Lesson 162 : systems of linear equations, elementary operations, algorithmic aspects and theoretical consequences.

I Generalities about systems of linear equations

I.1 Definitions

Definition 1. — We place ourselves in a commutative body $\mathbb{K}$. We call *linear system with p equations and n unknowns* a system :

$$\begin{cases} a_{1,1}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ \vdots & & & & \vdots & \vdots & \vdots \\ a_{p,1}x_1 & + & \cdots & + & a_{p,n}x_n & = & b_p \end{cases}$$

where $X = (x_1, ...x_n)^T$ is a *solution* if and only if the x_i check the equations of S.

Definition 2. — It is said that a system S is *consistent* if and only if there exists a solution.

Example 3. — $\begin{cases} x & + & y & + & z & = & 0 \\ 2x & - & y & + & z & = & 3 \end{cases}$ is consistent, $\begin{cases} x & + & y & = & 0 \\ 2x & + & 2y & = & 4 \end{cases}$ is not.

Remark 4. — A system can be written in :
- matrix form : $AX = B$ where $A = (a_{i,j})_{(i,j)\in[\![1;p]\!]\times[\![1;n]\!]}$ and $B = (b_i)_{i\in[\![1;p]\!]}$
- vector form : $x_1C_1 + \cdots + x_nC_n = B$ where $C_i = (a_{1,i}, ...a_{p,i})^T$.

I.2 Cramer system

Definition 5. — It is said that S is a Cramer system if it is written $AX = B$ in its matrix form, with A a square invertible matrix.

Theorem 6. — A Cramer system always admits one solution $X = A^{-1}B = (\frac{1}{det(A)}det([C1; ...C_{i-1}; B; C_{i+1}; ...C_n]))_{i=1}^n$. The complexity of this method depends on the determinant calculus method.

Example 7. —

$$\begin{cases} x+y & = & 2 \\ x-y & = & 0 \end{cases} \text{ if and only if } \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \text{ if and only if } X = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Theorem 8. — More generally, S is consistent if and only if B is in the vector space generated by $C_1, ...C_n$.

Theorem 9 (Rouché-Fontené). — If the rank of A is r, and $\begin{pmatrix} a_{1,1} & \cdots a_{1,r} & \\ \vdots & \ddots & \vdots \\ a_{r,1} & \cdots & a_{r,r} \end{pmatrix}$ is invertible, then S is consistent if and only if for all k in $[\![r+1; n]\!]$, $\begin{vmatrix} a_{1,1} & \cdots a_{1,r} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{r,1} & \cdots & a_{r,r} & b_r \\ a_{k,1} & \cdots & a_{k,r} & b_k \end{vmatrix}$.

Remark 10. — In this case, we say that $x_1, ...x_r$ are principal variables, and $x_{r+1}, ...x_n$ are free variables.

Corollary 11. — The set of solutions of S (solution set) form an affine space of dimension $n - r$.

I.3 Homogenous case $AX = 0$

Definition 12. — A system S is said to be *homogenous* if $B = 0$.

Remark 13. — By theorem 7, 0 is the only solution to S when A is invertible.

Theorem 14. — Under the same conditions as theorem 9, the solution set is a $(n - r)$-dimensional space.

Corollary 15. — If $n > p$, then S admits a non-zero solution.

Application 16. — If A entries are integers, then there exists a solution $X \in \mathbb{N}^n$ to $AX = 0$ if and only if $0_{\mathbb{R}^p}$ is in the $\mathbb{R}$-convex hull of $(C_1; ...C_p)$.

II Gaussian elimination

II.1 Elementary operations and Gaussian elimination

Proposition 17. — The solution set of S is stable under the following elementary operations :
- equations permutations (denoted $L_i \leftrightarrow L_j$),
- multiplication by a non-zero scalar (denoted $L_i \leftarrow \lambda L_i$)
- adding one equation to another (denoted $L_i \leftarrow L_i + L_j, i \neq j$).

Remark 18. — These basic operations apply to both first and second members.

Proposition 19. —
- $L_i \leftrightarrow L_j$ is equivalent to multiplying $(A \mid B)$ on the left by the permutation matrix P_φ with $\varphi = (ij)$.
- $L_i \leftarrow \lambda L_j$ is equivalent to multiplying $(A \mid B)$ on the left by $\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$ with the λ at coordinates (i, i).
- $L_i \leftarrow L_i + L_j$ is equivalent to multiplying $(A \mid B)$ on the left by $\begin{pmatrix} 1 & & & & \\ & \ddots & & 1 & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$ with the additional 1 at coordinates (i, j). We note $M_{i,j,\lambda}$ the matrix equivalent to $L_i \leftarrow L_i + \lambda L_j$.

Remark 20. — The same matrices can be used on the right to perform these operations on the columns.

Definition 21. — A system S is said to be in echelon form if each line of A contains more zeros to the left than the previous one.

Example 22. — $\begin{pmatrix}1 & 4 & \pi\\ 0 & 0 & \sqrt{2}\\ 0 & 0 & 0\end{pmatrix}$ is in echelon form, and $\begin{pmatrix}3 & 5 & 1\\ 0 & 8 & 2\\ 0 & 4 & 7\end{pmatrix}$ is not.

Definition 23 (Gaussian elimination algorithm). — Here we present the Gauss pivot algorithm, which to a system S returns an equivalent system in echelon form.

- If needed, we swap two lines so that $a_{1,1} \neq 0$,
- Apply $L_i \leftarrow L_i - \frac{a_{i,1}}{a_{1,i}} L_j$ for i in $[\![2;n]\!]$,
- Then the system contains zeros on the first row (except a_{11}). Iterate the same algorithm on $\begin{pmatrix}a_{2,2} & \cdots & a_{2,n}\\ \vdots & \ddots & \vdots\\ a_{p,2} & \cdots & a_{p,n}\end{pmatrix} = A'$

Remark 24. — If $C_1 = 0_{\mathbb{R}^p}$, then iterate on $A' = \begin{pmatrix}a_{1,2} & \cdots & a_{1,n}\\ \vdots & \ddots & \vdots\\ a_{p,2} & \cdots & a_{p,n}\end{pmatrix}$

II.2 Applications

Application 25 (Immediate applications). —

- Finding the solution set of $AX = B$,
- Determining if a family is linearly independent, spans the whole space or is a basis of $\mathbb{R}^p$,
- Determining if a vector is in the span of a given family,
- Determining the intersection of two vector subspaces,
- Determining the rank, the image of A.

Application 26. — The transvections generate the group $SL(E)$. The directionnal scalings and the transvections generate the group $GL(E)$.

Remark 27. — The Gaussian elimination is used in real life to compute the solution set of $AX = B$, when we don't have further information about A (for example, a dense matrix).

III Matrix decomposition

Remark 28. — The aim is to factorize A into the product of matrices with good properties (diagonal, triangular, easy to invert...). We suppose $n = p$.

III.1 LU decomposition

Remark 29. — If we never use permutations in the Gauss elimination, we get : $\prod_{(i,j,\lambda} M_{i,j,\lambda} A = U$ upper triangular, so $A = LU$ with L lower triangular.

Theorem 30. — If for all k in $[\![1;n]\!]$, $\begin{vmatrix}a_{1,1} & \cdots & a_{1,k}\\ \vdots & \ddots & \vdots\\ a_{k,1} & \cdots & a_{k,k}\end{vmatrix} \neq 0$, then there exists a unique decomposition $A = LU$ such that L is lower triangular with ones on the diagonal and U is upper triangular.

Example 31. —

$$\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix} = \begin{pmatrix}1 & 0\\ 3 & 1\end{pmatrix}\begin{pmatrix}1 & 2\\ 0 & -2\end{pmatrix}$$

Corollary 32. — Under the condition A invertible, there exists a decomposition $A = PLU$ with P a permutation matrix.

Remark 33. — If $A = LU$, then $AX = B$ if and only if $UX = L^{-1}B$ which is easy to compute and solve.

III.2 QR decomposition

Theorem 34. — If A is a complex matrix, there exists a unitary matrix Q and an upper triangular matrix R. If we suppose that the diagonal entries are positive reals, then this decomposition is unique.

Proposition 35. —

$$\begin{array}{cccccc} f & : & U_n(\mathbb{C}) & \times & T_n^+(\mathbb{C}) & \rightarrow & GL_n(\mathbb{C}) \\ & & (Q & , & R) & \mapsto & QR \end{array}$$

is a homeomorphism.

Application 36 (Development). — If A is an invertible matrix such that $A = P^{-1}DP$, $D = \begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$ $|\lambda_1| > ... > |\lambda_n| > 0$ and P^{-1} admits a LU decomposition, then the diagonal of the sequence defined by :

$$\begin{cases} A_1 & = & A \\ A_{k+1} & = & R_k Q_k \text{ where } Q_k R_k \text{ is the QR decomposition of } A_k \end{cases}$$

converges to $(\lambda_1, ...\lambda_n)$.

IV Approximation methods to solve $AX = B$

Remark 37. — The methods we saw to compute a solution of $AX = B$ (Cramer method and Gaussian elimination method) have a big time complexity. We will now see approximation methods.

IV.1 Iterative methods

Remark 38. — In iterative methods, we decompose $A = M - N$ with M easy to invert. Then, we have $AX = B$ if and only if $MX = NX + B$ if and only if $X = M^{-1}NX + M^{-1}B$.

Theorem 39. — We associate to a $M - N$ decomposition an iterative method, where we define the sequence : $\begin{cases} X_0 & \text{arbitrary choosed} \\ X_{k+1} = & M^{-1}NX_k + M^{-1}B \end{cases}$, which converges when the spectral radius of $M^{-1}N$ is strictly less than 1.

Definition 40 (two methods). —

- The Jacobi method is defined by $M = Diag(A)$, and converges when A is a strictly diagonally dominant matrix.
- The Gauss-Seidel method is defined by M being A with zeros strictly above the diagonal. It converges when A is a positive definite symmetric matrix, or when it is strictly diagonally dominant.

IV.2 Gradient descent method

Remark 41. — Gradient descent is an iterative optimization algorithm for finding the local minimum of a differentiable function. It consists of descending the slope of f graph by the steepest direction. We can use it on the function

$$\begin{array}{cccc} f & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & X & \mapsto & \frac{1}{2}\langle AX, X\rangle - \langle B, X\rangle \end{array}$$

where we suppose that A is a positive definite symmetric matrix.

Lemma 42 (Kantorovich inequality). — If μ_1 and μ_n are the lowest and biggest eigenvalues of A, then we have :

$$\frac{\|X\|^4}{\|X\|_A^2\|X\|_{A^{-1}}^2} \geq 4\frac{\mu_1\mu_n}{(\mu_1+\mu_n)^2}$$

Theorem 43. — The gradient descent algorithm defines three sequences :

$$\left\{\begin{array}{ccl} X_0 & \in & \mathbb{R}^n \text{ arbitrary,} \\ R_0 & = & AX_0 - B \\ \alpha_{k+1} & = & \frac{\|R_k\|^2}{\langle AR_k, R_k\rangle} \\ X_{k+1} & = & AX_k - \alpha_{k+1}R_k \\ R_{k+1} & = & AX_{k+1} - B \end{array}\right.$$

, X_k converges to $\overline{X}$ the solution of $AX = B$, and :

$$\|X_k - \overline{X}\|_2 \leq \sqrt{Cond_2(A)}\frac{Cond_2(A)-1}{Cond_2(A)+1}^k\|X_0 - \overline{X}\|_2$$

.