▷ **Exercice 2.** Le but de cet exercice est de démontrer le théorème

Théorème 1. *s^* est solution du problème*

$$(P^{rc}) \begin{cases} \min q(s) = f + g^\top s + \frac{1}{2} s^\top H s \\ ||s||^2 \leq \delta, \end{cases}$$

si et seulement si $||s^*||^2 \leq \delta$ et il existe $\mu^* \geq 0$ tel que

1. *$(H + 2\mu^* I)s^* = -g$;*
2. *$\mu^*(||s^*||^2 - \delta) = 0$;*
3. *$H + 2\mu^* I$ est semi-définie positive.*

2.1. Démontrer le lemme

Lemme 1. *Soit q la forme quadratique $q(s) = g^\top s + \frac{1}{2} s^\top H s$, H symétrique, alors les assertions suivantes sont vraies :*

1. *q atteint un minimum si et seulement si H est semi-définie positive et $g \in \operatorname{Im} H$ et dans ce cas tout point solution de $Hs = -g$ est un minimum global de q.*
2. *q a un unique minimum si et seulement si H est définie positive.*

2.2. Démontrer le théorème.

Rque : $q(s) = f + g^T s + \frac{1}{2} s^T H s$ n'est pas nécᵗ convexe car on n'a pas fait l'hyp que H est semi-déf. pos.

$\mathcal{C} = \{ s / \|s\|^2 \leq \delta \}$ est boule fermée - compacte - convexe dans $\mathbb{R}^n$

par contre $q(s)$ est $\mathcal{C}^\infty$ sur $\mathbb{R}^n$, q admet un min global sur le compact $\mathcal{C}$. $(\mathcal{C}^0)$

Si HQC est validée en s^*, la CN1 (KKT) nous dit que

$$\nabla_s \mathcal{L}(s^*, \mu^*) = \nabla q(s^*) + \mu^* \nabla c(s^*) = (Hs^* + g) + \mu^*(2s^*) = 0$$

avec : $(s \in \mathbb{R}^n, \mu \in \mathbb{R}^+)$

et $c : \begin{array}{l} \mathbb{R}^n \longrightarrow \mathbb{R} \\ s \longmapsto \|s\|^2 - \delta \end{array}$

càd que $(H + 2\mu^* I) s^* = -g$ (pté ①)

et on a aussi les relations de compatibilité, à savoir

$$\mu^* (\|s^*\|^2 - \delta) = 0 , \quad \mu^* \geq 0 \quad (\text{pté ②})$$

Pour HQC. on $s \in \mathcal{C}$: càd que $\|s\|^2 \leq \delta$

$\nabla c(s) = 2s$ et c'est nul que si $s = 0$ mais dans ce cas là on est à l'intérieur du domaine $\mathcal{C}$

a) Dans le cas où s^* est à l'intérieur du domaine $\mathcal{C}$ on est sans contraintes, et alors le lemme s'applique. (H semi déf pos et $g \in \operatorname{Im} H$)

De plus, la CN1 (p.t. ②) implique que $\mu^* = 0$
et on retrouve bien l'équivalence avec les 3 relations.

① $H + 0I \geq 0$

② $\mu^* = 0$

③ $H s^* = -g$ $\quad$ ($g \in \mathrm{Im}\, H$, s^* réalise l'égalité)

b) si on est sur la frontière de $\mathcal{C}$, $\|s^*\|^2 = \delta \Rightarrow s \neq 0$
et HQC en s^* est validé.
et ① et ② sont validées (CN1 sous HQC)

La CN2 indique alors que

$$\langle \nabla^2_{ss} \mathcal{L}(s^*, \mu^*)\, d \mid d \rangle \geq 0 \quad \forall d \in T_L(\mathcal{C}, s^*)$$

$$T_L(\mathcal{C}, s^*) = \{ d \in \mathbb{R}^n \,/\, \langle \underbrace{\nabla c(s^*)}_{2s^*} \mid d \rangle = 0 \} = \{s^*\}^\perp$$

on a donc la semi définie positivité de $(H + 2\mu^* I) = \nabla^2_{ss}\mathcal{L}(s^*, \mu^*)$
sur le s.e.v $\{s^*\}^\perp$ (p.t. ③ mais en partie seul^t)

Vérifions alors, que $(H + 2\mu^* I)$ est nec^t semi-déf pos sur $\mathbb{R}^n$.

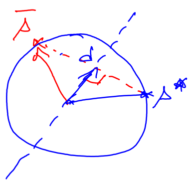
Considérons $d \in \mathbb{R}^n$ t.q. $d^T s^* \neq 0$.

et notons $t = -2\dfrac{d^T s^*}{\|d\|_2^2} \neq 0$

Posons $\bar{s} = s^* + td = s^* - \dfrac{2dd^T s^*}{\|d\|^2} = \underbrace{\left(I_n - 2\dfrac{dd^T}{\|d\|^2}\right)}s^*$

réflexion orthogonale au travers de l'hyperplan engendré par d

$$\|\bar{s}\|^2 = \left\| \left(I - \frac{2dd^T}{\|d\|^2}\right) s^* \right\|_2^2 = \|s^*\|^2 = \delta$$



$$q(\bar{s}) + \mu^* \|\bar{s}\|^2 = q(s^* + td) + \mu^* \|s^*\|^2$$
$$\geq q(s^*) + \mu^* \|s^*\|^2$$

puisque s^* réalise le minimum de $q(s)$ sur la sphère. (s^* est solution de P^{RC})

On développe :

$$q(s^*) + \langle \nabla q(s^*), td \rangle + \frac{1}{2}\langle \nabla^2 q(s^*)\, td, td \rangle + \mu^* \left(\|s^*\|^2 + 2\langle td, s^*\rangle + t^2 \|d\|^2 \right)$$

$$= q(s^*) + \langle \underbrace{Hs^* + g}, td\rangle + \frac{1}{2}t^2\langle Hd, d\rangle$$
$$+ \mu^* \underbrace{\|s^*\|^2}_{\delta} + \langle 2\mu^* s^*, td\rangle + \mu^* t^2\|d\|^2$$
$$= q(s^*) + \langle \underbrace{(H + 2\mu^* I)s^* + g}_{=0} \,|\, td\rangle + \underset{(CN1)}{\mu^* \delta} + \underbrace{\mu^* t^2\|d\|^2 + \frac{1}{2}t^2\langle Hd, d\rangle}$$
$$\geq q(s^*) + \mu^* \delta$$

On a donc $\qquad \mu^* t^2 \|d\|^2 + \frac{1}{2} t^2 d^T H d \geq 0$

ou encore : $\qquad \frac{1}{2} t^2 \langle \underbrace{(H + 2\mu^* I)} d, d\rangle \geq 0$

et on a donc vérifié la semi-définie positivité de

$\nabla^2_{ss} \mathcal{L}(s^*, \mu^*) = H + 2\mu^* I$ $\underline{\underline{\text{sur } \mathbb{R}^n}}$ c-à-d la p$^{\text{té}}$ ③.

$\Leftarrow$ Réciproquement, vérifions que si les 3 p$^{\text{tés}}$ sont validées en un couple (s^*, μ^*), $s^* \in \mathcal{C}$ et $\mu^* \geq 0$ alors s^* est une solution du problème P^{rc} $\left(\min_{\|s\|^2 \leq \delta} q(s)\right)$

Considérons $\tilde{q}(s) = q(s) + \mu^* \|s\|^2$

Si on applique le lemme à $\tilde{q}$ (forme quadratique)

$$\begin{cases} \nabla^2 \tilde{q}(s^*) = H + 2\mu^* I \geq 0 & \text{p}^{\text{té}} \text{ ③} \\ \nabla \tilde{q}(s^*) = (H + 2\mu^* I)s^* + g = 0 & \text{p}^{\text{té}} \text{ ①} \end{cases}$$

alors s^* est un min global de $\tilde{q}(s)$ sur $\mathbb{R}^n$

$\tilde{q}(s) - \tilde{q}(s^*) \geq 0 \qquad \forall s \in \mathbb{R}^n$

$\Rightarrow \quad q(s) - q(s^*) \geq \mu^* (\|s^*\|^2 - \|s\|^2)$

mais alors si $\|s^*\|^2 < \delta$, nec^t $\mu^* = 0$ | p$^{\text{té}}$ ②

et alors $q(s) \geq q(s^*)$

et si $\|s^*\|^2 = \delta$, $\mu^* \geq 0$

et dans ces conditions $\forall s \in \mathcal{C}$ $(\|s\|^2 \leq \delta)$

$\|s^*\|^2 - \|s\| \geq 0 \qquad (\mu^* \geq 0)$

ce qui implique que $q(s) \geq q(s^*)$

Donc dans tous les cas, on a vérifié que si $(s^*, \mu^*) \in \mathcal{C} \times \mathbb{R}^+$

vérifient les p^tés ① ② et ③ ; alors $q(s^*) \leq q(s), \forall s \in \mathcal{C}$
c.a.d que s^* est solution de P^{RC}.

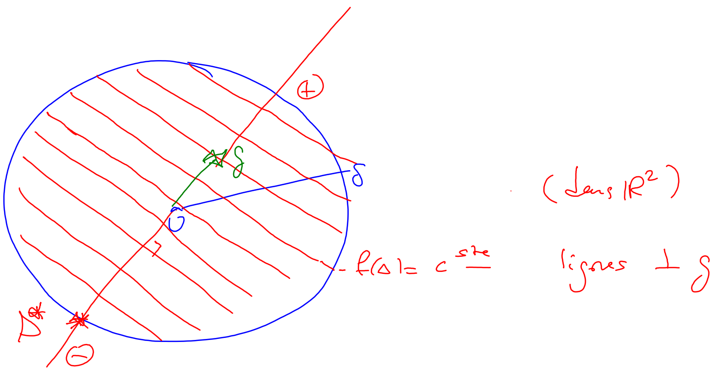
[d'où l'équivalence du Théorème]

Exo 1 (indications)

mardi 3 novembre 2020
11:20

$$\begin{cases} \operatorname{Min} \; f(s) = g^T s + c \\ \|s\|^2 \leq \delta \end{cases}$$

$$s^* = -\delta \frac{g}{\|g\|}$$



TD4 - Exo 1

mardi 3 novembre 2020
11:23

▷ **Exercice 1.** On s'intéresse ici à un cas simple des "Support Vector Machines". On considère dans $\mathbb{R}^n$ deux groupes de points $\mathcal{X}^1 = \{x^1_{1.}, \dots, x^1_{n_1.}\}$ et $\mathcal{X}^2 = \{x^2_{1.}, \dots, x^2_{n_2.}\}$ où $x^k_{i.} = (x^k_{i1}, \dots, x^k_{in}) \in \mathbb{R}^n$. On suppose que ces deux groupes de points sont séparables par un hyperplan affine de $\mathbb{R}^n$ (et non vides !). L'objectif est de trouver le "meilleur" hyperplan séparateur (cf. Figure 1). On désire donc trouver les hyperplans H_1 d'équation $\langle a\,,\,x\rangle = \alpha_1$ $(a \neq 0)$ et H_2 d'équation $\langle a\,,\,x\rangle = \alpha_2$ tels que :

- Pour tout $x \in \mathcal{X}^1$, $\langle a\,,\,x\rangle - \alpha_1 \geq 0$;
- Pour tout $x \in \mathcal{X}^2$, $\langle a\,,\,x\rangle - \alpha_2 \leq 0$;
- $d(H_1, H_2) = |\alpha_1 - \alpha_2|/||a||$ soit maximal.

On peut toujours en fait écrire les deux premières conditions de la façon suivante :

- H_1 d'équation $\langle \omega\,,\,x\rangle + b - 1 = 0$ et pour tout $x \in \mathcal{X}^1$, $\langle \omega\,,\,x\rangle + b \geq 1$;
- H_2 d'équation $\langle \omega\,,\,x\rangle + b + 1 = 0$ et pour tout $x \in \mathcal{X}^2$, $\langle \omega\,,\,x\rangle + b \leq -1$.

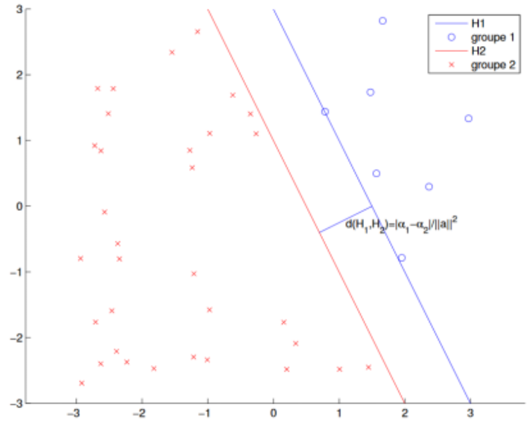


FIGURE 1 – *SMV pour* $n = 2$.

Le problème d'optimisation s'écrit alors

$$(P)\begin{cases}\text{Max}\frac{2}{||\omega||}\\ \langle \omega\,,\,x\rangle + b \geq 1, \forall x \in \mathcal{X}^1\\ \langle \omega\,,\,x\rangle + b \leq -1, \forall x \in \mathcal{X}^2\\ (\omega, b) \in \mathbb{R}^n \times \mathbb{R}\end{cases} \iff (P')\begin{cases}\text{Min}||\omega||^2\\ \langle \omega\,,\,x\rangle + b \geq 1, \forall x \in \mathcal{X}^1\\ \langle \omega\,,\,x\rangle + b \leq -1, \forall x \in \mathcal{X}^2\\ (\omega, b) \in \mathbb{R}^n \times \mathbb{R}\end{cases}$$

fct. quadratique en (ω, b)

} contraintes affines en (ω, b)

1.1. Montrer l'existence de solution.

1.2. Écrire le Lagrangien associé à ce problème

1.3. Donnez les conditions de (KKT) associés à ce problème. Ces conditions sont-elles ici des conditions nécessaires et suffisantes ?

On pose maintenant

$$D = \begin{pmatrix} -1 & & & & & \\ & \ddots & & & & \\ & & -1 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}, X = \begin{pmatrix} x^1_{11} & \dots & x^1_{1n} \\ \vdots & & \vdots \\ x^1_{n_1 1} & \dots & x^1_{n_1 n} \\ x^2_{11} & \dots & x^2_{1n} \\ \vdots & & \vdots \\ x^2_{n_2 1} & \dots & x^2_{n_2 n} \end{pmatrix}, e = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix},$$

D est de dimension $(n_1 + n_2, n_1 + n_2)$. Le problème (P') s'écrit alors

$$(P'')\begin{cases}\text{Min}||\omega||^2\\ DX\omega + bDe \leq -e\\ (\omega, b) \in \mathbb{R}^n \times \mathbb{R}.\end{cases}$$

1.4. Écrire le Lagrangien associé au problème (P'').

1.5. Écrire le problème dual du problème (P'')

1.1) existence de la solution :

• Vérifions que $f(\omega, b) = ||\omega||^2$

est croissante à l'∞ sur le domaine

$$\mathcal{C} = \left\{ (\omega, b) \;\middle/\; \begin{bmatrix} \langle \omega, x^i \rangle + b \geq 1 & \forall x^i \in \mathcal{X}^1 \\ \langle \omega, x^j \rangle + b \leq -1 & \forall x^j \in \mathcal{X}^2 \end{bmatrix} \right\}$$

- $\mathcal{C}$ est naturellement fermé et convexe en tant qu'intersection de demi-espaces affines

Pour $\|(\omega, b)\| \longrightarrow +\infty$

- soit $\|\omega\| \longrightarrow +\infty$ et naturellement $f(\omega,b) = \|\omega\|^2$ tend vers l'∞
- Soit $|b| \longrightarrow +\infty$ mais alors

① $b \geq 1 - \langle \omega, x^i \rangle \geq 1 - \|\omega\| \|x^i\|$ pour $x^i \in \mathcal{X}^1$ donné

② $b \leq -1 - \langle \omega, x^j \rangle \leq -1 + \|\omega\| \|x^j\|$ pour $x^j \in \mathcal{X}^2$ donné

d'où, si $b \longrightarrow -\infty$ ① $\Rightarrow \|\omega\| \longrightarrow +\infty$.
et si $b \longrightarrow +\infty$ ② $\Rightarrow \|\omega\| \longrightarrow +\infty$

et par conséquent $f(\omega, b) = \|\omega\|^2 \longrightarrow +\infty$.

On a donc la garantie de l'existence d'un min. global de f sur le fermé $\mathcal{C}$.

1.2) ⚠ mettre le pb sous forme canonique

$$\mathcal{L}(\omega, b, \mu^1, \mu^2) = \|\omega\|^2 - \sum_{i=1}^{n_1} \mu_i^1 (\langle \omega, x_i^1 \rangle + b - 1) + \sum_{j=1}^{n_2} \mu_j^2 (\langle \omega, x_j^2 \rangle + b + 1)$$

$$\mu^1 \in (\mathbb{R}^+)^{n_1}$$
$$\mu^2 \in (\mathbb{R}^+)^{n_2}$$

1.3) Les contraintes sont toutes de type affine, donc HQC est valide

$$(KKT) = \left[\begin{array}{lll} \nabla_{(\omega,b)} \mathcal{L}(\omega, b, \mu^1, \mu^2) = 0 & & \\ \langle \omega, x_i^1 \rangle + b \geq 1 & \forall i \in \{1 \dots n_1\} & \\ \langle \omega, x_j^2 \rangle + b \leq -1 & \forall j \in \{1 \dots n_2\} & \\ \mu_i^1 \geq 0 \quad , \quad \mu_j^2 \geq 0 & \forall i, j & \\ \mu_i^1 (\langle \omega, x_i^1 \rangle + b - 1) = 0 & \forall i & \multirow{2}{*}{\Big\} \text{ relation de compatibilité.}} \\ \mu_j^2 (\langle \omega, x_j^2 \rangle + b + 1) = 0 & \forall j & \end{array} \right.$$

$$\nabla_{(\omega,b)} \mathcal{L}(\omega, b, \mu^1, \mu^2) = \begin{array}{l} \partial_\omega \\ \partial_b \end{array} \begin{bmatrix} 2\omega - \sum_{i=1}^{n_1} \mu_i^1 x_i^1 + \sum_{j=1}^{n_2} \mu_j^2 x_j^2 \\ - \sum_{i=1}^{n_1} \mu_i^1 + \sum_{j=1}^{n_2} \mu_j^2 \end{bmatrix}$$

$$(P'')\begin{cases} \min \|\omega\|^2 \\ DX\omega + bDe \leq -e \\ (\omega,b)\in\mathbb{R}^n\times\mathbb{R}\end{cases} \qquad X = \begin{array}{c}\left(\begin{array}{c} \vdots \\ - x_i^{1\,T} - \\ \vdots \\ \hline - x_j^{2\,T} - \\ \vdots \end{array}\right)\begin{array}{l}\}①\\ \\ \}②\end{array}\end{array}$$

$$\mathcal{L}(\omega,b,\mu) = \|\omega\|^2 + \langle \mu, DX\omega + bDe + e\rangle$$

$$\mu\in(\mathbb{R}^+)^{n_1+n_2}$$

- $\mathcal{L}(\omega,b,\mu)$ est convexe en (ω,b)
et donc $\min_{(\omega,b)} \mathcal{L}(\omega,b,\mu) \Leftrightarrow \nabla_{(\omega,b)}\mathcal{L}(\omega,b,\mu) = 0$.
(μ fixé)

- Pour le Pb (P''). les contraintes sont affines, la fonctionnelle $\|\omega\|^2$ est convexe en (ω,b)
et donc HQC est valide et KKT deviennent une CNS d'optimalité globale
en outre, on a le th de dualité fort qui s'applique.

Regardons $\nabla_{(\omega,b)}\mathcal{L}(\omega,b,\mu) = \begin{array}{c}\partial\omega \\ \partial b\end{array}\begin{bmatrix} 2\omega + (DX)^T\mu \\ \langle\mu, De\rangle \end{bmatrix}$

Fonction Duale: $\varphi(\mu) = \operatorname{Inf}_{(\omega,b)} \mathcal{L}(\omega,b,\mu)$

- Si $\langle\mu, De\rangle = 0$ alors en posant $\omega = -(DX)^T\mu\times\frac{1}{2}$
on annule $\nabla_{\omega b}\mathcal{L}(\omega,b,\mu)$ et ce qui garantit que on réalise le min du Lagrangien

$$\varphi(\mu) = \left\| -\frac{(DX)^T\mu}{2}\right\|^2 + \left\langle \mu, DX\frac{(-DX)^T\mu}{2} + bDe + e\right\rangle$$

$$= -\frac{1}{4}\|(DX)^T\mu\|^2 + b\underbrace{\langle\mu, De\rangle}_{=0} + \langle\mu, e\rangle$$

$$= -\frac{1}{4}\|(DX)^T\mu\|^2 + \langle\mu, e\rangle$$

- Si $\langle\mu, De\rangle \neq 0$
alors le terme $b\langle\mu, De\rangle$ dans le Lagrangien,

on peut le faire tendre vers $-\infty$, avec $|b| \to +\infty$

et donc $\inf_{(w,b)} \mathcal{L}(w, b, \mu) = -\infty$.

c.à.d que $\Psi(\mu) = -\infty$.

Par conséquent, le PB Dual s'écrit :

$$\begin{cases} \text{Max } \Psi(\mu) = -\frac{1}{4} \| (DX)^T \mu \|^2 + \langle \mu, e \rangle \\ \langle De, \mu \rangle = 0 \\ \mu \geq 0 \end{cases}$$