Convex Optimization - Homework de Charrin Théotime

06/11

Exercice 1 - LP Duality

1) (P) On reconnaît la forme classique d'un problème de programmation linéaire.

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

$\mathcal{L}$ est linéaire en x, d'où :

$$\inf_x \mathcal{L}(x, \lambda, \nu) = g(\lambda, \nu) = \begin{cases} -\nu^T b & \text{si } c + A^T\nu - \lambda = 0 \\ -\infty & \text{sinon.} \end{cases}$$

g est bien concave.

Le problème dual revient donc à maximiser $g(\lambda, \nu)$ s.c. $\lambda \succeq 0$

i.e maximiser $-\nu^T b$ sc. $\begin{cases} A^T\nu + c - \lambda = 0 \\ \lambda \succeq 0 \end{cases} \iff A^T\nu + c \succeq 0$

on remplace ν par $-\nu$:

On veut maximiser $\nu^T b$ sc $-A^T\nu \succeq -c \iff A^T\nu \preceq c$

On retrouve (D) car maximiser $\nu^T b$ revient à maximiser $b^T \nu$

2) On a $\max_y b^T y$ s.c $A^T y \leq c$ équivalent à :

$$\min_y -b^T y \quad \text{sc} \quad A^T y - c \leq 0$$

D'où $\mathcal{L}(x, \lambda) = -b^T y + \lambda^T(A^T y - c)$

$$= -\lambda^T c + (A\lambda - b)^T y.$$

$\mathcal{L}$ est linéaire en y, d'où

$$g(\lambda) = \inf_y \mathcal{L}(x, \lambda) = \begin{cases} -\lambda^T c & \text{si } A\lambda - b = 0 \\ -\infty & \text{sinon.} \end{cases}$$

g est bien concave.

Le problème dual de (D) revient donc à

maximiser $g(\lambda)$ sc $\lambda \succcurlyeq 0$

soit $\begin{cases} \text{maximiser} & -\lambda^T c \\ \text{minimiser} & \lambda^T c \end{cases}$ sc $\begin{cases} A\lambda - b = 0 \\ \lambda \succcurlyeq 0 \end{cases} \Leftrightarrow \begin{cases} A\lambda = b \\ \lambda \succcurlyeq 0 \end{cases}$ (P)

On retrouve bien (P), le dual du dual de (P) est lui-même.

3. On cherche $\min_{x,y} c^T x - b^T y$ sc $\{Ax = b \,;\, x \geq 0 \,;\, A^T y \leq c\}$.

$\Leftrightarrow \{Ax - b = 0 \,;\, -x \leq 0 \,;\, A^T y - c \leq 0\}$

On a $\mathcal{L}(x, \lambda_1, \lambda_2, \nu) = c^T x - b^T y - \lambda_1^T x + \lambda_2^T (A^T y - c) + \nu^T (Ax - b)$

On regroupe : $\mathcal{L} = -\nu^T b + (c + A^T \nu - \lambda_1)^T x - \lambda_2^T c + (A\lambda_2 - b) y$

On reconnaît la somme des lagrangiens de (P) et (D), indépendants.

D'où $\inf_{x,y} \mathcal{L} = \begin{cases} -(\nu^T b + \lambda_2^T c) & \text{si } c + A^T \nu - \lambda_1 = 0 \text{ et } A\lambda_2 - b = 0 \\ -\infty & \text{sinon} \end{cases} = g(\lambda_1, \lambda_2, \nu)$

On maximise $g(\lambda_1, \lambda_2, \nu)$ sc $\lambda_1 \geq 0$, $\lambda_2 \geq 0$

ie : $\max_{\lambda_1, \lambda_2, \nu} -(\nu^T b + \lambda_2^T c)$ sc $\begin{cases} c + A^T \nu - \lambda_1 = 0 \\ A\lambda_2 - b = 0 \\ \lambda_1 \geq 0 \quad \lambda_2 \geq 0 \end{cases} \Leftrightarrow \begin{cases} A^T \nu + c \geq 0 \\ A\lambda_2 = b \\ \lambda_2 \geq 0 \end{cases}$

En posant $\nu = -\nu$

$\max_{\cancel{\lambda_1}, \lambda_2, \nu} -(\lambda_2^T c - \nu^T b)$ sc $\begin{cases} A^T \nu \leq c \\ A\lambda_2 = b \\ \lambda_2 \geq 0 \end{cases}$

$(\lambda_2^T c - \nu^T b)^T = c^T \lambda_2 - b^T \nu$

On a donc comme problème dual :

$\min_{\cancel{\lambda_1}, \lambda_2, \nu} c^T \lambda_2 - b^T \nu$ sc $\begin{cases} A^T \nu \leq c \\ A\lambda_2 = b \\ \lambda_2 \geq 0 \end{cases}$ ...

On retrouve bien notre problème initial avec λ_2 dans le rôle de x et ν dans le rôle de y.

On a bien un "self-dual".

4) • On sait que les contraintes du problèmes self-dual sont disjointes, soit :

$$\{(x,y) \mid Ax=b,\ x\geq 0,\ A^Ty\leq c\} = \{x \mid Ax=b,\ x\geq 0\} \cup \{y \mid A^Ty\leq c\}$$

La minimisation de $c^Tx - b^Ty$ revient donc à minimiser de manière disjointe c^Tx et $-b^Ty$ sous leurs contraintes respectives.

Soit

$$(\text{self-dual}) = \begin{cases} \min_{x,y} c^Tx - b^Ty \\ \text{sc } Ax=b,\ x\geq 0,\ A^Ty\leq c \end{cases} = \begin{cases} \min_x c^Tx \\ \text{sc } \begin{array}{l} Ax=b \\ x\geq 0 \end{array} \end{cases} + \begin{cases} \min_y -b^Ty \\ \text{sc } A^Ty\leq c \end{cases}$$

$$= (P) + \begin{cases} \max b^Ty \\ \text{sc } A^Ty\leq c \end{cases}$$

$$= (P) + (D).$$

• On sait que (P) et (D) sont convexes car des formes particulières de programmation linéaire.

On suppose (self-dual) faisable et borné, (x^*, y^*) la solution optimale.

On peut donc dire que (P) est faisable et borné, avec une solution $x^* \geq 0$ optimale.

Les contraintes sur (P) sont linéaires, les conditions de Slater sont vérifiées donc (P) a une dualité forte.

Il en vient que $p^* = q^*$.

On a vu que $p^* = c^Tx^*$ et $q^* = b^Ty^*$

D'où $c^T x^* - b^T y^* = p^* - q^* = 0$

La valeur optimale du self-dual est bien 0.

Exercice 2 (RLS)

1) La fonction conjuguée f^* de $f(\cdot) = \|\cdot\|_1$ est définie par :

$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^T x - f(x) \right)$$

Avec $f(x) = \|x\|_1 = \sum_{i=1}^{d} |x_i|$

- Si $\|y\|_\infty \leq 1$:

$$y^T x - f(x) = y^T x - \sum_1^d |x_i| = \sum_1^d y_i x_i - \sum_1^d |x_i|$$

Or $\forall i \in [\![1; d]\!]$: $x_i y_i \leq |x_i y_i| \leq |x_i|$ car $|y_i| \leq 1 \ \forall i$.

D'où $y^T x - f(x) \leq \sum_1^d |x_i| - \sum_1^d |x_i| \leq 0$

On a $f^*(y) = \sup (y^T x - f(x)) = 0$ si $\|y\|_\infty \leq 1$

- Si $\|y\|_\infty > 1$

- Si $y_j > 1 \ \forall i$ et x particulier tq $x_j = t > 0$, $x_k = 0 \ \forall k \neq j$:

Alors $y^T x - f(x) = y_j t - t = t(y_j - 1) > 0$

$\lim_{t \to \infty} t(y_j - 1) = +\infty$.

- Avec le même raisonnement pour $y_j < 1$ et $x_j = t < 0$:

$$y^T x - f(x) = \underset{<0}{t} \underbrace{(y_j - 1)}_{<0} > 0$$

$\lim_{t \to +\infty} y^T - f(x) = +\infty$.

Notre x particulier appartient bien au domaine de f.

On a donc le conjugué de $\|x\|_1$:

$$f^*(y) = \begin{cases} 0 & \text{si } \|y\|_\infty \leq 1 \\ +\infty & \text{sinon} \end{cases}$$

2) On reformule comme slide 28 du cours sur la dualité en introduisant une variable y contrainte :

$$\min_x \|Ax-b\|_2^2 + \|x\|_1 \iff \min_{x,y} \|y\|_2^2 + \|x\|_1 \quad \text{sc} \quad y = Ax-b$$

- D'où $\mathcal{L}(x,y,\nu) = \|y\|_2^2 + \|x\|_1 + \nu^T(y - Ax + b)$.

$$= \|y\|_2^2 + \nu^T y + \|x\|_1 - \nu^T Ax + \nu^T b$$

Pour calculer le problème dual on cherche $g(\nu) = \inf_{x,y} \mathcal{L}(x,y,\nu)$.

$$\nabla_y \mathcal{L}(x,y,\nu) = \nabla_y \|y\|_2^2 + \nu^T y = 2y^T + \nu^T$$

On veut $\nabla_y \mathcal{L}(x,y,\nu) = 0 \iff 2y^T + \nu^T = 0$

$$\iff 2y + \nu = 0$$

$$\iff y = -\frac{1}{2}\nu$$

Pour $y = -\frac{1}{2}\nu$ on a $\|y\|_2^2 + \nu^T y = -\frac{1}{4}\|\nu\|_2^2$.

On cherche $\inf_x \mathcal{L}(x,y,\nu) = \inf_x (\|x\|_1 - \nu^T Ax) + \|y\|_2^2 + \nu^T y + \nu^T b$.

$$\inf_x (\|x\|_1 - \nu^T Ax) = \sup_x ((A^T\nu)^T x - \|x\|_1)$$

On reconnaît la norme duale de $\|\cdot\|_1$.

D'où $\sup_x ((A^T\nu)^T x - \|x\|_1) = \|A^T\nu\|_1^* = \begin{cases} 0 & \text{si } \|A^T\nu\|_\infty \leq 1 \\ +\infty & \text{sinon} \end{cases}$

On a donc $\inf_{x,y} \mathcal{L}(x,y,\nu) = -\frac{1}{4}\|\nu\|_2^2 + \nu^T b + \begin{cases} 0 \text{ si } \|A^T\nu\|_\infty \leq 1 \\ +\infty \text{ sinon} \end{cases}$

Soit $g(\nu) = -\frac{1}{4}\|\nu\|_2^2 + \nu^T b$ sc $\|A^T\nu\|_\infty \leq 1$

Le problème dual de (~~RLS~~) consiste à maximiser $g(\nu)$ ie :

$$\max_{\nu} -\frac{1}{4}\|\nu\|_2^2 + \nu^T b \quad \text{sc} \begin{cases} \|A^T\nu\|_\infty \leq 1 \\ \nu \geq 0 \end{cases}$$

Exercice 3 (Data separation)

1) Le problème (Sep. 2) revient à réaliser séparément :

$$\min_{w} \frac{1}{2}\|w\|_2^2 \quad \text{et} \quad \min_{z} \frac{1}{n\tau} 1^T z.$$

Le problème sur z est linéaire en z.

On a vu que $1 - y_i(w^T x_i) \leq 0$ si x_i est bien classifié, et ≥ 0 sinon.

Il vient que $z_i = 0$ si x_i est bien classifié et doit être supérieur ou égal à $1 - y_i(w^T x_i)$.

La minimisation du problème revient à minimiser z_i soit prendre :

$$z_i = \max(0, 1 - y_i(w^T x_i)) \quad \forall i \in [\![1;n]\!]$$

En réécrivant $1^T z = \sum_1^n z_i$, on a :

$$\min_{z} \frac{1}{n\tau} 1^T z = \min_{z} \frac{1}{n\tau}\sum_1^n z_i = \frac{1}{n\tau}\sum_1^n \max(0, 1 - y_i(w^T x_i)).$$

(Sep. 2) revient donc à minimiser :

$$\min_{w} \frac{1}{2}\|w\|_2^2 + \frac{1}{n\tau}\sum_1^n \max(0, 1 - y_i(w^T x_i))$$

$$= \min_{w} \frac{1}{n\tau}\sum_1^n \mathcal{L}(w, x_i, z_i) + \frac{1}{2}\|w\|_2^2.$$

Ce qui revient à résoudre (Sep. 1), à un facteur τ près, qui est >0.

2) Pour (Sep. 2), on a :

$$\mathcal{L}(w, z, (\lambda_i)_{1\le i\le n}, \pi) = \frac{1}{n\tau} \mathbb{1}^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i \left(1 - y_i (w^T x_i) - z_i\right) - \pi^T z$$

$$= \frac{1}{n\tau} \mathbb{1}^T z \underbrace{- \sum_{i=1}^{n} \lambda_i z_i}_{= \lambda^T z} - \pi^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i \left(1 - y_i (w^T x_i)\right)$$

avec $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$.

D'où $\mathcal{L} = \left(\frac{1}{n\tau}\mathbb{1} - \lambda - \pi\right)^T z + \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i \left(1 - y_i (w^T x_i)\right)$.

$\mathcal{L}$ est linéaire en z, on a donc $\inf_{w,z} \mathcal{L} = \begin{cases} \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i (1 - y_i (w^T x_i)) \\ \quad \text{si } \frac{1}{n\tau}\mathbb{1} - \lambda - \pi = 0 \\ 0 \text{ sinon} \end{cases}$

On cherche $\nabla_w \mathcal{L}$.

$$\nabla_w \mathcal{L} = \nabla_w \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \lambda_i \left(1 - y_i (w^T x_i)\right)$$

$$= w - \sum_{i=1}^{n} \lambda_i y_i x_i$$

$\nabla_w \mathcal{L} = 0 \iff w = \sum_1^n \lambda_i y_i x_i$

En $w_{min} = \sum_1^n \lambda_i y_i x_i$, on a :

$$\frac{1}{2}\|w_{min}\|_2^2 + \sum_{i=1}^{n} \lambda_i \left(1 - y_i (w^T x_i)\right) = \frac{1}{2}\left\|\sum \lambda_i y_i x_i\right\|_2^2 - \sum \lambda_i y_i \left(\sum \lambda_i y_i x_i\right)^T x_i + \sum \lambda_i$$

$$= \frac{1}{2}\left\|\sum \lambda_i y_i x_i\right\|_2^2 - \left\|\sum \lambda_i y_i x_i\right\|_2^2 + \lambda^T \mathbb{1}$$

$$= -\frac{1}{2}\|w_{min}\|_2^2 + \lambda^T \mathbb{1}$$

Il vient donc en combinant :

$$\inf_{w,\xi} \mathcal{L} = \begin{cases} -\frac{1}{2} \|\sum \lambda_i y_i x_i\|_2^2 + \lambda^T \mathbb{1} & \text{si } \frac{1}{n\tau} \mathbb{1} - \lambda - \Pi = 0 \\ -\infty & \text{sinon.} \end{cases}$$

Le problème dual revient à maximiser $g(\lambda)$ soit :

$$\max_{\lambda_i} -\frac{1}{2} \|\sum \lambda_i y_i x_i\|_2^2 + \lambda^T \mathbb{1} \quad \text{sc} \begin{cases} \frac{1}{n\tau} \mathbb{1} - \lambda - \Pi = 0 \\ \lambda \geqslant 0 \quad \Pi \geqslant 0 \end{cases}$$

$$\Leftrightarrow \max_{\lambda_i} -\frac{1}{2} \|\sum \lambda_i y_i x_i\|_2^2 + \lambda^T \mathbb{1} \quad \text{sc} \begin{cases} \frac{1}{n\tau} \mathbb{1} - \lambda \geqslant 0 \\ \lambda \geqslant 0 \end{cases}$$

$$\Leftrightarrow \max_{\lambda_i} -\frac{1}{2} \|\sum \lambda_i y_i x_i\|_2^2 + \lambda^T \mathbb{1} \quad \text{sc} \left\{ \frac{1}{n\tau} \mathbb{1} \geqslant \lambda \geqslant 0 \right.$$