MATHÉMATIQUES
VISION
APPRENTISSAGE

Convex Optimization - Homework 2

Louis Martinez
louis.martinez@telecom-paris.fr

Exercise 1: LP Duality

1)

The Lagrangian of (P) is:

$$\begin{aligned}\mathcal{L}(x,\lambda,\mu) &= c^T x - \lambda^T x + \mu^T (Ax - b) \\ &= -\mu^T b + (c - \lambda + A^T \mu)^T x\end{aligned}$$

We deduce the dual function g:

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

So the dual problem of (P) is:

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

We then perform the change of variable $\mu = -y$ and eliminate λ:

$\mu = -y \Rightarrow A^T\mu + c = -A^T y + c$ and $\lambda \geq 0 \Rightarrow -A^T y + c \geq 0 \Leftrightarrow A^T y \leq c$

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

We notice that the dual problem of (P) is (D)

2)

The Lagrangian of (D) is:

$$\begin{aligned}\mathcal{L}(y,\nu) &= -b^T y + \nu^T (A^T y - c) \\ &= -\nu^T c + (A\nu - b)^T y\end{aligned}$$

The dual function g is:

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

So the dual of the problem (D) is:

$$\boxed{\begin{aligned} &\max_\nu -\nu^T c \\ \text{s.t.}\ & \nu \geq 0 \\ & A\nu = b \end{aligned}}$$

Again we notice that the dual of (D) is (P), as maximizing $-\nu^T c$ is equivalent to minimizing $\nu^T c$.

3)

The Lagrangian of (Self-Dual) is:

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

The dual function is:

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

We deduce the dual problem:

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

We then eliminate λ as previously and perform the change of variables $(x, y) = (\nu, -\mu)$ to finally get:

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

Finally by changing the sign of the function to optimize, the maximization is turned into a minimization over x and y.
So:

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

We find back the primal problem.

$\boxed{\text{So the problem is self-dual.}}$

4)

Let $\tilde{x}, \tilde{y}$ the respective optimal solutions of (P) and (D). Then for any feasible point x and y for problems (P) and (D) the following inequalities hold:

$$\begin{cases} c^T \tilde{x} \leq c^T x \\ b^T \tilde{y} \geq b^T y \end{cases} \Leftrightarrow \quad c^T \tilde{x} - b^T \tilde{y} \geq c^T x - b^T y$$

So $[\tilde{x}, \tilde{y}]$ is an optimal solution of (Self-Dual), hence $[\tilde{x}, \tilde{y}] = [x^*, y^*]$.

$\boxed{\text{By solving } (P) \text{ and } (D) \text{ we get the solution of (Self-Dual)}}$

As strong duality holds for linear programs and (D) is the dual of (P) we have:

$$p^* = d^* \Leftrightarrow c^T x^* = b^T y^* \Leftrightarrow c^T x^* - b^T y^* = 0$$

.

$\boxed{\text{So the optimal value of (Self-Dual) is exactly } 0.}$

Exercise 2: Regularized Least-Square

1)

$$\begin{aligned} f_*(y) &= \sup_{x \in \mathbb{R}^d} \left(y^T x - f(x)\right) && \text{By definition} \\ &= \sup_{x \in \mathbb{R}^d} \left(y^T x - \|x\|_1\right) \\ &= \sup_{x \in \mathbb{R}^d} \sum_{i=1}^{d} (y_i x_i - |x_i|) \\ &= \sum_{i=1}^{d} \sup_{x_i \in \mathbb{R}} (y_i x_i - |x_i|) && \text{Each element of the sum can be optimized separately} \end{aligned}$$

Let $i \in (1, ..., d)$. We now determine the value of x_i maximizing the i-th term of the sum:
If $x_i \geq 0, |x_i| = x_i$ and $y_i x_i - |x_i| = x_i(y_i - 1)$:
 If $y_i > 1$ then $x_i = +\infty$
 If $y_i \leq 1$ then $x_i = 0$

If $x_i \leq 0, |x_i| = -x_i$ and $y_i x_i - |x_i| = x_i(y_i + 1)$:
 If $y_i < -1$ then $x_i = +\infty$
 If $y_i \geq -1$ then $x_i = 0$

In order of the sum to have a finite value, we must have $y_i \in [-1, 1]$ for $i = 1, ...d$, or, equivalently, $\|y\|_\infty < 1$. In this case $f_*(y) = 0$. So we have:

$$f_*(y) = \begin{cases} 0 & \text{if } \|y\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

2)

We can rewrite (RLS) by introducing $y = Ax - b$ in the following way:

$$\begin{gathered} \min_{x,y} y^T y + \|x\|_1 \\ \text{s.t. } y = Ax - b \end{gathered}$$

The Lagrangian of the problem is:

$$\begin{aligned} \mathcal{L}(x, y, \lambda) &= y^T y + \|x\|_1 + \lambda^T (Ax - b - y) \\ &= y^T I y - \lambda^T y + \|x\|_1 + \lambda^T A x - \lambda^T b \end{aligned}$$

- On the one hand, $\mathcal{L}$ is a quadratic form with respect to y. Since $I \succeq 0$, it is bounded below and minimized by $y = -\frac{1}{2} I^{-1}(-\lambda) = \frac{1}{2}\lambda$.
- On the other hand,

$$\inf_x \left(\|x\|_1 - \left(-A^T \lambda\right)^T x\right) = -\sup_x \left(\left(-A^T \lambda\right)^T x - \|x\|_1\right) = -f_*(-A^T \lambda)$$

So we have:

$$
\begin{aligned}
g(\lambda) &= \inf_{x,y}(y^T I y - \lambda^T y + \|x\|_1 + \lambda^T A x - \lambda^T b) \\
&= \frac{1}{4}\lambda^T\lambda - \frac{1}{2}\lambda^T\lambda - \lambda^T b + \inf_x\Big(\|x\|_1 - \left(-A^T\lambda\right)^T x\Big) \\
&= -\frac{1}{4}\ \|\lambda\|_2^2 - \lambda^T b - f_*(-A^T\lambda) \\
&= \begin{cases} -\frac{1}{4}\ \|\lambda\|_2^2 - \lambda^T b & \text{if } \ \|A^T\lambda\|_\infty \leq 1 \\ -\infty & \text{otherwise} \end{cases}
\end{aligned}
$$

Finally the dual of (RLS) is:

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

Exercise 3: Data Separation

1)

Constraints on z in (Sep. 2) represent the loss function of (Sep. 1). If a data-point x_i is misclassified we have $1 - y_i(\omega^T x_i) < 0$. But the constraint $z \geq 0$ ensures that if sample x_i is a misclassified, then $z_i = 0$. On the other hand if x_i is well classified, $1 - y_i(\omega^T x_i) > 0$ and according to the constraints on z_i, we must have $z_i = 1 - y_i(\omega^T x_i)$ for the i-th to be minimal.

So by choosing $z_i = \max\{0; 1 - y_i(\omega^T x_i)\}$ we fall back to (Sep. 1). As τ is a constant, dividing (Sep. 1) by τ doesn't change the optimal value of ω.

2)

(For the sake of readability, we denote $\mathbf{1_n} = (1, ..., 1)^T \in \mathbb{R}^n$)
The Lagrangian of (Sep. 2) is:

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

- On the one hand, $\mathcal{G}$ is a quadric form bounded below with respect to ω. By setting the gradient with respect to ω to zero we get:

$$\begin{aligned}
\min_{\omega} \mathcal{G} &= \left(\frac{1}{n\tau}\mathbf{1_n} - \lambda - \pi\right)^T z + \frac{1}{2}\left\|\sum_{i=1}^{n} \lambda_i y_i x_i\right\|_2^2 - \left\|\sum_{i=1}^{n} \lambda_i y_i x_i\right\|_2^2 + \mathbf{1_n}^T \lambda \\
&= \left(\frac{1}{n\tau}\mathbf{1_n} - \lambda - \pi\right)^T z - \frac{1}{2}\|\sum_{i=1}^{n} \lambda_i y_i x_i\|_2^2 + \mathbf{1_n}^T \lambda
\end{aligned}$$

- On the other hand, $\mathcal{G}$ is linear with respect to z so:

$$\min_{z,\omega} \mathcal{G} = \begin{cases} -\frac{1}{2}\left\|\sum_{i=1}^{n} \lambda_i y_i x_i\right\|_2^2 + \mathbf{1_n}^T \lambda & \text{if } \frac{1}{n\tau}\mathbf{1_n} = \lambda + \pi \\ -\infty & \text{otherwise} \end{cases}$$

Finally the dual boils down to:

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

We can eliminate π, thus transforming the equality constraint into an inequality. We have:

$$\lambda, \pi \geq 0 \Rightarrow \begin{cases} \frac{1}{n\pi} \geq \lambda \\ \lambda \geq 0 \end{cases}$$

Finally the dual is:

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