) = \inf_{x,y} L(x, y, \lambda)$
We can minimize $||y||_2^2 - \lambda^T y$ then $||x||_1 + \lambda^T(Ax - b)$

$$||y||_2^2 - \lambda^T y = ||y - \frac{\lambda}{2}||_2^2 - \frac{\lambda^2}{4} \geq -\frac{\lambda^2}{4}$$

Reached for $y = \frac{\lambda}{2}$ So knowing this the dual Lagrangian can be written
$g(\lambda) = -\frac{\lambda^2}{4} + \inf\left(||x||_1 + \lambda^T(Ax - b)\right) = -\frac{\lambda^2}{4} - \lambda^T b + \inf\left(||x||_1 + \lambda^T Ax\right)$
but we know from the conjugate of $||x||_1$ that

$$\inf ||x||_1 + \lambda^T Ax = \sup -\lambda^T Ax - ||x||_1 = \begin{cases} 0, & \text{if } \|A^T\lambda\|_\infty \leq 1, \\ -\infty, & \text{if } \|A^T\lambda\|_\infty > 1. \end{cases}$$

And so when $\|A^T\lambda\|_\infty \leq 1$ the dual Lagrangian is not trivial ($-\infty$) and can be written

$$g(\lambda) = -\frac{\lambda^2}{4} - \lambda^T b$$

So the dual problem is

$$\begin{cases} \max_\lambda -\frac{\lambda^2}{4} - \lambda^T b \\ \|A^T\lambda\|_\infty \leq 1 \end{cases}$$

Exercice 3 (Data Separation)

1.) We are asked to explain why Problem (Sep. 2) is equivalent to Problem (Sep. 1). Problem (Sep. 2) is defined as:

$$\min_{\omega,z} \frac{1}{n\tau}1^T z + \frac{1}{2}\|\omega\|_2^2,$$

subject to:

$$z_i \geq 1 - y_i(\omega^T x_i), \quad \forall i = 1, \dots, n, \quad z \geq 0.$$

The loss $\mathcal{L}(\omega, x_i, y_i) = \max\{0, 1 - y_i(\omega^T x_i)\}$ can be expressed using an auxiliary variable z_i for each i. By setting $z_i \geq 1 - y_i(\omega^T x_i)$ and $z_i \geq 0$, we ensure that z_i captures the loss value:

$$z_i = \max\{0, 1 - y_i(\omega^T x_i)\}.$$

The objective in (Sep. 2) minimizes the average loss term $\frac{1}{n}\sum_{i=1}^n z_i$ scaled by $\frac{1}{\tau}$ and adds the regularization term $\frac{1}{2}\|\omega\|_2^2$. This matches the objective in (Sep. 1) when we rewrite the loss using z_i and scale it appropriately.
Therefore, Problem (Sep. 2) is equivalent to Problem (Sep. 1), as it captures both the original loss and regularization terms while ensuring the constraints define the hinge loss behavior.

2.) Let's compute the dual of (Sep. 2). We introduce dual variables $\lambda_i \geq 0$ for each constraint $z_i \geq 1 - y_i(\omega^T x_i)$ and $\pi_i \geq 0$ for the constraint $z_i \geq 0$.

The Lagrangian $\mathcal{L}$ is:

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

we first minimize $\mathcal{L}$ with respect to z. We take the gradient search when it's equal to zero components by components:

$$\frac{\partial \mathcal{L}}{\partial z_i} = \frac{1}{n\tau} - \lambda_i - \pi_i = 0 \Rightarrow \lambda_i = \frac{1}{n\tau} - \pi_i.$$

Since $\lambda_i \geq 0$ and $\pi_i \geq 0$, this imposes the constraint $0 \leq \lambda_i \leq \frac{1}{n\tau}$. So to recap, if we want to minimize $\mathcal{L}$, we minimize it over z. That's possible when $\lambda_i = \frac{1}{n\tau} - \pi_i$ by the way which bonds λ and π.

This condition given, we can re write our Lagrangian like that

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

Then we minimize with respect to ω:

$$\frac{\partial \mathcal{L}}{\partial \omega} = \omega - \sum_{i=1}^n \lambda_i y_i x_i = 0 \Rightarrow \omega = \sum_{i=1}^n \lambda_i y_i x_i.$$

Now we will substitute $\omega = \sum_{i=1}^n \lambda_i y_i x_i$ into the Lagrangian.
Before that let's just remember that $y_i \in \{-1, 1\}$, so

$$\begin{aligned}\sum_{i=1}^n \lambda_i y_i(\omega^T x_i) &= \omega^T \sum_{i=1}^n \lambda_i y_i x_i) \\ &= \omega^T \omega \\ &= ||\omega||_2^2\end{aligned}$$

Now we put back the new expression of ω. We finally get

$$\mathcal{L}(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}||\omega||_2^2$$

So

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

The dual problem is then

$$\max_{\lambda} \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\left\|\sum_{i=1}^{n} \lambda_i y_i x_i\right\|_2^2,$$
$$\text{subject to} \quad 0 \leq \lambda_i \leq \frac{1}{n\tau} \quad \forall i.$$

Exercice 4 Robust Linear Programming

First, let's use the hint and see that the constraint $\sup_{a\in\mathcal{P}} a^T x \leq b$ can be actually interpreted as first trying to maximize the $a^T x$ if x is fixed and play on a afterwards; which is similar to maximize $x^T a$ for x fixed. Written like that we can see another LP appearing. Indeed it's exactly the LP (let's note it (2))

$$\begin{aligned}\max x^T a\\ C^T a \leq d\end{aligned}$$

Let's write it down.
Let's denote $f(x)$ the optimal solution of (2). The original problem (let's call it (1)) can be written

$$\begin{aligned}\min c^T x\\ f(x) \leq b\end{aligned}$$

Let's compute the Lagrange dual of (2). We done it in the question Ex I.2 (**??**) So this dual is

$$\begin{aligned}\min\ d^T z\\ \text{st } c^T z = x\\ z \geq 0\end{aligned}$$

$f(x)$ is the optimal value of this LP so we have $f(x) \leq b$ if and only if we can find z such as $d^T z \leq b$, $C^T z = x$, $z \geq 0$