Convex Optimization

Homework n°2

Exercise 1 (LP Duality)

1. The Lagrangian of (P) is given by
$\mathcal{L}(x;\lambda,\nu) = c^T x + \lambda^T(-x) + \nu^T(Ax-b) = (c-\lambda+A^T\nu)^T x - b^T\nu$ where $\lambda \in \mathbb{R}^d$, $\nu \in \mathbb{R}^n$.

We have the dual Lagrangian function:
$\mathcal{G}(\lambda,\nu) = \inf_{x\in\mathbb{R}^d} \mathcal{L}(x;\lambda,\nu) = \inf_{x\in\mathbb{R}^d} \left((c-\lambda+A^T\nu)^T x - b^T\nu\right) = \begin{cases} -b^T\nu & \text{if } A^T\nu + c - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$ (an affine function is bounded over $\mathbb{R}^d$ is bounded if and only if it is constant).
The dual problem of (P) is then

$$\begin{cases} \max\limits_{\lambda,\nu} \; -b^T\nu \\ \text{s.t. } A^T\nu + c - \lambda = 0 \text{ and } \lambda \geq 0 \end{cases} \iff \begin{cases} \max\limits_{\nu} \; -b^T\nu \\ \text{s.t. } A^T\nu \geq -c \end{cases} \overset{\nu\leftrightarrow-\nu}{\iff} \begin{cases} \max\limits_{\nu} \; b^T\nu \\ \text{s.t. } A^T\nu \leq c, \end{cases}$$

so $\boxed{\text{the dual of } (P) \text{ is (equivalent to) } (D)}$.

2. (D) is a maximization problem, which can be converted into a minimization problem by considering the opposite of the objective function.
The Lagrangian is then given by $\mathcal{L}(y;\lambda) = -b^T y + \lambda^T(A^T y - c) = (A\lambda - b)^T y - \lambda^T c$.

We have the dual Lagrangian function:
$\mathcal{G}(\lambda) = \inf_{y\in\mathbb{R}^n} \mathcal{L}(y;\lambda) = \inf_{y\in\mathbb{R}^n} \left((A\lambda-b)^T y - \lambda^T c\right) = \begin{cases} -\lambda^T c & \text{if } A\lambda = b \\ -\infty & \text{otherwise.} \end{cases}$

The dual problem of (D) is then equivalent to:

$$\begin{cases} \max\limits_{\lambda\in\mathbb{R}^d} \; -c^T\lambda \\ \text{s.t. } A\lambda = b \text{ and } \lambda \geq 0 \end{cases} \iff \begin{cases} \min\limits_{\lambda\in\mathbb{R}^d} \; c^T\lambda \\ \text{s.t. } A\lambda = b \text{ and } \lambda \geq 0. \end{cases}$$

so $\boxed{\text{the dual of } (D) \text{ is (equivalent to) } (P)}$.

3. The Lagrangian of (Self-Dual) is:

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

We have the dual Lagrangian function:

$$\begin{aligned} \mathcal{G}(\lambda_1,\lambda_2,\nu) &= \inf_{x,y} \mathcal{L}(x,y,\lambda_1,\lambda_2,\nu) \\ &= \inf_{x,y} \left((A^T\nu + c - \lambda_1)^T x + (A\lambda_2 - b)^T y - b^T\nu - c^T\lambda_2\right) \\ &= \begin{cases} -b^T\nu - c^T\lambda_2 & \text{if } A^T\nu + c - \lambda_1 = 0 \text{ and } A\lambda_2 - b = 0 \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The dual problem of (Self-Dual) is then:

$$\left\{\begin{array}{l}\max\limits_{\lambda_1,\lambda_2,\nu} \ -b^T\nu - c^T\lambda_2 \\ \text{s.t. } A^T\nu + c - \lambda_1 = 0 \text{ and } A\lambda_2 = b \\ \lambda_1 \geq 0, \lambda_2 \geq 0\end{array}\right. \iff \left\{\begin{array}{l}\min \ b^T\nu + c^T\lambda_2 \\ \text{s.t. } A^T\nu + c - \lambda_1 = 0 \text{ and } A\lambda_2 = b \\ \lambda_1 \geq 0, \lambda_2 \geq 0\end{array}\right.$$

$$\overset{\nu \leftrightarrow -\nu}{\iff} \left\{\begin{array}{l}\min \ c^T\lambda_2 - b^T\nu \\ \text{s.t. } A^T\nu = c - \lambda_1 \text{ and } A\lambda_2 = b \\ \lambda_1 \geq 0, \lambda_2 \geq 0\end{array}\right.$$

$$\iff \left\{\begin{array}{l}\min \ c^T\lambda_2 - b^T\nu \\ \text{s.t. } A^T\nu \leq c \text{ and } A\lambda_2 = b \\ \lambda_2 \geq 0.\end{array}\right.$$

So the dual problem of (Self-Dual) is itself: $\boxed{\text{(Self-Dual) is self-dual}}$.

4.

- We first show that x^* is solution of (P). Let x be such that $Ax = b$, $x \geq 0$ (x exists because (Self-Dual) is feasible).

  Then (x, y^*) is feasible for (Self-Dual), therefore $c^Tx^* - b^Ty^* \leq c^Tx - b^Ty^*$, or equivalently $c^Tx^* \leq c^Tx$.

  Since this holds for any x feasible for (P), we then have that $\boxed{x^* \text{ is solution of } (P)}$.

  Similarly, $\boxed{y^* \text{ is solution of } (D)}$.

  The converse also holds: if $\tilde{x}$ and $\tilde{y}$ solve (P) and (D), then they are feasible for (Self-Dual), and for any (x, y) feasible for (Self-Dual), x is feasible for (P) so $c^T\tilde{x} \leq c^Tx$, and y is feasible for (D) so $b^T\tilde{y} \geq b^Ty$, from which we deduce that $c^T\tilde{x} - b^T\tilde{y} \leq c^Tx - b^Ty$, so $(\tilde{x}, \tilde{y})$ solves (Self-Dual).

  So $\boxed{(x^*, y^*) \text{ can be obtained by solving } (P) \text{ and } (D)}$.

- (P) is a linear problem (the constraints are affine, and so is the objective function), which is feasible, because (Self-Dual) is. And (D) is the dual of (P).

  (D) is also feasible, but we do not need it to apply strong duality.

  So by strong duality (which holds for any feasible LP, which can be seen as a consequence of Slater's theorem), (P) and (D) have the same value.

  Therefore since x^* is optimal for (P) and y^* is optimal for (D), we have $c^Tx^* = \text{val}(P) = \text{val}(D) = b^Ty^*$, or equivalently $c^Tx^* - b^Ty^* = 0$.

  And (x^*, y^*) is optimal for (Self-Dual), so $\boxed{\text{the optimal value for (Self-Dual) is } 0}$.

Exercise 2 (Regularized Least-Square)

1. Let $\boxed{f(x) = \|x\|_1}$ for $x \in \mathbb{R}^d$.

By definition, for $y \in \mathbb{R}^d$, $f^*(y) = \sup\limits_{x \in \mathbb{R}^d} \big(y^T x - f(x)\big) = \sup\limits_{x \in \mathbb{R}^d} \sum\limits_{i=1}^d (y_i x_i - |x_i|)$.

- First case: there exists $i \in [\![1, d]\!]$ such that $|y_i| > 1$.

  Take $x_i = \mathrm{sign}(y_i) t$ $(t \geq 0)$ and $x_k = 0$ for $k \neq i$.

  Then $y^T x - f(x) = |y_i|\, t - t = t\underbrace{(|y_i| - 1)}_{>0} \underset{t \to +\infty}{\longrightarrow} +\infty$, so $f^*(y) = +\infty$.

- Second case: $\forall i \in [\![1, d]\!], |y_i| \leq 1$.

  We then have, for every $x \in \mathbb{R}^d$ and every $i \in [\![1, d]\!]$ $y_i x_i \leq |y_i x_i| \leq |x_i|$ so $y_i x_i - |x_i| \leq 0$, and therefore $y^T x - f(x) \leq 0$, with equality when for instance $x = 0$. Therefore $f^*(y) = 0$.

So we have shown:

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

2. (RLS) is equivalent to $\min\limits_{x,y} \|y\|_2^2 + \|x\|_1$ s.t. $y = Ax - b$.

Then the Lagrangian is $\mathcal{L}(x, y; \nu) = \|y\|_2^2 + \|x\|_1 + \nu^T(Ax - y - b)$.

The dual Lagrangian function is:

$$\begin{aligned} \mathcal{G}(\nu) &= \inf_{x \in \mathbb{R}^d, y \in \mathbb{R}^n} \big(\|y\|_2^2 + \|x\|_1 + \nu^T(Ax - y - b)\big) \\ &= \inf_{x,y} \big((\|x\|_1 + \nu^T Ax) + (\|y\|_2^2 - \nu^T y) - \nu^T b\big) \\ &= \inf_x \big(\|x\|_1 + \nu^T Ax\big) + \inf_y \big(\|y\|_2^2 - \nu^T y\big) - \nu^T b \end{aligned}$$

The function $y \mapsto \|y\|_2^2 - \nu^T y$ is convex (for instance because its hessian is $2I_d$), with gradient $2y - \nu$ which is 0 for $y = \nu/2$, so $\inf\limits_y \big(\|y\|_2^2 - \nu^T y\big) = \|\nu\|_2^2/4 - \|\nu\|_2^2/2 = -\|\nu\|_2^2/4$.

And using the previous question:

$\inf\limits_x \big(\|x\|_1 + \nu^T Ax\big) = -\sup\limits_x \big(-(A^T\nu)^T x - \|x\|_1\big) = -f^*(-A^T\nu) = \begin{cases} 0 & \text{if } \|A^T\nu\|_\infty \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$

So $\mathcal{G}(\nu) = \begin{cases} -\|\nu\|_2^2/4 - \nu^T b & \text{if } \|A^T\nu\|_\infty \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$

Then the dual problem of (RLS) is

$$\begin{aligned} &\max_\nu \; -\|\nu\|^2/4 - \nu^T b \\ &\text{s.t.} \;\; \|A^T\nu\|_\infty \leq 1, \end{aligned}$$

or equivalently:

$$\boxed{\begin{aligned} &\min_\nu \; \|\nu\|^2/4 + \nu^T b \\ &\text{s.t.} \;\; \|A^T\nu\|_\infty \leq 1. \end{aligned}}$$

Exercise 3 (Data Separation)

1. (Sep. 1) is of course equivalent (in terms of optimal ω) to:

$$\min_{\omega} \frac{1}{n\tau}\sum_{i=1}^{n} \max\left\{0; 1 - y_i(\omega^T x_i)\right\} + \frac{1}{2}\|\omega\|_2^2.$$

• Given ω, and setting $z_i = \max\left\{0; 1 - y_i(\omega^T x_i)\right\}$ for each i, we have

$$\frac{1}{n\tau}\sum_{i=1}^{n} \max\left\{0; 1 - y_i(\omega^T x_i)\right\} + \frac{1}{2}\|\omega\|_2^2 = \frac{1}{n\tau}\mathbf{1}^T z + \frac{1}{2}\|\omega\|_2^2$$

and ω, z are feasible for (Sep. 2). Therefore the optimal value for (Sep. 2) is no larger than the optimal value for (Sep. 1).

• Conversely, if (ω, z) is feasible for (Sep. 2), then (using (λ_i) and (π)) we have $\forall i, z_i \geq \max\left\{0; 1 - y_i(\omega^T x_i)\right\}$, and

$$\frac{1}{n\tau}\sum_{i=1}^{n} \underbrace{\max\left\{0; 1 - y_i(\omega^T x_i)\right\}}_{\leq z_i} + \frac{1}{2}\|\omega\|_2^2 \leq \frac{1}{n\tau}\mathbf{1}^T z + \frac{1}{2}\|\omega\|_2^2,$$

which shows that the optimal value for (Sep. 1) is no larger than the optimal value for (Sep. 2).

So $\boxed{\text{(Sep. 2) solves (Sep. 1)}}$ in the sense that (ω, z) is solution to (Sep. 2) if and only if ω is solution to (Sep. 1) and $\forall i, z_i = \max\left\{0; 1 - y_i(\omega^T x_i)\right\}$.

Remark: Going from (Sep. 2) to (Sep. 1) is essentially just using the **epigraph formulation**.

2. The Lagrangian of (Sep. 2) is

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

If $\frac{1}{n\tau}\mathbf{1} - \lambda - \pi \neq 0$ then the first term is a non-zero linear function of z, so the dual function satisfies $\mathcal{G}(\lambda, \pi) = -\infty$.

Otherwise, the Lagrangian is a convex function, and we have $\nabla_\omega \mathcal{L}(\omega, z; \lambda, \pi) = \omega - \sum_{i=1}^{n} \lambda_i y_i x_i$ which is 0 when $\omega = \sum_{i=1}^{n} \lambda_i y_i x_i$, so:

$$\mathcal{G}(\lambda, \pi) = -\frac{1}{2}\left\|\sum_{i=1}^{n} \lambda_i y_i x_i\right\|^2 + \lambda^T \mathbf{1} = -\frac{1}{2}\sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T x_j + \lambda^T \mathbf{1}.$$

Therefore

$$\mathcal{G}(\lambda, \pi) = \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T x_j + \lambda^T \mathbf{1} & \text{if } \frac{1}{n\tau}\mathbf{1} - \lambda - \pi = 0 \\ -\infty & \text{otherwise,} \end{cases}$$

so the dual of (Sep. 2) is equivalent to

$$
\begin{aligned}
&\max_{\lambda,\pi} -\frac{1}{2}\sum_{i,j=1}^{n} \lambda_i\lambda_j y_i y_j x_i^T x_j + \lambda^T \mathbf{1} \\
\text{s.t.}\quad & \lambda + \pi = \frac{1}{n\tau}\mathbf{1}, \lambda \succeq 0, \pi \succeq 0,
\end{aligned}
$$

or equivalently:

$$
\boxed{
\begin{aligned}
&\max_{\lambda} \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{n} \lambda_i\lambda_j y_i y_j x_i^T x_j \\
\text{s.t.}\quad & \frac{1}{n\tau}\mathbf{1} \succeq \lambda \succeq 0.
\end{aligned}
}
$$

Optional Exercise 4 (Robust linear programming)

I may be mistaken, but I think it should be C instead of C^T in the definition of $\mathcal{P}$ if we want to find the LP given in the exercise.

We want to find a more explicit formulation of $\sup_{a \in \mathcal{P}} a^T x \leq b$.
Fix x, and consider the LP

$$\begin{aligned} \max_{a} \ & a^T x \\ \text{s.t.} \ & Ca \preceq d \end{aligned} \tag{P}$$

This is a maximization problem, with Lagrangian
$\mathcal{L}(a; \lambda) = -a^T x + \lambda^T (Ca - d) = \left(C^T \lambda - x\right)^T a - \lambda^T d$, so the dual function is given by

$$\mathcal{G}(\lambda) = \begin{cases} -\lambda^T d & \text{if } C^T \lambda = x \\ -\infty & \text{otherwise,} \end{cases}$$

hence the dual problem

$$\begin{aligned} \min_{\lambda} \ & d^T \lambda \\ \text{s.t.} \ & C^T \lambda = x, \lambda \succeq 0. \end{aligned} \tag{D}$$

And if we assume that $\mathcal{P}$ is not empty, then (P) has a feasible a, so by strong duality for linear programs, (P) and (D) have same value, and the optimal value of (D) is attained.

Therefore $\sup_{a \in \mathcal{P}} a^T x \leq b \iff \text{val}(P) \leq b \iff \text{val}(D) \leq b \iff \exists \lambda \succeq 0, C^T \lambda = x, d^T \lambda \leq b$.
So the original problem is equivalent to

$$\boxed{\begin{aligned} \min_{x,\lambda} \ & c^T x \\ \text{s.t.} \ & d^T \lambda \leq b \\ & C^T \lambda = x \\ & \lambda \succeq 0. \end{aligned}}$$

Optional Exercise 5 (Boolean LP)

1. The Lagrangian is given by
$\mathcal{L}(x;\lambda,\nu) = c^Tx + \lambda^T(Ax-b) - \sum_{i=1}^n \nu_i x_i(1-x_i) = (c+A^T\lambda-\nu)^Tx + x^T\,\mathbf{diag}(\nu)x - \lambda^Tb$ (we chose a sign $-$ for ν for convenience).

We can also write $\mathcal{L}(x;\lambda,\nu) = \sum_{i=1}^n \left((c+A^T\lambda-\nu)_i x_i + \nu_i x_i^2\right) - \lambda^T b$.

- If $\exists i \in [n], \nu_i < 0$, then $\inf_x \mathcal{L}(x;\lambda,\nu) = -\infty$ (set $x_k = 0$ for $k \neq i$ and take the limit $x_i \to +\infty$).
- If $\exists i \in [n], (\nu_i = 0) \wedge (c+A^T\lambda-\nu)_i \neq 0$, then $\inf_x \mathcal{L}(x;\lambda,\nu) = -\infty$ (take $x_k = 0$ for $k \neq i$, and take the limit $x_i \to +\infty$ if $(c+A^T\lambda-\nu)_i < 0$, $x_i \to -\infty$ if $(c+A^T\lambda-\nu)_i > 0$).
- Otherwise, $\nabla_x^2\mathcal{L}(x;\lambda,\nu) = \mathbf{diag}(\nu) \succ 0$, so $\mathcal{L}$ is (strictly) convex in x for fixed λ,ν, and $\nabla_x\mathcal{L}(x;\lambda,\nu) = c + A^T\lambda - \nu + 2\,\mathbf{diag}(\nu)x$ which is 0 when $x_i = \dfrac{\left(\nu - c - A^T\lambda\right)_i}{2\nu_i}$ for i such that $\nu_i \neq 0$ (and the other components do not matter, because in the case we study, if $\nu_i = 0$ then $(c+A^T\lambda-c)_i = 0$, so $\left(\nabla_x\mathcal{L}(x;\lambda,\nu)\right)_i = 0$).

  Therefore, in this third case:
  $$\inf_x \mathcal{L}(x;\lambda,\nu) = \sum_{\substack{i\in[n]\\ \nu_i\neq 0}} \left(-\frac{(c+A^T\lambda-\nu)_i^2}{2\nu_i} + \frac{(c+A^T\lambda-\nu)_i^2}{4\nu_i}\right) - \lambda^Tb = -\lambda^Tb - \sum_{\substack{i\in[n]\\ \nu_i\neq 0}} \frac{(c+A^T\lambda-\nu)_i^2}{4\nu_i}.$$

So the dual Lagrangian is given by

$$\mathcal{G}(\lambda,\nu) = \begin{cases} -\lambda^Tb - \sum\limits_{\substack{i\in[n]\\ \nu_i\neq 0}} \dfrac{(c+A^T\lambda-\nu)_i^2}{4\nu_i} & \text{if } \nu \succeq 0 \text{ and } \forall i \in [n], \nu_i = 0 \Rightarrow (c+A^T\lambda-\nu)_i = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

We rewrite it an easier way, by adopting the convention that $x/0 = +\infty$ if $x > 0$, and $x/0 = 0$ if $x = 0$. Then the dual becomes:

$$\begin{aligned} \max_{\lambda,\nu} \quad & -\lambda^Tb - \sum_{i=1}^n \frac{(c+A^T\lambda-\nu)_i^2}{4\nu_i} \\ \text{s.t.} \quad & \nu \succeq 0, \lambda \succeq 0. \end{aligned} \tag{D}$$

We can optimize in ν for each coordinate, for a fixed λ.
For $a \in \mathbb{R}$, the derivative of the function $x \in \mathbb{R}_{++} \mapsto -\dfrac{(a-x)^2}{4x}$ is $x \mapsto \dfrac{a^2-x^2}{4x^2}$. If $a > 0$, the maximum of the function on $\mathbb{R}_{++}$ is reached in a, and is 0, which is still true for $a = 0$. Otherwise, the maximum of the function is in $x = -a$, and is a. So the maximum is $\min(0,a)$.

So (D) becomes

$$\boxed{\begin{aligned} \max_{\lambda} \quad & -\lambda^Tb + \sum_{i=1}^n \min\left(0, (c+A^T\lambda)_i\right) \\ \text{s.t.} \quad & \lambda \succeq 0. \end{aligned}}$$

2. For the LP relaxation, the Lagrangian is
$\mathcal{L}(x;\lambda,\nu,\omega) = c^T x + \lambda^T(Ax-b) - \nu^T x + \omega^T(x-\mathbf{1}) = (c + A^T\lambda - \nu + \omega)^T x - \lambda^T b - \omega^T \mathbf{1}$,
and the dual function is

$$\mathcal{G}(\lambda,\nu,\omega) = \begin{cases} -\lambda^T b - \omega^T \mathbf{1} & \text{if } c + A^T\lambda - \nu + \omega = 0 \\ -\infty & \text{otherwise,} \end{cases}$$

so the dual problem is

$$\begin{aligned} \max_{\lambda,\nu,\omega} \quad & -\lambda^T b - \omega^T \mathbf{1} \\ \text{s.t.} \quad & c + A^T\lambda - \nu + \omega = 0 \\ & \lambda \succeq 0, \nu \succeq 0, \omega \succeq 0, \end{aligned}$$

which can be rewritten

$$\begin{aligned} \max_{\lambda,\omega} \quad & -\lambda^T b - \omega^T \mathbf{1} \\ \text{s.t.} \quad & \omega \succeq -c - A^T\lambda, \omega \succeq 0 \\ & \lambda \succeq 0. \end{aligned}$$

The two conditions on ω are equivalent to $\forall i, \omega_i \geq \max(0, -(c + A^T\lambda)_i)$ or
$\forall i, \omega_i \geq -\min(0, (c + A^T\lambda)_i)$. And $-\omega_i$ (which appears in the objective function) is maximal when the inequality becomes an equality (it is basically an epigraph formulation):

$$\boxed{\begin{aligned} \max_{\lambda} \quad & -\lambda^T b + \sum_{i=1}^{n} \min(0, (c + A^T\lambda)_i) \\ \text{s.t.} \quad & \lambda \succeq 0. \end{aligned}}$$

We recover the same lower bound as previously.