DAix Théo
VERHOYE Victor

TP n° 2 d'Analyse numérique

Nous nous interessons à la résolution de $Ax = b$ lorsque $A \in \mathbb{R}^{M\times N}$ avec rang $A = N < M$. Dans le cas, nous allons rechercher la meilleure solution au sens des moindres carrés.

Question 1 :

Rechercher la meilleure solution au sens des moindres carrés consiste à chercher la valeur de x qui minimise la fonction $\tilde{f} : \begin{cases} \mathbb{R}^N \longrightarrow \mathbb{R} \\ x \longmapsto |Ax-b|_2^2 \end{cases}$. (1).

Cette question consiste à montrer que x réalise le minimum de (1) ssi x vérifie l'équation $A^T Ax = A^T b$.

$\Leftarrow$ On suppose $A^T Ax = A^T b$ pour un $x \in \mathbb{R}^N$.

Soit $f : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ application linéaire telle que A est la matrice de f dans les bases canoniques de $\mathbb{R}^N$ et $\mathbb{R}^M$.

L'hypothèse devient : $f^t \circ f(x) = f^t(b)$ (*)

où f^t est la transposée de f.

$(*) \Rightarrow \forall y \in \mathbb{R}^N, \langle f^t \circ f(x) - f^t(b), y \rangle = 0.$

$\Leftrightarrow \forall y \in \mathbb{R}^N, \langle f^t(f(x) - b), y \rangle = 0$

$\Leftrightarrow \forall y \in \mathbb{R}^N, \langle f(x) - b, f(y) \rangle = 0$ (définition de transposée).

$\Leftrightarrow \forall \bar{y} \in \mathrm{Im}(f), \langle f(x) - b, \bar{y} \rangle = 0.$

Donc, on a que $f(x) - b \in (\mathrm{Im}(f))^\perp$. (**)

Montrons que x minimise $\tilde{f}$:

Soit $y \in \mathbb{R}^N$, $|f(x)-b|_2^2 \leq |f(x)-b|_2^2 + \overbrace{|f(y)-f(x)|_2^2}^{\geq 0}$

par (**) car $f(y)-f(x) = f(y-x) \in \mathrm{Im}(f)$ et par le théorème de Pythagore $= |f(x)-b+f(y)-f(x)|_2^2$

$= |f(y)-b|_2^2.$

$\Rightarrow$) On suppose que x minimise $\tilde{f}$.

Posons $x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$, $A = (a_{i,j})_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}$.

On a : $\forall 1 \leq k \leq N$, $\partial_{x_k} \tilde{f}(x) = 0$ (par le cours).

Or, $\tilde{f}(x) = \sum_{i=1}^{M} \left(\left(\sum_{j=1}^{N} a_{i,j} . x_j \right) - b_i \right)^2$.

Donc, soit $k \in \{1, ..., N\}$,

$$\begin{aligned} \partial_{x_k} \tilde{f}(x) &= \sum_{i=1}^{M} 2 . \left(\sum_{j=1}^{N} a_{i,j} . x_j - b_i \right) . a_{i,k} \\ &= 2 . \sum_{i=1}^{M} a_{i,k} . \left(\sum_{j=1}^{N} a_{i,j} . x_j - b_i \right) \\ &= 2 . \left[\sum_{i=1}^{M} a_{i,k} . \sum_{j=1}^{N} a_{i,j} . x_j - \sum_{i=1}^{M} a_{i,k} . b_i \right] \end{aligned}$$

Donc, $\partial_{x_k} \tilde{f}(x) = 0$

$$\Leftrightarrow \sum_{i=1}^{M} a_{i,k} \sum_{j=1}^{N} a_{i,j} . x_j = \sum_{i=1}^{M} a_{i,k} . b_i .$$

Or, on a que $A^T . A . x = \left(\sum_{i=1}^{M} a_{i,k} \sum_{j=1}^{N} a_{i,j} . x_j \right)_{1 \leq k \leq N}$.

et $A^T . b = \left(\sum_{i=1}^{M} a_{i,k} . b_i \right)_{1 \leq k \leq N}$.

Donc on a bien l'égalité. □

Question 2 :

On considère $f : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ comme dans la question précédente.

Si on montre que $f^t \circ f$ est bijective, alors $A^T . A$ sera inversible et donc le système aura une unique solution.

Par hypothèse, on sait que rang $A = N$.

Donc, par le théorème du rang,

$$\dim \operatorname{Ker}(f) = \dim \mathbb{R}^N - \dim \operatorname{rang} A = N - N = 0 .$$

On en déduit que f est injective.

Montrons que $f^t \circ f$ est injective.

Si l'on montre que $\operatorname{Ker}(f^t \circ f) = \operatorname{Ker}(f)$, on aura ce qu'on veut (car $f^t \circ f : \mathbb{R}^N \to \mathbb{R}^N$: dimensions égales $\Rightarrow f^t \circ f$ bijective)

$\supseteq$) : Soit $x \in \mathbb{R}^N$ t.q. $f(x) = 0$.

Clairement, vu que f^t linéaire, $f^t \circ f(x) = f^t(0) = 0$.

$\subseteq$) : Soit $x \in \mathbb{R}^N$ t.q. $f^t \circ f(x) = 0$.

Par bilinéarité du produit scalaire, $\langle f^t \circ f(x), x \rangle = 0$.

Or, on a : $\langle f^t \circ f(x), x \rangle = \langle f(x), f(x) \rangle$ (définition de transposée).

Vu qu'on utilise le produit scalaire usuel (il est non-dégénéré), on a donc que $f(x) = 0$, i.e. $x \in \mathrm{Ker}(f)$. $\square$

Question 3 :

Voir le programme q1 qui utilise la routine DGESV.

Question 4 :

On considère à présent les fonctions $A : \mathbb{R} \to \mathbb{R}^{M\times N} : t \mapsto A(t)$ et $b : \mathbb{R} \to \mathbb{R}^M : t \mapsto b(t)$.

Soit t fixé.

On désire calculer $\partial_t x(t)$.

On a que $A(t) \in \mathbb{R}^{M\times N}$ et $b(t) \in \mathbb{R}^M$.

On peut donc trouver $x(t)$ **tel que** $A^T(t).A(t)x(t) = A^T(t).b(t)$.

On a vu au cours que, si on considère $A \in \mathbb{R}^{M\times N}$ et $B \in \mathbb{R}^{N\times P}$,

$$\partial(A.B) = \partial A.B + A.\partial B$$ (car le produit matriciel est bilinéaire).

On a l'égalité suivante : $A^T(t).A(t).x(t) = A^T(t).b(t)$.

En passant à la dérivée, on a :

$$\partial_t(A^T(t).A(t).x(t)) = \partial_t(A^T(t).b(t)).$$

$$\Leftrightarrow \partial_t(A^T(t).A(t)).x(t) + A^T(t).A(t).\partial_t x(t) = \partial_t A^T(t).b(t) + A^T(t).\partial_t(b(t))$$

$$\Leftrightarrow [\partial_t(A^T(t)).A(t) + A^T(t).\partial_t(A(t))].x(t) + A^T(t).A(t).\partial_t x(t) = \partial_t A^T(t)b(t) + A^T(t).\partial_t b(t).$$

Donc, on a : $\underbrace{A^T(t).A(t)}_{A_1}.\partial_t x(t) = \underbrace{\partial_t(A^T(t)).(b(t) - A(t)x(t)) + A^T(t).(\partial_t b(t) - \partial_t A(t)x(t))}_{B_1}$.

On peut donc trouver $\partial_t x(t)$ en résolvant le système

$$A_1.\partial_t x(t) = B_1.$$

Question 5 :

Voir q2, celui-ci utilise les idées de la Question 4.

Question 6 :

Pour retourner une approximation de $\partial_t x(t)$, t fixé, nous voulons utiliser notre programme en Question 5.

Pour que celui-ci fonctionne, il faut être sous les bonnes hypothèses, càd rang $A(t) = 2$.

Nous devons donc montrer que $\dim \mathrm{Im}(A(t)) = 2$.

Pour ce faire, nous allons montrer que les vecteurs

$v = (\sin(t), t, \arctan(t))$ et $v' = (2t, e^t, t)$

sont linéairement indépendants pour tout t non nul :

Soit $t \neq 0$, soient $a, b \in \mathbb{R}$ t.q. $a.v = bv'$.

i.e. $\begin{cases} a\sin(t) = 2bt & (1) \\ a.t = b.e^t & (2) \\ a.\arctan(t) = b.t & (3) \end{cases}$.

Nous devons montrer que $a = 0$ et $b = 0$.

Supposons que $a \neq 0$ et $b \neq 0$.

$(1) - 2.(3) \Rightarrow a.\sin(t) = 2a.\arctan(t) \underset{a\neq 0}{\Rightarrow} \sin(t) = 2.\arctan(t)$.

Or, pour tout $t \neq 0$, $\sin(t) \neq 2.\arctan(t)$.

Montrons-le :

Posons $p(t) = \sin(t) - 2.\arctan(t)$.

Il nous faut donc montrer que pour tout $t \neq 0$, $p(t) \neq 0$.

Soit $t \in \mathbb{R}\setminus\{0\}$. On a toujours : $-1 \leq \sin(t) \leq 1$.

1°) si $t \geq 1$:

Vu que la fonction $\arctan(t)$ est croissante, on a

$\arctan(t) \geq \frac{\pi}{4} \Rightarrow p(t) = \sin(t) - 2\arctan(t) \leq 1 - \frac{\pi}{2} < 0$.

2°) si $t \leq 1$:

On a : $\arctan(t) \leq -\frac{\pi}{2} \Rightarrow p(t) = \sin(t) - 2\arctan(t) \geq -1 + \frac{\pi}{2} > 0$.

3°) si $t \in]-1, 1[$:

Montrons que p est strictement décroissante sur $]-1, 1[$.

Vu qu'on a $p(0) = 0$, on aura alors $\forall t \neq 0$, $p(t) \neq 0$.

Regardons la dérivée de p :

$\partial_t p(t) = \cos(t) - \frac{2}{1+t^2}$.

On a : $t^2 < 1 \Rightarrow 1+t^2 < 2 \Rightarrow \frac{1}{1+t^2} > \frac{1}{2} \Rightarrow \frac{2}{1+t^2} > 1$

$\Rightarrow -\frac{2}{1+t^2} < -1$.

Donc, $\partial_t p(t) < 1 - 1 = 0$. (car $\cos(t) \leq 1$).

Vu que la dérivée est strictement négative, on a que p est strictement décroissante sur $]-1,1[$

On arrive donc à une contradiction pour tout $t \neq 0$.

Donc, on a que $a = 0$ ou $b = 0$.

Or, par l'équation (2), si $a = 0$, on a $b = \frac{0}{e^t} = 0$,

et si $b = 0$, $a = \frac{0}{t} = 0$.

Donc, on a bien montré que les vecteurs v et v' sont linéairement indépendants, et donc $\dim \mathrm{Im}(A(t)) = 2$ (pour t non nul).

Remarques:

- Quand t est nul, $A(0) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

  Donc, $A^T(0) . A(0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} . \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ de déterminant nul.

  On n'a, de ce fait, pas l'unicité de la solution $x(t)$.

- Notre fonction deriv dans q2.py prend en argument les dérivées de A et b.

Pour la question 6, on a :

$$\partial_t \begin{pmatrix} \sin(t) & 2t \\ t & e^t \\ \arctan(t) & t \end{pmatrix} = \begin{pmatrix} \cos(t) & 2 \\ 1 & e^t \\ \frac{1}{1+t^2} & 1 \end{pmatrix}$$

$$\text{et } \partial_t \begin{pmatrix} t \\ t^2 \\ \frac{1}{1+t^2} \end{pmatrix} = \begin{pmatrix} 1 \\ 2t \\ \frac{-2t}{(1+t^2)^2} \end{pmatrix}.$$