P. Maurer
ENS Rennes

Leçon 162 : Systèmes d'équations linéaires, opérations élémentaires, aspects algorithmiques et conséquences théoriques.

Devs :

- Décomposition de Bruhat
- Algorithme du gradient à pas optimal

Références :

1. Gourdon, Algèbre
2. Griffone, Algèbre linéaire
3. Caldero, H2G2
4. Dos Santos, Groupes finis et représentations (Poly M1 Jussieu)
5. Hirriat, Optimisation et analyse convexe
6. Ciarlet, Introduction à l'analyse numérique matricielle

Dans tout le plan, $\mathbb{K}$ désigne un corps de caractéristique nulle (généralement, $\mathbb{K}=\mathbb{R}$ ou $\mathbb{C}$).

1 Systèmes d'équations linéaires sur $\mathbb{K}^n$

1.1 Définitions, première propriétés.

Définition 1. *On appelle système d'équations linéaires un système du type :*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n = b_p \end{cases}$$

Où les a_{ij} et les b_i sont des éléments de $\mathbb{K}$. Les x_i sont appelés « inconnues », et résoudre le système signifie déterminer les $x_i \in \mathbb{K}$, s'il y en a, qui vérifient les équations ci-dessus.

Le système est dit compatible s'il admet au moins une solution $(x_1, \ldots, x_n) \in \mathbb{K}^n$.

Exemple 2. $\begin{cases} x+y=1 \\ x=0 \end{cases}$ est compatible. $\begin{cases} x+y=1 \\ x+y=0 \end{cases}$ ne l'est pas.

Proposition 3. *Notons $A=(a_{ij})_{1\leq i,j\leq n}$ la matrice $A=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$, $B=(b_i)_{1\leq i\leq p}$ la matrice colone $B=\begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix}$, et $X=(x_i)_{1\leq i\leq n}$ la matrice colone $X=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Alors le système (1) s'écrit sous la forme $AX=B$.*

Définition 4. *On appelle rang du système linéaire le rang de la matrice A associée.*

Remarque 5. Le système est compatible si et seulement si $B \in \mathrm{Vect}(A_1, \ldots, A_n)$, où $A_1, \ldots, A_n$ désignent les colones de A (en tant que vecteurs colones).

1.2 Systèmes de Cramer et cas général.

Définition 6. *On appelle système de Cramer un système linéaire dont la matrice A associée et carrée et inversible. Il s'agit donc d'un système de n équations à n inconnues de rang n :*

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = b_n \end{cases} \quad \text{avec } \det(A) \neq 0$$

Proposition 7. *Un système de Cramer admet toujours une unique solution donnée par $X=A^{-1}B$.*

Théorème 8. *(Formules de Cramer)*

Soit (S): $AX=B$ un système de Cramer avec $A \in \mathrm{GL}_n(\mathbb{K})$. On note $A_1, \ldots, A_n$ les colones de A, et $B_0=(e_1, \ldots, e_n)$ la base canonique de $\mathbb{K}^n$. Alors l'unique solution $X=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ de (S) vérifie :

$$\forall i \in [\![1,n]\!] \quad x_i = \frac{\det_{B_0}(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_n)}{\det(A)}$$

Définition 9. *Soit $A \in \mathcal{M}_{p,q}(\mathbb{K})$ de rang r. Il existe un déterminant Δ d'ordre r non nul extrait de A. On l'appelle déterminant principal de A (il n'est en général pas unique).*

Les équations dont les indices sont ceux des lignes de Δ s'appellent les équations principales.

Les inconnues dont les indices sont les colones de Δ s'appellent les inconnues principales.

En notant $\Delta=(a_{ij})_{i\in I,j\in J}$, on appelle déterminants caractéristiques de A les déterminants d'ordre $r+1$ de la forme :

$$\Delta_k := \left|\begin{array}{c|c} (a_{ij})_{i\in I,j\in J} & (b_i)_{i\in I} \\ \hline (a_{k,j})_{j\in J} & b_k \end{array}\right| \quad , \ k\notin J$$

Théorème 10. *(Rouché-Fontené)*

Le système $AX=B$ avec $A\in\mathcal{M}_{p,q}(\mathbb{K})$ de rang r admet des solutions si et seulement si $p=r$ ou si les $p-r$ déterminants caractéristiques sont nuls. Le système est alors équivalent au système des équations principales, les inconnues principales étant déterminées par un système de Cramer à l'aide des inconnues non principales.

Exemple 11. Soit (S) le système :

$$\left\{\begin{array}{rrrrcl} x_1 & +2x_2 & -x_3 & +x_4 & = & 1 \\ x_1 & & -x_3 & -x_4 & = & 1 \\ -x_1 & +x_2 & +x_3 & +2x_4 & = & m \end{array}\right. , \quad m\in\mathbb{R}$$

On a $A=\begin{pmatrix} 1 & 2 & -1 & 1 \\ 1 & 0 & -1 & -1 \\ -1 & 1 & 1 & 2 \end{pmatrix}$, qui est de rang 2. On choisit le déterminant principal $\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix}$. Il n'y a qu'un seul déterminant caractéristique, qui est $\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & m \end{vmatrix} = -2(m+1)$. On déduit du théorème de Rouché-Fontené que (S) admet des solutions si et seulement si $m=-1$, et dans ce cas, (S) est équivalent à :

$$\left\{\begin{array}{l} x+2x_2=1+x_3-x_4 \\ x_1=1+x_3+x_4 \end{array}\right. \Longleftrightarrow \left\{\begin{array}{l} x_1=1+x_3+x_4 \\ x_2=-x_4 \end{array}\right.$$

1.3 Solution des moindres carrés

Dans cette partie, E et E' désignent deux espaces euclidiens et $f: E\to E'$ une application linéaire.

Définition 12. *Il existe une unique application linéaire $f^*: E'\to E$ vérifiant*

$$\forall (x,y)\in E'\times E \qquad \langle f(x),y\rangle_E = \langle x, f^*(y)\rangle_{E'}$$

f^* est appelée adjoint de f.

On suppose dans la suite $\dim(E)=q\leq n=\dim(E')$, et que f est injective.

Proposition 13. *Sous ces hypothèses, on a $\det(f^*\circ f)\neq 0$ et la projection orthogonale p sur $\mathrm{Im}(f)$ vérifie $p=f\circ(f^*\circ f)^{-1}\circ f^*$.*

Définition 14. *On appelle inverse généralisé de f l'application $f^\bullet := (f^*\circ f)^{-1}\circ f^*$. Elle vérifie $f^\bullet\circ f=\mathrm{id}_E$ et $f\circ f^\bullet = p$, et coïncide avec l'inverse de f si f est bijective.*

Définition 15. *Soit $f(x)=b$ un système linéaire, avec f injective. On appelle solution des moindres carrés le vecteur $x_0\in E$ tel que :*

$$\|f(x_0)-b\| = \inf_{x\in E}\|f(x)-b\|$$

Théorème 16. *La solution x_0 des moindres carrés de $f(x)=b$ vérifie $x_0=f^\bullet(b)$.*

Théorème 17. *Soit $A\in\mathcal{M}_n(\mathbb{R})$, et $B\in\mathbb{R}^n$.*

Le vecteur X_0 est solution des moindres carrés de $AX=B$ si et seulement si $A^TAX_0=A^TB$.

2 Opérations élémentaires. Etude de $\mathrm{GL}_n(\mathbb{K})$.

2.1 Opérations élémentaires et pivot de Gauss.

Définition 18. *Pour $i\neq j$, $n\geq 2$ et $\lambda,\alpha\neq 0$, on définit les matrices élémentaires :*

	Dilatation	Transvection	Permutation
Matrice	$D_i(\lambda)=I_n+(\alpha-1)E_{ii}$	$T_{ij}(\lambda)=I_n+\lambda E_{ij},\ i\neq j$	$P_{ij}=I_n-E_{ii}-E_{jj}+E_{ij}+E_{ji}$
Determinant	α	1	-1
Action à gauche	$L_i\leftarrow L_i+\alpha L_i$	$L_i\leftarrow L_i+\lambda L_j$	$L_i\leftrightarrow L_j$
Action à droite	$C_i\leftarrow C_i+\alpha C_i$	$C_j\leftarrow C_j+\lambda C_i$	$C_i\leftrightarrow C_j$
Inverse	$D_i(\alpha^{-1})$	$T_{ij}(-\lambda)$	P_{ij}

Définition 19. *On appelle pivot d'une ligne non nulle le coefficient non nul situé dans la colone la plus à gauche. Une matrice est dite échelonnée en lignes lorsqu'elle vérifie les conditions suivantes :*

- *Si une ligne est nulle, toute les lignes suivantes sont nulles.*
- *Le pivot d'une ligne est strictement plus à droite que les pivots des lignes précédentes.*

Une matrice échelonnée est dite réduite si, de plus, tous les pivots sont égaux à 1 et les pivots sont les seuls coefficients non nuls de leur colone.

Théorème 20. *Soit $m,n\in\mathbb{N}$. On considère l'action de $\mathrm{GL}_n(\mathbb{K})$ par multiplication à gauche sur l'espace $\mathcal{M}_n(\mathbb{K})$. Alors :*

- *Deux matrices A et A' de $\mathcal{M}_n(\mathbb{K})$ ont la même orbite si et seulement si elles ont le même noyau.*
- *Toute matrice est dans l'orbite d'une unique matrice échelonnée en ligne réduite : on a la réunion disjoints suivante :*

$$\bigcup_{E\in\mathcal{E}_n}\mathrm{GL}_n(\mathbb{K})\cdot E$$

Où $\mathcal{E}_n$ désigne l'ensemble des matrices échelonnées réduites de taille $n\times n$.

Remarque 21. Le théorème précédent se démontre via l'algorithme du pivot de Gauss. Partant d'une matrice $A\in\mathcal{M}_n(\mathbb{K})$, on la multiplie à gauche par des matrices élémentaires pour obtenir une matrice d'abord échelonnée en lignes, puis échelonnée en lignes réduite en annulant les coefficients éventuels au-dessus des pivots. On trouve alors P inversible telle que PA soit échelonnée réduite.

Proposition 22. *La méthode du pivot de Gauss a une complexité algorithmique en $O(n^3)$.*

Par comparaison, le calcul des formules de Cramer nécessite $O(n!)$ opérations.

Exemple 23. $\begin{cases} 2x+y-4z=1\\ 2x+2y-z=4\\ 2x+3y+3z=5\end{cases} \iff \begin{cases} 2x+y-4z=1\\ -y+3z=3\\ 2y+7z=4\end{cases} \iff \begin{cases} 2x+y-4z=1\\ -y+3z=3\\ 13z=10\end{cases} \iff \begin{cases} x=\frac{1}{2}(1-y+4z)\\ y=3-3z\\ z=10/13\end{cases}$.

Application 24. La méthode du pivot de Gauss s'applique :

- Au calcul du rang d'une matrice.
- Au calcul de l'inverse d'une matrice.
- A la recherche d'un système d'équation d'un sous-espace vectoriel défini par une famille génératrice.
- A la recherche d'une base d'un sous-espace vectoriel défini par un système d'équations.

2.2 Décomposition de $\mathrm{GL}_n(\mathbb{K})$

On présente ici un résultat plus théorique lié aux opérations élémentaires.

Définition 25. *On appelle drapeau de k^n toute suite $\{0\}=F_0\subset\cdots\subset F_n$ de sous-espaces vectoriels de k^n tels que les inclusions soient strictes. Si de plus $\dim(F_i)=i$ pour tout i, on dit que le drapeau $(F_0,\ldots,F_n)$ est complet.*

Exemple 26. Soit $(e_1,\ldots,e_n)$ la base canonique de k^n. On définit $F_i=\mathrm{Vect}(e_1,\ldots,e_i)$ pour $i\geq 1$ et $F_0=\{0\}$. Alors $\mathcal{C}=(F_0,\ldots,F_n)$ est un drapeau complet, appelé le drapeau complet canonique.

Définition 27. *On note $B_n(k)$ l'ensemble des matrices triangulaires inversibles de $\mathrm{GL}_n(k)$.*

Proposition 28. *$B_n(k)$ est le stabilisateur du drapeau complet canonique $\mathcal{C}$ pour l'action naturelle de $\mathrm{GL}_n(k)$ sur les drapeaux. En particulier, $B_n(k)$ est un sous-groupe de $\mathrm{GL}_n(k)$.*

Proposition 29. *Pour $i<j$, $T_{ij}(\lambda)\in B_n(k)$ et pour $\alpha\neq 0$, $D_i(\alpha)\in B_n(k)$.*

Proposition 30. *Soit $(e_1,\ldots,e_n)$ la base canonique de k^n. Pour $\sigma\in\mathcal{S}_n$, on note w_σ l'application linéaire donnée par $w_\sigma(e_i)=e_{\sigma(i)}$.*

Alors $w\colon\mathcal{S}_n\to\mathrm{GL}_n(k)$ est un morphisme de groupes injectif.

Développement 1 :

Théorème 31. *(Décomposition de Bruhat)*

On a la décomposition suivante :

$$\mathrm{GL}_n(k)=\bigsqcup_{\sigma\in\mathcal{S}_n} B_n(k)\,w_\sigma\,B_n(k)$$

Corollaire 32. *$\mathrm{GL}_n(k)$ est engendré par les transvections et les matrices diagonales inversibles.*

3 Autres méthodes de résolution

3.1 Factorisation LU

Théorème 33. *Soit $A\in\mathcal{M}_n(\mathbb{K})$ telles que les n sous-matrices diagonales :*

$$\Delta_k=\begin{pmatrix} a_{11} & \cdots & a_{1k}\\ \vdots & & \vdots\\ a_{k1} & \cdots & a_{kk}\end{pmatrix},\ k\in[\![1,n]\!]$$

Soient inversibles. Alors il existe un unique couple de matrices (L,U) avec U triangulaire supérieure et L triangulaire inférieure de coefficients diagonaux tous égaux à 1, tel que $A=LU$.

Remarque 34. Si la condition sur les mineurs n'est pas respectée, on s'y ramène par permutations.

Application 35. Le système linéaire $AX=B$ équivaut à $LUX=B$. On résout $LY=B$ et $UX=Y$.

Si on connaît la décomposition LU de A, on peut alors résoudre $AX=B$ pour tout $B\in\mathbb{R}^n$.

3.2 Factorisation de Cholesky

Théorème 36. *Soit A une matrice symétrique définie positive. Il existe une matrice réelle triangulaire inférieure B dont tous les éléments diagonaux sont positifs et telle que $A=BB^T$.*

Application 37. Le système linéaire $AX=b$ équivaut à $BB^TX=b$. On résout $BY=b$, puis $B^TY=b$. La méthode de Cholesky a une complexité comparable au pivot de Gauss, en $O(n^3)$ opérations, mais il est a noté qu'elle est plus efficace que cette dernière lorsque A est symétrique définie positive.

3.3 Factorisation QR

Théorème 38. *Soit $A\in\mathcal{M}_n(\mathbb{R})$. Il existe un unique couple (Q,R) avec $Q\in O_n(\mathbb{R})$ et R triangulaire supérieure de coefficients diagonaux strictement positifs tel que $A=QR$.*

Application 39. On a $AX=B\Longleftrightarrow QRX=B\Longleftrightarrow RX=Q^TB$, ce qui donne un système échelonné.

3.4 Méthode de gradient

On se place sur l'espace euclidien $\mathbb{R}^n$ muni de son produit scalaire habituel $\langle .,.\rangle$. L'objectif est de trouver une solution (algorithmique) au système $Ax=b$, où $A\in\mathcal{S}_n^{++}(\mathbb{R})$.

Lemme 40. *(Inégalité de Kantarovitch)*

Soit $A\in\mathcal{S}_n^{++}(\mathbb{R})$. On note $\lambda_1\geq\cdots\geq\lambda_n>0$ ses valeurs propres. Alors :

$$\forall x\in\mathbb{R}^n\quad\langle Ax,x\rangle\langle A^{-1},x\rangle\leq\frac{1}{4}\left[\sqrt{\frac{\lambda_1}{\lambda_n}}+\sqrt{\frac{\lambda_n}{\lambda_1}}\right]^2\|x\|^4$$

Développement 2 :

Théorème 41. *(Gradient à pas optimal)*

Soit $f:\left\{\begin{array}{lll}\mathbb{R}^n & \to & \mathbb{R}\\ x & \mapsto & \frac{1}{2}\langle Ax,x\rangle-\langle b,x\rangle\end{array}\right.$, *où $A\in\mathcal{S}_n^{++}(\mathbb{R})$ et $b\in\mathbb{R}^n$. Pour tout $x_0\in\mathbb{R}^n$, on considère la suite définie par :*

$$x_{k+1}=x_k-t_k\nabla f(x_k)$$

Où t_k est l'unique réel tel que $f(x_k-t_k\nabla f(x_k))=\min_{t\in\mathbb{R}}\{f(x_k-t\nabla f(x_k))\}$.
Alors $(x_n)_{n\in\mathbb{N}}$ converge vers x^*, l'unique solution de $Ax^*=b$. Plus précisément, il existe $C>0$ et $0<\lambda<1$ tels que :

$$\forall k\in\mathbb{N}\quad\|x_k-x^*\|\leq C\lambda^k$$