Séance 18 : minimisation d'une fonctionnelle quadratique sous contraintes linéaires

(voir séance 8 : minimisation d'une fonctionnelle quadratique sur $\mathbb{R}^n$)

1) Définition. Existence et unicité d'un minimum

On s'intéresse à la fonctionnelle :

$J : \mathbb{R}^n \to \mathbb{R}$ tq

$$\boxed{J(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle}$$

où A est symétrique définie positive ($A \gg 0$) et $b \in \mathbb{R}^n$.

On cherche à minimiser J sur

$$\boxed{\Omega = \{ x \in \mathbb{R}^n \text{ tq } Cx = d \}}$$

où C est une matrice $p \times n$ avec $p < n$ et de rang maximal : $rg(C) = p$, et $d \in \mathbb{R}^p$.

$$p \updownarrow \underbrace{\begin{pmatrix} & & \\ & C & \\ & & \end{pmatrix}}_{n} \begin{pmatrix} x \\ \\ \\ \end{pmatrix} = \begin{pmatrix} d \end{pmatrix} \updownarrow p$$

On a le résultat suivant d'existence et d'unicité d'un minimum de J sur Ω :

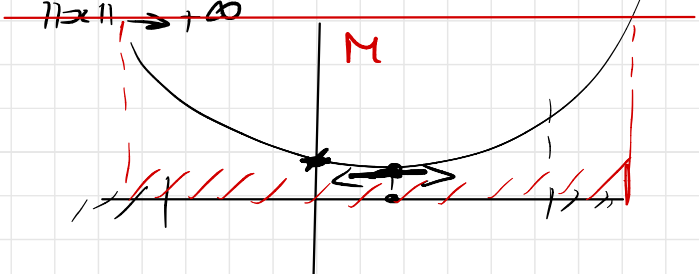
Théorème : J admet un unique minimum sur Ω, noté x^*. De plus, il existe un unique $\lambda^* \in \mathbb{R}^p$ tel que

$$\boxed{\begin{cases} Ax^* + {}^tC\lambda^* = b \\ Cx^* = d \end{cases}} \quad (*)$$

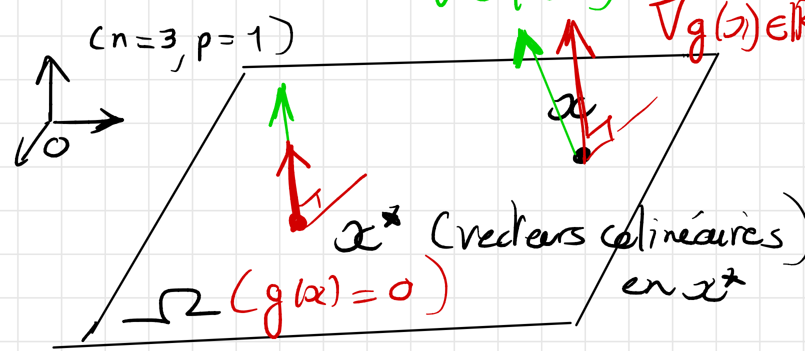
preuve :

* **Existence** : J est coercive sur $\mathbb{R}^n$:

$$\lim_{\|x\| \to +\infty} J(x) = +\infty$$



* J est continue sur $\mathbb{R}^n$.
* De plus Ω est fermé. La preuve sur $\mathbb{R}^n$ de l'existence d'un minimum de J s'étend sur Ω fermé.
* **Unicité** : J est strictement convexe et Ω est convexe, d'où l'unicité.
* **Caractérisation** : on la propriété des extremas liés.



* Dans le cas d'une seule contrainte linéaire ($p=1$), au point de minimum, on a

$$\nabla J(x^*) = \overbrace{\lambda}^{\text{multiplicateur Lagrange}} \nabla g(x)$$

où $g : \mathbb{R}^n \to \mathbb{R}$ linéaire
et $\Omega = \{ x \in \mathbb{R}^n / g(x) = 0 \}$

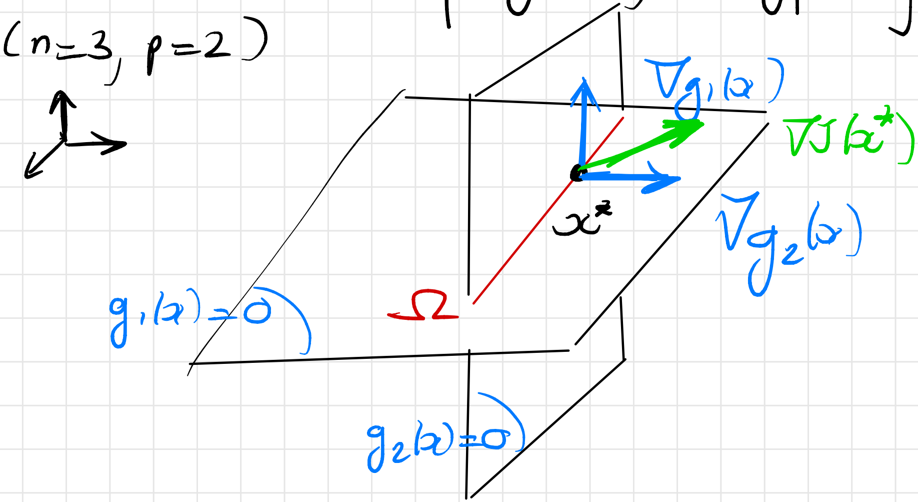
* Dans le cas de p contraintes linéaires, avec $p \geq 2$:

$$\Omega = \left\{ x \in \mathbb{R}^n, \begin{cases} g_1(x) = 0 \\ \vdots \\ g_p(x) = 0 \end{cases} \right\}$$

au point de minimum x^*,

$$\nabla J(x^*) \in \mathrm{Vect}\{ \nabla g_1(x^*), \ldots, \nabla g_p(x^*) \}$$

($n=3$, $p=2$)



soit $\nabla J(x^*) = \lambda_1 \nabla g_1(x^*) + \ldots + \lambda_p \nabla g_p(x^*)$

multiplicateurs de Lagrange

Ici $\Omega = \{ x \in \mathbb{R}^n \text{ tq } Cx = d \}$

Par exemple

$g_1(x) = c_{11}x_1 + \dots + c_{1n}x_n - d_1$

et $\nabla g_1(x) = \begin{pmatrix} c_{11} \\ \vdots \\ \vdots \\ c_{1n} \end{pmatrix}$

D'où

$\nabla J(x^*) \in \text{Vect}\left\{ \begin{pmatrix} c_{11} \\ \vdots \\ \vdots \\ c_{1n} \end{pmatrix}, \dots, \begin{pmatrix} c_{n1} \\ \vdots \\ \vdots \\ c_{nn} \end{pmatrix} \right\}$

soit : $\nabla J(x^*) + {}^tC\lambda^* = 0$

avec $\nabla J(x^*) = Ax^* - b$

Par l'unicité de la caractérisation :

on réécrit le système en taille $n+p$:

$$\begin{array}{c} \begin{array}{c} n \\ p \end{array} \begin{pmatrix} A & {}^tC \\ C & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} \begin{array}{c} n \\ p \end{array} = \begin{array}{c} n \\ p \end{array} \begin{pmatrix} b \\ d \end{pmatrix} \\ \quad\;\; n \qquad p \end{array}$$

Or $M = \left(\begin{array}{c|c} A & {}^tC \\ \hline C & 0 \end{array} \right)$ est de rang $n+p$ donc inversible.

et (x^*, λ^*) est unique.

On appelle λ^* les multiplicateurs de Lagrange associés au minimum x^* de J sur Ω

Il existe trois algorithmes permettant d'approcher la solution x^* du problème :

- Méthode de pénalisation
- Méthode du gradient projeté
- Méthode d'Uzawa

2) Méthode de pénalisation

(voir Ex 2, TD 13)

On se ramène à la résolution d'une suite de problèmes sans contraintes.

On note $J_\rho : \mathbb{R}^n \to \mathbb{R}$ la fonctionnelle :

$$\boxed{J_\rho(x) = J(x) + \frac{\rho}{2} \underbrace{\|Cx - d\|^2}_{\geq 0}}$$

$\rho > 0$

où $\| \ \|$: norme euclidienne dans $\mathbb{R}^p$.

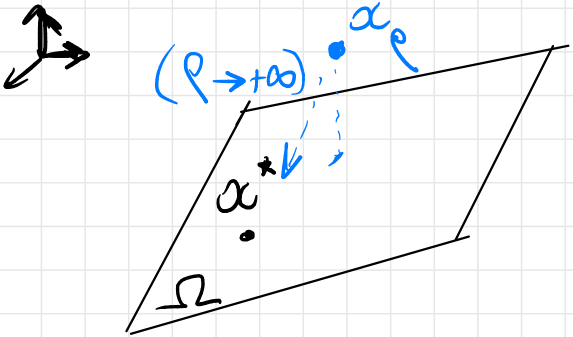
→ J_ρ est une fonctionnelle quadratique sur $\mathbb{R}^n$ car

$$\|Cx-d\|^2 = \langle Cx-d, Cx-d \rangle$$
$$= \langle \underbrace{{}^tCC}_{n \times n} x, x \rangle - 2 \langle x, \underbrace{{}^tCd}_{n \times 1} \rangle$$

et ${}^tCC >> 0$ sur $\mathbb{R}^n$ ($rg(C) = p$)

J_ρ possède donc un unique minimum x_ρ sur $\mathbb{R}^n$.

6



On considère la fonction pénalisée :

$$J_\rho(x) = J(x) + \frac{\rho}{2}||Cx - d||^2$$

o $\rho > 0$ est un paramètre de pénalisation.

1. Montrer que J_ρ possède un unique minimum x_ρ sur $\mathbb{R}^n$.
2. Montrer que pour tout $\rho > 0$, on a $J(x_\rho) \leq J(x^*)$.
3. En déduire que la famille $(x_\rho)_{\rho>0}$ est bornée.
4. Montrer que la seule valeur d'adhérence possible de toute suite $(x_{\rho_n})_{n\in\mathbb{N}}$ avec $\rho_n \to +\infty$ est égale à x^*.
5. En déduire que toute suite $(x_{\rho_n})_{n\in\mathbb{N}}$ avec $\rho_n \to +\infty$ tend vers x^*.

(à faire en exercice)

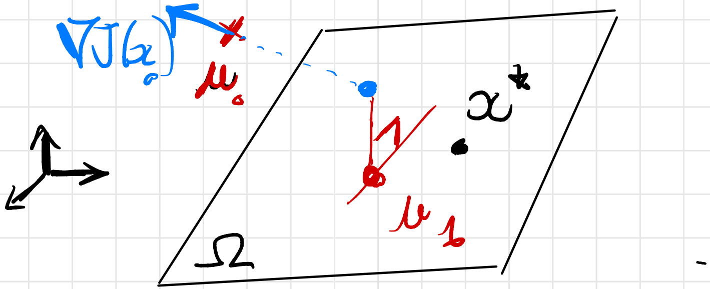
On montre que $\lim_{\rho \to +\infty} x_\rho = x^*$

(voir Ex 2)

On dispose alors d'un premier algorithme (voir Ex 3, TD 13) consistant à rechercher x_ρ par une méthode de gradient (voir séance 8) pour ρ grand (ou en faisant tendre ρ vers $+\infty$ par itérations successives)

3) Méthode du gradient projeté

L'idée consiste à utiliser une méthode de gradient pour minimiser J et à projeter à chaque itération le point obtenu sur Ω



Une itération de l'algorithme est la suivante :

$$\boxed{u_{k+1} = P_\Omega\left(u_k - \rho \nabla J(u_k)\right)}$$

où $\rho > 0$: pas de descente

et $P_\Omega : \begin{pmatrix} \mathbb{R}^n \longrightarrow \Omega \\ x \longmapsto P_\Omega(x) \end{pmatrix}$

projection orthogonale de x sur Ω (car Ω convexe, fermé ds $\mathbb{R}^n$ complet)

On montre que pour tout $u_0 \in \mathbb{R}^n$

$$\lim_{k \to +\infty} u_k = x^* \quad (\text{Ex3, TD 13})$$

si $\rho < \dfrac{2}{\rho(A)}$.

(où x^* est le minimum de J sur Ω)

1. Montrer que x^* est tel que

$$\forall v \in \Omega, \quad < \nabla J(x^*), v - x^* > \geq 0$$

En déduire que pour tout $\rho > 0$,

$$x^* = P_\Omega(x^* - \rho \nabla J(x^*))$$

2. On considère l'algorithme du gradient projeté à pas constant : partant de $u_0 \in \mathbb{R}^n$, on construit la suite $(u_k)_{k\in\mathbb{N}}$ telle que

$$u_{k+1} = P_\Omega(u_k - \rho \nabla J(u_k))$$

Montrer que si $\rho < \frac{2}{\lambda_n}$, alors la suite $(u_k)_{k\in\mathbb{N}}$ converge vers x^*.

(à faire en exercice)

Cette méthode n'est pas facilement implémentable car on doit exprimer P_Ω.

3) Méthode d'Uzawa

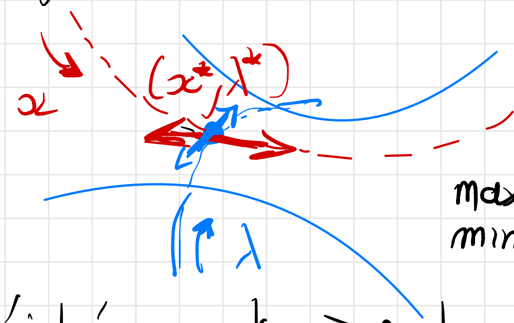
Cette méthode consiste à rechercher une solution du système (*) en exploitant une propriété sur le Lagrangien $\mathcal{L} : \mathbb{R}^n_x \times \mathbb{R}^p_\lambda \to \mathbb{R}$

$$\mathcal{L}(x,\lambda) = J(x) + \langle \underbrace{Cx-d}_{\mathbb{R}^p}, \underbrace{\lambda}_{\mathbb{R}^p} \rangle$$

On peut montrer que le point (x^*,λ^*) est un point selle du Lagrangien :

$$\forall (x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^p$$

$$\mathcal{L}(x^*,\lambda) \leq \mathcal{L}(x^*,\lambda^*) \leq \mathcal{L}(x,\lambda^*)$$



maximum en λ
minimum en x

L'idée consiste à rechercher ce point selle en effectuant successivement

→ une itération descendante en x

→ une itération montante en λ

On considère l'algorithme suivant : soit $\lambda_0 \in \mathbb{R}^p$ et $\rho > 0$. On construit les suites $(\lambda_k)_{k\in\mathbb{N}}$ et $(u_k)_{k\in\mathbb{N}}$ ainsi : $Au_0 = b - {}^tC\lambda_0$ puis

$$\begin{cases} \lambda_{k+1} = \lambda_k + \rho(Cu_k - d) \\ Au_{k+1} = b - {}^tC\lambda_{k+1} \end{cases}$$

1. Montrer que

$$||\lambda_{k+1} - \lambda^*||^2 \leq ||\lambda_k - \lambda^*||^2 + (\rho^2||C||^2 - 2\rho\lambda_1)||u_k - x^*||^2$$

2. En déduire qu'il existe $\rho_1 > 0$ tel que si $\rho < \rho_1$, la suite $(u_k)_{k\in\mathbb{N}}$ converge vers x^*.

(à faire en exercice)

5) Retour sur les exercices

Ex 2 : pénalisation

$$J_\varepsilon(x) = J(x) + \frac{\varepsilon}{2}||Cx - d||^2$$
$(\varepsilon > 0)$

1) x_ε : unique minimum de J_ε sur $\mathbb{R}^n$;

2) $J(x_\varepsilon) \leq J(x^*)$?

On a en effet

$J_\varepsilon(x_\varepsilon) \leq J_\varepsilon(x^*)$ (car x_ε minimum de J_ε)

Or $J_\varepsilon(x^*) = J(x^*) + 0$ (car $x^* \in \Omega$)

D'où :

$J_\varepsilon(x_\varepsilon) \leq J(x^*)$

De plus :

$$J(x_\varepsilon) \leq \underbrace{J(x_\varepsilon) + \frac{\varepsilon}{2}||Cx_\varepsilon - d||^2}_{} \leq J(x^*)$$

3) $(x_\varepsilon)_{\varepsilon>0}$ est bornée. En effet, si on on suppose qu'il existe ε_n tq $||x_{\varepsilon_n}|| \xrightarrow[n\to+\infty]{} +\infty$

Alors $J(x_{\varepsilon_n}) \to +\infty$ car

J est coercive ce qui est absurde avec

$$J(x_{\rho_n}) \leq J(x^*).$$

4) Soit $\hat{x}$ une valeur d'adhérence de (x_{ρ_n}) avec $\rho_n \to +\infty$.

On doit montrer : $\hat{x} = x^*$.

On a

$$J(x_{\rho_n}) \leq J(x^*) \Rightarrow$$

$$J(\hat{x}) \leq J(x^*)$$

Il reste à montrer que $\hat{x} \in \Omega$:
(ce qui implique $\hat{x} = x^*$)

On a

$$J_{\rho_n}(x_{\rho_n}) = J(x_{\rho_n}) + \frac{\rho_n}{2}\|Cx_{\rho_n} - d\|^2$$

puis :

$$\|Cx_{\rho_n} - d\|^2 = \frac{2}{\rho_n}\left(J_{\rho_n}(x_{\rho_n}) - J(x_{\rho_n})\right)$$

$$= \frac{2}{\rho_n} J_{\rho_n}(x_{\rho_n}) - \frac{2J(x_{\rho_n})}{\rho_n}$$

$\to 0$ (car $J(x_{\rho_n}) \leq J(x^*)$)

Or $J_{\rho_n}(x_{\rho_n}) \leq J_{\rho_n}(x^*) = J(x^*)$

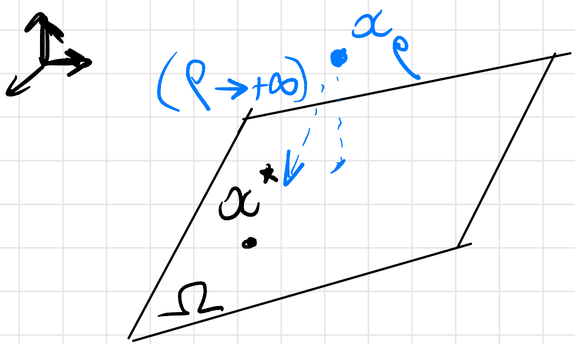
De même $\frac{J_{\rho_n}(x_{\rho_n})}{\rho_n} \to 0$ $(n \to +\infty)$

et par passage à limite

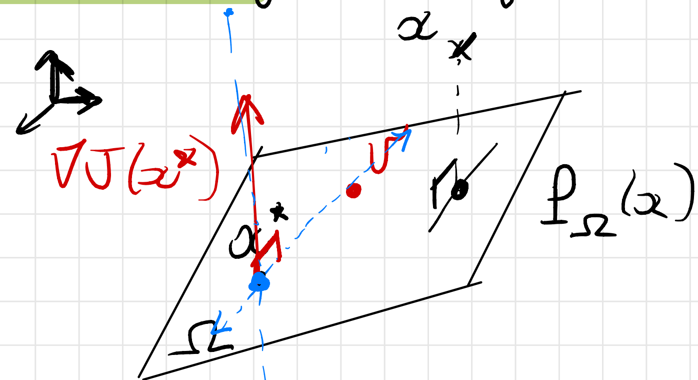
$\|C\hat{x} - d\| = 0$ soit : $\hat{x} \in \Omega$ //

5) En conclusion, comme une suite

bornée qui admet une seule valeur d'adhérence est convergente, on a la convergence de toute suite (x_{φ_n}) avec $\varphi_n \to +\infty$ vers x^* //



Exercice 3 (gradient projeté)



1) $\forall v \in \Omega, \langle \nabla J(x^*), v - x^* \rangle \geq 0$ //

Taylor à l'ordre 2 par J entre x^* et $x^* + t(v - x^*) = x_t$

$$J(x_t) = J(x^*) + t\langle \nabla J(x^*), v - x\rangle$$

$$\geq J(x^*) + \underbrace{\frac{1}{2} t^2 \langle \underbrace{\nabla^2 J(x^*)}_{A\,:\,\text{Hessien}} .(v-x), v-x \rangle}_{\geq 0 \ (A \gg 0)}$$

On en déduit

$$t\langle \nabla J(x^*), v-x^*\rangle + \frac{1}{2}t^2\langle Ah, h\rangle \leq 0$$

$\forall t \in \mathbb{R}$

Ceci implique que en réalité, on a :

$$\langle \nabla J(x^*), v-x^*\rangle = 0$$

2) On montre

$$x^* = P_\Omega(\overbrace{x^* - \rho\nabla J(x^*)}^{\tilde{x}})$$

On peut montrer par cela que x^* vérifie la caractérisation de la projection P_Ω :

$x^* = P_\Omega(\tilde{x})$ ssi :

$$\forall v \in \Omega \quad \langle x^* - \tilde{x}, v - x^*\rangle = 0$$

$$(\hat{x} - P_\Omega(\hat{x}) \perp v - x^*)$$

soit :

$$\langle \rho\nabla J(x^*), v-x^*\rangle = 0$$

ce qui a été démontré en 1).

Soit :

x^* point fixe de l'application

$$g : \left(\begin{array}{ccc} \Omega & \longrightarrow & \Omega \\ x & \longmapsto & P_\Omega(x - \rho\nabla J(x)) \end{array}\right)$$

3) On montre que si $\rho \in]0, \frac{2}{\rho(A)}[$

g est strictement contractante :

$$\|g(x)-g(y)\| \leq \|(x-\rho\nabla J(x))-(y-\rho\nabla J(y))\|$$

(car P_Ω est 1-Lipschitz)

$$\leq \|\underbrace{x-y-\rho(Ax-Ay)}_{(I-\rho A)(x-y)}\|$$

$$\leq |||I-\rho A|||\ \|x-y\|$$

Or, les v.p. de $I-\rho A$ sont $\{1-\rho\lambda_1, \dots, 1-\rho\lambda_n\}$ et de valeur absolue strictement inférieure à 1.

* Ω est fermé et donc complet. g possède donc par unique point fixe x^*. De plus toute suite telle que

$$\begin{cases} u_0 \in \Omega \\ u_{k+1} = g(u_k) \end{cases}$$

converge vers x^* (Picard).

Implémentation Scilab/Python possible (pénalisation ou Uzawa)

Complément

Ex4) (Uzawa) $\quad {}^t({}^tC\lambda)x - {}^t\lambda d$

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\,{}^t x A x - {}^t b x + \underbrace{{}^t\lambda(Cx-d)}$$

* (x^*, λ^*) est solution de KKT :

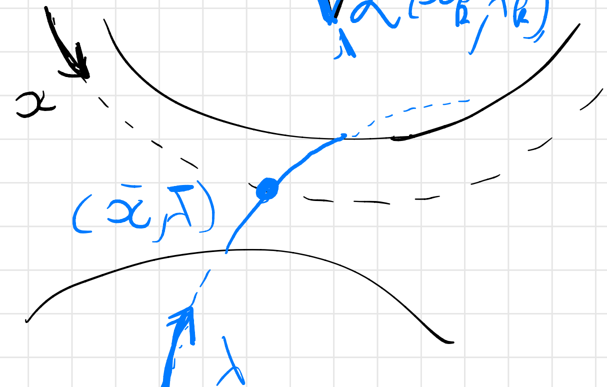
$$\begin{cases} Ax^* + {}^tC\lambda^* = b \\ Cx^* = d \end{cases}$$

* Algorithme proposé : (λ_0, u_0) fixés $\in \mathbb{R}^p \times \mathbb{R}^n$

$$\boxed{\begin{cases} \lambda_{k+1} = \lambda_k + \rho(Cu_k - d) \\ Au_{k+1} = b - {}^tC\lambda_{k+1} \end{cases}}$$

S'agit-il bien d'un algorithme de recherche d'un point selle de $\mathcal{L}$?

* $\lambda_{k+1} = \lambda_k + \rho\underbrace{(Cu_k - d)}_{\nabla_\lambda \mathcal{L}(x_k, \lambda_k)}$



→ itération d'une méthode de gradient par maximisation de $\mathcal{L}(\lambda)$

* $Au_{k+1} = b - {}^tC\lambda_{k+1}$

$\begin{pmatrix} \mathbb{R}^n \to \mathbb{R} \\ x \mapsto \mathcal{L}(x, \lambda_{k+1}) \end{pmatrix}$ a pour minimum u tq $Au = b - {}^tC\lambda$: résolution exacte de minimisation de $\mathcal{L}(x)$

1) On doit montrer :

$$\boxed{\|\lambda_{k+1}-\lambda^*\|^2 \le \|\lambda_k-\lambda^*\|^2 + (\rho^2\|C\|^2 - 2\rho\tilde{\lambda}_1)\|u_k-x^*\|^2}$$

où $\|C\| = \sup\limits_{x\neq 0} \dfrac{\|Cx\|}{\|x\|}$ et $\tilde{\lambda}_1$: plus petite valeur propre de A.

On a :

$$\begin{cases} \lambda_{k+1} = \lambda_k + \rho(Cu_k - d) \\ Au_{k+1} = b - {}^tC\lambda_{k+1} \end{cases}$$

On calcule

$$\lambda_{k+1}-\lambda^* = \lambda_k-\lambda^* + \rho(\underbrace{Cu_k - d}_{C(u_k-x^*)})$$

On élève au carré et on développe :

$$\|\lambda_{k+1}-\lambda^*\|^2 = \|\lambda_k-\lambda^*\|^2 + \rho^2\|C(u_k-x^*)\|^2 + 2\rho\langle \lambda_k-\lambda^*, C(u_k-x^*)\rangle$$

puis

$$\|\lambda_{k+1}-\lambda^*\|^2 \le \|\lambda_k-\lambda^*\|^2 + \rho^2\|C\|^2\|u_k-x^*\|^2 + 2\rho\langle {}^tC(\lambda_k-\lambda^*), (u_k-x^*)\rangle$$

Or, $Ax^* = b - {}^tC\lambda^*$, on a donc : $A(u_k-x^*) = -{}^tC(\lambda_k-\lambda^*)$ et :

$$\langle {}^tC(\lambda_k-\lambda^*), u_k-x^*\rangle = -\langle A(u_k-x^*), u_k-x^*\rangle.$$

Comme les valeurs propres de A sont $0<\tilde{\lambda}_1 \le \dots \le \tilde{\lambda}_n$, il vient :

$$-\langle A(u_k-x^*), u_k-x^*\rangle \le -\tilde{\lambda}_1\|u_k-x^*\|^2$$

2) On prend ρ tq $\rho^2\|C\|^2 - 2\rho\tilde{\lambda}_1 < 0$ //

soit $0 < \rho < \dfrac{2\tilde{\lambda}_1}{\|C\|^2}$, alors, en notant $\beta = \rho^2\|C\|^2 - 2\rho\tilde{\lambda}_1 < 0$, on a,

$$\|\lambda_{k+1}-\lambda^*\|^2 \le \|\lambda_k-\lambda^*\|^2 + \underbrace{\underbrace{\beta}_{<0}\|u_k-u^*\|^2}_{\le 0}$$

On en déduit donc que

$\|\lambda_{k+1}-\lambda^*\| \le \|\lambda_k-\lambda^*\|$ et que la suite $k \mapsto \|\lambda_k-\lambda^*\|$ est décroissante minorée donc convergente vers $\ell \ge 0$.

Or

$$0 \le -\beta\|u_k-u^*\|^2 \le \underbrace{\|\lambda_k-\lambda^*\|^2}_{\to \ell} - \underbrace{\|\lambda_{k+1}-\lambda^*\|^2}_{\to \ell}$$

Par encadrement, on a $\lim_{k\to+\infty}\|u_k-u^*\|^2 = 0$

soit $\boxed{\lim_{k\to+\infty} u_k = u^*}$

Remarque : comme C est de rang maximal (injective) on a également

$${}^tC(\lambda_k-\lambda^*) = \overbrace{-A(u_k-x^*)}^{0}$$

et $\lambda_k \to \lambda^*$